DOI: 10.24143/2072-9502-2021-4-76-81
УДК 81'33

POTENTIAL APPLICATIONS OF LINGUISTIC STEGANOGRAPHIC TRIGGER-CONTAINERS FOR TEXT WATERMARKING PURPOSES[1]

R. K. Potapova, A. V. Dzhunkovskiy

Moscow State Linguistic University,
Moscow, Russian Federation

Abstract. The paper gives analysis of feasibility of using linguistic steganographic trigger-containers as means of linguistic-based text watermarking. The proposed approah is based on the previous experimental research in the context of Russian native speaker text juncture perception. It was posited that specific minimal text modifications addressed in the paper may be used as means of text watermarking with the aim of tracking the leak of information for the purposes of taking legal actions, enforcing non-disclosure agreements, and testing for internal vulnerabilities. There is analyzed the viability of altering paragraph juncture points in Russian texts and, through the use of the corresponding trigger-containers, usage of this alteration as a means of linguistic watermarking.

Key words: paragraph, leak source, experimental steganography, joining words and paragraphs, steganography, trigger-containers, text watermarks.

For citation: Potapova R. K., Dzhunkovskiy A. V. Potential applications of linguistic steganographic trigger-containers for text watermarking purposes. *Vestnik of Astrakhan State Technical University. Series: Management, Computer Science and Informatics*. 2021;4:76-81. (In Russ.) DOI: 10.24143/2072-9502-2021-4-76-81.

Introduction

There are two main approaches to protecting sensitive information: cryptography and steganography. The basis of cryptography is altering a message in such a way as to encrypt it using various ciphers. In this way the message is altered and protected, but the fact is that an attempt to conceal or protect information is evident. Decrypting these ciphers means restoring the encrypted message into its initial state [1].

On the opposite end there exists steganography: the practice of concealing the fact of transferring confidential information itself. While steganography may on some level use cryptographic ciphers, the core of steganographic information protection is the subterfuge-based approach aimed at hiding the hidden message within plain text in such a manner that the altered text becomes virtually indistinguishable from the original unaltered text [2].

While the practice of steganography has existed throughout human history, there is little research that determines the efficacy of different methods thereof. In our work we set out to create a strong, data-driven, scientific groundwork based on an experimental approach that involves native speakers. While at this stage we analyze Russian written texts, the methodology may be expanded for other languages.

Our previous research has allowed us to gain and interpret data pertaining to how Russian native speakers choose to re-separate into words and paragraphs a sensical text that has been altered by deleting spaces between words, removing capital letters, punctuation symbols, and paragraph breaks.

Initially this data has been gathered for the purposes of trigger-container based steganographic inquiry. Trigger-containers are minimal, distinct text alterations that refer to a previously agreed-upon message that is to be received upon discovering the corresponding alteration [3].

The results of our perception experiment have indicated that altering the location of paragraph junctures (at least for Russian texts) appears to be a suitable variable for these purposes. We are now ready to propose a new application for these trigger-containers in text watermarking.

Background

Our research approach is based on using experimental perceptual methods and statistically analyzing the resulting data. As mentioned in the introduction, the stimulus was a sensical Russian text that has been transformed into a continuous string of letters. The initial text was a test text for the purposes of text companding [4]. The participants (n = 102) were tasked with restoring the text by separating it into paragraphs, phrases, syntagmas and words. For the purposes of this paper, we shall ignore

[1] The research was carried out within the state assignment of Moscow State Linguistic University (theme No. FSFU-2020-0020).

the phrase and syntagma juncture data and focus on paragraph and word junctures. Effectively, the resulting data is an approximation and the range of how Russian native speakers view the range of possible juncture points for words and paragraphs in Russian texts. In other words, the data allows us to gleam insight into how far we may manipulate a linguistic variable in the form of word and paragraph juncture without raising suspicion.

The data has been processed and we have calculated the probability of a certain juncture being restored (if it corresponded with one in the initial text presentation) or inserted (if it was not present in the initial text) by the participants of our experiment. This probability shows, in essence, how Russian native speakers perceive a certain juncture to be, which is valuable data for the purpose of discovering optimal trigger-containers for specific applications.

Results, Proposed Approach and Discussion

In Table 1 we present the statistical data for paragraph and word junctures.

