Journal of the Korea Industrial Information Systems Research Vol. 19 No.5, Oct. 2014
http://dx.doi.org/10.9723/jksiis.2014.19.5.051

# 텍스트 스테가노그래프의 개선된 접근과 연구

# (A Study and improved Approach of Text Steganography)

지 선 수[1)]
(Seon-Su Ji)

**요 약** 인터넷의 디지털 세상에서 스테가노그래피는 의심스럽지 않은 커버 매체 안에 비밀 메시지를 숨겨서 비밀 통신의 존재를 은닉하기 위해 도입되었다. 제 3자는 비밀 메시지가 전달되는 사실을 인식하지 못한다. 텍스트 기반 스테가노그래피 기법은 다양하게 적용할 수 있다. 이 논문에서는 존재하는 각각의 텍스트 스테가노그래픽 기법의 장점과 단점을 분석하고, 효율적인 접근 방법을 제시한다. 외부적 공격으로부터 비밀 메시지를 안전하게 숨기기 위해 재배열 순서키에 의한 방법을 제안한다.

**핵심주제어 :** 삽입 용량, 스테가노그래피, 텍스트 스테가노그래피, 텍스트 은닉

**Abstract** In the digital world of the internet, steganography is introduced to hide the existence of the secret communication by concealing a secret message inside another unsuspicious cover medium. The third parties are unaware that a stego medium is being communicated. There exists a large variety of steganography methods based on texts. In this paper, analyzed the advantages and significant disadvantages of each existing text steganography method and how new approach could be proposed as a solution. The objective of this paper is to propose a method for hiding the secret messages in safer manner from external attacks by encryption rearrangement key.

**Key Words :** Capacity, Steganography, Text hiding, Text steganography

## 1. 서 론

최근의 인터넷은 하루가 빠르게 성장하고, 사람들에게 멈출 수 없는 욕구와 매력을 제공하고 있으며 이곳을 통해 주고 받는 정보의 량과 영역은 무한하다. 그리고 정보의 무결성을 보호하기 위해 정보보안 및 보호는 가장 중요하고 핵심적 가치로 자림 매김 되었다. 보안성과 견고성 측면에서 전달하려는 비밀 메시지를 암호화하여 숨기려는 스테가노그래피는 암호화보다 더 많은 관심을 받는 효과적인 기술이다. 스테가노그래피의 목적은 텍스트, 이미지, 오디오 및 비디오 매개체에 비밀 메시지가 숨겨진 스테고 매체의 존재 자체를 은폐하여 통신 채널을 통해 허가된 수신자에게만 무결성 자료를 효율적이고, 안전하게 전송하는 것이다. 스테가노그래피 기법은 텍스트, 이미지, 오디오, 비디오 등의 네 가지 기본 유형이 있다. 다른 매개체보다 텍스트는 텍스트 파일에서 중복된 여분의 정보가 부족하기 때문에 적용면에서 가장 어려운 기법중의 하나이다. 텍스트 스테가노그래피는 정보를 숨기는 기술의 하나로써 언어적 스테가노그래피와 기술

1) 강릉원주대학교 정보기술공학과

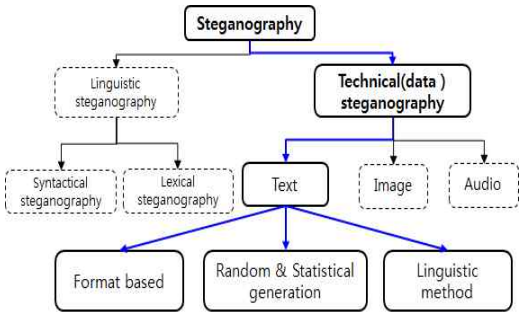
적인 스테가노그래피로 분류된다. 언어적인 스테가노그래피는 비밀 메시지를 숨기기 위해 자연언어를 커버 매체로 이용하는 기술이다. 기술적인 스테가노그래피는 다른 물리적인 매개체로서 나타낼 수 있는 텍스트보다 전송 매체로 설명된다[1-2]. 정보를 숨기는 과정에서 다루어야 할 중요한 3가지 요소는 삽입 용량, 보안성, 공격자가 숨겨진 정보를 파괴하기 전에 스테고 매체가 견딜 수 있는 수정되는 양으로 설명될 수는 견고성 등이다.

