3-8　改行位置の調整によるドキュメントへの情報ハイディング

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

滝澤　修　　松本　勉　　中川裕志　　村瀬一郎　　牧野京子
TAKIZAWA Osamu, MATSUMOTO Tsutomu, NAKAGAWA Hiroshi, MURASE Ichiro, and MAKINO Kyoko

要旨

情報が秘匿されている事実自体を隠す技術である情報ハイディングは、情報の埋め込み媒体が持つ情報の冗長性を利用するため、画像や音響信号など冗長度の高い媒体について多く研究されてきた。それに対して本論文では、デジタルドキュメントを埋め込み媒体とし、文書内に挿入された改行コードの位置を秘匿情報とする情報ハイディングについて述べる。本手法はドキュメントのレイアウト情報に依存しないため、電子メールのようなプレーンテキストに対しても秘匿情報の埋め込みが可能で、文字通信においてプライバシーを保つ手段などに利用できる。

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.

[キーワード]
情報ハイディング，電子透かし，ステガノグラフィ，ドキュメント，自然言語処理
Information hiding, Digital watermarking, Steganography, Document, Natural language processing

1　まえがき

計算機ネットワークの利用拡大に伴い、ネットワーク上で情報を安全に伝送する情報セキュリティ技術が重要になってきている。情報セキュリティ技術の一つである暗号は、攻撃者に解読されないように、あるいは改ざんされたらその検出ができるように、情報を加工・復元する技術であり、情報が隠されているという事実を隠すことは必ずしも目的としていない。そのため、通信路上で暗号通信を見つけることは容易であり、攻撃者は暗号を解読できなくても、重要と思われる暗号通信(暗号化されていることは、その通信内容が重要なものであることを示唆する)を見つけ出して妨害することは可能である。そのような不正行為への対抗策としては、情報が埋め込まれていること自体を隠す技術である情報ハイディングが有効である。情報ハイディングは、情報伝送に際してのカムフラージュ手段だけでなく、画像や音楽などの著作物に著作権情報や配布先情報を埋め込む手段としても利用できる。本論文では、デジタルドキュメントを埋め込み媒体とし、文書内に挿入された改行コードの位置を秘匿情報とする情報ハイディング手法について述べる。

2　ドキュメントに対する情報ハイディング[1]

2.1　情報ハイディングとは

情報ハイディングは、情報伝送に際しての秘匿通信すなわちカムフラージュ手段として、あるいは画像や音楽などの著作物(コンテンツ)に著作権情報や配布先情報などの権利主張のための情報を埋め込む手段としての応用が考えられる。秘匿通信としての用途の場合は「ステガノグラフィ」(steganography)、権利主張の用途の場合は「電子透かし」(digital watermarking)と呼ぶ。

情報ハイディングは、コンテンツ(カバーデータと呼ぶ)に対して、秘匿メッセージや著作権情報など(エンベデッドデータと呼ぶ)を埋め込み、埋め込み済のコンテンツ(ステゴデータ)を作る処理である。伝送されるのはステゴデータで、受信者はステゴデータからエンベデッドデータを取り出して利用することになる。ステガノグラフィの場合、エンベデッドデータが主体であり、カバーデータは秘匿通信のためのカムフラージュに過ぎない場合が多い。それに対して電子透かしの場合、主体はカバーデータ(著作物)であり、カバーデータに関する付帯情報をエンベデッドデータとして埋め込む。したがってステガノグラフィの場合は、多くのエンベデッドデータを埋め込むことに力点が置かれ、電子透かしの場合は、カバーデータとステゴデータの差異ができるだけ小さいこと(すなわちコンテンツの変質が少ないこと)に力点が置かれる。

2.2 ドキュメントに対する情報ハイディングの特徴と分類

カバーデータとしてドキュメント(文書)を用いる情報ハイディングは、第三者が気づかない作為をドキュメントに施すことによって情報を埋め込み、正当な者のみがそのドキュメントから秘匿情報を抽出できるようにすることを目指す。

有史以来使われてきた古典的な情報ハイディングは、もともとドキュメントを媒体とするものが多くを占めていた。現代においては、第三者による傍受や検閲等の脅威に対抗することなどを想定したステガノグラフィ(秘匿通信)としての応用がまず考えられる。ドキュメントにおけるステガノグラフィは、第三者が通常の通信とみなすデータに実は第三者の目を逃れる秘密の情報が埋め込まれているモデルである。ステガノグラフィと共に、ドキュメントへの情報ハイディングの有力な応用分野と考えられているのが、電子的コンテンツに対して著作権情報やフィンガープリントを埋め込む電子透かしである。これは、コンテンツを正当に入手した人や組織を特定できる情報などをコンテンツに埋め込んでおくもので、不正な2次配布をした場合に流出元を特定できることによって、海賊版の流布に対する抑止効果が期待できる。

ドキュメントへの情報ハイディングにおいて考慮しなければならないのは、カバーテキスト(カバーデータとしてのテキスト)の改変をどれくらい許容するかである。小説などカバーテキストそのものが著作物である場合、改変は全く許容できないと言っていい。一方、著作権を主張する主な対象がソフトウェアや画像あるいはビデオなどであり、その付属物であるドキュメントに著作権情報を埋め込む利用法の場合、例えば、ソフトウェアならマニュアルや使用許諾書のような添付文書をカバーテキストとして情報を埋め込む場合には、ステゴテキスト(ステゴデータとしてのテキスト)はカバーテキストの意味が保存されていればよく、したがって文面に若干の変更が加えられても許容できる場合がある。さらに、秘匿されている情報に重点が置かれ、ステゴテキストがカムフラージュに過ぎないステガノグラフィの場合、機械による自動的な検閲を逃れる目的であれば、ステゴテキストは意味のある文書になっていなくても構文的に正しくさえあればいい場合もある。