Table 1

Comparison of paradigm statistical data

Parameters	Paragraphs	Words	Word Anomalies
Mean	17,10	**94,07**	76,51
Standard Error	4,36	1,45	5,09
Median	4,90	100,00	97,06
Mode	3,92	100,00	98,04
Standard Deviation	22,67	21,28	37,41
Sample Variance	**513,82**	**452,93**	**1 399,82**
Kurtosis	1,71	13,23	–0,02
Skewness	1,65	–3,84	–1,36
Range	78,43	98,04	97,06
Minimum	0,98	1,96	1,96
Maximum	79,41	100,00	99,02
Sum	461,69	20 131,37	4 131,37
Count	**27,00**	**214,00**	**54,00**
Confidence Level (95,0%)	8,97	2,87	10,21

The data for “Word Anomalies” represents all data for word junctures with sub-100% restoration/insertion rate. In other words, if all participants unanimously agreed that a juncture between words should exist at a certain point, that juncture would not be counted in the “Word Anomalies” as a unanimously agreed-upon juncture does not leave any ambivalence necessary for trigger-containers to function (Table 1).

We highlight sample variance, count and mean parameters as noteworthy. We have discovered a trend according to which the rise of sample variance inversely corresponds to the size of an element into which the string of characters is separated while restoring a text to its initial form. While the data shows that in “Words”, sample variance is lower than in “paragraphs”, the initial assessment is proven to be likely when we discard the junctures in the “Words” paradigm universally inserted by our participants (Fig. 1).

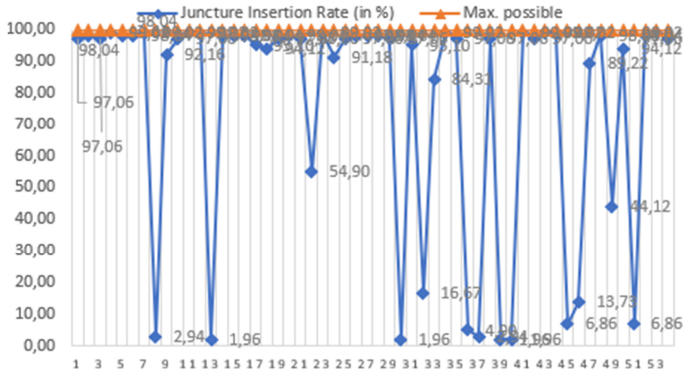

Fig. 1. Word anomalies paradigm juncture insertion rate analysis

The x-axis lists the word juncture by order of appearance in the text. The y-axis shows the insertion rate. As is evident, the mean for juncture insertion in the initial “Words” paradigm was higher than 94, while the altered “Word Anomalies” selection’s mean drops to 76, presenting us with enough ambivalence to consider certain unique cases of separating texts into words as a viable steganographic trigger-container.

Paragraph junctures, however, may be even more promising. The initial stimulus text contained only 5 paragraphs, yet the count parameter for “Paragraphs” clearly shows, that Russian native speakers determined 27 possible junctures for separating the text into paragraphs (Fig. 2).

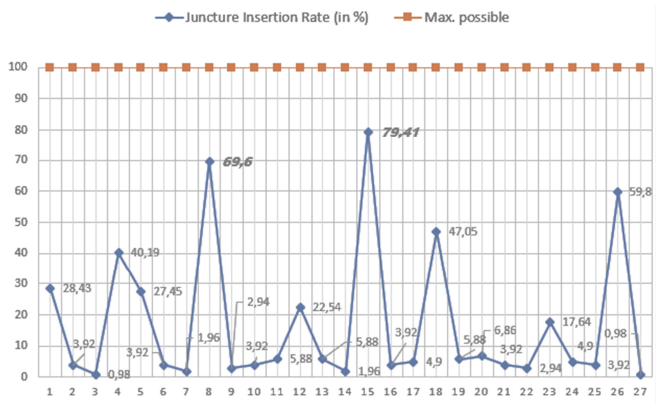

Fig. 2. Paragraph paradigm juncture insertion rate analysis

The x-axis lists the paragraph junctures by order of appearance in the text. The y-axis shows the insertion rate. As we can see, no paragraph juncture had an insertion rate higher that 80%. Some lower-scoring cases might be suboptimal for the purposes of steganography, but those in the 20-80% range demonstrate by the virtue of their existence the inherent variability of the range of possibility when separating Russian texts into paragraphs. Therefore we posit that manipulating paragraph junctures is a promising method for steganographic information encryption in the context of trigger-containers, as this method can be used by implanting a single alteration into a text.