이 논문에서는 텍스트 스테가노그래피의 적용 형태별 장점과 단점을 분석하고, 비트화되고 암호화된 비밀 메시지를 재배열키에 따라 배치 순서를 다르게 하여 비어있는 공간에 은닉하여 보안성을 강화하는 개선된 방법을 제시한다.

## 2. 텍스트 스테가노그래피

텍스트는 스테가노그래피에서 이용되는 가장 오래된 매개체 중의 하나로써 전자적 문서화 정보와 언어 분석으로 구현된다. 텍스트 스테가노그래피는 텍스트 매체에 비밀 자료의 조각을 숨기는 것으로써 전송된 텍스트의 의미가 바뀌지 않는 범위 내에서 문서의 구조를 조금씩 변경하여 비밀 메시지를 은닉하는 기법이다.

### 2.1 기본 형태




<Fig. 1>Type of text steganograohy

<Fig. 1>에서와 같이 텍스트 의미의 변화를 기반으로 하는 형식 기반, 임의 및 통계적 생성, 언어적 방법으로 설명된다[2-4].

#### 2.1.1 Format based method

형식 기반의 방법은 정보를 숨길 수 있는 매개체로 텍스트의 물리적인 문서형태(서식)를 사용한다. 이러한 방법은 메시지를 숨기기 위해 기존 텍스트 일부를 수정한다. 삽입 공간, 글꼴 크기, 텍스트 전체에 분산된 맞춤법 오류 등은 텍스트 스테가노그래피에서 사용되는 다양한 형식 기반 방법 중의 하나이다[4-5]. 과거에는 인간의 가시력 한계 때문에 속일 수 있었지만 컴퓨터가 대중화되어 쉽게 노출될 수 있다.

#### 2.1.2 Random and Statistical generation method

임의 및 통계적 생성은 통계적 특성에 따라 커버 매체(텍스트)를 생성한다. 이 방법은 문자와 단어의 시퀀스를 기반으로 한다. 문자 시퀀스 안에 정보를 숨기는 것으로 문자에 임의의 시퀀스에서 나타나는 정보를 끼워 넣는 방법이다. 문자 생성에 대한 두 번째 접근법은 주어진 언어에서 실제 단어와 같은 통계적 속성이 나타나는 단어를 만들기 위해 단어 길이와 문자 주파수의 통계적 특성을 갖는다. 워드 시퀀스 안으로 정보를 숨기는 것은 실제 사전 항목이 어휘 항목 및 비트 시퀀스 사이에 코드북 매핑을 이용하는 단어마다 더 많은 정보의 비트를 인코딩하거나 단어 자체가 숨겨진 정보를 인코딩할 수 있다[4-5].

#### 2.1.3 Linguistic method

컴퓨터의 발달은 복잡한 언어학적 구조를 분석하게 하며, 수정된 언어적 특성을 고려한 경우 생성된 언어적 특성과 수정된 텍스트는 메시지를 숨기기 위한 장소로써 언어적 구조를 이용한다. 스테가노그래픽 자료는 구문 구조 자체에  숨길 수 있다. 이때 어휘, 문법, 의미적인 검사에 견딜 수 있는 견고성을 만족해야 한다[5].

### 2.2 Available method

일반적으로 텍스트 스테가노그래피는 텍스트 형태의 변화와 텍스트 의미의 변화로 나누어 분류된다. 텍스트 형태의 변화를 기반으로 Line Shifting, Word Shifting 등이 있다. 텍스트 의미의 변화를 기반으로

Syntactic Method, Semantic Method, Abbreviations, Change Spelling 등이 있다[6-7].

이밖에 웹페이지 코딩의 HTML과 CSS(cascading style sheets) 부분에 텍스트를 숨기는 새로운 기법 즉, 웹페이지는 많은 양의 정보를 포함할 수 있고, 외부적으로 감지하기가 어려우며, 태그에 비밀 메시지를 숨길 수 있다는 것을 제시하였지만 보안성을 크게 강화하지는 못하였다. 단어 사이의 white/null space를 이용하여 비트 단위의 비밀 메시지를 숨기는 방법을 제시하였으며, 삽입 용량 면에서 효율적이지는 못하다[8].

