**Scientific Journals**
**Maritime University of Szczecin**

**Zeszyty Naukowe**
**Akademia Morska w Szczecinie**

**2014, 40(112) pp. 95–99**
**ISSN 1733-8670**

# The 2D-steganography method based on analysis of two adjacent LSB

**Larisa Dobryakova[1], Evgeny Ochin[2]**

[1] West Pomeranian University of Technology, Faculty of Computer Science and Information Technology
71-210 Szczecin, ul. Żołnierska 49, e-mail: ldobryakova@wi.zut.edu.pl
[2] Maritime University of Szczecin, Faculty of Navigation
70-500 Szczecin, ul. Wały Chrobrego 1–2, e-mail: e.ochin@am.szczecin.pl

**Key words:** steganography, container, embedding of dates, stego-message, Least Significant Bit, LSB

**Abstract**

In computing, the Least Significant Bit (LSB) is the bit position in a binary integer giving the units value, that is, determining whether the number is even or odd. We do not use this format. In conventional Intel bit ordering, the Most Significant Bit (MSB) is numbered 7 and the least significant bit (LSB) is numbered 0:

| | $2^7$ (MSB) | $2^6$ | $2^5$ | $2^4$ | $2^3$ | $2^2$ | $2^1$ | $2^0$ (LSB) |
|---|---|---|---|---|---|---|---|---|
| $D =$ | $d7$ | $d6$ | $d5$ | $d4$ | $d3$ | $d2$ | $d1$ | $d0$ |

For example, if on the record of number $D = 131 = 10000011$ we invert the MSB = 1 into MSB = 0, then obtain $D = 00000011 = 3$, but if we invert LSB = 1 into LSB = 0, then distortions are considerably fewer: $D = 10000010 = 130$. It is common to assign each bit a position number, ranging from zero to $N – 1$, where $N$ is the number of bits in the binary representation used. Normally, this is simply the exponent for the corresponding bit weight in base-2 (such as in $2^7, 2^6, \ldots, 2^1, 2^0$). In this article to reduce the stego-errors we use a pair of numbers $D_1$ and $D_2$ and a pair of corresponding LSB $d_{1,0}$ and $d_{2,0}$.

## Introduction

Due to the advent of computer networks ease and speed of access to information greatly increased threat of a security breach in the absence of data protection.

Security of information provided in digital form is important due with the intensive development and distribution of computer technology. Steganography methods are among the organizational, methodological and technical methods of information protection. The purpose of steganography is the hiding the fact of existence sensitive data during transmission, storage and processing. The process of information hiding is realized by different methods. Common to these methods is that the hidden message is embedded in the object, which not draws attention to themselves, which is then transported (sent to the addressee) open way [1, 2, 3].

Basic concepts of steganography were identified in 1996 at the first international conference Information Workshop on Information Hiding.

In computer steganography exists two main types of files: the stego-message (the hidden message) and the container – a file used for concealment in it of the message. The container initial state when it yet does not contain the hidden information is named as the container-original, and a finite state of the container when it already contains the stego--message, is named as the container-result. In the form of the container-original various media files, for example, maps, audio files, video files and others are often used. Stego-message embedding in the container it is attended by distortions of media data, however character of distortions should be minimum, that the listener (or the spectator) has not noted deterioration of the data containing in a media file. Sight and hearing of the person do not mark minor alterations in colours of the map or quality of a sound [4, 5]. In computing, the Least Significant Bit (LSB) is the bit position in a binary integer giving the units value, that is, determining whether the number is even or odd. We do not use

this format. In conventional Intel bit ordering, the Most Significant Bit (MSB[1]) is numbered 7 and the least significant bit (LSB) is numbered 0:

| | $2^7$ (MSB) | $2^6$ | $2^5$ | $2^4$ | $2^3$ | $2^2$ | $2^1$ | $2^0$ (LSB) |
|---|---|---|---|---|---|---|---|---|
| $D =$ | $d_7$ | $d_6$ | $d_5$ | $d_4$ | $d_3$ | $d_2$ | $d_1$ | $d_0$ |

