한국정보전자통신기술학회논문지(jkiiect)'18-06, Vol.11 No.3
논문 18-11-03-277 http://dx.doi.org/10.17661/jkiiect.2018.11.3.277

유전적 알고리즘과 LSB를 이용한 스테가노그래피의 정보은닉 기법

지선수

A Techniques for Information Hiding in the Steganography using LSB and Genetic Algorithm

Seon-Su Ji*

요 약 인터넷 상에서 비밀 메시지의 통신 목표는 비인지성과 기밀성을 유지하는 것이다. 디지털 스테가노그래피는 메시지 존재 자체를 제3자가 감지하지 못하게 커버 매체에 비밀 메시지를 삽입하여 목적지에 전송하는 기법이다. 스테가노그래피는 암호화 기법과 혼합되어 기밀성과 무결성을 함께 보장하기 위한 효율적인 방법이다. 비밀(한글) 메시지를 삽입하기 위해 비밀 문자를 분리하고 암호표를 참고로 하여 이진화 코드로 변환하며, 커버 이미지를 두 영역으로 분할하며, 비밀 메시지와 두 번째 영역의 오른쪽 l-LSB 정보를 암호화와 교차 과정을 진행한 후 첫 번째 영역의 k-LSB에 은닉하여 스테고 이미지를 구성하는 방법을 제안한다. 제안된 방법의 실험결과는 PSNR 값이 52.62로 허용 이미지 품질 수준임을 보여준다.

Abstract The goal of the secret message communication on the internet is to maintain invisibility and confidentiality. Digital steganography is a technique in which a secret message is inserted in a cover medium and transmitted to a destination so that a third party can not perceive the existence of the message itself. Steganography is an efficient method for ensuring confidentiality and integrity together with encryption techniques. In order to insert a secret (Hangul) message, I propose a image steganography method that the secret character is separated and converted into binary code with reference to the encryption table, the cover image is divided into two areas, and the secret message and the right l-LSB information of the second area are encrypted and crossed, concealing the k-LSB of the first region. The experimental results of the proposed method show that the PSNR value is 52.62 and the acceptable image quality level.

Key Words : Crossover, Hangul text - Secret Message, Image Steganography, Information Hiding, LSB technique

1. 서 론

지능정보사회에서 통신 네트워크를 통해, 정보의 편리성과 보안성을 기반으로, 디지털 정보가 소통되고 있으며, 미래사회에서 의존도는 확대되고, 집중화 될 것이다. 인터넷을 통한 메시지의 송수신은 정보의 무결성과 기밀성, 저작권 보호 등의 문제를 가지며, 이를 해결하기 위해 암호화와 정보 은닉 등의 방법을 사용한다. 이를 위해 일반적으로 다음과 같이 분류한다. 이미지에 비밀 메시지를 숨기는 방법은 제3자가 이해할 수 없도록 정보를 암호화 한다. 즉 수신자와 송신자가 공통키를 사용하여 암호화 및 복호화를 하는데, 알고리즘의 보안성 수준에 의해 정보의 기밀성을 구현한다. 다른 방법으로 정보의 비밀성을 유지하며, 비밀 메시지 존재 자체를 숨기는 스테가노그래피가 있다. 세 번째 방법으로 저작권 보호와 추적을 위한 워터마킹 기법이 있다.

디지털 스테가노그래피는 커버 매체에 비밀 메시지를 삽입하는 기법으로 스테고 키가 없는 순수 스테가노그래피,

Department of software, Gangneung Wonju National University
Received April 25, 2018 Revised May 03, 2018 Accepted May 29 2018

통신 스테고 키를 이용한 비밀키 스테가노그래피, 공개키 기반 스테가노그래피로 나누어 설명된다. 또한 커버 매체의 형태에 따라 이미지, 네트워크, 비디오, 오디오, 텍스트 형태의 스테가노그래피가 있으며, 인터넷에서 송수신되는 매체로서 이미지의 사용 빈도 증가가 가장 높으며, 실제 보안 분야에서 많은 적용 예를 가지고 있다. 디지털 이미지의 이용은 인간 시각 시스템의 약점, 예를 들어 인간의 시력은 적색 및 청색 등의 색상, 형태 등에서 미세한 변화를 구별하는 능력이 취약하다는 것을 역으로 이용한다[1-2].

