3-8 Information Hiding on Digital Documents by Adjustment of New-line Positions

TAKIZAWA Osamu, MATSUMOTO Tsutomu, NAKAGAWA Hiroshi, MURASE Ichiro, and MAKINO Kyoko

In the usual information hiding applied to digital documents, secret messages are embedded in the layout information (e.g., the space between lines or characters) because character codes have no redundancy. This paper describes a new method for hiding information in plain text without using any layout information. It enables a secret message to be embedded as binary digits that are related to the number of characters in each line of the cover text.

Keywords
Information hiding, Digital watermarking, Steganography, Document, Natural language processing

1 Introduction

With the expanding use of computer networks, information security techniques for transmitting information safely over a network are becoming increasingly important. Ciphers form one of these techniques, and are used in processing and decrypting information to hide it from attackers or to detect traces of tampering. Ciphers do not necessarily conceal their roles in carrying hidden information. Thus, it is easy to find cipher communications along the communication route, and an attacker, despite an inability to decode the cipher, can nevertheless find and interfere with important cipher communications. (That the communication is encrypted suggests to the attacker that the content of the communication is important.) An effective means of addressing such attacks is to hide the information, concealing the fact that secret information is embedded in the communication. Information hiding can be used not only as a means of camouflage but also as a means of embedding copyright information or distribution destination information in content, including images and music. This paper discusses an information hiding technique that uses a digital document as the cover medium and embeds secret information within the new-line codes inserted in the document.

2 Information hiding for documents [1]

2.1 What is information hiding?

Information hiding may be applied as a means of secret communication — as camouflage, in other words — when transmitting information. It may also be used as a means of embedding proprietary information, such as copyright and distribution destination details, in content such as images and music. When this approach is applied to secret communications it is referred to as “gsteganography”, and when it is applied to intellectual property rights it is referred to as “gdigital watermarking”.

Information hiding is a process of embedding secret messages or copyright information (referred to as the embedded data) into content (referred to as the cover data) to create content embedded with information (referred to as the stego data). The stego data is transmitted to the recipient, and the recipient extracts the embedded data from the stego data for use. The main subject of steganography is the embedded data, and the cover data is often used for camouflage only. On the other hand, the main subject of digital watermarking is the cover data (the content), and additional information concerning the cover data is hidden as the embedded data. Thus, steganography focuses on embedding as much data as possible, while digital watermarking focuses on minimizing the difference between the cover data and the stego data (in other words, minimizing the change in content).

2.2 Information hiding for documents; classifications

Information hiding that uses documents as cover data embeds information into the document adding an artificial component unrecognizable as such by third parties; the aim is to allow only the rightful recipient to extract the secret information from the document.

Classical information-hiding techniques used throughout history have employed documents as the cover media. Today, steganography (secret communication) is the first known case in which these techniques are primarily used against threats such as electronic eavesdropping and filtering. Steganography in documents embeds secret information in data that a third party would regard as comprising ordinary communication. Along with steganography, digital watermarking, which embeds copyright information and “fingerprints” into digital content, is another important application of information-hiding in documents. Digital watermarking adds information to the content to allow identification of the people or organizations that are the rightful holders of the content. This process can identify the source of illegal redistribution and is thus expected to have a deterrent effect on the distribution of pirated files.

In terms of hiding information in documents, one must consider the amount of acceptable modification to the cover text data. If the cover text itself forms the content, such as a novel, in principle no modification is acceptable. On the other hand, when the main subject of the copyright claim is an item of software, an image, or a video, and the copyright information is embedded in the document attached to the content, the stego text data should simply maintain the meaning of the cover text; slight changes in the expression of this data may be acceptable. An example of such a case is seen when embedding information using a software package insert as the cover text, such as the manual or the license agreement. Further, in steganography, in which the embedded information is the focus and the stego text is merely camouflage, if the purpose of information hiding is to avoid automatic filtering, the stego text may not need to carry meaning as long as the structure is basically textual.

Information hiding is a technique of hiding information using redundant cover data. Thus, the technique can be classified into several categories according to the type of document redundancy employed. To make this classification easier to understand, it is best to divide information-hiding methods roughly into two groups: those methods in which the artificiality remains in the hard copy (considered here and below as including screen display) and those in which it does not. Whether the artificiality remains in the hard copy depends on the output system; this is therefore not a strict classification. Nevertheless, this is a convenient division for explanation. Below, these methods are outlined assuming generic output systems.

(1) Information hiding in which artificiality remains in the hard copy

The methods in which the artificiality remains in the hard copy are based on the premise that it is visually possible but difficult to recognize the artificiality. Thus, these methods can be used not only for distribution of electronic data but also for distribution of hard copies. On the other hand, the artificiality must be implemented carefully so that it is not discovered. This category is further classified into the following two types according to the principles used to avoid recognition.

(i) Methods that use visually concealed artificiality

This type of information hiding tries to embed information unrecognized, through subtle artificiality that cannot be detected even if the cover text and the stego text are compared side-by-side with the naked eye. Some implement this effect by adding artificiality to the layout of the document. The basic procedure adds subtle artificiality to the document layout using post-script or other functions, and then the secret information is extracted by scanning the stego text printed as a hard copy. The content of the text is not important either in embedding or extracting the secret information. This technique makes use of visual differences between the cover and stego text. Thus, it can also be considered a special form of information hiding within images. When applying this method with hard copies, a weakness is found in that the secret information deteriorates and is lost as the images are repeatedly copied, reducing image quality. It is possible to dispense with hard copies and to receive and extract the secret information entirely in electronic data form. However, it is not then necessary to add artificiality to the layout in such cases, and thus these methods can be regarded as of the same type as information hiding within XML and LaTeX documents, discussed later.

Different methods have been proposed for adding artificiality to layout: scaling of the line spacing or word spacing, scaling of character widths, or rotation of the characters. For example, the standard number of bits between the lines is specified in advance, and the spacing is increased when bit 1 is embedded and decreased when bit 0 is embedded. Thus, the accuracy of extracting the secret information depends on the resolution of the scanner. Less scaling would render the artificiality more difficult to recognize, but then again, extraction error will also increase. The difficulty of recognition of the selected artificiality depends on the language. For example, the scaling of word spacing is said to be more advantageous in European languages (such as English) and the scaling and rotation of fonts is said to be more useful in languages that do not insert spaces between words that use many ideographic characters, such as Japanese[2]. Some methods require collation between the stego text and the original cover text and some do not. Reference[3] describes a number of methods that add artificiality to the layout of the document.

In addition to artificiality within layout, some methods hide small characters and marks in the periphery of the document or within ruled lines. These methods also belong to the current category. Handwritten steganography[4], which hides information in artificiality within the coordinates of the writing or in the tool force, may also belong to this category, to the extent this is regarded as document-based information hiding.