For example, if on the record of number 131 = 1000 0011 we invert the MSB = $d_7$, then obtain 0000 0011 = 3, but if we invert LSB = $d_0$, then distortions are considerably fewer: 1000 0010 = 130. It is common to assign each bit a position number, ranging from zero to $N-1$, where $N$ is the number of bits in the binary representation used. Normally, this is simply the exponent for the corresponding bit weight in base-2 (such as in $2^{31}$, $2^{30}$, …, $2^1$, $2^0$).

In this article to reduce the stego-errors we use a pair of numbers $D_1$ and $D_2$ and a pair of corresponding LSB $d_{1,0}$ and $d_{2,0}$.

| | $MSB_1$ | | | | | | | $LSB_1$ |
|---|---|---|---|---|---|---|---|---|
| $D_1 =$ | $d_{1,7}$ | $d_{1,6}$ | $d_{1,5}$ | $d_{1,4}$ | $d_{1,3}$ | $d_{1,2}$ | $d_{1,1}$ | $d_{1,0}$ |
| | $MSB_2$ | | | | | | | $LSB_2$ |
| $D_2 =$ | $d_{2,7}$ | $d_{2,6}$ | $d_{2,5}$ | $d_{2,4}$ | $d_{2,3}$ | $d_{2,2}$ | $d_{2,1}$ | $d_{2,0}$ |

Since the pair of these numbers is extracted from the matrix of numbers $D_{i,j}$, we use the following notation:

| | $MSB_{i,j}$ | | | | | | | $LSB_{i,j}$ |
|---|---|---|---|---|---|---|---|---|
| $D_1 =$ | $d_{i,j,7}$ | $d_{i,j,6}$ | $d_{i,j,5}$ | $d_{i,j,4}$ | $d_{i,j,3}$ | $d_{i,j,2}$ | $d_{i,j,1}$ | $d_{i,j,0}$ |
| | $MSB_{i+1,j}$ | | | | | | | $LSB_{i+1,j}$ |
| $D_2 =$ | $d_{i+1,j,7}$ | $d_{i+1,j,6}$ | $d_{i+1,j,5}$ | $d_{i+1,j,4}$ | $d_{i+1,j,3}$ | $d_{i+1,j,2}$ | $d_{i+1,j,1}$ | $d_{i+1,j,0}$ |

## The algorithm of LSB replacement

One of the most well-known methods using steganography media files is a method of replacing the least significant bits (LSB) of the container to the suitable binary message [1, 4].

This method can be explained as follows: let there is a container original image in the form of a matrix $I \times J$ of $N$ digit integer binary numbers:

$$D_{i,j} = \sum_{n=0}^{N-1} d_{i,j,n} \cdot 2^n, \; i = 0,1,\ldots,I-1, \; j = 0,1,\ldots,J-1 \tag{1}$$

where: $d_{i,j,n} \in \{0,1\}$ – $n$ binary digit of $i, j$ sample of the container-original. The range of representable numbers is equal $0 \le D_{i,j} \le (2^N - 1)$. For example, if $N = 8$, the range is equal $0 \le D_{i,j} \le 255$.

Always it is possible to present the text stego-message in the form of bit string length $K$ of bits:

$$\{s_0 s_1 \cdots s_k \cdots s_{K-1}, s_k \in \{0,1\}\} \tag{2}$$

Define container-result as:

$$D^+_{i,j} = d^+_{i,j,0} + \sum_{n=1}^{N-1} d_{i,j,n} \cdot 2^n$$
$$i = 0,1,\ldots,I-1, \; j = 0,1,\ldots,J-1 \tag{3}$$

where: $d^+_{i,j,n} \in \{0,1\}$ – $n$ binary digit of $i, j$ sample of the container-result.