비밀(한글) 메시지에서 각각의 문자를 (1)초성, 중성, 종성으로 분리한 후 변환 암호표를 적용하여 재배열한 후(l'), (2)커버 매체를 두 개의 영역으로 분리[2]하며, (3)두 번째 영역에서 비트화 된 정보의 l'-LSB 참고로 암호화와 l을 기반으로 교차기법[3]을 적용하며, 이를 기반으로 하는 정보를 (4)첫 번째 영역의 k-LSB에 은닉하는 기법을 제안한다.

2. 관련 연구

스테가노그래피에서 일반적으로 사용하는 정보은닉 기법은 순차적 혹은 임의로 선택한 픽셀의 최하위 k-비트에 메시지를 삽입하는 것이다. 암호화와 위치 등을 함께 사용하여 보안성과 저항성을 높이는 연구가 진행되고 있다.

Al-Farraji는 커버 이미지를 두 영역으로 분리하여 비밀 메시지와 첫 번째 영역의 2-LSB를 비교하며, 일치할 경우의 정보를 두 번째 영역에 은닉하는 기법을 제안하여 보안성을 높였다[2]. Wang 등은 유전 알고리즘, 무작위 추출 방법과 최적의 LSB 치환 방법을 이용하여 정보를 은닉하는 새로운 기법을 제안하였으며[3], Nag 등은 아핀암호의 개인키와 스테가노그래피의 두 개의 층을 이용하여 커버 매체로 이미지를 이용할 때가 매우 효과적임을 보였다[4]. Tavoli 등은 특정 스캔을 적용한 상태에서 적절한 마스크 사용의 혼합, 은닉 자료 압축 상태를 암호화 한 후 이미지 픽셀 공간에 정보를 삽입하며, 주파수 영역을 이용하는 것보다 삽입용량이 향상되는 것을 보였다[5].

Chan 등과 Wang 등은 스테고 매체와 커버 매체 간의 최악의 평균 제곱 오차(MSE, mean square error)를 유도하였으며, 최적의 픽셀 조정 과정을 이용한 LSB 대체에 의한 데이터 은닉 방법을 보였으며, 스테고 매체의 최악의 이미지 품질 기준을 제시하였다. 즉 $M_c \times N_c$ 픽셀의 커버 이미지를 다음과 같이 표시할 수 있으며[6-8],

$$C=\{x_{ij} \mid 0 \le i \le M_c,\ 0 \le j \le N_c,\ x_{ij} \in \{0,1,2,\dots,255\}\} \quad (1)$$

여기에서 x_{ij}는 (i,j)위치에서 픽셀값을 의미하며, M은 n-비밀 메시지를 나타낸다.

$$M=\{m_i \mid 0 \le i < n,\ m_i \in \{0,1\}\} \quad (2)$$

n 비트 비밀 메시지 M이 커버 이미지 C의 최하위 비트의 처음 오른쪽 k비트에 삽입된다고 가정한다. 비밀 메시지 M을 재배치하여 k비트의 허상 m'을 형성한다.

$$M'=\{m_i' \mid 0 \le i < n',\ m'_i \in \{0,1,\dots,2^k-1\}\} \quad (3)$$

여기에서 $n'=M_c \times N_c$이며, n 비트 비밀 메시지 $M=\{m_i\}$과 삽입된 메시지 $M'=\{m'_i\}$ 사이의 매핑은 다음과 같이 정의될 수 있다.