情報ハイディングは、カバーデータの冗長性を利用して情報を埋め込む技術であるので、ドキュメントのどの側面の冗長性を利用するかによって、幾つかの方式に分類できる。分類の際、情報秘匿のための作為がハードコピー上(あるいはディスプレイの画面、以下同じ)に残る方式と、残らない方式の二つに大別して考えると理解しやすい。ハードコピー上に残るか残らないかの違いは出力系に依存するので、必ずしも厳密な分類とはいえないが、説明上は都合がいいので、以下では一般的な出力系を想定して両方式を概観する。

(1) 作為がハードコピー上に残る情報ハイディング

ハードコピー上に残る方式は、作為が目視確認できるはずだが気付きにくいことを利用するもので、電子データとしてだけでなくハードコピーとしての流通にも使える特長がある一方、見破られないように作為の仕方を工夫する必要がある。この方式は、気付きにくさの原理によって、更に以下の二つのタイプに細分類できる。

① 作為が見えにくいことを利用するタイプ

カバーテキストとステゴテキストとを並べて目視比較しても見分けられない程度の、微小な作為を施すことによって、見破られることなく情報を埋め込むことを目指す。その実現方法として、文書レイアウトへの作為が考えられる。これは、ポストスクリプト機能等を活用して、文書レイアウトに微小な作為を施し、ハードコピーとして印刷されたステゴテキストをスキャナで読み取って秘匿情報を抽出する手順が基本となる。文字情報そのものは埋め込みと抽出の両場面で重要ではなく、ドキュメントの画像としての情報の差異を利用する。そのため、これは画像への情報ハイディングの一特殊形とみなすこともできる。ハードコピーとして利用する場合、複写を繰り返して画像が劣化することにより、秘匿情報も劣化消失することが、このタイプの弱点といえる。ハードコピーを介さずデータ内に秘匿した情報を電子データのまま受け取って抽出するモデルもあり得るが、その場合はそもそもレイアウトに作為を施す必要はなく、したがって後述する XML や LaTeX 文書への情報ハイディングなどと同類とみなせる。

レイアウトへの作為の施し方としては、行間隔あるいは語間隔の拡大縮小や、文字幅の拡大縮小あるいは文字の回転などが提案されている。例えば行間隔の標準画素数を定めておき、ビット“1”を埋め込むと間隔が拡大し、“0”を埋め込むと狭まるとする方法である。したがって秘匿情報の抽出性能はスキャナの読み取り解像度に依存することになるので、拡大縮小の程度を小さくすればより作為に気付かれにくいが、半面、抽出エラーも増加することになる。どの作為が気付かれにくいかは言語に依存し、例えば英語などの欧文では、語間隔の拡大縮小が有利で、日本語のような語間にスペースが挿入されない言語では、フォントの拡大縮小と回転への作為が有利とされている[2]。また、秘匿情報の抽出に際して、原本であるカバーテキストとの比較照合を必要とする手法としない手法とがある。レイアウトへの作為を用いる各種手法については、文献[3]に多数紹介されている。

レイアウトへの作為以外に、ドキュメントの周辺や罫線などに極めて小さい文字や記号を隠しておく方式も、このタイプに属する。また、筆跡の座標や筆圧に作為を施して情報を秘匿する手書きステガノグラフィ[4]も、ドキュメントへの情報ハイディングとみなすならば、このタイプに属すると言えよう。

② 作為が自然なため気付かないことを利用するタイプ

デジタルドキュメントは基本的に、文字列とレイアウト情報とからなる。文字はそれ自体が意味の一部を成すため、デジタル情報としての文字へ無配慮に作為を行うと、わずかな作為であっても文字化けなどを起こし、意味にまで波及することでドキュメントの品質が大きく損なわれ、また作為が露見する恐れが高まる。そのため、ドキュメントへの情報ハイディングは従来、前述した文書レイアウトへの作為に分類される手法が多く提案されてきた。しかし電子メールのようにレイアウト情報を持たないプレーンテキストに情報を埋め込む場合は、文字への作為のみに頼る必要がある。この場合、作為の見えにくさによるカムフラージュを目指すことは断念し、ステゴテキストだけを観察した場合に不自然さに気付かれないことでよしとする戦略をとる。この方式ではカバーテキストとステゴテキストとを比較すると作為が露見するため、カバーテキストを公開しない利用モデルが想定される。本タイプは作為がかなり大きいため、ハードコピーとして利用する場合に複写を繰り返しても秘匿情報が劣化消失しにくい特長がある。

文字への作為において、ドキュメントの変質を避けるためには、単語の置き換えなど自然言語処理を応用した方法と、文面に影響しない文字あるいは文字コードを挿入する方法とが考えられる。前者の手法については文献[5]などの研究がある。後者の手法に属するのが、本論文で述べる手法であり、次節以降で説明する。

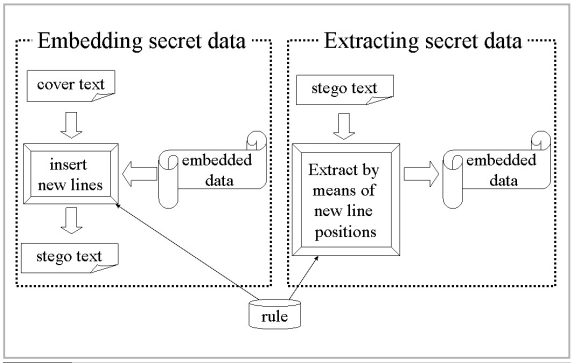


図1 改行位置の調整による情報ハイディングにおける処理の流れ

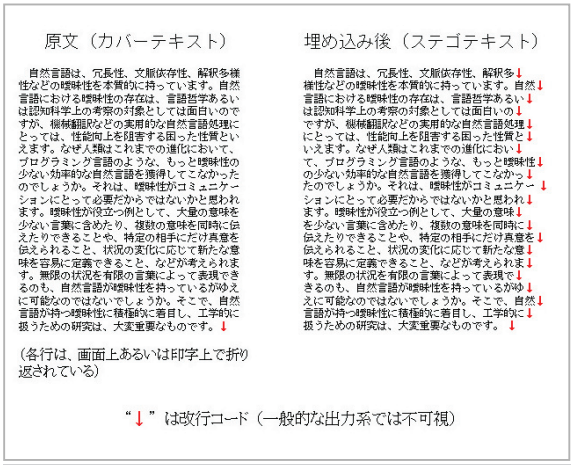


図2 改行位置の調整による情報ハイディングにおけるカバーテキストとステゴテキストの例

なお、元となるカバーテキストが存在せず、埋め込みデータに従ってステゴテキストを無から生成する方式も、このタイプに属する。そのような方式として、uuencode ファイルや PGP メッセージを、あたかも詩のような英文に変換する“Texto”や、バイナリデータを指定された文体の英文に変換する“NICETEXT”などのツールが提案されている[6]。