The algorithm of embedding stego-message is the replacement of the first to the least significant bits of the container $d_{i,j,0}$ to the message length $K$ bits. While the remaining $(I \cdot J - K)$ bits of the container-original remain unchanged:

$$\left\{\begin{array}{l} k = 0; \\ \text{for } i = 0,1,\ldots,I-1: \\ \quad \text{for } j = 0,1,\ldots,J-1: \\ \quad\quad \text{if } k < K, \text{ then } D^+_{i,j} = s_k + \sum_{n=1}^{N-1} d_{i,j,n} \cdot 2^n \\ \quad\quad\quad \text{else } D^+_{i,j} = D_{i,j}; \\ \quad\quad k = k + 1; \\ \quad \text{end}; \\ \text{end}; \\ \text{if } k < K, \text{ then Goto Error\_Short\_container} \end{array}\right\} \tag{4}$$

## Embedding a message based on the analysis of the two adjacent samples

In this article to reduce the stego-errors[2] reviewed by embedding messages in two-dimensional data based on the analysis and on the modification of data samples pairs LSB $d_{i,j,0}d_{i+1,j,0}$.

The method can be explained using of the table 1.

Algorithm of embedding can be explained as follows:

[1] IBM SNA Formats Bit Ordering is Opposite of Intel Convention/ http://support.microsoft.com/kb/130861

[2] Stego-errors is a distortion of the container.

Table 1. Embedding of the data in least significant bits

| Embedding $s_k = 0$ | | | | Embedding $s_k = 1$ | | | | Extraction $s_k$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $d_{i,j,0}$ | $d_{i+1,j,0}$ | $d^+_{i+1,j,0}$ | $k=$ | $d_{i,j,0}$ | $d_{i+1,j,0}$ | $d^+_{i+1,j,0}$ | $k=$ | $d^+_{i,j,0}$ | $d^+_{i+1,j,0}$ | $s_k=$ | $k=$ |
| 0 | 0 | $d_{i+1,0}$ | $k+1$ | 0 | 0 | $\bar{d}_{i+1,0}$ | $k$ | 0 | 0 | 0 | $k+1$ |
| 0 | 1 | $\bar{d}_{i+1,0}$ | $k+1$ | 0 | 1 | $d_{i+1,0}$ | $k$ | 0 | 1 | – | $k$ |
| 1 | 0 | $d_{i+1,0}$ | $k$ | 1 | 0 | $\bar{d}_{i+1,0}$ | $k+1$ | 1 | 0 | – | $k$ |
| 1 | 1 | $\bar{d}_{i+1,0}$ | $k$ | 1 | 1 | $d_{i+1,0}$ | $k+1$ | 1 | 1 | 1 | $k+1$ |

$$
\left\{\begin{array}{l}
k=0;\\
\text{for } i=0,1,\ldots,I-1:\\
\quad \text{for } j=0,1,\ldots,J-1:\\
\qquad \text{if } k<K, \text{ then } \left[\begin{array}{l}
\text{if } s_k d_{i,j,0} d_{i,j+1,0}=000\\
\qquad \text{then } \{D^+_{i,j}=D_{i,j};\ k=k+1\};\\
\text{if } s_k d_{i,j,0} d_{i,j+1,0}=001\\
\qquad \text{then } \{d_{i,j+1,0}=0;\ k=k+1\};\\
\text{if } s_k d_{i,j,0} d_{i,j+1,0}=011\\
\qquad \text{then } d_{i,j+1,0}=0;\\
\text{if } s_k d_{i,j,0} d_{i,j+1,0}=111\\
\qquad \text{then } \{D^+_{i,j}=D_{i,j};\ k=k+1\};\\
\text{if } s_k d_{i,j,0} d_{i,j+1,0}=110\\
\qquad \text{then } \{d_{i,j+1,0}=1;\ k=k+1\};\\
\text{if } s_k d_{i,j,0} d_{i,j+1,0}=100\\
\qquad \text{then } d_{i,j+1,0}=1;
\end{array}\right.\\
\qquad \text{else } D^+_{i,j}=D_{i,j};\\
\quad \text{else if } k=K\\
\qquad \text{then Goto End\_Effective\_embedding}\\
\text{end;end;}\\
\text{if } k<K, \text{ then Goto Error\_Short\_container}
\end{array}\right. \tag{5}
$$