$$m'_i = \sum_{j=0}^{k-1} m_{i\times k+j} \times 2^{k-1-j} \quad (4)$$

커버 이미지 C로부터 차례로 픽셀값 $\{c_1,c_2,\dots,c_n\}$을 획득한다. c_i의 오른쪽 최하위 k비트에 m'_i로 대체하는 과정으로 삽입이 이루어진다. 수학적으로 k비트 메시지 m'_i를 저장하기 위해 커버 이미지의 (i,j) 위치에서 선택한 픽셀의 값 c_i는 스테고 픽셀 s_i의 형태로 나타낼 수 있다[6-8].

$$s_i = c_i - c_i \bmod 2^k + m'_i \quad (5)$$

삽입된 메시지 m'_i는 스테고 이미지의 (i,j) 위치에서 다음 식에 의해 복원할 수 있다.

$$m'_i = s_i \bmod 2^k \quad (6)$$

스테고 이미지와 커버 이미지 사이의 최악의 평균제곱 오류는 최하위 비트를 이용하는 기법에서 얻은 것의 $\frac{4}{9}$가 됨을 보였으며[6-8], 최적의 삽입용량을 얻기 위한 간단한 방법은 각 대체에 대한 PSNR(peak signal to noise ratio)을 계산하고, 최적의 결과로 최대 PSNR을 선택함을 제시하였다.

$$PSNR_* = 10 \cdot \log_{10} \frac{255^2}{(2^{k-1})^2} \quad (dB) \qquad (7)$$

이를 기준으로 PSNR의 최저 기준을 표 1과 같이 설정하였으며, $PSNR_*$는 (7)식에 의해 계산된 결과이다.

표 1. k값이 1,2,3,4,5 경우 $PSNR_*$
Table 1. $PSNR_*$ for k=1,2,3,4,5

k	1	2	3	4	5
$PSNR_*$	48.13	42.11	36.09	30.07	24.05

Wu와 Tsai는 비트화된 비밀 메시지를 얼마만큼 삽입하는지를 결정하기 위해 두 이웃한 픽셀 사이의 차이를 이용하는 새로운 스테가노그래픽 기법을 제안한다. 여기에서 삽입되는 자료의 비트와 유사한 값으로 커버 이미지의 2 픽셀 블록의 차이 값을 대체하는 것에 의해 커버 이미지로 비밀자료를 삽입한다[9]. Liao 등은 삽입용량을 개선하기 위해 4-픽셀 차이와 수정된 LSB의 대체를 기반으로 하는 스테가노그래피 기법을 제안하였으며, 삽입용량과 비인지성 측면에서 우수함을 제시하며, 이때 삽입용량/이미지 품질을 높이기 위해 공격 저항성을 낮추는 상충관계(trade off)의 타당성을 보였다[10]. Yang 등은 테두리(edge) 영역과 평활화(smooth) 영역 사이를 구별하기 위한 두 개의 연속된 픽셀의 차이 값을 이용하여 수정된 LSB 스테가노그래피 기법을 제안하였다. 여기에서 테두리 영역에 위치한 픽셀은 평활화 영역에 있는 픽셀보다 k 값 이상을 가진 k LSB에 삽입하는 방법으로 이미지 품질, 비인지성과 삽입용량이 Wu 등의 연구에서 보다 개선됨을 보였다[11].

3. 제안된 방법

비밀 (한글) 메시지를 커버 매체에 삽입하는 방법을 제안한다. 조합형 한글을 은닉할 경우 삽입용량과 저항성 측면에서 효율성을 개선하기 위해 다음의 방법을 고려한다. 선택된 비밀 문자를 초성, 중성, 종성으로 분리한 후 변환표에 의해 이진화 코드로 재정렬한 후 커버 매체를 2개의 영역으로 분리한다. 암호화를 적용한 후 l을 기반으로 하는 유전 알고리즘의 교차와 k-LSB를 기반으로 하는 비트 정보를 은닉하는 과정을 적용한다. 표 2는 한글 사용 빈도수를 기반으로 초성, 중성, 종성 문자를 재배치한 이진화 코드표이다.