### 2.3 분석도구

다양한 텍스트 스테가노그래피 기법에서 각각의 장점과 약점이 존재하므로 기본적인 방법과 성능을 분석하며, 효율적으로 적용을 할 필요가 있다. 일반적으로 Linguistic 기법과 space를 이용하여 비밀 메시지를 은닉하는 방법을 사용한다[2,5,9]. 텍스트 스테가노그래피에서 왜곡 정도를 알아보기 위해 유사성을 측정하며, 효율성을 위해 삽입 용량을 확인할 필요가 있다. 텍스트 스테가노그래피에서 비밀 메시지를 삽입할 수 있는 삽입 용량은 수식 (1)로 계산할 수 있다[5].

$$pacity = \frac{Nmber\ of\ hidden\ bits}{Cover\ file\ size} \cdot 100 \quad (1)$$

커버 매체와 비밀 메시지가 삽입된 스테고 매체 사이의 유사성을 비교하기 위한[4] 상관계수를 수식 (2)를 사용하여 측정할 수 있다. 여기에서 $X$는 커버 매체, $Y$는 스테고 매체 자료이며, $x \quad X$, $y_i \in Y$이다. $x$와 $y$는 $X$와 $Y$의 각각의 표본 평균을 의미한다.

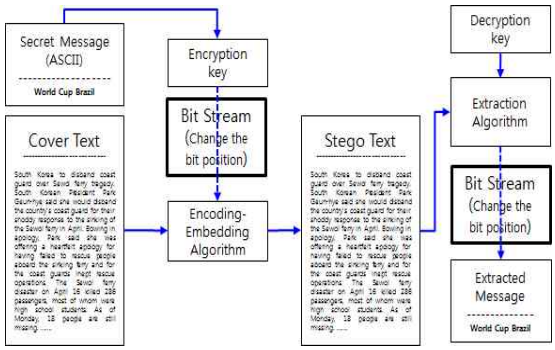
$$Corr = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2 \cdot \sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}} \quad (2)$$

## 3. 개선된 텍스트 스테가노그래피

사용자의 적용 목적, 커버 매개체의 형태, 숨겨진 메시지의 수용 능력과 견고성에 따라 효율적인 스테가노그래피가 결정되며, 견고성과 효율성 측면에서 개선되어야 한다. 커버 텍스트 매체의 빈 공간에 비밀 메시지를 삽입할 때 삽입하고자 하는 자료를 비트 패턴으로 변환한 후 암호화하며, 재배열키를 이용하여 은닉 시점에 따라 배치 순서가 다르게 텍스트 매체에 비밀 메시지를 숨기는 개선된 텍스트 스테가노그래피를 제안한다.

### 3.1 제안된 방법

커버 텍스트 매체에 비밀 메시지를 암호화하여 텍스트의 단어와 단어 사이의 공간에 비트 단위로 삽입된 스테고 매체를 만들어 허가된 수신자에게만 송신하는 기법을 적용한다. 이때 '0'과 '1'의 비트 정보를 빈 공간(null space)의 형태에 따라 다르게 삽입하는 방법[10]을 참고한다. <Fig. 2>는 커버 텍스트에 비트 단위로 변환된 비밀 메시지 정보가 삽입되는 과정과 스테고 텍스트에서 은닉된 비밀 메시지를 추출하는 과정을 보여준다.



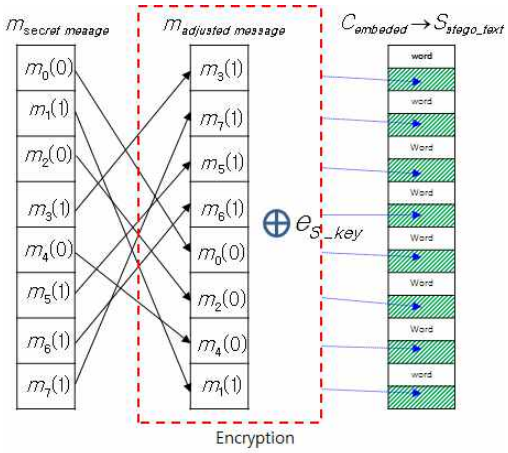

<Fig. 2> Mechanism of text steganography (implementation process)

이 논문에서는 <Fig. 3>과 같이 비트 단위로 변환된 비밀 메시지, 예를 들어 문자 'W', '01010111', 위치 변경키(37560241)를 참고하여 비트 정보의 위치를 바꾸고, 암호화 과정을 적용하여 커버 텍스트 매체의 빈 공간에 숨김으로써 보안성을 강화할 수 있음을 제시한다. 논문에서 제시되는 것은 <Fig. 2>에서 굵은 사각형 부분과 <Fig. 3>의 점선 부분에서 표시하였으며,