To further increase the ambivalence of a juncture, we superimpose one paradigm over the other. For this method we only consider the junctures that appear in both paradigms for overlaying purposes. One item of not we have to adress immediately is juncture 21, where we can observe an anomaly. More native speakers believe that this juncture is that of different paragraph than that of different words. This is possible in the following way. One cannot start a new paragraph in the middle of a word in Russian. It may be different in some other languages, but we are currently unaware of such cases. What is notable here that the word with this anomalous juncture in the middle is a complex word consisting of two sensical roots. This allows one to realistically separate a word into paragraphs seemingly in the middle of a word (Fig. 3).

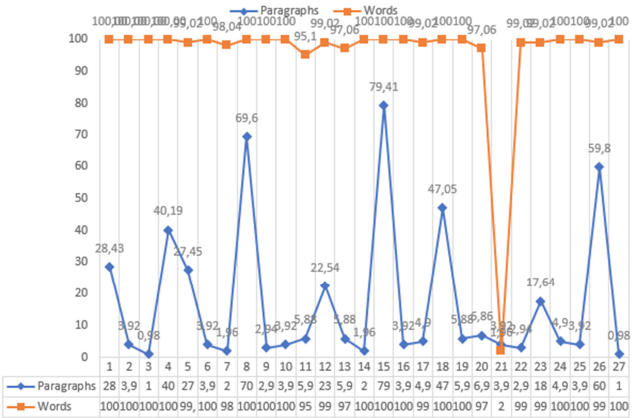

Fig. 3. Comaprison of paragraph and word junctures

Having discussed the data and the findings, we may finally proceed to how they relate to text watermarking. The goal of watermarking is to insert an alteration into a text or image before sending or publishing it. For most applications, this is done to protect copyright. In broad terms, the performance of a watermarking can be described by four factors, which are imperceptibility, robustness, capacity and security (see, e.g. [5]). Imperceptibility is the visual similarity of the watermarked text to the non-watermarked original. Robustness is how resistant the watermark is to attacks. Capacity is how much data the watermark potentially contains. Finally, security relates to watermark resistance to manipulation.

One of the approaches in watermarking is linguistic-based, where natural language methods are used to embed watermarks into a text. This method is useful for watermarking texts. Usually it is prudent to separate syntactic and semantic-based approaches in linguistic watermarking. The former deal with altering the syntax of a text, whereas the latter mean altering semantic elements of a text. Ultimately, both approaches aim to alter the linguistic features of a text without altering its meaning.

We propose using paragraph juncture alteration as a method of linguistic watermarking. Our data shows that the location of paragraph junctures is easily shifted and is connected to a high degree of detection ambivalence. There is no consensus among Russian native speakers on the topic of which locations within the text must be used when separating a text into distinct paragraphs. Therefore, altering this variable satisfies the factor of imperceptibility.

We believe that the optimal application of our proposed watermarking method would be source tracking. This watermarking application is based on embedding different watermarks based on the same technique into each distributed copy of a sensitive confidential text. In praxis, this could be useful for tracking parties that break the non-disclosure agreements by sharing the original text with third parties or the general public or for the purpose of internal security drills and discovering unreliable assets within the company by sending out falsified watermarked pseudo-confidential information and tracking any data leaks that follow. The former application is especially useful when the parties in question need to be brought into the court of law, as watermark personalization provides a reliable way to establish connection between the leaked document and the source of the leak.

The robustness of the proposed method would depend on how many copies of a sensitive document a third party has in their possession. If a third party would manage to gain access to multiple variations of the same watermarked document, it would become fairly trivial for them to recognize that paragraph junctures may be a watermark and promptly destroy them by removing paragraph junctures altogether or, in an even more troubling scenario, altering them, with intent or by chance, in such a way that draws suspicion to another innocent party. If this method is to function, it is imperative that security measures be in place that prevent any chance of any single party ever possessing more than one copy of a confidential document with watermarks embedded in this manner. A possible solution would be the embedding of a second false watermark that uses a different embedding method that is deliberately detectable but technically meaningless, which may misdirect the attacker.

The capacity is an interesting factor to consider. In our proposed application, the capacity plays little role by virtue of our method of using trigger-containers. By their very nature, trigger containers may transmit a technically unlimited volume of data due to the fact that the message is predetermined and is simply waiting for the trigger-container to appear. This is further bolstered by the fact that in our proposed method, the watermark only contains information about the source of the leak and the variation of the trigger-container watermark may be easily correlated to the source.