표 2. 한글 자음과 모음에서 이용되는 코드
Table 2. Code used in Hangul consonants and vowels

Chosung (initial)	Jungsung (medial)	Jongsung (final)	Binary code
ㅇ ㅁ ㅆ	ㅏ ㅔ ㅖ	null ㅅ ㄹㄱ ㄹㅂ	000
ㄱ ㅂ ㅉ	ㅣ ㅟ ㅐ	ㄴ ㄴㅎ ㄲ ㄹㅌ	001
ㄷ ㅊ ㅃ	ㅡ ㅘ ㅖ	ㄹ ㅂㅅ ㄹㅁ ㄹㅍ	010
ㅈ ㅌ	ㅓ ㅚ ㅒ	ㅇ ㅌ ㄹㅎ ㄹㅅ	011
ㅅ ㅍ	ㅗ ㅛ	ㄱ ㅎ ㄴㅈ	100
ㅎ ㄸ	ㅜ ㅑ	ㅁ ㄷ ㅈ	101
ㄹ ㅋ	ㅕ ㅠ	ㅆ ㅍ ㄱㅅ	110
ㄴ ㄲ	ㅒ ㅝ	ㅂ ㅊ ㅋ	111

3.1 삽입 과정

커버 이미지에 비밀 메시지를 삽입하는 과정은 다음과 같다.

1. 은닉하려는 비밀 메시지(M)로부터 한 문자씩 읽어 들여 초성, 중성, 종성으로 분리한 후 한글 문자 변환표(Table 2)에 의해 이진화 코드를 획득하여 저장한다. (array1)

2. 커버 이미지를 읽어 들여 두 개의 영역으로 분리하여 저장한다. (array2, array3)

3. 비밀 메시지(array1)와 커버 이미지(array3) 기반으로 암호화와 l-LSB를 참고로 교차 과정을 진행한다.

3.1 비밀 메시지의 9비트(m_i)와 두 번째 영역(array3)의 최하위 비트(LSB) 처음 9개의 비트(c_i) 정보를

추출하여 XOR 연산한다. 여기에서 $l' = 9$이다.

$$cipher_i = c_i \oplus m_i$$

$$1 \le i \le \le ngth\ of\ (M)\ in\ block-bits$$

3.2 3.1에서 획득한 정보에서 최상위 l비트 정보와 최하위 l비트 정보를 교차시켜 예비공간인 array4에 저장한다. Fig 1에서 $l = 3$일 경우에 교차와 돌연변이 과정을 보여준다.

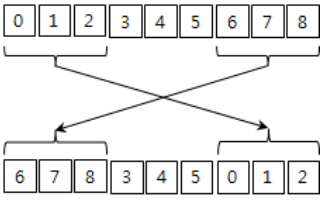

그림1. 돌연변이 과정의 결과
Fig 1. Result of mutation operation

3.3 비밀 메시지의 끝이 나타날 때까지 3.1단계부터 3.2단계를 반복한다.

4. array4 정보를 array2에 은닉한다. 최초 은닉시점을 확인한다.

4.1 array4의 이진화 코드에서 k비트씩 선택한 후 array2의 오른쪽 최하위 k비트에 대체한다.

4.2 array4에 저장된 정보 끝까지 4.1 단계를 반복한다.

5. 은닉시점과 종점의 위치를 저장한다. (key1, key2)

6. 분할된 두 영역을 결합하여 스테고 이미지를 완성한다.

7. 스테고 이미지와 키 정보를 전송한다.

3.2 추출 과정

스테고 이미지와 키 정보를 이용하여 삽입된 메시지를 추출하는 과정은 다음과 같다.

1. 수신된 정보로부터 스테고 매체와 키 정보를 획득한다.

2. 스테고 매체를 두 영역으로 분할한다. (array2, array3)

3. 은닉 위치키(key1, key2)를 기반으로 array2에서 k LSB 정보를 순차적으로 예비공간인 array4에 저장한다.

