Journal of Engineering and Applied Sciences 12 (Special Issue 8): 8251-8255, 2017
ISSN: 1816-949X

An Approach for Hiding Steganography Data Within Web Applications

Stanimir Zhelezov, Boryna Uzunova-Dimitrova and Hristo Paraskevov
Faculty of Mathematics and Computer Science, Shumen University, Universitetska Street 115, 9700 Shumen, Bulgaria

Abstract: The objective of the present study is the use of web programming tools for the realization of a hidden channel for data transmission. Steganography algorithm based on the embedding of text information within the white spaces of the HTML code is proposed. The addition of extra white space will not affect the visualization of the web-document but will enhance the effectiveness of embedding. This will increase the amount of the transmitted hidden information.

Key words: Steganography, stego object, stego methods, data hiding, tools, objective

INTRODUCTION

The prevalence of the internet and the possibilities for utilizing the services provided by any kind of desktop and mobile devices at any point of the world leads to a higher level of threat to the information transmitted within the global network. The number of incidents related to unsanctioned access to confidential data transmitted among internet users is increasing. This in turn requires focus on the improvement of the existing and development of new, more efficient methods for protection (Boyanov, 2014).

Aspects of information protection: The existing methods for protection of information that is exchanged between two subjects can be divided into two main groups:

- Hiding the content of the transmitted information
- Hiding the fact of the information transmission

The cryptography methods occupy a key role in the first type. Both the increase of the computing power and the easy access to huge computing resources (parallel computing systems and cloud systems) give a strong impetus to the development of cryptographic algorithms. This is why in recent years for the creation of cryptographic methods and algorithms more mathematical tools from the chaos theory, abstract algebra, number theory, combinatorics and others are being used. The research related to creating pseudo random generators for data encryption is increasingly widespread (Stoyanov and Kordov, 2014; Malchev and Ibryam, 2015). The main task of all methods of this group is to make sure that the resources needed to break the information protection repeatedly exceed its value.

The second group of methods for information protection are designed to hide the presence of communication between the subjects and are the subject of study of the steganology and steganography in particular (Stanev *et al*., 2013). In recent years, there has been a sharp increase in the number of publications in prestigious scientific publications related to the application of the computer steganography methods. The prevalence and free access to steganography programs within which a variety of methods and algorithms are embedded require constant research and development of new protection methods.

It is known that almost all file formats without compression which possess a high degree of information surplus may be used to hide steganography messages (Stanev *et al*., 2013). Graphic and audio files have such a surplus and this is why they are the most commonly used "carrying" file formats which makes them a primary objective for the steganalysts. Even the very fact of suspicion about the presence of covert communication is already a security breach of the stego channel for data exchange. For this purpose it is necessary to look for such carriers of hidden information that do not arouse suspicion (Stanev *et al*., 2013).

Text steganography: Steganography can be classified based on the carrier media used to hide data. The four types of carriers that are typically used in steganography are audio, video, text and images. Carriers can be a text, a text from web pages or a binary file.

Figure 1 shows the classification of the types of text steganography and its methods. Text steganography can be categorized into two groups: change of the text format technical and change in the meaning of the text-linguistic (Ilchev and Ilcheva, 2012).

Corresponding Author: Stanimir Zhelezov, Faculty of Mathematics and Computer Science, Shumen University, Universitetska Street 115, 9700 Shumen, Bulgaria

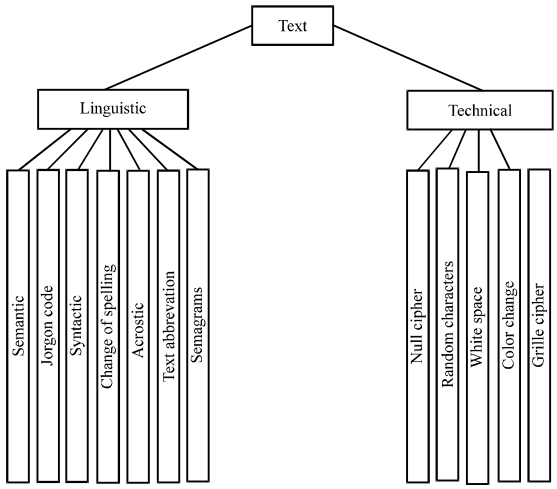

Fig.1: Classification of the types of text steganography

The object of this study is the second strand of text steganography-the technical text steganography. There could be assigned the following main methods for steganographic embedding of hidden information: the null cipher method, the random characters method, the colour change method, the grille cipher and the white space method. What they have in common is the manipulation of the technical characteristics of the containers and not the linguistic transformations of the information. A major drawback of most of these methods is the serious restriction on the amount of information that can be embedded into the containers. From studies carried out one can conclude that only the white space method indicates the potential for embedding greater amounts of data (Chang and Clark, 2010; Natthawut, 2009).