(ii) Methods that use natural-appearing artificiality

Digital documents basically consist of character sequences and layout information. As the characters constitute part of the meaning of the document, indiscriminate digital artificiality in the characters, however slight, may garble them and perhaps significantly damage meaning (and thus reduce the quality of the document). This will also increase the possibility of detection of the artificiality. For this reason, many methods traditionally proposed for information hiding in documents use artificiality in the layout of the document, as described above. However, plain text such as that found in an email does not feature layout information. When hiding information in plain text, one needs to rely on the artificiality added to the characters themselves. In this case, the strategy is to abandon the effort to camouflage the artificiality and instead to rely on the apparent authenticity of directly observed stego text. With this method, artificiality would only be detected with a cover text for comparison. Thus, the assumed utility model does not include any cover text. The artificiality in this category is large, and the secret information is not easily degenerated or lost even with repeated copying in hard-copy format.

To avoid deterioration in documents when adding artificiality, two different approaches are possible: to apply natural language processing (such as word replacement) or to insert characters or character codes that do not influence the outward meaning of the document. Reference[5] presents examples of the former method. The current paper discusses the latter method, which will be explained in the next and subsequent sections.

Some methods do not require an original cover text, generating the stego text from scratch. These methods are also classified within this category. Two examples of proposed tools of this type are "Texto", which converts uuencode files or PGP messages into English sentences resembling poetry, and "NICETEXT", which converts binary data into English sentences of a specified style[6].

(2) Information hiding in which artificiality does not remain in the hard copy

With methods in which artificiality does not remain in the hard copy, the artificiality cannot be recognized visually, and thus is not easily detected. However, the secret information is eliminated when the document is converted from electronic data into the display media (paper or monitor screen). Thus, the methods in this category are applied under the assumption that the document is treated as electronic data until the secret information is extracted.

Among proposed methods of this type is "SNOW", which uses English sentences as the cover text and embeds information by inserting null characters at the end of each line[6]. SNOW first encodes the secret information by compressing the data with Huffman coding, and then inserts up to seven null characters at the end of each line, corresponding to three bits of embedded information per line. Another example is the FFEncode tool[6], which distributes null characters within text data according to Morse code. Another method uses an English LaTeX document as the cover text and embeds information by controlling the positions of line feeds in the main body of the document source file[7]. The methods that embed information in structured documents such as XML documents are also classified into this category, as these also leave no traces of artificiality in hard copies[8].

2.3 Information hiding through new-line position control

Here we discuss an information-hiding technique in which information is embedded by controlling the positions of line feeds in a document[9]. This method is intended for an agglutinative language such as Japanese, in which new lines may be started relatively freely. This technique assumes the use of a filler text as the cover text, with new-line codes only at the ends of paragraphs, such as those prepared by a word processor. Figure 1 shows the flow of the embedding and extraction processes for the embedded data in this method. Figure 2 shows examples of cover and stego texts. Embedding data by providing line feeds at appropriate intervals produces a document with many line feeds (the stego text). Two strategies are used when inserting line feeds: (1) reduction in line-length variation (the sum of the widths of the characters in each line) in order to preserve the apparent artificiality of the document and (2) avoidance

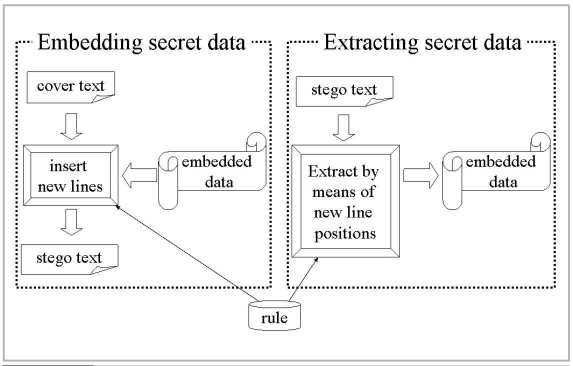

Fig.1 *Process flow in information hiding through control of the number of characters in each line*

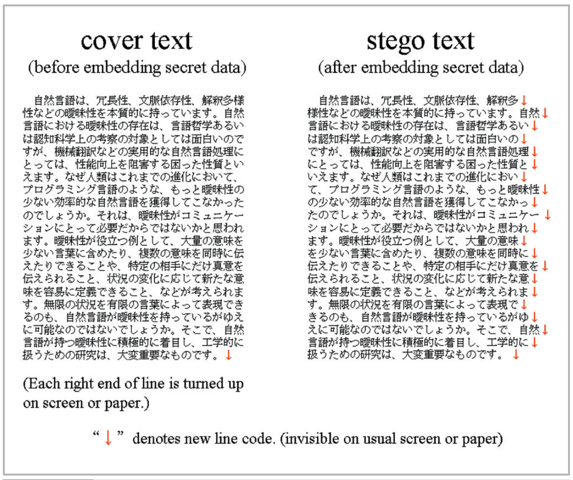

Fig.2 *Examples of cover and stego texts in information hiding through control of the number of characters in each line*

of unnatural line feeds (such as those in the middle of a word). It is necessary to consider the tradeoffs between these two strategies and to determine the positions of the line feeds to make the document appear as natural as possible.

Information hiding by controlling the positions of line feeds does not influence the content of the document. Thus, it can also be applied when the cover text cannot be easily modified. This method adds artificiality to plain text at the character level and also to the positions of line feeds, which form part of the document layout.

3 Information hiding through newline position control

3.1 Introduction

In information hiding in which the number of characters in each line is controlled, the correlation between the position of the line feeds and the embedded data (in other words, the rule illustrated in Fig.1) is essential. This rule may be based on the positions of the line feeds within words or on the number of characters in each line. These approaches are described in detail below.

3.2 Method based on the positions of line feeds within words

In the method based on the positions of line feeds within words, information is embedded within the entry words of a morphological dictionary according to the relationship between the position of the line feed in each word (morpheme) and the embedded information bit (either 1 or 0). Figure 3 shows examples. It is specified in advance that the line feed in "suru" (the Japanese verb meaning "to do") as "sulru" corresponds to "1" ("l" indicates the position of the line feed.). To maintain a natural appearance in the stego text, this

bit "0"	bit "1"	meaning in English
する↓	す↓る	do
プログラミング↓	プログラミン↓グ	programming
プロ↓グラミング	プログ↓ラミング	programming
獲得↓	獲↓得	obtain
コミュニケーション↓	コミュニケーショ↓ン	communication
コミュニケ↓ーション	コミュニケー↓ション	communication
コミュ↓ニケーション	コ↓ミュニケーション	communication
役立つ↓	役↓立つ	useful
と↓して	として↓	as
同時に↓	同時↓に	at the same time
こと↓	こ↓と	thing
考↓え	考え↓	opinion
言語↓	言↓語	language
そこで↓	そこ↓で	therefore
研↓究	研究↓	research

Fig.3 *Example bit assignment tables for each morpheme*
(The morphemes are based on the attached dictionary of Reference [10].)