⑵ 作為がハードコピー上に残らない情報ハイディング

ハードコピー上に残らない方式は、見た目に作為が全く識別できないため見破られにくいものの、電子データから表示メディア（紙、画面など）に変換された時点で秘匿情報が消去されるので、秘匿情報の抽出時まで電子データのままで扱う利用法が前提となる。

この方式としては、英文をカバーテキストとし、複数の空白文字を各行末に挿入することにより情報を埋め込む“SNOW”と呼ばれる手法[6]などが提案されている。SNOW では、秘匿情報はハフマン符号化により圧縮して暗号化した後、行末に 0〜7 個の空白を挿入することによって 1 行当たり 3 ビットの情報を埋め込む。また、ヌルキャラクタをモールス信号にのっとってテキストデータの中に配する FFEncode というツールも提案されている[6]。さらに、英文の LaTeX 文書をカバーテキストとし、ソース中の本文の各行の単語の個数を加減することにより、情報を埋め込む手法も提案されている[7]。XML などの構造化文書への埋め込みも、基本的に作為がハードコピー上に残らない方式に属する[8]。

2.3 改行位置の調整による情報ハイディングの位置付け

本論文で扱うのは、ドキュメントの改行する場所をコントロールすることにより情報を埋め込む方式である[9]。この方式は、日本語のように、改行する位置が比較的自由な言語を対象としている。本手法は、ワープロ文書のように、段落（パラグラフ）の末尾にのみ改行コードが入ったベタテキストを埋め込み媒体（カバーテキスト）とすることを想定している。この方式によるエンベデッドデータの埋め込み及び抽出処理の流れを図 1 に示し、図 2 にカバーテキストとステゴテキストの例を示す。適当な長さごとに改行コードを入れることによってエンベデッドデータを埋め込んだ結果、改行が多数挿入された文書（ステゴテキスト）が生成されるというものである。改行を挿入する際に、1 行当たりの行幅（各文字の字幅の合計）の変動を小さくしてドキュメントの見た目の不自然さを少なくする戦略と、単語の途中などの不自然な位置での改行をなるべく回避する戦略とのトレードオフを考慮し、最も不自然さが少ない方法を考える必要がある。

改行位置の調整による情報ハイディングは、文面に全く影響を及ぼさないため、改変困難な著作物をカバーテキストにする場合にも適用できる。またこの方式は、プレーンテキストへの文字レベルでの作為であると同時に、改行位置というレイアウト上の作為にもなっているといえる。

3　改行位置の調整による情報ハイディング

3.1　はじめに

改行位置の調整による情報ハイディングでは、改行位置とエンベデッドデータとの対応付け、すなわち図1におけるルールが重要である。このルールについては、単語中の改行位置による方法と、1行文字数による方法とが考えられる。以下ではそれぞれについて述べる。

3.2　単語中の改行位置による方法

単語中の改行位置による方法では、形態素解析辞書の見出し単語を対象に、各単語（形態素）中の改行位置と、埋め込み情報のビット（0又は1）との対応関係に基づき情報を埋め込む。例えば図3に例示するように、形態素「する」を「す｜る」と改行したら“1”などとあらかじめ定めておく（"|"は改行位置）。その際に、ステゴテキストの見た目の自然さを保つために、各行の文字密度の均一さを重視し、1行当たりの幅（行幅すなわち各文字の字幅の合計）がなるべく均一になるようにする。そのため、各文字の幅について、1バイト文字を1、かな漢字などの2バイト文字を2と定義する。埋め込み処理時に指定する基準行幅に従い、行末の近傍に来た単語を埋め込み対象とする。図3に示すように、「プログラミング」や「コミュニケーション」などの長い単語は、複数の改行位置を0, 1に対応させておき、どれを選んでもいいようにしておく。こうすることで基準行幅から大きくかけ離れない文字数で改行できる。

図3の対応表を用いて情報を埋め込んだ例を図4に示す。情報を埋め込んだ単語（形態素）を下線で示している（下線は実際には非表示）。図4は均等割付をしたものであるが、行幅のバラツキはほとんど気づかれない程度であることが分かる。図4の例では、“01111101011…”が埋め込まれた情報（エンベデッドデータ）となる。

本節で説明した方法は、以下の特長を持っている。

⑴　字種（ひらがな/カタカナ/漢字）による切り分けを行えば、形態素解析を使わず軽い処理が可能。
⑵　単語単位で埋め込み方を定義できるため、後述する1行当たりの文字数による方法と比較して、埋め込み情報のビットと改行との対応関係の法則性を見破ることが困難であり、したがって抽出攻撃に強い。
⑶　単語ごとに改行位置を定義できるため、不自然な位置での改行を回避することが可能。