MATERIALS AND METHODS

White space method: The objective of the present study is the use of the tools for web programming for the realization of a hidden channel for data transmission.

In contrast to the multimedia containers that can be manipulated at the binary level, the files created by means of web programming do not allow this type of manipulation. Regardless of the tools used to create a web page in the end it all comes down to generating an HTML code that browsers could interpret. Standard steganography methods for embedding information, like LSB are inapplicable for such containers. For this purpose it is necessary to look for other solutions.

Since, the HTML code is insensitive to "white spaces" using them to hide information allows the realization of a hidden channel for data transmission. For this purpose, the methods of text steganography can be used which is one of the basic branches of steganography.

For the realization of the steganography method based on the embedding of text information within the white spaces of the HTML code the following algorithm has been used:

- The uncovered text is converted into a binary format
- All the characters of the HTML code are being inspected to find a "white space" (spaces or tabs)
- The current element of the uncovered text is embedded as the space character is used for the value of 0 and the tab character-for 1. After embedding it is moved on to the next item from the information for embedding
- Items 2 and 3 are repeated until the embedding of the entire message or till the end of the containter

One of the key indicators in assessing the effectiveness of steganographic methods and algorithms is the effectiveness of embedding $E_{eâ}$:

$$E_e = V_{max}/V_k$$

Where:
V_{max} = The maximum amount of hidden information
V_k = The volume of the container

Modification of the method: To improve the efficiency of embedding it is necessary to increase the number of positions that can embed hidden information. The proposed method is again based on the fact that the HTML code is insensitive to "white spaces". With it during embedding an additional embedding of a character (space or tab) is used before the character"<"of the HTML tag and the same is used after the character">". This embedding does not affect the visualization via. the browser and allows to significantly increase the capacity of the container.

For the realization of the modified steganography method based on embedding text information within the white spaces of the HTML code the following algorithm is being used:

- The uncovered text is converted into a binary format
- All the characters of the HTML code are being inspected to find a "white space" (spaces or tabs) or the characters "<" and">"
- The current element of the uncovered text is embedded as the space character is used for the value of 0 and the tab character for 1. After embedding it is moved on to the next item from the information for embedding

- Items 2 and 3 are repeated until the embedding of the entire message or till the end of the containter

RESULTS AND DISCUSSION

For carrying out of a comparative analysis of the white space method and the proposed modification there were conducted over 1,000 experiments and for this purpose there was established a base of 500 containers. Half of them were created manually and the other half were downloaded from the internet.

During the experiments carried out it was found that with the white space method this coefficient E_e depends on the container (Table 1) and varies within the range of 0.007252-0.509713. From here it can be concluded that in order to hide a large amount of information it is necessary for this method to be modified.

For each container from the base there was defined the total volume and the quantity of the positions "available for embedding". On their base was defined the coefficient of efficiency of embedding the proposed modified method (Table 2).

Table 1:Efficiency of embedding of the "white space" method

Name	No. of characters (V_{max})	No. of the free positions (V_k)	Coefficient of the efficiency of the "white spaces"
html/24.html	1361	165	0.121234386
html/27.html	1381	68	0.049239681
../stego_hide/1.html	2696	287	0.106454006
html/1.html	2696	288	0.106824926
html/32.html	4262	1036	0.243078367
html/34.html	5955	709	0.119059614
html/14.html	15325	2422	0.158042414
html/42.html	15899	2928	0.184162526
html/66.html	16438	1305	0.079389220
html/73.html	16979	1387	0.081689145
html/51.html	17246	1738	0.100776992
html/37.html	20774	2689	0.129440647
html/80.html	22583	710	0.031439578
html/35.html	22629	2247	0.099297362
html/38.html	24773	1776	0.071690954
html/16.html	25706	5830	0.226795301
html/17.html	26331	4847	0.184079602
html/19.html	26715	3559	0.133221037
html/75.html	27014	5607	0.207559043
../stego_hide/12.html	27283	1635	0.059927427
html/12.html	27283	1636	0.059964080
../stego_hide/16.html	27948	2650	0.094818949
html/39.html	28091	3861	0.137446157
../stego_hide/3.html	28703	2539	0.088457653
html/3.html	28703	2540	0.088492492
../stego_hide/17.html	29469	2718	0.092232516
html/79.html	31049	4618	0.148732648
html/72.html	34448	4377	0.127061078
../stego_hide/4.html	35179	9256	0.263111515
html/4.html	35179	9257	0.263139941
html/36.html	36256	3666	0.101114298
../stego_hide/8.html	42816	4939	0.115354073
html/8.html	42816	4940	0.115377429

Table 1 and 2 represent the obtained coefficients of effectiveness of embedding for one and the same files from the database, so that, they can be compared and analyzed. The total Volume of the file V_{max} and the number of items for embedding V_k for each file are represented as well. The coefficients are being calculated on that base.