　自然言語は、冗長性、文脈依存性、解釈多様性などの曖昧性を本
質的に持っています。自然言語における曖昧性の存在は、言語哲
学あるいは認知科学上の考察の対象としては面白いのですが、機械
翻訳などの実用的な自然言語処理にとっては、性能向上を阻害する　0
困った性質といえます。なぜ人類はこれまでの進化において、プログ　1
ラミング言語のような、もっと曖昧性の少ない効率的な自然言語を獲　1
得してこなかったのでしょうか。それは、曖昧性がコミュニケー　1
ションにとって必要だからではないかと思われます。曖昧性が役　1
立つ例として、大量の意味を少ない言葉に含めたり、複数の意味を同時　1
に伝えたりできることや、特定の相手にだけ真意を伝えられること　0
、状況の変化に応じて新たな意味を容易に定義できること、などが考え　1
られます。無限の状況を有限の言葉によって表現できるのも、自然言語　0
が曖昧性を持っているがゆえに可能なのではないでしょうか。そこ　1
で、自然言語が持つ曖昧性に積極的に着目し、工学的に扱うための研究　1
は、大変重要なものです。
....

Fig.4 Example of information embedding with the proposed method
(The number on the right edge is the embedded data (not shown in actual text).)

method pays particular attention to the evenness of character density in each line, and makes the length of each line (the sum of the widths of the characters in the line) as uniform as possible. For this purpose, we define the width of a one-byte character as “1” and the width of a two-byte character such as a kana or kanji as “2”. According to the standard line length specified at the start of the embedding process, the word at the end of a line is subject to embedding. As shown in Fig.3, for long words such as “puroguramingu” (programming) or “comyunikeshon” (communication), 0 or 1 values are ascribed to two or more new-line positions; any of these positions may be used. In this manner, line feed encoding is possible without deviating too far from the standard line length.

Figure 4 shows an example of information embedding using the assignment table shown in Fig.3. The words with embedded data (morphemes) are underlined. (The underlines are not shown in the actual text.) The text shown in Fig.4 is equally spaced. It is clear that the variation in line length is almost undetectable. In the example shown in Fig.4, “01111101011...” is the embedded data.

The technique described in this section has the following characteristics:

(1) Distinction between the types of characters (hiragana/katakana/kanji) enables processing with a lighter computational load without using morphological analysis.
(2) As the embedding method can be defined for each word, the rules for the correlation between the bit of the embedded information and the new-line position are more difficult to detect than in a method based on the number of words in each line (discussed later). Thus, this method is more resistant to extraction attacks.
(3) As the new-line position can be defined for each word, unnatural line feeding can be avoided.

On the other hand, there are a number of problems with this technique, involving the handling of errors in morphological analysis and the handling of single-character morphemes.

3.3 Method based on the number of characters in each line

The method described in this section defines in advance an assignment table linking the number of characters in each line and an embedded bit. A new-line code is inserted where the number of characters in the line corresponds to the embedded data bit. The new-line codes are inserted in such a way that the standard line length remains as uniform as possible. When extracting the embedded data, the number of characters in each line is counted and the embedded data are extracted using the same assignment table. In other words, this method embeds a single bit per line. Figure 5 shows an example of information embedding using an assignment table correlating the number of characters in each line to an embedded

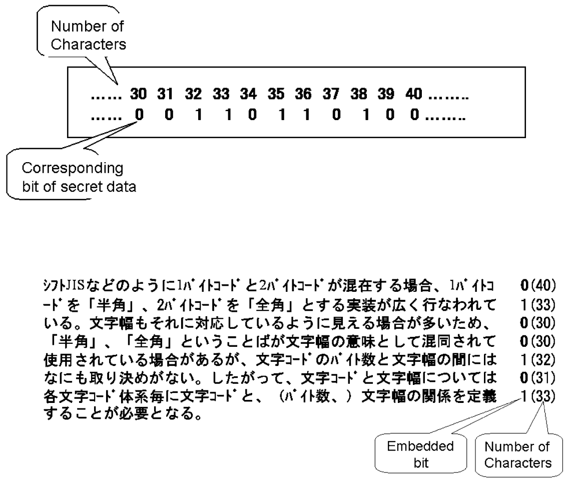

Fig.5 *Assignment table for the embedded bit and example stego text*

(The thick numbers on the right edge are the embedded bits. The numbers in parentheses are the number of characters in the line.)

bit.

To render the line length as uniform as possible, the example in Fig.5 uses 40 characters in the first line, 33 characters in the second line, and so on. to embed "0100101..."

This method does not require collation with the bit assignment table for each morpheme, as is the case with the method based on the new-line positions within words. Thus processing is rapid, with little need for error handling. On the other hand, the embedding rules are simple, which leads to a higher risk of extraction attacks.

4 Implementation

4.1 Introduction

This section discusses the results obtained through the implementation of information hiding tools for embedding one-bit data corresponding to the number of characters per line, as discussed in Section **3.3** Two tools are used here. One uses plain text as the cover text and inserts line feeds according to the bit sequence to be embedded consisting of zeros and ones (the embedded data containing the encrypted secret information) to create a document containing numerous line feeds (stego text). The other tool extracts the secret information from the stego text. The JAVA language is used in development, in consideration of the most appropriate development environment, future extensibility, and the use of encryption algorithms. The embedded data consists of secret information encrypted with RC4 (40-bit key length) to prevent decoding attacks. To thwart guesses as to the key assignment table for embedding information, the tool can create a table based on random numbers to prevent extraction attacks. The tool uses the random number generator Random(), provided by JAVA.

4.2 Embedding method

The implemented tool selects from two types of methods for arranging the embedded data and three types of methods for determining the new-line positions. Combined, the tool offers six embedding methods. The following describes the details of each embedding method.

(A) Arrangement of the embedded data

With this tool, which embeds secret information in a document by mapping information to the number of characters in each line, it is necessary to implement a mechanism to identify the line containing the embedded data in the stego text when extracting the secret information. The authors have implemented two types of embedding methods: A1, which uses flags to indicate the embedded range, and A2, which embeds the data in sequence from the

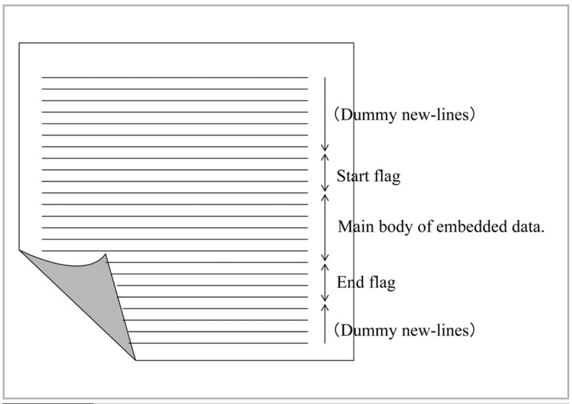

Fig.6 Embedding technique for Method A1

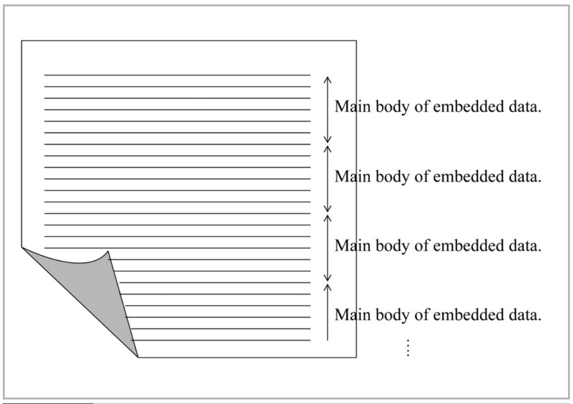

Fig.7 Embedding procedure based on Method A2

top of the cover text. These methods are described below.