一方、課題としては、形態素解析処理の誤りへの対処、一文字形態素への対処などがある。

3.3　1行当たりの文字数による方法

本節で説明する方法では、各行の文字数と埋

0	1	
する｜	す｜る	（動詞-自立 サ変・スル）
プログラミング｜	プログラミン｜グ	（名詞-サ変接続）
プロ｜グラミング	プログ｜ラミング	（名詞-サ変接続）
言｜語	言語｜	（名詞-一般）
獲得｜	獲｜得	（名詞-サ変接続）
コミュニケーション｜	コミュニケーショ｜ン	（名詞-一般）
コミュニケ｜ーション	コミュニケー｜ション	（名詞-一般）
コミュ｜ニケーション	コ｜ミュニケーション	（名詞-一般）
役立つ｜	役｜立つ	（動詞-自立 五段・タ行）
と｜して	として｜	（助詞-格助詞-連語）
同時に｜	同時｜に	（副詞-一般）
こと｜	こ｜と	（名詞-非自立-一般）
考｜え	考え｜	（動詞-自立 一段）
言語｜	言｜語	（名詞-一般）
そこで｜	そこ｜で	（接続詞）
研｜究	研究｜	（名詞-サ変接続）

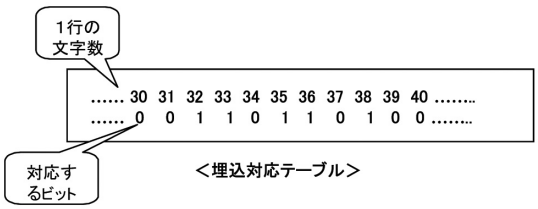
図3　形態素ごとのビット対応表の例

（形態素は参考文献[10]の付属辞書に基づく）

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。
....

図4　本方法により情報を埋め込んだ例

（右端の数字は埋め込まれたエンベデッドデータ（実際は非表示））



ｼﾌﾄJISなどのように1ﾊﾞｲﾄｺｰﾄﾞと2ﾊﾞｲﾄｺｰﾄﾞが混在する場合、1ﾊﾞｲﾄｺ　0(40)
ｰﾄﾞを「半角」、2ﾊﾞｲﾄｺｰﾄﾞを「全角」とする実装が広く行なわれて　1(33)
いる。文字幅もそれに対応しているように見える場合が多いため、　0(30)
「半角」、「全角」ということばが文字幅の意味として混同されて　0(30)
使用されている場合があるが、文字ｺｰﾄﾞのﾊﾞｲﾄ数と文字幅の間には　1(32)
なにも取り決めがない。したがって、文字ｺｰﾄﾞと文字幅については　0(31)
各文字ｺｰﾄﾞ体系毎に文字ｺｰﾄﾞと、（ﾊﾞｲﾄ数、）文字幅の関係を定義　1(33)
することが必要となる。

<出力結果>

図5　埋め込みビットとの対応表と、ステゴテキストの例

（右端の太数字は埋め込まれたビット、括弧数字は各行の文字数）

め込みビットとの対応表を定義しておく。そして、埋め込もうとするエンベデッドデータのビットに対応する文字数になる位置で改行コードを挿入していく。その際に、基準行幅がなるべく均一になるように処理する。抽出時には、各行の文字数をカウントし、同じ対応表を用いてエンベデッドデータを抽出する。つまりこの方式では1行につき1ビットの情報を埋め込むことになる。図5に、各行の文字数と埋め込みビットとの対応表を用いて情報を埋め込んだ例を示す。

図5の例は、行幅を均一にするため、1行目を40文字、2行目を33文字などとして、“0100101…”を埋め込んだ例である。

本手法は、前節で述べた単語中の改行位置による方法のような形態素ごとのビット対応表との照合を必要としないため、処理が速く誤処理が少ない。反面、埋め込み方の法則性が平易なので、抽出攻撃の危険性が高い問題がある。

4　実装

4.1　はじめに

本章では、**3.3**で述べた、1行当たりの文字数に応じて1ビットのエンベデッドデータを埋め込む情報ハイディングツールを実装した結果について述べる。実装したツールは、プレーンテキストをカバーテキストとし、埋め込もうとする0か1のビット列（秘匿情報を暗号化したエンベデッドデータ）に従って改行コードを入れてい

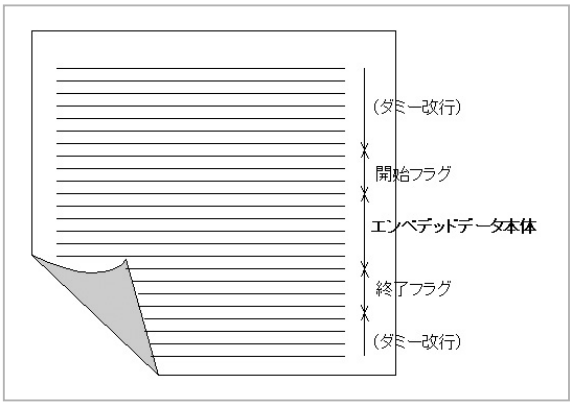


図6　方式 A1 の場合の埋め込み方

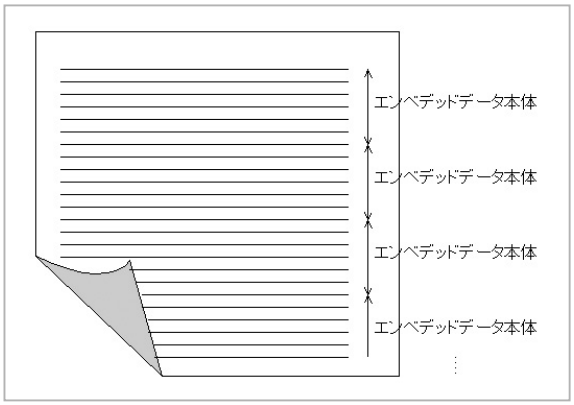


図7　方式 A2 による埋め込み方

き、改行が多数挿入された文書（ステゴテキスト）を生成するツール及びその文書から秘匿情報を抽出するツールである。開発言語は、開発環境、今後の拡張性、暗号化アルゴリズムの利用などを考慮し、JAVA 言語を用いた。エンベデッドデータは、秘匿情報を RC4（鍵長 40ビット）で暗号化したものとし、解読攻撃への対処を図っている。また、鍵である埋め込み対応テーブルが容易に推測されないよう、乱数によりテーブルを作成する機能を持ち、抽出攻撃への対処を図っている。乱数は、JAVA が提供する乱数生成ジェネレータ Random () を利用している。