Algorithm to extract the stego-messages from the container:

$$
\left\{\begin{array}{l}
k=0;\\
\text{for } i=0,1,\ldots,I-1:\\
\quad \text{for } j=0,1,\ldots,J-1:\\
\qquad \text{if } d^+_{i,j,0}=d^+_{i,j+1,0}, \text{ then } \{s_k=d^+_{i,j,0};\ k=k+1\};\\
\qquad \text{if } k=K, \text{ then Goto End\_of\_destego}\\
\quad \text{end;}\\
\text{end;}\\
\text{End\_of\_destego:}
\end{array}\right. \tag{6}
$$

Look at an example: let the test stego-message looks as follows: $s_k = \{1\ 0\ 0\ 1\}$, $k = 0,1,2,3$, and the appropriately:

$$
d_{i,j,0}=\begin{Bmatrix} 1 & 1 & 1 & 1\\ 0 & 0 & 1 & 1\\ 1 & 0 & 1 & 0 \end{Bmatrix},\ i=0,1,2\ \ j=0,1,2,3.
$$

As a result of applying the algorithm (5) was built container-result:

$$
d^+_{i,j,0}=\begin{Bmatrix} 1 & 1 & 0 & 0\\ 0 & 0 & 1 & 1\\ 1 & 0 & 1 & 0 \end{Bmatrix}.
$$

Algorithm of stego-message embedding into container can be shown in a table 2.

Table 1. Embedding and extraction of test-message into/from container

| | | | Embedding $s_k$ | | | | Extraction $s_k$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $k$ | $i$ | $j$ | $s_k$ | $d_{i,j,0}$ | $d_{i,j+1,0}$ | $d^+_{i,j+1,0}$ | $d^+_{i,j,0}$ | $d^+_{i,j+1,0}$ | $s_k$ |
| 0 | 0 | 0 | 1 | **1** | 1 | **1** | 1 | 1 | 1 |
| 1 | 0 | 1 | 0 | 1 | 1 | **0** | 1 | 0 | – |
| 1 | 0 | 2 | 0 | 0 | 1 | **0** | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 | **0** | 0 | **0** | 0 | 0 | 0 |
| 3 | 1 | 1 | 1 | 0 | 1 | **1** | 0 | 1 | – |
| 3 | 1 | 2 | 1 | 1 | 1 | **1** | 1 | 1 | 1 |

The figure 1 shows the test container-original

$$
d_{i,j,0}=\begin{Bmatrix} 1 & 1 & 1 & 1\\ 0 & 0 & 1 & 1\\ 1 & 0 & 1 & 0 \end{Bmatrix},
$$

the figure 2 shows the test stego-message $s_k = \{1\ 0\ 0\ 1\}$. Marked in gray are the elements whose value is 1, and in white colour – the elements of the value 1.

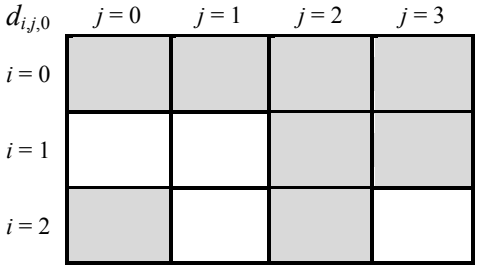


Fig. 1. The test container

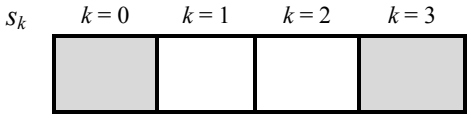


Fig. 2. The test stego-message

Figure 3 shows the results of embedding of the test stego-message in the test container-original using the algorithm (5).