4. 3단계에서 획득한 정보를 기반으로 은닉문자를 추출한다.

4.1 9비트를 선택한다.

4.2 최상위 l비트 정보와 최하위 l비트 정보를 교차시킨다.

4.3 9비트와 두 번째 영역(array3)의 최하위 비트(LSB) 처음 9개의 비트 정보를 추출하여 XOR 연산한다.

4.4 연산결과를 3영역으로 분리한 후 한글 문자 변환표(표 2)를 참고로 하여 초성, 중성, 종성 자에 해당되는 정보를 기반으로 은닉 문자를 완성한다.

5. 은닉 종점이 나타날 때까지 3단계부터 4단계를 반복한다.

6. 완성된 비밀 메시지를 확인한다.

제안된 커버 매체에 대한 스테고 매체의 이미지 품질은 (8)식을 이용하여 평가한다[3,6,7].

$$PSNR = 10 \cdot \log_{10} \frac{(2^h - 1)^2}{MSE}\ (dB) \qquad (8)$$

$$MSE = \frac{1}{h} \sum_{i=0}^{h} (\hat{X}_i - X_i)^2 \qquad (9)$$

커버 매체와 비밀 메시지가 삽입된 스테고 매체 사이의 유사성을 비교하기 위한 상관계수는 수식 (10)을 사용하여 계산할 수 있다[12].

$$Corr = \frac{\sum_{i=1}^{h} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{h} (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{h} (y_i - \bar{y})^2}} \qquad (10)$$

여기에서 h는 이미지의 픽셀 수를 나타내며, $\hat{X}_i$는 커버 매체정보를 나타내고, X_i는 스테고 매체정보를 나타낸다. $x_i \in X$, $y_i \in Y$이다. $\bar{x}$와 $\bar{y}$는 X와 Y의 각각의 평균을 의미한다.

4. 적용 및 결과

논문에서 사용된 비밀 메시지의 크기는 2, 6, 12, 18 바이트이며 커버 매체의 크기는 11,772 바이트이다. 3.1에

서 제안된 방법으로 비밀 메시지를 은닉하였다. 은닉된 문자를 추출할 경우 3.2의 추출 과정을 이용할 수 있다. 알고리즘을 구현하는 과정은 J2SE와 Matlab을 이용하였다. 예를 들어 '한' 글자를 은닉하기 위해 'ㅎ:101'+'ㅏ:000'+'ㄴ:001'으로 정보를 분리한 후 표 2의 이진화 코드를 참고로 하여 3.1절의 3과 4단계에 따라 이진화된 코드 정보를 은닉한다.

삽입되는 비밀 메시지의 크기를 기반으로 k값이 1, 2, 3에 따라 커버 매체와 스테고 매체의 유사성을 비교하기 위해 (10)식을 이용하였다. 또한 스테고 매체의 이미지 품질을 위해 (8)식을 이용하였으며, (7)식에 의해 계산된 수치와 비교하였다.

표 3에서 k값에 따라 커버 매체와 스테고 매체의 유사성과 삽입 용량을 확인할 수 있다. 또한 k가 1일 때 커버 매체와 스테고 매체의 중복률은 48.1%이며, 2일 때 17.5%, 3일 때 11.1% 내외임을 보였으며, 유사성 측면에서도 0.9870으로 우수함을 알 수 있다. 또한 이미지 품질 측면에서 표 1의 최저 기준을 만족함을 확인하였다.