[Method A1] Secret information is embedded between the start and end flags.

Method A1 embeds the secret information somewhere within the cover text only once, placing the start flag, embedded data, and end flag in this order. The line feeds from the start of the cover text to the start flag are dummies containing no information. The start position for embedding and the positions of line feeds up to the start flag are determined using random numbers. Thus, the same input produces a different result in each run, to prevent extraction attacks. Figure 6 shows a conceptual diagram of embedding in the cover text using this method.

This method specifies the following parameters for the embedding process: the assignment table, standard line length, minimum line length, cipher (decoding) key, start flag (eight-bit binary), end flag (eight-bit binary), and maximum starting line. When extracting the secret information, this method specifies the same assignment table, minimum line length, cipher (decoding) key, beginning flag, and end flag used for embedding. A minimum line length is specified to prevent the embedding of information in lines that are deemed too short. This is necessary to exclude lines with lengths that differ significantly from the others, as at the end of a paragraph or in captions, as embedding targets. The minimum line length is a parameter required both in embedding and extraction. The maximum start line specifies the maximum number of dummy line feeds up to the start flag. In embedding, the start line is placed after a number of lines that is randomly chosen within this maximum start-line value. The maximum starting line is a required parameter only in embedding.

With this method, attackers cannot easily determine the location of embedded information. However, the data is embedded only once, so that resistance to attack (the conservation of embedded data) when the stego text is partially deleted for editing is low. When extracting secret information, information is also required for start and end flags as well as for the assignment table, which is the common key, and the cipher (decoding) key.

This method is suitable when the embedded data is relatively large compared to the cover text and repeated embedding is difficult, or when partial deletion of the stego text for editing is unlikely.

[Method A2] Secret information is embedded repeatedly

Method A2 repeatedly embeds data from the beginning of the cover text in all line feeds. Thus, it requires no dummy line feeds, start flag, or end flag. Figure 7 shows a conceptual diagram of embedding in a cover text using this method.

This method specifies the following parameters when embedding information: the assignment table, standard line length, minimum line length, and cipher (decoding) key. When extracting the secret information, this method specifies the same assignment table, minimum line length, and cipher (decoding) key used for embedding. This method embeds data redundantly, so that if a means is provided to identify the start of the data for extraction, it is highly probable that the embedded data is correctly extracted even if the stego text is partially deleted for editing. However, this method poses a potentially high risk that the assignment table will be discovered from the repeated patterns.

(B) Method for determining the new-line positions

Three methods are implemented to determine the new-line positions, in consideration of the tradeoff between uniformity of line length and the natural appearance of the new-line positions. These three methods are explained below. Although the examples below all use Method A1 for the arrangement of the embedded data, these three methods can also be combined with Method A2.

[Method B1] Emphasis on uniformity in line length

Method B1 places line feeds near the standard line length while minimizing variation in lengths, subject to Japanese hyphenation and other punctuation restrictions. Japanese hyphenation is in accordance with standard MS-Word Japanese hyphenation rules for line heads and tails. Figure 8 shows an example of output using this method.

In this method, the variation in line length is small, so the document appears natural in terms of page design. However, many unnatural line feeds result, as in the middle of a word; the stego text thus may give readers the impression that something is awry.

[Method B2] Line feeds for particular types of characters are restricted

Method B2 applies additional restrictions to Method B1 and avoids line feeds in particular types of character sequences (numbers and alphabets). Figure 9 shows an example of output with this method.

In Fig.9, an alphabetical string such as “representation” is not broken into two lines, so the line with this word is slightly longer than other lines. Thus, Method B2 allows greater variation in line length than Method B1.

[Method B3] Significant emphasis on character-type boundaries

Method B3 adds further constraints to Method B2 to avoid line feeds in kanji, hiragana, and katakana sequences and to restrict line feeds in parentheses. (With five or fewer characters within a pair of parentheses, line feed is avoided.) Thus, the line feed is primarily inserted between different types of characters (kanji/hiragana/katakana/alphabet). In Japanese, the boundary between different types of characters (such as between hiragana and kanji, or between katakana and hiragana) is often the boundary between clauses. Thus, this method increases the natural line feeds between clauses. Figure 10 shows an example of output with this method.

In Fig.10, the document appears to feature clause-based line feeds, which makes the document easy to read. Nevertheless, the deviation in line length is even greater than in Method B2.

5 Evaluation

5.1 Introduction

Information hiding methods should be evaluated in light of (1) the amount of information that can be embedded, (2) the difficulty of detecting information embedding, (3) the difficulty of extracting the embedded data, and (4) the difficulty of destroying the embedded data. With respect to criterion (1), the embedding rate can be quantitatively evaluated. However, criteria (2), (3), and (4) involve evaluation of the behavior of attackers, thus requiring subjective evaluation using actual subjects. This section considers the criteria (2), (3) and (4) in more detail.

In the subjective evaluation for (2), (3), and (4), it is our opinion that criterion (2), the difficulty of detecting that information has

4 XML 情報ハイディングの検討

構造化文書であるXML に適した情報ハイディング手法
を検討する。特に文書の論理構造に着目した埋め込み手
法は、これまでの研究事例には見られないが、XML 等
の構造化文書に適用可能な情報ハイディング手法として
有望である。

4.1 XML の特徴

XML やSGML のような構造化文書は、基本的には文書には
論理構造のみを持たせ、物理構造(体裁)は必要に応じ
て外部から付与する。XML では文書の内容(content)、構
造(structure)、体裁(style)は個別に扱われ、実際の文
書は複数のテキストデータを組み合わせて表現される。
内容をタグによってマークアップされたテキストをXML
 文書(XML document)と呼び、文書構造を表現するための
要素と属性をDTD において定義する。XML 文書のマー
クアップ部分の表記(representation)は文書の内容と
は別に扱う。スタイルはCSS やXSL のようなスタイル
シートに定義し、必要に応じて文書と組み合わせて用い
る。

4.2 内容の表し方のバリエーションの利用

XML 文書中の要素の内容を同義語で表せるならば、3.1に
挙げたような内容に関するテキスト情報ハイディング
手法が適用できる。同義語で表された文書がアプリケー
ションで全く同じ処理を受けることが前提となる。

4.3 スタイル指定のバリエーションの利用

スタイルシートをstego-text とすれば、文書の論理構造
を変更せずに、物理構造に関する記述のみの変更で情
報ハイディングが行える。印刷・表示された文書の見
た目に関してはアプリケーションへの依存度が高いた
め、3.2 のテクニックを応用した手法を構成するには、
想定アプリケーション環境を限定する必要がある。

Fig.8 *Example of stego text based on Method B1*

been embedded, is equivalent to an assessment of the naturalness of the stego text. We also equate (3), the difficulty of extracting the embedded data, with the issue of security in information hiding, and (4), the difficulty of destroying embedded data (resistance to destructive attacks) with the strength of information hiding. As such, the subjective evaluation here in fact examines two aspects: one is the naturalness of the stego text, and the other is the security and tamper-proofing of the information hiding. Thus it would be reasonable to perform subjective evaluation experiments and subsequent analysis based on these two aspects. The experiments should vary the combination of implementation methods discussed in Section 4, (A) in the arrangement of embedded data or (B) in the determination of new-line positions, and the types of cover text should also be varied, as shown in Table 2. Table 1 summarizes the applicable classifications. The difference in the arrangement of the embedded data is considered to have an effect only when extracting or destroying the embedded information. Thus, this variable is included only in the evaluation of the security and tamper-proofing of information hiding. We also describe the details of the experimental procedure to evaluate the naturalness of the stego text based on different cover text genres (Section **5.3.2**).