4.2　埋め込み方式

実装したツールは、エンベデッドデータの配置方式により 2 種及び改行位置の決定方式により 3 種の組合せで合計 6 通りの埋め込み方を選択できるようになっている。以下では、それぞれの埋め込み方の詳細を説明する。

（A）エンベデッドデータの配置方式

秘匿情報を各行の文字数に対応させて文書に埋め込む当ツールでは、ステゴテキストの中でエンベデッドデータが埋め込まれている行を、抽出に際して同定する仕組みを講じる必要がある。筆者らの実装では、埋め込まれている範囲を示すフラグを用いる方式 A1 と、カバーテキストの冒頭から繰り返し埋め込む方式 A2 の 2 種類を実装した。それぞれについて以下で説明する。

【方式 A1】開始・終了フラグに挟んで埋め込む

方式 A1 では、カバーテキストの途中から、開始フラグ＋エンベデッドデータ＋終了フラグの順番で 1 回だけ埋め込む。カバーテキストの冒頭から開始フラグまでの改行はダミー改行とし、情報は埋め込まない。埋込開始位置及び埋込開始行以前の改行位置は処理ごとに乱数で決めているため、同じ入力であっても出力結果は処理ごとに異なるようになっており、抽出攻撃への対処を図っている。図 6 に、本方式によるカバーテキストへの埋め込み方の概念図を示す。

本方式では、埋め込み処理時に、埋め込み対応テーブル、基準行幅、最小行幅、暗号（復号）鍵、開始フラグ（8bit バイナリ）、終了フラグ（8bit バイナリ）及び最大埋め込み開始行をパラメータとして指定する。また抽出処理時には、埋め込み処理時に使用したものと同じ埋め込み対応テーブル、最小行幅、暗号（復号）鍵、開始フラグ及び終了フラグを指定する。最小行幅は、指定した行幅未満の行には情報を埋め込まないようにするために指定するもので、これは、段落の末尾やキャプションのように、他の行とは行幅が著しく異なる部位を埋め込み対象外とするために必要である。最小行幅は埋め込みと抽出の両処理において必要なパラメータとなる。また最大埋め込み開始行は、開始フラグまでのダミー改行の最大行数を指定するもので、埋め込み処理に際してはこの値以下のランダムな行数が自動的に指定される。最大埋め込み開始行は、埋め込み処理においてのみ必要なパラメータとなる。

本方式によれば、情報が埋め込まれている場所を攻撃者が検出することは困難と考えられる。しかし、エンベデッドデータは 1 回だけしか埋め込まれていないため、ステゴテキストに対す

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。

図8 方式 B1 によるステゴテキストの例

る部分的な切り出し編集が行われた場合の耐性(エンベデッドデータの保存性)は弱いといえる。また抽出の際には、共通鍵である埋め込み対応テーブルや暗号(復号)鍵のほかに、開始・終了フラグの情報も必要とする。

本方式は、カバーテキストのサイズと比較してエンベデッドデータが相対的に大きくて繰り返しの埋め込みが難しい場合や、文書の切り出しが行われる可能性の少ない文書への埋め込みに適していると言える。

【方式 A2】繰り返し埋め込む

方式 A2 では、カバーテキストの冒頭からすべての改行にエンベデッドデータを繰り返し埋め込む。そのためダミー改行はなく、開始・終了フラグも不要である。図 7 に、本方式によるカバーテキストへの埋め込み方の概念図を示す。

本方式では、埋め込み処理時に、埋め込み対応テーブル、基準行幅、最小行幅及び暗号(復号)鍵をパラメータとして指定する。また抽出処理時には、埋め込み処理時に使用したものと同じ埋め込み対応テーブル、最小行幅及び暗号(復号)鍵を指定する。本方式はエンベデッドデータを冗長に埋め込むため、検出時にデータの冒頭を正しく同定する手立てを講じれば、ステゴテキストが切り出し編集されてもエンベデッドデータを正しく抽出できる確率が高いと考えられる。しかし繰り返しパターンを手がかりとして、埋め込み対応テーブルを見破られる可能性が高い

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。

図9 方式 B2 によるステゴテキストの例

懸念がある。

(B) 改行位置の決定方式

改行位置の決定方式については、行幅の均一性と改行位置の自然性とのトレードオフを考慮し、3 通り実装した。それぞれについて以下で説明する。なお、以下に示す例ではいずれも、エンベデッドデータの配置方式として方式 A1 を用いているが、方式 A2 と組み合わせることも可能である。

【方式 B1】行幅の均一性を重視

方式 B1 は、句読点等の禁則処理の制約以外は、基準行幅の付近で、できるだけ幅のばらつきが少ないように改行する方式である。禁則処理は、MS-Word における標準的な行頭及び行末の禁則処理のルールに準じている。図 8 に、本方式による出力結果の例を示す。

この方式では、各行の行幅のばらつきが小さいため、ページデザイン的には自然に見える。しかし単語の途中など不自然な位置での改行が多いため、文章的には違和感を生じさせる場合がある。

【方式 B2】特定の文字種における改行を制限

本方式は、方式 B1 における制約に加えて、特定の文字列内(数字及びアルファベット)での改行を避ける方式である。図 9 に出力結果の例を示す。

図 9 の例では、“representation”のようなアルファベット列は途中で改行されず、他の行より

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。

図10 方式 B3 によるステゴテキストの例

もやや行幅が長くなっていることが分かる。そのため方式 B1 よりも行幅のばらつきが大きくなっている。

【方式 B3】文字種の境目を強く重視

方式 B3 は、方式 B2 の制約を更に強め、漢字、ひらがな及びカタカナの文字列内での改行も避け、その上、括弧内の改行位置も制限する方式である(括弧に囲まれた文字が 5 字以下の場合はその間に改行を入れない)。したがって、改行位置の大部分は文字種(漢字/ひらがな/カタカナ/数字/アルファベット)の境目になる。日本語では、文字種の境目(ひらがなと漢字、カタカナとひらがな等)が文節の境界であることが多いため、本方式により、文節単位での自然な改行が多くなる。図 10 に、本方式による出力結果の例を示す。