The comparison of the coefficients obtained by the two methods and the reporting of their difference allows to analyze the positive aspects of the proposed modification (Table 3).

From the results obtained from the experiments it is clear that the coefficient of the proposed modified white space method is increased by 0.000546-0.086719 (Fig. 2), i.e. in some cases the increase is more than 198% (Fig. 3).

From the research findings it is clear that the proposed method for hiding steganography information is fully applicable to the majority of scripting languages for web programming (CSS, JavaScript, etc.).

Table 2: Efficiency of embedding of the modified method

Name	No. of characters (V_{max})	No. of the free positions (V_k)	Coefficient of the efficiency of the modified method
html/24.html	1361	209	0.153564
html/27.html	1381	134	0.097031
../stego_hide/1.html	2696	501	0.185831
html/1.html	2696	502	0.186202
html/32.html	4262	1155	0.271000
html/34.html	5955	955	0.160369
html/14.html	15325	2900	0.189233
html/42.html	15899	3476	0.218630
html/66.html	16438	2184	0.132863
html/73.html	16979	2447	0.144119
html/51.html	17246	2468	0.143106
html/37.html	20774	3839	0.184798
html/80.html	22583	810	0.035868
html/35.html	22629	2391	0.105661
html/38.html	24773	2851	0.115085
html/16.html	25706	6754	0.262740
html/17.html	26331	5843	0.221906
html/19.html	26715	4465	0.167135
html/75.html	27014	7054	0.261124
../stego_hide/12.html	27283	2549	0.093428
html/12.html	27283	2550	0.093465
../stego_hide/16.html	27948	3846	0.137613
html/39.html	28091	4944	0.175999
../stego_hide/3.html	28703	4157	0.144828
html/3.html	28703	4158	0.144863
../stego_hide/17.html	29469	3892	0.132071
html/79.html	31049	6420	0.206770
html/72.html	34448	6435	0.186803
../stego_hide/4.html	35179	10258	0.291594
html/4.html	35179	10259	0.291623
html/36.html	36256	5850	0.161353
../stego_hide/8.html	42816	7383	0.172436
html/8.html	42816	7384	0.172459

Table 3: Comparison of the coefficients for both methods

Name	Coefficient for the white space method	Coefficient for the modified method	Difference	Increase (%)
html/24.html	0.121234	0.153564	0.032329	26.6666667
html/27.html	0.049240	0.097031	0.047791	97.0588235
../stego_hide/1.html	0.106454	0.185831	0.079377	74.5644599
html/1.html	0.106825	0.186202	0.079377	74.3055556
html/32.html	0.243078	0.271000	0.027921	11.4864865
html/34.html	0.119060	0.160369	0.041310	34.6967560
html/14.html	0.158042	0.189233	0.031191	19.7357556
html/42.html	0.184163	0.218630	0.034468	18.7158470
html/66.html	0.079389	0.132863	0.053474	67.3563218
html/73.html	0.081689	0.144119	0.062430	76.4239366
html/51.html	0.100777	0.143106	0.042329	42.0023015
html/37.html	0.129441	0.184798	0.055358	42.7668278
html/80.html	0.031440	0.035868	0.004428	14.0845070
html/35.html	0.099297	0.105661	0.006364	6.40854473
html/38.html	0.071691	0.115085	0.043394	60.5292793
html/16.html	0.226795	0.262740	0.035945	15.8490566
html/17.html	0.184080	0.221906	0.037826	20.5487931
html/19.html	0.133221	0.167135	0.033914	25.4565889
html/75.html	0.207559	0.261124	0.053565	25.8070269
../stego_hide/12.html	0.059927	0.093428	0.033501	55.9021407
html/12.html	0.059964	0.093465	0.033501	55.8679707
../stego_hide/16.html	0.094819	0.137613	0.042794	45.1320755
html/39.html	0.137446	0.175999	0.038553	28.0497280
../stego_hide/3.html	0.088458	0.144828	0.056370	63.7258763
html/3.html	0.088492	0.144863	0.056370	63.7007874
../stego_hide/17.html	0.092233	0.132071	0.039838	43.1935247
html/79.html	0.148733	0.206770	0.058037	39.0212213
html/72.html	0.127061	0.186803	0.059742	47.0185058
../stego_hide/4.html	0.263112	0.291594	0.028483	10.8254105
html/4.html	0.263140	0.291623	0.028483	10.8242411
html/36.html	0.101114	0.161353	0.060238	59.5744681
../stego_hide/8.html	0.115354	0.172436	0.057081	49.4837012
html/8.html	0.115377	0.172459	0.057081	49.4736842