We nevertheless consider that the subjec-

4 XML 情報ハイディングの検討

構造化文書であるXML に適した情報ハイディング手法を
検討する。特に文書の論理構造に着目した埋め込み手法
は、これまでの研究事例には見られないが、XML 等の
構造化文書に適用可能な情報ハイディング手法として有望
である。

4.1 XML の特徴

XML やSGML のような構造化文書は、基本的には文書に
は論理構造のみを持たせ、物理構造(体裁)は必要
に応じて外部から付与する。XML では文書の内容(content)、
構造(structure)、体裁(style)は個別に扱
われ、実際の文書は複数のテキストデータを組み合わせ
て表現される。内容をタグによってマークアップされ
たテキストをXML 文書(XML document)と呼び、文
書構造を表現するための要素と属性をDTD において定義
する。XML 文書のマークアップ部分の表記(representation)
は文書の内容とは別に扱う。スタイルはCSS やXSL のよ
うなスタイルシートに定義し、必要に応じて文書と組み
合わせて用いる。

4.2 内容の表し方のバリエーションの利用

XML 文書中の要素の内容を同義語で表せるならば、3.1
に挙げたような内容に関するテキスト情報ハイディング
手法が適用できる。同義語で表された文書がアプリケー
ションで全く同じ処理を受けることが前提となる。

4.3 スタイル指定のバリエーションの利用

スタイルシートをstego-text とすれば、文書の論理構造
を変更せずに、物理構造に関する記述のみの変更で情報ハイ
ディングが行える。印刷・表示された文書の見た目に関し
てはアプリケーションへの依存度が高いため、3.2 のテク
ニックを応用した手法を構成するには、想定アプリケーシ
ョン環境を限定する必要がある。

Fig.9 Example of stego text based on Method B2

tive evaluation experiments require future elaboration and improvements. Thus, in Sections **5.3** and **5.4** below we present only an overview of the subjective evaluation experiments.

5.2 Cover texts used for evaluation

Table 2 shows the cover texts used in the evaluation. The characteristics of the cover texts will affect the results of subject evaluation. Thus, various texts are prepared including news articles, technical papers, and literary works.

5.3 Subjective evaluation of difficulty in detecting information embedding

5.3.1 Evaluation of the naturalness of the stego text based on method of determining new-line positions

This test evaluates the effect of the differences among the three methods of determining new-line positions, as discussed in Section **4.2** (B), on the naturalness of the generated stego text. The subject group, consisting of 5 to 10 people, is selected with no particular conditions. Stego texts are generated with the same cover data and different methods for determining new-line positions; the data is then provided to subjects in the form of paper or electronic documents. The subjects review each stego

4 XML 情報ハイディングの検討

構造化文書であるXML に適した情報ハイディング手法を
検討する。特に文書の論理構造に着目した埋め込み手法
は、これまでの研究事例には見られないが、XML 等の
構造化文書に適用可能な情報ハイディング手法として有望
である。

4.1 XML の特徴

XML やSGML のような構造化文書は、基本的には文書には論理構造のみを
持たせ、物理構造(体裁)は必要に応じて外部から
付与する。XML では文書の内容(content)、
構造(structure)、体裁(style)は個別に扱
われ、実際の文書は複数のテキストデータを組み合
わせて表現される。内容をタグによってマークアップ
されたテキストをXML 文書(XML document)と呼
び、文書構造を表現するための要素と属性をDTD において
定義する。XML 文書のマークアップ部分の表記(representation)
は文書の内容とは別に扱う。スタイルはCSS やXSL のような
スタイルシートに定義し、必要に応じて文書と組み
合わせて用いる。

4.2 内容の表し方のバリエーションの利用

XML 文書中の要素の内容を同義語で表せるならば、3.1
に挙げたような内容に関するテキスト情報ハイディング
手法が適用できる。同義語で表された文書がアプリケーションで
全く同じ処理を受けることが前提となる。

4.3 スタイル指定のバリエーションの利用

スタイルシートをstego-text とすれば、文書の論理構造
を変更せずに、物理構造に関する記述のみの変更で情報
ハイディングが行える。印刷・表示された文書の見た
目に関してはアプリケーションへの依存度が高いため、
3.2 のテクニックを応用した手法を構成するには、想定アプリケーション環境を
限定する必要がある。

Fig.10 Example stego text based on Method B3

text and rate it using the five-point scale shown in Fig.11.

5.3.2 Evaluation of the naturalness of the stego text based on the cover-text type

This test evaluates the effect of the type of cover text on the naturalness of the stego text. As in Section ***5.3.1***, the subject group is selected with no specific conditions to include five to 10 people. Stego texts are generated with a single method and different types of cover data; the data is then provided to subjects in the form of paper or electronic documents. The subjects review each stego text and rate it using the five-point scale shown in Fig.11.

The following are the details of the experimental procedure.

(1) Preparation

For the cover texts in Table 2, stego texts are generated using the tool discussed in Section 4, with the same embedded data. In this experiment, the method for determining newline positions is Method B1, which emphasizes uniformity in line length, and the A2 method is used to arrange the embedded data (repeated data embedding). The methods are each limited to a single type to highlight the effect of cover-text type on the evaluation of naturalness among the defined number of subjects. The set relationship between the number of characters in a line and the bit value of the embedded data is simple: the bit value is “1” when the number of characters is even and “0”

Table 1 *Classification of the subjective evaluation experiment*

主観評価対象	条件
ステゴテキストの自然性 …（２） （情報が埋め込まれていることの見破られにくさ）	改行位置の決定方式(5.3.1節)
	カバーテキストのジャンル(5.3.2節)
情報秘匿の安全性及び強度 （埋め込まれた情報の抽出されにくさ）…（３） （埋め込まれた情報の破壊されにくさ）…（４）	改行位置の決定方式あるいはエンベデッドデータの配置方式(5.4.1節)
	カバーテキストのジャンル(5.4.2節)