図 10 では、おおむね文節単位で改行されているように見え、文章として読みやすい改行になっていることが分かる。その代わり、方式 B2 よりも更に行幅がばらついている。

5 評価

5.1 はじめに

情報ハイディング手法の評価の観点としては、(1)埋め込める情報量の多さ、(2)情報が埋め込まれていることの見破られにくさ、(3)埋め込まれた情報の抽出されにくさ、(4)埋め込まれた情報

表1　主観評価実験の分類

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)

表2　評価に用いたカバーテキスト

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」

の破壊されにくさ、などが考えられる。(1)については、埋め込み率の定量的な評価が行えるが、(2)(3)(4)については、攻撃者の行動に関する評価であるため、被験者を用いた主観評価を行う必要がある。本章では、(2)(3)(4)の主観評価についてそれぞれ検討する。

(2)(3)(4)の主観評価において、筆者らは、(2)情報が埋め込まれていることの見破られにくさはステゴテキストの自然性と等価である、と考えている。また、(3)埋め込まれた情報の抽出されにくさと、(4)埋め込まれた情報の破壊されにくさ(無効化攻撃への耐性)は、それぞれ情報秘匿の安全性と強度によって評価されるものとする。以上より、主観評価ではステゴテキストの自然性及び情報秘匿の安全性・強度の2通りを対象とし、それぞれについて**4**で延べた(A)エンベデッドデータの配置方式あるいは(B)改行位置の決定方式と、表2に示したカバーテキストのジャンルの2通りを組み合わせることによって主観評価実験を行い、それぞれの結果を比較することが妥当と考えられる。それぞれの分類を表1に整理する。エンベデッドデータの配置方式の違いは、埋め込まれた情報の抽出あるいは破壊に際してのみ影響を与えるものと考えられるので、情報秘匿の安全性及び強度の評価においてのみ条件として加える。また、カバーテキストのジャンルによるステゴテキストの自然性の比較評価(**5.3.2**)については、実験の手順についても詳細に述べることにする。

ただし、主観評価実験については今後更に詳細化及び改良を加える必要があると考えている。そのため**5.3**及び**5.4**の主観評価実験については、本論文では実験手順の概要を述べるにとどめることにする。

5.2　評価に用いたカバーテキスト

評価に用いたカバーテキストを表2に示す。カバーテキストの性質によって主観評価に影響

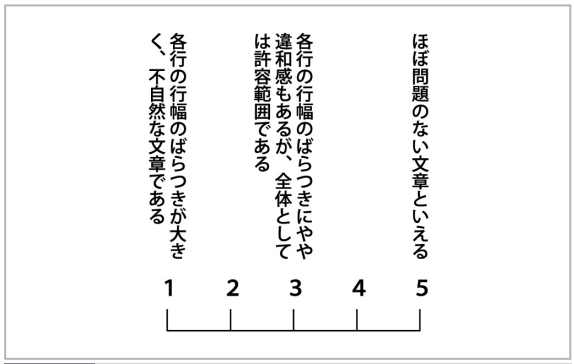


図11 評価基準

を与えることが予想されたので、ニュース記事、論文、文学作品など、多様なテキストを評価対象とした。

5.3 情報が埋め込まれていることの見破られにくさの主観評価

5.3.1 改行位置の決定方式の違いによるステゴテキストの自然性の評価

4.2(B)で述べた3種類の改行位置の決定方式の違いによって、生成されたステゴテキストの自然性が受ける影響について評価する。被験者群については特に条件を付けず、人数は5〜10人程度とする。そして、同一のカバーデータに対して異なる改行位置決定方式で生成したステゴテキストを被験者に紙又は電子媒体で配布し査読してもらい、それぞれのステゴテキストに対してその評価を、図11に示す5段階の評価基準から選択してもらう。

5.3.2 カバーテキストのジャンルによるステゴテキストの自然性の評価

カバーテキストのジャンルの違いが、生成されたステゴテキストの自然性に与える影響について評価する。被験者群は**5.3.1**と同じく、特に条件を付けず、人数 は5〜10人程度とする。そして、ジャンルの異なるカバーデータに対して同一の方式で生成したステゴテキストを被験者に紙又は電子媒体で配布し査読してもらい、それぞれのステゴテキストに対してその評価を図11に示した5段階の評価基準から選択してもらう。

以下では、実験手順の詳細について述べる。

⑴ 事前準備

表2のカバーテキストに対して、**4**で述べたツールを用いて、同一のエンベデッドデータを埋め込んだステゴテキストを生成する。本実験では、改行位置の決定方式については、行幅の均一性を重視する方式B1に固定し、エンベッドデータの配置方式については、エンベデッドデータを繰り返し埋め込む方式A2に固定する。それぞれ1種類ずつに限定する理由は、限られた被験者数の中で、ジャンルが自然性評価に与える影響のみを浮き彫りにするためである。また、1行当たりの文字数とエンベデッドデータのビット値との対応関係は、偶数の場合は1、奇数の場合は0とする単純な関係とする。