일반적인 적용 과정은 (3) 식으로 표현할 수 있다.

$$_{bedded\_Algorithm}\ Ch_{bit\_po}(m_i())\oplus e_{s\_key})\rightarrow S_{stego\_text}$$
$$D_{Extracted\_Algorithm}(S_{stego\_text})\rightarrow m_{\text{secret message}} \quad (3)$$

여기에서 ()은 비트 단위로 변환된 비밀 메시지의 $i$번째 비트 정보를 나타내며, $e_{s_key}$는 암호화키, $Ch_{bit_po}()$는 변경키에 따라 비트 정보 위치를 바꾸는 함수, ⊕는 XOR 함수를 의미한다. $_{stego_text}$는 스테고 텍스트 매체이며, 빗금친 부분은 빈 공간을 표시한다. $m_{sereet_message}$는 비밀 메시지를 의미한다.




<Fig. 3> Change the bit position in secret message

### 3.2 적용

논문에서 사용된 비밀 메시지의 크기는 9, 12, 33, 101, 112바이트로 사용하였으며. 비밀 메시지를 비트 패턴으로 변환한 후 스트림 암호를 이용하여 암호화하였다. 여기에서 사용한 영문 커버 텍스트의 크기는 5,048바이트이다. 알고리즘을 구현하는 과정은 J2SE을 이용하였다.

<Table 1>에서는 스테고 파일 생성 여부(①), 적용 파일 형태(②), 비밀 메시지의 가시성 여부(③), 암호화 형태(④), 암호키 사용 여부(⑤), 이중 암호화키(재배열키) 사용 여부(⑥)에 따라 텍스트 스테가노그래피에서 사용하는 도구를 보여준다. 일반적으로 SNOW보다는 wbStego가 성능과 삽입 용량 면에서 우수하지만

<Table 1> Text steganograohy techniques

| | ① | ② | ③ | ④ | ⑤ | ⑥ |
|---|---|---|---|---|---|---|
| GATS | Yes | txt | No | Playfair | Yes | - |
| WbStego | No | image, txt | No | Various | Y/N | - |
| SNOW | No | - | Yes | ICE | Y/N | No |
| Stego | No | - | No | - | Yes | - |
| proposed method | Yes | txt | No | LFSR | Yes | Yes |

[3,10] 논문에서 제안된 수정된 방법이 재배열키를 사용한다는 측면에서 이중으로 암호화하는 효과를 가져옴으로써 견고성을 높였다고 볼 수 있다.

비밀 메시지의 은닉 장소를 단어 사이의 공간으로 사용하기 때문에 스테고 텍스트에서 왜곡의 흔적을 찾을 수가 없다. 텍스트 스테가노그래피에서 왜곡 정도를 알아보기 위해 유사성을 측정하고, 효율성을 알아보기 위해 각각 사용하는 상관계수와 삽입 용량 값을 비밀 메시지의 크기에 따라 계산한 결과를 <Table 2>에서 보여준다.

<Table 2> Correlation value of cover and stego along to message size(byte)

| messag size | correlation | capacity | |
|---|---|---|---|
| | | proposed method | SNOW |
| 9 | 0.9888 | 1.43 | 4.19 |
| 12 | 0.9831 | 1.90 | 6.59 |
| 17 | 0.9784 | 2.69 | 9.64 |
| 33 | 0.9535 | 5.23 | - |
| 101 | 0.8553 | 16.01 | - |
| 112 | 0.8382 | 17.75 | - |

<Table 2>에서와 같이 비밀 메시지의 크기가 작을 경우 가시성과 삽입 용량에서 SNOW 기법이 우수하지만 은닉 메시지의 크기가 클 경우 커버 매체를 초과하여 삽입되는 단점이 있다. 본 논문에서 제시한 방법은 커버 매체에 비밀 메시지가 삽입된 전과 후를 비교할 때 비밀 메시지의 크기가 커짐에 따라 급격하게 상관계수가 감소하지만 삽입 용량을 1.9%이하로

할 때는 1.0에 근접하는 매우 높은 상관성을 보여주고 있다. 즉 통계학적으로 왜곡이 발생했다고 판단할 수 없으며, 통신의 허용범위 안에서 커버 텍스트 파일의 크기가 커질 경우 적절한 삽입 용량을 제공한다. 빈 공간에 삽입된 위치의 비밀 정보를 확인할 수 없기 때문에 가시성면에서 매우 효율적이다. 텍스트의 공간을 이용하여 비밀정보를 삽입할 때 삽입 용량은 언어별로 100회 이상 측정한 결과 영문 및 아랍어 19.2%, 일본어 0%, 독일어 14.4%, 한글인 경우 28.8% 내외의 삽입 용량을 갖는다는 것을 확인하였다.