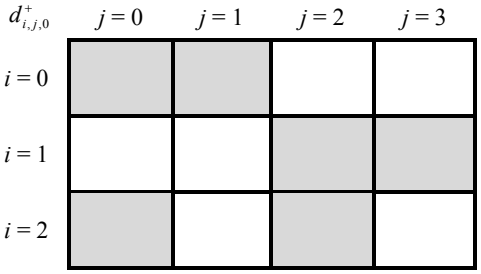


Fig. 3. Embedding of the message into container using the algorithm (5)

Figure 4 shows the results of embedding of the test stego-message in the test container-original using the algorithm LSB (4).

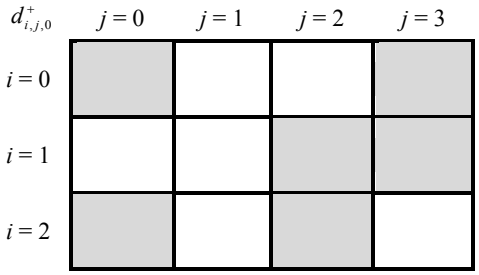


Fig. 4. Embedding of the message into container using the algorithm (4)

As can be seen the least significant bits of container-result after embedding messages based on the algorithm (4) precisely repeat the message (Fig. 4), what does not have the embedding messages on the basis of an algorithm developed (Fig. 3).

From these figures it can be concluded that the extraction of the messages from the container-result in the case of use the algorithm (4) is easier than in the case of use the algorithm (5), because anyone can allocate LSB of container-result and to combine the message from the received bit line. In the case of the use of developed algorithm (5) cannot be directly read a message from each least significant bits, because not all bits of the embedding messages were distorted. This is the main advantage of the proposed method. An additional advantage of the method is its newness: each new developed method enables to embed stego-messages in the media data in a way unknown to strangers, which increases its resistance to reading through these people.

Embedding information's resistance is a measure of assurance the correct read hidden information after performing sequence defined transformation of the signal container. Embedding information's is resistant only when all the information will be read correctly.

## Experimental evaluation of the container distortions

Quantitative estimation of firmness of a protection system to exterior distortions represents enough challenge which usually in practice is realized by methods of the systems analysis, mathematical modelling or experimental research [6, 7].

As a rule, professionally developed stego--system ensures three-level model of the privacy, solving two cores of tasks:

- concealment of the fact of presence of the protected information (the first security clearance);
- unauthorized access blocking to the information, realized by election of an appropriate method of concealment of the information (the second security clearance).

Also can exist and the third security clearance – preliminary cryptography enciphering.

To compare the quality of steganographic tools used various measures, giving a quantitative assessment. This article uses two criteria:

- Mean Square Error (MSE);
- Normalized Cross-Correlation (NC).

In the capacity of estimations of distortions of the container-original the root-mean-square error (MSE) is often used, although it does not always correlate with subjective perceptions of media data.

$$\mathrm{MSE} = \frac{1}{I \cdot J} \sum_{i=0}^{I-1} \sum_{j=0}^{J-1} \left( D_{i,j} - D^{+}_{i,j} \right)^2 \qquad (7)$$

In the implementation the algorithm (5) using application Mat Lab, found that the distortion of container-original value is MSE ≈ 0.5, that is equals the value of distortion of the container with the embedding stego-message by using the replace algorithm of least significant bits (4).

It allows conclude that the developed method is not worse than the existing method from the viewpoint distortion of the container.

For the analysis the resistance before reading the message by foreign persons also can use Normalized Cross-Correlation as a measure of proximity to the embedding message $s_k$ from container-result $d^+_{i,j,0}$:

$$\mathrm{NC} = \frac{\sum_{k=0}^{K-1}\left(s_k \cdot d^+_{k,0}\right)}{\sum_{k=0}^{K-1}\left(s_k\right)^2} \tag{8}$$

where: $d^+_{k,0}$ – the bit value of the container-result $d^+_{i,j,0}$, in a message that has been built.