⑵ 実施手順

①実験シート及び評価シートの配布

被験者に実験シート及び評価シートを紙媒体又は電子媒体で配布する。実験シートの例を図12に、評価シートの例を図13に示す。

②評価作業マニュアルの配布

実験管理担当者が被験者に対し、図14に示す「評価作業マニュアル」を配布し、その内容について説明を行う。さらに、評価作業前に一読するよう被験者に指示する。

③被験者による評価作業の実施

配布された評価作業マニュアルに従って被験者が評価作業を行う。

④実験データの回収

実験管理担当者が、評価作業を終えた被験者より実験シート、評価シート及び評価作業マニュアルを回収する。

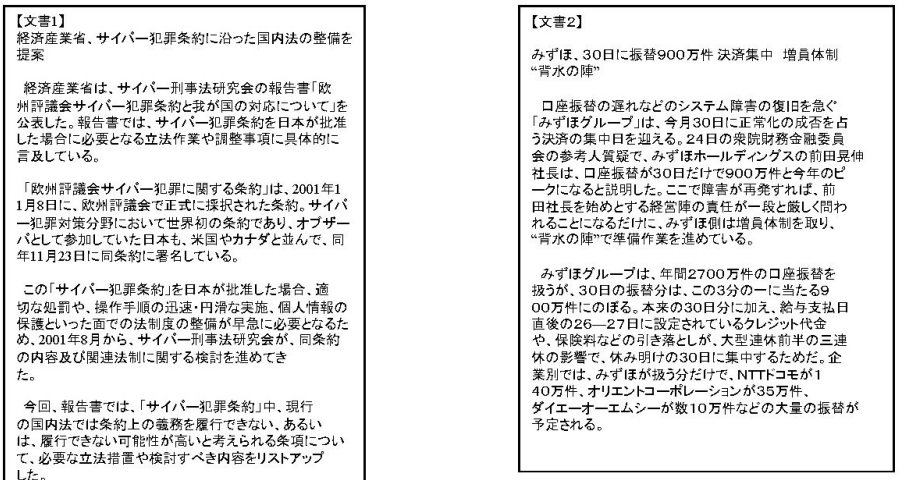
⑶ 実験結果の分析と評価

表2に示した(子供向けニュースを除く)カバーテキストの各ジャンルにつき複数のドキュメントを用いて実験を行う。したがって、ジャンルごとの評価点数の差異と、ドキュメントごとの評価点数の差異とを切り分けた分析が可能になる。評価点数の集計は以下のとおりとする。

(Ⅰ) ジャンルごとの評価分布及び平均評価点数

(Ⅱ) ドキュメントごとの評価分布及び平均評価点数

上記の集計結果に基づき、Ⅰの結果から、ジャンルの違いがステゴテキストの自然性の評価



【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11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。

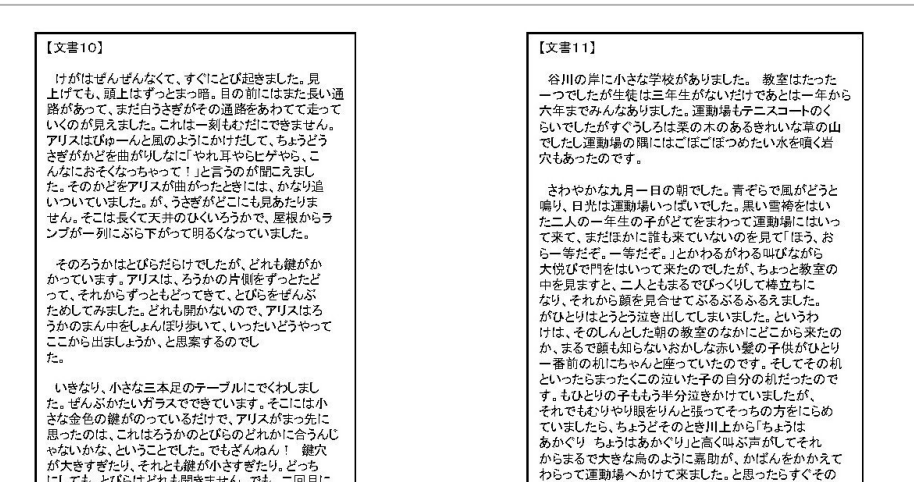
【文書2】

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900万件にのぼる。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140万件、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。

図12a　実験シート例（シート番号１　一般ニュース）



【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センチぐ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、おらー等だぞ。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って来ました。

「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。

図12b　実験シート例（シート番号６　子供向け文学）

評価シート

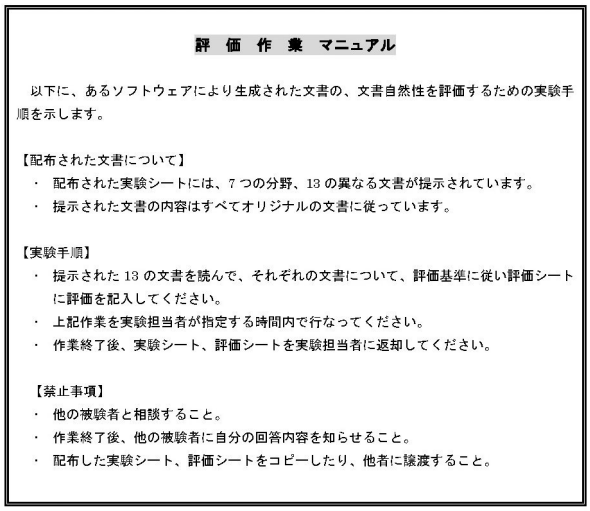
実験シート（その１）～（その７）の１３の文書について、それぞれの文書の自然性に対する主観評価を評価基準を参考に1～５の数字で記入してください。

【評価基準】

1 各行の行頭のばらつきが大きく、不自然な文章である
2
3 各行の行頭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	

図13 評価シート例



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。

図14 評価作業マニュアル

に与える影響について分析し、またⅡの結果から、ドキュメントの個体差がステゴテキストの自然性の評価に与える影響について分析する。

5.4 情報秘匿の安全性及び強度に関する主観評価

5.4.1 エンベッドデータの配置方式あるいは改行位置の決定方式の違いによる情報秘匿の安全性及び強度の評価

4.2(A)(B)で述べた 2 種類のエンベッドデータの配置方式の違い又は 3 種類の改行位置の決定方式の違いによって、改ざんに対する耐性が受ける影響について評価する。被験者群は、暗号技術への関心度が高いと思われる情報工学系の大学生及び大学院生とし、人数は 5～10 人程度とする。そして、エンベッドデータの配置方