Table 2 *Cover texts used for evaluation*

テキスト種類		テキスト名	テキストサイズ（バイト）	特　徴	備考
ニュース	一般	A	1,929	漢字の長い文字列多い	
		B	1,751	全角文字のみ	
	専門分野	C	2,258	半角英数文字、半角カナあり	暗号関連記事
		D	2,433	半角英数文字、英単語あり	Windows 関連記事
	子供向け	E	3,765	ひらがな多い	子供向けニュース解説
論文	専門分野	F	2,290	半角英数文字、英単語あり	SCIS 論文
		G	3,336	半角記号文字あり	SCIS 論文
文　学	古典	H	3,789	全角文字のみ、読点・ひらがな多い	「枕草子」
		I	6,353	全角文字のみ、ひらがな多い	「源氏物語」
	子供向け	J	3,606	全角文字のみ、ひらがな・話ことば多い	「不思議の国のアリス」
		K	5,418	全角文字のみ、ひらがな・話ことば多い	「風の又三郎」
	一般	L	5,640	全角文字のみ	「我輩は猫である」
		M	1,866	全角文字のみ	「羅生門」

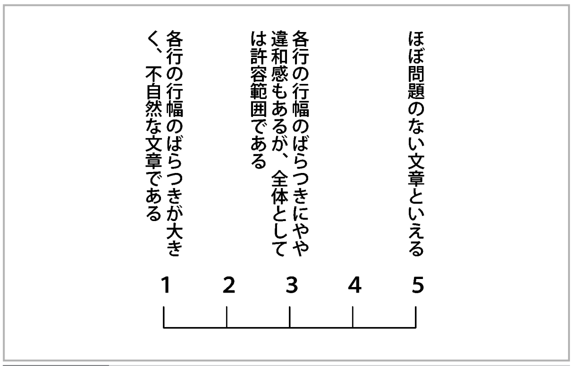

Fig.11 *Standards for evaluation*

when it is odd.

(2) Experiment procedure

(i) Distribution of experiment sheet and evaluation sheet

The experiment sheet and evaluation sheet are distributed to the subjects on paper or as electronic documents. Figure 12 shows examples of the experiment sheet, and Figure 13 shows an example of the evaluation sheet.

(ii) Distribution of evaluation manual

The experiment leader distributes the "evaluation manual" shown in Fig.14 to the subjects and explains its contents. He or she then instructs the subjects to read the manual before beginning the evaluation.

(iii) Evaluation by subjects

The subjects evaluate the stego texts according to the distributed evaluation manual.

(iv) Collection of experimental data

The experiment leader collects the experiment sheet, the evaluation sheet, and the evaluation manual from the subjects after evaluation.

(3) Analysis of the experimental results and evaluations

Experiments are performed using two or more documents of the different cover-text types indicated in Table 2 ("children's news" text not used). Thus, it is possible to assess evaluations in terms of genre and in terms of different documents. The evaluation marks are

【文書1】
経済産業省、サイバー犯罪条約に沿った国内法の整備を
提案

経済産業省は、サイバー刑事法研究会の報告書「欧
州評議会サイバー犯罪条約と我が国の対応について」を
公表した。報告書では、サイバー犯罪条約を日本が批准
した場合に必要となる立法作業や調整事項に具体的に
言及している。

「欧州評議会サイバー犯罪に関する条約」は、2001年1
1月8日に、欧州評議会で正式に採択された条約。サイバ
ー犯罪対策分野において世界初の条約であり、オブザー
バとして参加していた日本も、米国やカナダと並んで、同
年11月23日に同条約に署名している。

この「サイバー犯罪条約」を日本が批准した場合、適
切な処罰や、操作手順の迅速・円滑な実施、個人情報の
保護といった面での法制度の整備が早急に必要となるた
め、2001年8月から、サイバー刑事法研究会が、同条約
の内容及び関連法制に関する検討を進めてき
た。

今回、報告書では、「サイバー犯罪条約」中、現行
の国内法では条約上の義務を履行できない、あるい
は、履行できない可能性が高いと考えられる条項につい
て、必要な立法措置や検討すべき内容をリストアップ
した。

例えば、「条約義務を履行できない行為」とし
て、「スタンドアロン・コンピュータを対象とする妨
害や、データ妨害を行うために、システムにアクセス
するためのパスワードなどを、製造、提供、販売、譲
渡、貸し渡し、輸入する行為」があり、この行為に対
しては、電磁的記録毀棄罪、電子計算機等使用業務妨
害罪について準備罪を新設することが提言されている。

また、「条約義務を履行できない可能性が高い行為」
として、「児童ポルノ画像自体をインターネットを通
じて送信する行為」があり、この行為に対しては、児
童買春・児童ポルノ禁止法第2条第3項における「児童ポ
ルノ」の定義規定を改正し、児童ポルノが含まれるこ
とを明文で追加するか、または、児童ポルノデータを
コンピュータ・システムを通じて送信することを処罰す
る規定を創設すべきと提言している。

【文書2】

みずほ、30日に振替900万件 決済集中　増員体制
“背水の陣”

口座振替の遅れなどのシステム障害の復旧を急ぐ
「みずほグループ」は、今月30日に正常化の成否を占
う決済の集中日を迎える。24日の衆院財務金融委員
会の参考人質疑で、みずほホールディングスの前田晃伸
社長は、口座振替が30日だけで900万件と今年のピ
ークになると説明した。ここで障害が再発すれば、前
田社長を始めとする経営陣の責任が一段と厳しく問わ
れることになるだけに、みずほ側は増員体制を取り、
“背水の陣”で準備作業を進めている。

みずほグループは、年間2700万件の口座振替を
扱うが、30日の振替分は、この3分の一に当たる9
00万件にのぼる。本来の30日分に加え、給与支払日
直後の26―27日に設定されているクレジット代金
や、保険料などの引き落としが、大型連休前半の三連
休の影響で、休み明けの30日に集中するためだ。企
業別では、みずほが扱う分だけで、NTTドコモが1
40万件、オリエントコーポレーションが35万件、
ダイエーオーエムシーが数10万件などの大量の振替が
予定される。

Fig.12a *Example experiment sheet (Sheet number 1: General news)*

calculated as follows:

(Ⅰ) Evaluation distribution and average evaluation mark for each genre

(Ⅱ) Evaluation distribution and average evaluation mark for each document

Based on the results of the above calculations, the results of (Ⅰ) are used to analyze the effect of cover-text type on the evaluation of the naturalness of the stego text, and the results of (Ⅱ) are used to analyze the effect of individual document choice on the evaluation of the naturalness of the stego text.

5.4 Subjective evaluation of security and tamper-proofing of information hiding

5.4.1 Evaluation of security and tamper-proofing of information hiding based on methods of arranging embedded data or of determining new-line positions

This test evaluates the effects on tamper-proofing of the two methods of arranging embedded data or the effects of the three methods of determining new-line positions, as discussed in Section **4.2** (A) and (B). The subject group consists of five to 10 undergraduate and graduate students in information engineering, presumed to have significant interest in cipher techniques. Stego texts are generated in six different ways by combining the two methods of arranging the embedded data and the three methods of determining new-line positions, and are distributed as electronic documents. The subjects are requested to modify freely any texts that they consider to hold embedded information, while maintaining textual meaning.