표 3. 커버 이미지에 삽입된 비밀메시지의 결과
Table 3. The result of inserting a secret message into the cover image

secret messages	k	PSNR	Capacity	Corr.
2	1	56.68	11.1	0.9974
	2	43.65	20.0	0.9807
	3	37.22	33.3	0.9814
6	1	51.30	11.1	0.9998
	2	42.18	21.4	0.9773
	3	39.34	33.1	0.9747
12	1	50.68	10.34	0.9997
	2	42.67	22.2	0.9991
	3	37.23	33.2	0.9816
18	1	51.85	11.1	0.9997
	2	44.48	21.95	0.9986
	3	39.00	33.3	0.9962

5. 결 론

비밀 메시지를 조합형 문자의 유니코드를 이용하는 것보다 초성, 중성, 종성으로 분리하여 삽입하면 43.8%의 메모리를 절약할 수 있다. 커버 매체의 분리와 키를 이용한 암호화, l을 기반으로 하는 교차와 $k-LSB$를 적용함으로 저항성과 기밀성 측면에서 효율성을 높일 수 있으며, 유사성과 스테고 매체의 이미지 품질이 각각 0.9896과 최저 수준을 모두 만족함을 확인하였다.

REFERENCES

[1]S. S. Ji, "A Study of Hangul Text Steganography based on Genetic Algorithm", KIISC, Vol. 21, No. 3, pp. 7-12, 2016.

[2]Orooba Ismaeel Ibraheem Al-Farraji, "New Technique of Steganography based on Locations of LSB", International Journal of Information Research and Review, Vol. 4, Issue 1, pp. 3549-3553, 2017.

[3]R. Z. Wang, C. F. Lin and J. C. Lin, "Image Hiding by Optimal LSB Substitution and Genetic Algorithm", Pattern Recognition, Vol. 34, No. 3, pp. 671-683, 2001.

[4]A. Nag, J. P. Singh, S. Khan, Ghosh S., Biswas S., Sarkar D., and Sarkar P. P., "A Weighted Location Based LSB Image Steganography Technique", ACC 2011, Part II, CCIS 191, pp. 620-627, 2011.

[5]Reza tavoli, Maryam bakhshi and Fatemeh salehian, "A New Method for Text Hiding in the Image by Using LSB", International Journal of Advanced Computer Science and Applications, Vol. 7, No. 4, pp. 126-132, 2016.

[6]C. K. Chan and L. M. Cheng, "Hiding Data in Images by Simple LSB Substitution", Pattern Recognition, Vol. 37, No. 3, pp. 469-474, 2004.

[7]Marghny Mohamed, Fadwa Al-Afari and Mohamed Bamatraf, "Data Hiding by LSB Substitution Using Genetic Optimal Key-Permutation", International Arab Journal of e-Technology, Vol. 2, No. 1, pp. 11-17, 2011.

[8]R. Z. Wang, C. F. Lin and J. C. Lin, "Hiding Data in Images by Optimal Moderately Significant-bit Replacement", IEE Electron. Lett. Vol. 36, No. 25, pp. 2069-2070, 2000.

[9]Da Chun Wu and Wen Hsiang Tsai, "A Steganographic Method for Images by Pixel Value Differencing", Pattern Recognition Letters, Vol. 24, No. 9-10, pp. 1613-1626, 2003.

[10]Xin Liao, Qiao-yan Wen, and Jie Zhang, "A Steganographic Method for Digital Images with Four-pixel Differencing and Modified LSB Substitution", Journal Visual Communication and Image Representation, Vol.22, pp. 1-8, 2011.

[11]C. H. Yang, C. Y. Weng, S. J. Wang and H. M. Sun, "Adaptive Data Hiding in Edge Areas of Images with Spatial LSB Domain Systems", IEEE Transactions on Information Forensics and Security, Vol. 3, No. 3, pp. 488-497, 2008.
[12]G. Swain and S. K. Lenka, "Classification of Image Steganography Techniques in Spatial Domain: A Study", International Journal of Computer Science& Engineering Technology, Vol. 5, No. 3, pp. 219-232, 2014.

저자약력

지 선 수(Seon-Su Ji) **[종신회원]**

- 충남대학교 계산통계학과 (학사)
- 중앙대학교 응용통계학과(석사)
- 중앙대학교 응용통계학과(박사)
- 명지대학교 컴퓨터공학과(박사수료)
- (현)강릉원주대학교 소프트웨어학과 교수

<관심분야>

정보보호(암호키, 정보은닉), 스테가노그래피