## 4. 결 론

이미지나 오디오 스테가노그래피에 비해 효율성이 떨어짐에도 암호화가 적용된 텍스트 스테가노그래피는 다양한 언어로 구성된 문서와 통신 채널을 통해 가장 많이 송수신되는 매개체라는 측면에서 매우 중요한 은닉 도구이다. 단어와 단어 사이의 공간에 비트 단위 비밀 정보를 삽입하여 정보를 은닉하는 기법은 유사성 측면에서 매우 효율적인 스테가노그래피 기법임을 보였으며, 특히 비트 단위로 변환된 비밀 메시지 정보의 위치를 재지정하고 암호화함으로서 경고성면에서 효율적이다. 아라비아와 인도어의 특정 부분에 비밀 메시지를 삽입하는 텍스트 스테가노그래피와 같이[11] 한글 문서 속에 비밀 메시지를 숨기는 텍스트 스테가노그래피의 연구는 향후 좀 더 보완해야할 부분이다.

## Reference


[1] Seonsu Ji, "Locating and Searching Hidden Messages in Stego-Images", KIISC, Vol. 14, No. 3, pp. 37-43, 2009.

[2] L. Y. POR and B. Delina, "Information Hiding: A New Approach in Text Steganography", Applied Computer&Applied Computational Science(ACACOS 2008) Hangzhou, China, pp. 689-695, 2008.

[3] Indradip Banerjee, Souvik Bhattacharyya and Gautam Sanyal, "Study and Analysis of Text Steganography Tools", I. J. Computer Network and Information Security, Vol. 12, pp. 45-52, 2013.

[4] Indradip Banerjee, Souvik Bhattacharyya and Gautam Sanyal, "A Procedure of Text Steganography Using Indian Regional Language", I. J. Computer Network and Information Security, pp. 65-73, 2012.

[5] Krista Bennett, "Linguistic Steganography: Survey, Analysis, and Robustness Concerns for Hiding Information in Text", CERIAS Tech Report 2004-13.

[6] Swati Gupta and Deepti Gupta, "Text - Steganography: Review Study & Comparative Analysis", International Journal of Computer Science and Information Technologies, Vol. 2 (5), pp. 2060-2062, 2011.

[7] S. R. Govada, B. S. Kumar, M. Devarakonda and M. J. Stephen, "Text Steganography with Multi level Shielding", IJCSI International Journal of Computer Science, Vol. 9, Issue 4, No 3, pp. 401-405, 2012.

[8] Neha Rani and Jyoti Chaudhary, "Text Steganography Techniques: A Review", International Journal of Engineering Trends and Technology, Vo. 4 Issue 7, pp. 3013-3015, 2013.

[9] Prem Singh, Rajat Chaudhary and Ambika Agarwal, "A Novel Approach of Text Steganography based on null spaces", IOSR Journal of Computer Engineering, Vol. 3, Issue 4, pp. 11-17, 2012.

[10] Anandaprova Majumder and Suvamoy Changder, "A Novel Approach for Text Steganography: Generating Text Summary using Reflection Symmetry", International Conference on Computational Intelligence: Modeling Techniques and Applications, Procedia Technology 10, pp. 112-120, 2013.

[11] Ammar Odeh and Khaled Elleithy, "Steganography in Arabic Text using Zero with and Ashidha Letters", International Journal of Computer Science & Information Technology,

**지 선 수**(Seon-Su Ji)

- 정회원
- 1984년 충남대학교 계산통계학과 (학사)
- 1986년 중앙대학교 응용통계학과 (석사)
- 1993년 중앙대학교 응용통계학과(박사)
- 2006년 명지대학교 컴퓨터공학과(박사수료)
- (현)강릉원주대학교 정보기술공학과 교수
- 관심분야 : 혼잡제어, 정보보안(암호키, 정보은닉), 스테가노그래피

논 문 접 수 일 : 2014년 07월 09일
1 차 수 정 완 료 일 : 2014년 08월 29일
2 차 수 정 완 료 일 : 2014년 09월 25일
게 재 확 정 일 : 2014년 09월 30일