Regarding the final factor, security, it is evident that upon examination of different variations of the same watermark based on this method side by side, the attacker would have little trouble altering or destroying the watermark. Furthermore, it is possible that the watermark may be destroyed by simply changing the formatting of the text [6]. However, this is when we have to consider the data received from the native speakers. The range of the possible locations of paragraph junctures in the Russian texts is vast and may be readily manipulated without raising suspicion.

One possible implementation of the proposed trigger-container based paragraph-juncture watermark could be represented in the following manner (Table 2).

Table 2

Paragraph juncture watermarking example

Receiving party ID	Watermark code
John Adams	15 : 15 : 15
Microcosm Ltd.	13 : 16 : 16
Friends Inc.	16 : 13 : 16
McJill and McWright	14 : 16 : 15
Dr. A. Connors	15 : 16 : 14

In Table 2, the “Receiving party ID” contains 5 fictional receiving parties of a hypothetical sensitive document. Let us imagine that the confidential document contains 45 sentences. We employ sentences in this method because paragraph junctures are simultaneously in all cases sentence junctures in Russian. The “Watermark code” contains a simplified visualization of how the watermark is individualized for each party. For example, for “John Adams” the text would be separated into three paragraphs each consisting of 15 sentences. For “Microcosm Ltd.” the first paragraph would contain 13 sentences whereas the second and third would both have 16 sentences, etc.

The visualization clearly shows how simple this method would be to implement. It could feasibly be automated for larger-scale use, but such an algorithm would need to consider extreme cases. Some sort of limits on the variation would possibly need to be imposed upon it if this embedding process is to be automated. One solution would be to have a human expert check the results for feasibility. For additional security, the expert can be only provided with the watermark code and not the ID of the receiving party. The number of possible combinations can be manipulated via altering the initial number of paragraphs. However, it should be noted that the entropy of such a way of watermarking would be finite and at a certain number of recipients may require increasing the text size to accommodate more paragraph junctures.

Conclusion

Visual analysis perception experiment carried out on Russian native speakers enabled us to determine that paragraph junctures in Russian written texts can be manipulated by changing their location. This in conjunction with using trigger-containers steganographically protects the sense information in written texts by inserting a trigger - a minimal alteration - into the text with the meaning of the secret message agreed upon beforehand and allows to design a possible method of watermarking for using the paragraph junction-based trigger containers in future. The proposed approach is envisaged to be most useful in source tracking in scenarios where the source of a leak needs to be determined. The method based on this approach is to move paragraph junctures in the text and to create a table of corresponding paragraph location coordinates in the text and the identificators of a receiving party for every variation of the watermark.

Our preliminary analysis indicates that this method will have high capacity and imperceptibility. As for robustness and security, the main threats when employing this watermarking technique include the attacker gaining multiple copies of the watermarked document intended for different receiving parties. If that happens, it is possible to determine that the paragraph juncture is the watermark and to destroy or modify it. To prevent this, it might be possible to insert and additional false watermark with higher detection rates to prevent the destruction of the linguistic trigger-container watermark.

Currently, the method based on the proposed approach presupposes the use in Russian texts that are written, printed or electronic. It is unclear if it will prove effective in other languages. To determine that, the additional research needs to be carried out for examining how the native speakers of other languages perceive necessary junctures for paragraphs in their target languages.

REFERENCES

1. Alferov A. P., Zubov A. K., Kuz'min A. S., Cheremushkin A. V. *Osnovy kriptografii: uchebnoe posobie* [Principles of cryptography: teaching aids]. Moscow, Gelios ARB Publ., 2002. 480 p.

2. Wayner P. Strong theoretical steganography. *Cryptologia*, 1995, vol. XIX/3, pp. 285-299.

3. Potapova R., Dzhunkovskiy A. Preliminary Investigation of Potential Steganographic Container Localization. *Lecture Notes in In Computer Science*. Springer, 2020. Vol. 12335LNAI. Pp. 389-398.

4. Potapova R. K. *Rech': kommunikatsiia, informatsiia, kibernetika: uchebnoe posobie* [Speech: communication, information, cybernetics: tutorial]. Moscow, URSS Publ., 2015. 594 p.

5. Kamaruddin N. S., Kamsin A., Por L. Y., Rahman H. A Review of Text Watermarking: Theory, Methods, and Applications. *IEEE Access*, 2018, vol. 6, pp. 8011-8028.