5.4.2 Evaluation of security and tamper-proofing of information hiding based on the cover-text type

This test evaluates the effect of cover-text type on the resistance to tampering. The experiment is planned as follows. As in Section ***5.4.1***, the subject group consists of five to 10 undergraduate and graduate students in

【文書10】

けがはぜんぜんなくて、すぐにとび起きました。見上げても、頭上はずっとまっ暗。目の前にはまた長い通路があって、まだ白うさぎがその通路をあわてて走っていくのが見えました。これは一刻もむだにできません。アリスはぴゅーんと風のようにかけだして、ちょうどうさぎがかどを曲がりしなに「やれ耳やらヒゲやら、こんなにおそくなっちゃって！」と言うのが聞こえました。そのかどをアリスが曲がったときには、かなり追いついていました。が、うさぎがどこにも見あたりません。そこは長くて天井のひくいろうかで、屋根からランプが一列にぶら下がって明るくなっていました。

そのろうかはとびらだらけでしたが、どれも鍵がかかっています。アリスは、ろうかの片側をずっとたどって、それからずっともどってきて、とびらをぜんぶためしてみました。どれも開かないので、アリスはろうかのまん中をしょんぼり歩いて、いったいどうやってここから出ましょうか、と思案するのでした。

いきなり、小さな三本足のテーブルにでくわしました。ぜんぶかたいガラスでできています。そこには小さな金色の鍵がのっているだけで、アリスがまっ先に思ったのは、これはろうかのとびらのどれかに合うんじゃないかな、ということでした。でもざんねん！　鍵穴が大きすぎたり、それとも鍵が小さすぎたり。どっちにしても、とびらはどれも開きません。でも、二回目にぐるっとまわってみたところ、さっきは気がつかなかったひくいカーテンがみつかりました。そしてそのむこうに、高さ40センチくらいの小さなとびらがあります。さっきの小さな金色の鍵を、鍵穴に入れてためしてみると、うれしいことにぴったりじゃないですか！

あけてみると、小さな通路になっていました。ネズミの穴くらいの大きさしかありません。ひざをついてのぞいてみると、それは見たこともないようなきれいなお庭につづいています。こんな暗いろうかを出て、あのまばゆい花だんやつめたいふん水の間を歩きたいなぁ、とアリスは心から思いました。でも、その戸口には、頭さえとおらないのです。「それに頭はとおったにしても、かたがないとあんまり使いものにならないわ」とかわいそうなアリスは考えました。「ああ、望遠鏡みたいにちぢまれたらな！　できると思うんだ、やりかたさえわかれば」というのも、近ごろいろいろへんてこりんなことが起こりすぎたので、アリスとしては、ほんとうにできないことなんて、じつはほとんどないんだと思いはじめていたのです。

【文書11】

谷川の岸に小さな学校がありました。　教室はたった一つでしたが生徒は三年生がないだけであとは一年から六年までみんなありました。運動場もテニスコートのくらいでしたがすぐうしろは栗の木のあるきれいな草の山でしたし運動場の隅にはごぼごぼつめたい水を噴く岩穴もあったのです。

さわやかな九月一日の朝でした。青ぞらで風がどうと鳴り、日光は運動場いっぱいでした。黒い雪袴をはいた二人の一年生の子がどてをまわって運動場にはいって来て、まだほかに誰も来ていないのを見て「ほう、おら一等だぞ。一等だぞ。」とかわるがわる叫びながら大悦びで門をはいって来たのでしたが、ちょっと教室の中を見ますと、二人ともまるでびっくりして棒立ちになり、それから顔を見合せてぶるぶるふるえました。がひとりはとうとう泣き出してしまいました。というわけは、そのしんとした朝の教室のなかにどこから来たのか、まるで顔も知らないおかしな赤い髪の子供がひとり一番前の机にちゃんと座っていたのです。そしてその机といったらまったくこの泣いた子の自分の机だったのです。もひとりの子ももう半分泣きかけていましたが、それでもむりやり眼をりんと張ってそっちの方をにらめていましたら、ちょうどそのとき川上から「ちょうはあかぐり　ちょうはあかぐり」と高く叫ぶ声がしてそれからまるで大きな鳥のように嘉助が、かばんをかかえてわらって運動場へかけて来ました。と思ったらすぐそのあとから佐太郎だの耕助だのどやどややってきました。

「なして泣いでら、うなかもたのが。」嘉助が泣かないこどもの肩をつかまえて云いました。するとその子もわあと泣いてしまいました。おかしいとおもってみんながあたりを見ると教室の中にあの赤毛のおかしな子がすましてしゃんとすわっているのが目につきました。みんなはしんとなってしまいました。だんだんみんな女の子たちも集って来ましたが誰も何とも云えませんでした。

赤毛の子どもは一向こわがる風もなくやっぱりちゃんと座ってじっと黒板を見ています。　すると六年生の一郎が来ました。一郎はまるでおとなのようにゆっくり大股にやってきてみんなを見て「何した」とききました。みんなははじめてがやがや声をたててその教室の中の変な子を指しました。一郎はしばらくそっちを見ていましたがやがて鞄をしっかりかかえてさっさと窓の下へ行きました。　みんなもすっかり元気になってついて行きました。

Fig.12b Example experiment sheet (Sheet number 6: Children's literature)

評価シート

実験シート（その１）～（その７）の１３の文書について、それぞれの文書の自然性に対する主観評価を評価基準を参考に1～５の数字で記入してください。

【評価基準】

1　各行の行幅のばらつきが大きく、不自然な文章である
2
3　各行の行幅のばらつきにやや違和感もあるが、全体としては許容範囲である
4
5　ほぼ問題のない文章といえる

文書番号	評　価
1	
2	
3	
4	
5	
6	
7	
8	
9	
１０	
１１	
１２	
１３	

Fig.13 Example evaluation sheet

評 価 作 業 マニュアル

以下に、あるソフトウェアにより生成された文書の、文書自然性を評価するための実験手順を示します。

【配布された文書について】

- 配布された実験シートには、7 つの分野、13 の異なる文書が提示されています。
- 提示された文書の内容はすべてオリジナルの文書に従っています。

【実験手順】

- 提示された 13 の文書を読んで、それぞれの文書について、評価基準に従い評価シートに評価を記入してください。
- 上記作業を実験担当者が指定する時間内で行なってください。
- 作業終了後、実験シート、評価シートを実験担当者に返却してください。

【禁止事項】

- 他の被験者と相談すること。
- 作業終了後、他の被験者に自分の回答内容を知らせること。
- 配布した実験シート、評価シートをコピーしたり、他者に譲渡すること。

Fig.14 *Evaluation manual*

information engineering, presumed to have significant interest in cipher techniques. Stego texts are generated using cover texts of different types and are distributed as electronic documents. The subjects are requested to modify freely any texts that they consider to hold embedded information, while maintaining textual meaning.