Since in the method (4) embeds the message in each subsequent least significant bit of the container, then we can write $s_k = d^+_{k,0}$, $s_k \in \{0,1\}$, that is:

$$\mathrm{NC}_{(4)} = \frac{\sum_{k=0}^{K-1}\left(s_k \cdot d^+_{k,0}\right)}{\sum_{k=0}^{K-1}\left(s_k\right)^2} = \frac{\sum_{k=0}^{K-1}\left(s_k\right)^2}{\sum_{k=0}^{K-1}\left(s_k\right)^2} = 1 \tag{9}$$

$$\mathrm{NC}_{(5)} = \frac{\sum_{k=0}^{K-1}\left(s_k \cdot d^+_{k,0}\right)}{\sum_{k=0}^{K-1}\left(s_k\right)^2} = \frac{\sum_{k=0}^{K-1}\left(s_k \cdot d^+_{k,0}\right)}{\sum_{k=0}^{K-1} s_k} \tag{10}$$

When calculating the value of the Normalized Cross-Correlation algorithm (5) using Mat Lab application, we found that this value is $\mathrm{NC} = 0.5 \pm 0.1$ which demonstrate that the proposed method is more resistant to trying to read the messages by unauthorized persons.

A more detailed study of resistance to reading messages from a container-result relates to the domain stego-analysis and in this article is not considered.

## Conclusions

The article is described a new method of embedding of the text-message, which based on analysis of least significant bits and their partial distortion. As a result of the experiments can see that the Mean Square Error this method is not greater than in the method of replacement of the first $K$ samples container-original. However, the developed method is more robust against reading messages through strangers than the existing method of embedding of stego-message in media data (4), as indicated by the Normalized Cross-Correlation coefficients, which are for developed method (5) two times lower than in the method (4).

Cause of this are the advantages of the method listed in the article that not all bits of the embedding messages were distorted and it is impossible to directly read the message from the LSB container-result.

It should be stressed that the complexity of the proposed method is comparable to the complexity of the classical method with one LSB. A numerical estimate of the complexity (number of arithmetic and logical operations with fixed-point, data transfer operations, the area of the matching circuit) makes sense for the hardware implementation of steganography algorithms and analysis of performance (productivity) in on-line mode. All these issues (and a property of the parallelizable) are outside the scope of our study.

The novelty of the proposed algorithm is that we use to encode both a pair of numbers of the container, which certainly increases stego stability**,** because such methods of steganography are not studied**.**

## References


1. OCHIN E., DOBRYAKOVA L., PIETRZYKOWSKI Z., BORKOWSKI P.: The application of cryptography and steganography in the integration of seaport security subsystems. Scientific Journals Maritime University of Szczecin 26(98), 2011, 80–87.
2. BENDER W., GRUHL B., MORIMOTO N., LU A.: Techniques for data hiding. IBM Systems Journal 35, 3, 1996.
3. OCHIN E., DOBRYAKOVA L.: Metoda steganograficzna na podstawie analizy LSB dwóch sąsiednich próbek danych jednowymiarowych. Metody Informatyki Stosowanej 4, 2011.
4. Gribunin W., Okov I., Turincev I.: Cifrovaâ steganografiâ. SOLON-Press, Moskva, 2002.
5. PETITCOLAS F., ANDERSON R., KUHN M.: Information Hiding – A Survey. Proceedings IEEE, Special Issue on Identification and Protection of Multimedia Information, Vol. 87, No. 7, 1999. Pfitzmann B., Information Hiding Terminology. In: Information Hiding. Springer Lecture Notes in Computer Science. V. 1174, 1996.
6. KONAHOVIČ G., PUZYRENKO A.: Komp'ûternaâ stenografiâ. Teoriâ i praktika. MK-Press, Kuev, 2006.
7. OSBORNE C., VAN SCHYNDEL R., TIRKEL A.: A Digital Watermark. IEEE International Conference on Image Processing, 1994.