6. Su J. K., Eggers J. J., Girod B. Capacity of Digital Watermarks Subjected to an Optimal Collusion Attack. *Proceedings of European Signal Processing Conference, IEEE (Conference Rroceedings of EUSIPCO, Tampere, Finland, 2000).* Tampere, 2000. Pp. 1-4.

The article submitted to the editors 27.09.2021

INFORMATION ABOUT THE AUTHORS

Rodmonga K. Potapova – Doctor of Philological Sciences, Professor; Head of the Department of Applied and Experimental Linguistics; Moscow State Linguistic University; Russia, 119034, Moscow; RKPotapova@yandex.ru.

Andrey V. Dzhunkovskiy – Junior Researcher of the Laboratory of Experimental Phonetics and Forensic Linguistics; Moscow State Linguistic University; Russia, 119034, Moscow; Vetinari01@gmail.com.

ПОТЕНЦИАЛЬНЫЕ ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ЛИНГВИСТИЧЕСКИХ СТЕГАНОГРАФИЧЕСКИХ ТРИГГЕР-КОНТЕЙНЕРОВ ДЛЯ СОЗДАНИЯ ТЕКСТОВЫХ ВОДЯНЫХ ЗНАКОВ

Р. К. Потапова, А. В. Джунковский

Московский государственный лингвистический университет,
Москва, Российская Федерация

Анализируется возможность использования лингвистических стеганографических триггеров в качестве средства лингвистической маркировки текста водяными знаками. Предлагаемый подход основан на предыдущих экспериментальных исследованиях по теме восприятия сочленений текста носителями русского языка. Было высказано предположение, что определенные минимальные текстовые модификации, рассматриваемые в статье, могут быть использованы в качестве средства нанесения водяных знаков на текст с целью обнаружения и отслеживания источника утечки информации в целях принятия правовых мер, обеспечения соблюдения договора о неразглашении информации и тестирования внутренних уязвимостей. Анализируется возможность изменения точек соединения абзацев в русских текстах и, с помощью соответствующих триггер-контейнеров, использование этого изменения в качестве лингвистических водяных знаков.

Ключевые слова: абзац, источник утечки, соединение слов и абзацев, стеганография, триггер-контейнеры, текстовые водяные знаки.

Для цитирования: *Потапова Р. К., Джунковский А. В.* Потенциальные возможности использования лингвистических стеганографических триггер-контейнеров для создания текстовых водяных знаков // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2021. № 4. С. 76–81. DOI: 10.24143/2072-9502-2021-4-76-81.

СПИСОК ЛИТЕРАТУРЫ

1. *Алферов А. П., Зубов А. К., Кузьмин А. С., Черемушкин А. В*. Основы криптографии: учеб. пособие. М.: Гелиос АРБ, 2002. 480 с.

2. *Wayner P.* Strong theoretical steganography // Cryptologia. 1995. V. XIX/3. P. 285–299.

3. *Potapova R., Dzhunkovskiy A*. Preliminary Investigation of Potential Steganographic Container Localization // Lecture Notes in In Computer Science. Cham: Springer, 2020. V. 12335LNAI. P. 389–398.

4. *Потапова Р. К*. Речь: коммуникация, информация, кибернетика: учеб. пособие. М.: URSS, 2015. 594 с.

5. *Kamaruddin N. S., Kamsin A., Por L. Y., Rahman H*. A Review of Text Watermarking: Theory, Methods, and Applications // IEEE Access. 2018. V. 6. P. 8011–8028.

6. *Su J. K., Eggers J. J., Girod B*. Capacity of Digital Watermarks Subjected to an Optimal Collusion Attack // Proceedings of European Signal Processing Conference, IEEE (Conference Rroceedings of EUSIPCO, Tampere, Finland, 2000). Tampere, 2000. P. 1–4.

Статья поступила в редакцию 27.09.2021

ИНФОРМАЦИЯ ОБ АВТОРАХ

Родмонга Кондратьевна Потапова – д-р филол. наук, профессор; зав. кафедрой прикладной и экспериментальной лингвистики; Московский государственный лингвистический университет; Россия, 119034, Москва; RKPotapova@yandex.ru.

Андрей Владимирович Джунковский – младший научный сотрудник экспериментально-фонетической лаборатории речеведения по криминалистике; Московский государственный лингвистический университет; Россия, 119034, Москва; Vetinari01@gmail.com.