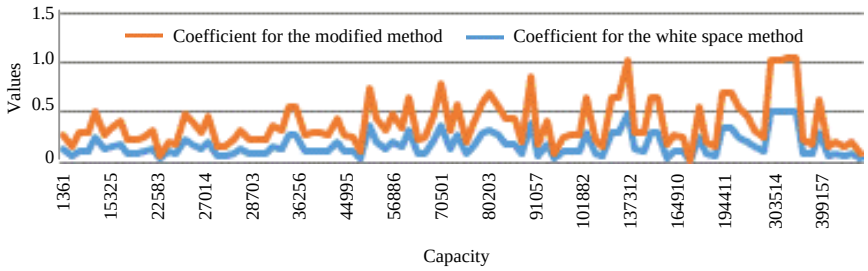

Fig. 2: Coefficients of embedding

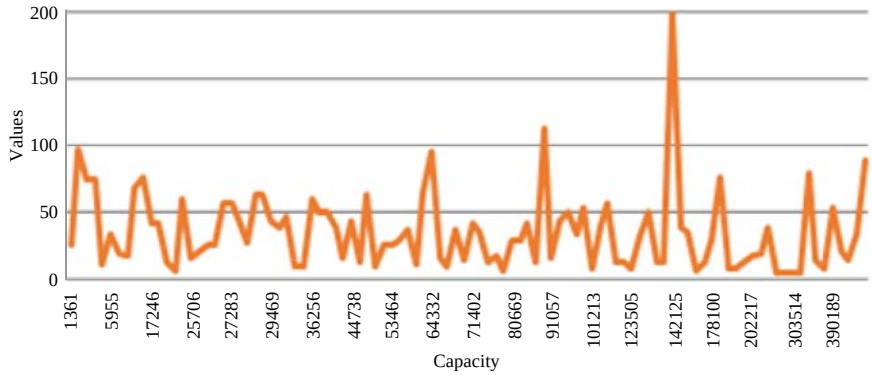

Fig. 3: Capacity increase (%)

CONCLUSION

Nowadays because of the great threat to security, data protection is mandatory. Steganography provides information security by hiding it in a carrier. This study includes the classification of steganography techniques that have already been introduced to hide information in web documents. Hiding data in a web document is less suspicious than other carriers because HTML web pages are now a routine part of everyone's life and HTML pages contain a significant number of tags, attributes and other elements where data could be hidden.

In terms of steganalysis the standard steganalytic techniques for multimedia objects (Meghanathan and Nayak, 2010) and any modifications of theirs (Zhelezov, 2016) are not applicable.

The proposed modified method can be extended through the use of pseudo-random sequence generators, based on which dispersed message embedding is being carried out (Kardov, 2015). Thus, enhancing the protection of the information transmitted and possible steganalytic attacks are being impeded.

ACKNOWLEDGEMENTS

This research was supported in part by Project RD-08-119/2016 "Steganography in mobile devices and 3-dimensional modeling". The Project is realized by the financial support of the Konstantin Preslavski University of Shumen, Bulgaria.

REFERENCES

Boyanov, P., 2014. Using HTTP filter to analyze and monitor the vulnerability and security states in determined computer network. J. Sci. Educ. Innovation, 2: 45-51.

Chang, C.Y. and S. Clark, 2010. Linguistic steganography using automatically generated paraphrases. Proceedings of the Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, June 2-4, 2010, Association for Computational Linguistics, Los Angeles, California, USA., pp: 591-599.

Ilchev, S. and Z. Ilcheva, 2012. Modular data hiding as an alternative of classic data hiding for web-based applications. Inf. Technol. Control, 1: 9-15.

Kordov, K., 2015. Signature attractor based pseudorandom generation algorithm. Adv. Stud. Theor. Phys., 9: 287-293.

Malchev, D. and I. Ibryam, 2015. Construction of pseudorandom binary sequences using chaotic maps. Appl. Math. Sci., 9: 3847-3853.

Meghanathan, N. and L. Nayak, 2010. Steganalysis algorithms for detecting the hidden information in image audio and video cover media. Int. J. Netw. Secur. Appl., 2: 43-55.

Natthawut, S., 2009. Steganography Via Running Short Text Messages. Springer, Berlin, Germany.

Stanev, S., H. Hristov and D. Dimanova, 2013. Aproaches for stego defense of sensitive information from inside leakage. Assoc. Sci. Appl. Res., 1: 126-132.

Stoyanov, B. and K. Kordov, 2014. Pseudorandom Bit Generator with Parallel Implementation. In: Large Scale Scientific Computing, Ivan, L., M. Svetozar and W. Jerzy (Eds.). Springer, Berlin, Germany, ISBN: 978-3-662-43879-4, pp: 557-564.

Zhelezov, S., 2016. Modified algorithm for steganalysis. Math. Software Eng., 1: 31-36.