6 Discussion

As stated at the beginning of this paper, information hiding can be applied in two major applications: "digital watermarking", which embeds copyright information or "fingerprints" (information for identifying the distribution destination) into electronic contents, and "steganography" (secret communication), intended to counter threats such as electronic eavesdropping and filtering by a third party. Information hiding in documents as discussed in this paper is considered best applied in cases in which a third party cannot easily modify the new-line positions—for example, in direct document exchanges between two people (as with email and printed documents). For example, when distributing a confidential printed document among concerned parties, "fingerprints" may be embedded based on the number of words in each line throughout the document without modifying the content. Then, as a person intending to leak the document cannot easily produce a paper copy that can hide the source of the leak, this method prevents easy leaks. When printed documents are used as the media, the secret information is extracted using an OCR (Optical Character Reader), as when information is hidden in the document layout; many related methods have traditionally been proposed. However, it should be recognized that the embedded data is not subtly conveyed, as in the size of line spacing, character spacing, or miniature characters, but instead corresponds to each line length (the sum of the widths of each character), which is relatively conspicuous. This method is nevertheless superior in that the secret information is not easily lost even if the document is repeatedly copied with low-quality reproduction.

Let us consider the points to keep in mind when applying this technique to steganography or digital watermarking. Steganography focuses on communicating secret information and uses the stego text only for camouflage. Thus, if the purpose of information hiding is to avoid automatic filtering by machines in the course of distribution as electronic data, a composition resembling natural language may be sufficient as the stego text, even in the absence of any logical meaning in the document. On the other hand, when applying the technique to digital watermarking, the cover text must have meaning. If content with significant meaning even in subtle expressions, as in novels, is to be used as the cover text, the text cannot be modified in any way. Even if the document emphasizes the basic meaning of the content, as in confidential documents and manuals, only subtle modification is allowed within a range that does not change the meaning of the document. In this respect, the developed tool will only modify the document in the position of the line feeds. Thus, this tool can be used for both steganography and digital watermarking.

When using the technique for steganography, it is particularly important to hide the fact that information is embedded in the document. Thus, it is necessary to devise methods that can maintain the visual naturalness of the stego text, i.e., the uniformity of line lengths and the naturalness of the new-line positions. To this end, it is effective to optimize the method of determining the new-line positions. It is also effective to use layout functions when displaying or printing the document, such as justification.

Whether a technique is used for steganography or for digital watermarking, we must consider measures against decoding, extraction, tampering, and spoofing. The technique discussed in this paper uses randomizing of the assignment table and encoding of secret information. Error correction may also be used as an additional measure. When considering the distribution of the stego text as electronic data, it is also essential to take measures against destructive attacks through partial deletion of the stego text for editing and modification of new-line positions. The technique provides two methods for arranging the embedded data—redundant embedding in Method A2, and randomized selection of embedding position in Method A1, both of which are effective to an extent.

The technique discussed in this paper may be applied not only to information hiding but also to detection of tampered documents. In other words, the hash value or message authentication codes (MAC) can be embedded into a text document according to this method as verification data; this data is then extracted for comparison with the stego text in verification. Any tampering can thus be detected[11].

7 Conclusions

This paper discusses an information hiding technique that uses a digital document as the embedding medium and the new-line positions inserted in the document as the secret information. Even in our present society, in which multimedia technology continues to advance, text-based information such as e-mail, is still the most important means of information exchange. Information hiding in documents is therefore likely to remain important, and many applications will continue to arise that lend themselves to related techniques.

Acknowledgements

This study is being conducted in the context of regular discussions with members of Prof. Tsutomu Matsumoto's laboratory at Yokohama National University, members of Prof. Hiroshi Nakagawa's laboratory at the University of Tokyo, and members of the Mitsubishi Research Institute, Inc. We appreciate their useful advice.

References

1 Hirosho Nakagawa, Osamu Takizawa and Shingo Inoue, “Information Hiding on Digital Documents”, IPSJ Magazine, Vol.44, No.3, pp.248-253, 2003. (In Japanese)

2 Kineo Matsui, “Primer of Digital Watermarking”, Morikita Publishing, 1998. (In Japanese)

3 R.J.Anderson and F.A.P.Petitcolas, “Information Hiding -An Annotated Bibliography”, http://www.cl.cam.ac.uk/~fapp2/steganography/bibliography/Annotated_Bibliography.pdf, 1999.

4 Norihisa Segawa, Yuko Murayama and Masatoshi Miyazaki, “The Proposal of a Handwriting Steganography with the Characteristic of Handwriting Input Equipment”, Computer Security Symposium 2002, pp.215-219, 2002. (In Japanese)

5 Hiroshi Nakagawa, Koji Sanpei, Tsutomu Matsumoto, Takeshi Kashiwagi, Shuji Kawaguchi, Kyoto Makino and Ichiro Murase, “Meaning Preserving Information Hiding _Japanese text Case”, IPSJ Journal, Vol.42, No.9, pp. 2339 - 2350, 2001. (In Japanese)

6 Information-Technology Promotion Agency, “Technical Research Report of Information Hiding”, http://www.ipa.go.jp/security/fy10/contents/crypto/report/Information-Hiding.htm, 1998. (In Japanese)

7 Tsutomu Matsumoto, Hiroshi Itoyama, “Can Bypassing Lawful Access be Always Detected?”, Technical Report of IEICE, ISEC96-79, pp. 159-164, 1997. (In Japanese)

8 Shingo Inoue, Ichiro Murase, Osamu Takizawa, Tsutomu Matsumoto and Hiroshi Nakagawa, “A Proposal on Steganography Methods using XML”, The 2002 Symposium on Cryptography and Information Security, IEICE, pp.301-306, 2002. (In Japanese)

9 Osamu Takizawa, Tsutomu Matsumoto, Hiroshi Nakagawa, Ichiro Murase and Kyoko Makino, “Steganography on Digital Documents by Adjustment of New-line Positions”, IPSJ Journal, Vol.45, No.8, pp. 1977 - 1979, 2004. (In Japanese)

10 “ChaSen -A morphological analysis system”, version 2.0 for Windows, Computational Linguistics Laboratory, Graduate School of Information Science, Nara Institute of Science and Technology, 1999. (In Japanese)

11 Tsutomu Matsumoto, Katsunari Yoshioka, Masataka Suzuki, Ken' ichiro Akai, Osamu Takizawa, Kyoko Makino and Hiroshi Nakagawa, “Text Alteration Detection by New-Line Positions”, The 2004 Symposium on Cryptography and Information Security, IEICE, pp.983-988, 2004. (In Japanese)

TAKIZAWA Osamu, *Ph.D.*

Senior Researcher, Security Advancement Group, Information and Network Systems Department

Contents Security, Telecommunication Technology for Disaster Relief

MATSUMOTO Tsutomu, *Dr. Eng.*

Professor, Yokohama National University

Information Security

NAKAGAWA Hiroshi, *Dr. Eng.*

Professor, University of Tokyo

Natural Language Processing

MURASE Ichiro

Mitsubishi Research Institute, Inc.

Information Security

MAKINO Kyoko

Mitsubishi Research Institute, Inc.

Information Security