式(2 通り)と改行位置の決定方式(3 通り)の各組合せによる計 6 通りによって作成された複数のステゴテキストを電子媒体で配布し、情報が埋め込まれていると思われるテキスト(複数可)に対して意味を損なわない範囲で自由に改ざんを行ってもらう。

5.4.2　カバーテキストのジャンルによる情報秘匿の安全性及び強度の評価

カバーテキストのジャンルの違いが、改ざんに対する耐性に与える影響について評価する。実験を以下のとおり計画した。被験者群は、5.4.1 と同じく、暗号技術への関心度が高いと思われる情報工学系の大学生及び大学院生とし、人数は 5〜10 人程度とする。そして、被験者にカバーテキストのジャンルが異なる複数のステゴテキストを電子媒体で配布し、情報が埋め込まれていると思われるテキスト(複数可)に対して意味を損なわない範囲で自由に改ざんを行ってもらう。

6　考察

本論文の冒頭で述べたように、情報ハイディングは、電子的コンテンツに対して著作権情報やフィンガープリント(配布先の個人識別情報)を埋め込む「電子透かし」と、第三者による傍受や検閲等の脅威に対抗することなどを想定した「ステガノグラフィ」(秘匿通信)の、大きく二つの応用が考えられる。本論文で述べたドキュメントへの情報ハイディングは、第三者による改行位置の付け替えが施されにくい、電子メールのような 2 者間の直接の文書交換や、あるいは印字文書を応用として想定するのが良いと考えられる。例えば、取扱注意文書を印字して関係者限定で配布する際に、文章の内容を全く改変することなく改行位置だけによって全文にわたってフィンガープリントを埋め込むと、流出を意図する者にとって流出元を隠す加工が紙上では容易ではないことから、安易な流出を阻止できる。印字文書を媒体とする場合、秘匿情報の抽出には、従来から多く提案されているレイアウトへの情報ハイディングと同じく OCR を用いることになるが、認識すべきなのは、従来のような行間サイズ、字間サイズ、マイクロ文字などの微細な情報ではなく、各行の行幅(各文字の字幅の合計)という目立つ情報なので、度重なる品質の悪いコピーを経ても秘匿情報が消える恐れが少ない点で優れているといえる。

ステガノグラフィあるいは電子透かしとして利用する場合の留意点について考察する。ステガノグラフィの場合、秘匿情報の伝送に主目的が置かれるので、ステゴテキストはカムフラージュに過ぎず、したがって電子データとしての流通に際して機械による自動的な検閲を逃れる目的であれば、ステゴテキストは自然言語的な文章が連なってさえいればよく、文書としての意味を持っていなくてもいい場合もある。それに対して電子透かしとして利用する場合、小説など微細な表現にも重要な意味を持つ著作物をカバーテキストにするならば、テキストの改変は全く許容できないし、取扱注意文書やマニュアルのような意味内容重視の文書の場合にも、意味を変えない程度の微細な改変しかできない。その点、開発したツールは、改行位置以外の改変を全く施さないものであるため、ステガノグラフィとしても電子透かしとしても利用できる。

ステガノグラフィとして利用する場合、情報が埋め込まれていること自体を隠しておくことが特に重要な要件になるので、ステゴテキストの見た目の自然性、すなわち行幅の均一性や改行位置の自然性を確保する手立てを講じる必要がある。そのためには、改行位置の決定方式の最適化や、均等割付など表示あるいは印刷時のレイアウト機能の利用が有効である。

ステガノグラフィあるいは電子透かしのどちらとして利用する場合であっても、解読、抽出、改ざん、なりすましへの対処は必要である。本論文で述べた方法では、埋め込み対応テーブルのランダム化や秘匿情報の暗号化を講じているが、さらに誤り訂正の利用なども考えられる。また、電子データとしての流通を想定する場合、ステゴテキストの部分的な切り出しや改行位置の付け替えによる無効化攻撃への対処も講じておく必要がある。エンベデッドデータの配置方式については2方式を用意しているので、エンベデッドデータを冗長に埋め込んでおくこと(方式 A2)や、埋込位置をランダム化しておくこと(方式 A1)などの手立てが、ある程度は有効である。

本論文で述べた手法は、情報ハイディングだけでなく、文書の改ざん検出にも応用できる。すなわち、テキスト文書のハッシュ値やメッセージ認証子(MAC)などの検証用データをエンベデッドデータとして本手法によってテキスト文書に埋込み、検証時にこれを抽出してステゴテキストと照合することで、改ざんの有無を検出できる[11]。

7 むすび

本論文では、デジタルドキュメントを埋め込み媒体とし、文書内に挿入された改行コードの位置を秘匿情報とする情報ハイディングについて述べた。マルチメディア化が進んでいる現代においても、電子メールなどテキストによる情報交換はいまだ主流の位置を占めており、情報伝達手段としてのドキュメントの重要性は今後も変わらないと考えられる。したがって、ドキュメントへの情報ハイディングには、今後多くの応用が期待できる。

謝辞

本研究は、横浜国立大学の松本勉研究室、東京大学の中川裕志研究室、株式会社三菱総合研究所のメンバーとの定期的な意見交換により進められている。有益な助言を頂いた各位に感謝する。

参考文献

1 中川裕志，滝澤修，井上信吾，“ドキュメントへのインフォメーションハイディング”，情報処理，Vol.44，No.3，pp.248-253，2003.
2 松井甲子雄，“電子透かしの基礎”，森北出版，1998.
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 瀬川典久，村山優子，宮崎正俊，“手書き入力装置の特性を利用した手書きステガノグラフィの提案”，情報処理学会コンピュータセキュリティシンポジウム(CSS2002)，pp.215-219，2002.
5 中川裕志，三瓶光司，松本勉，柏木健志，川口修司，牧野京子，村瀬一郎，“意味保存型の情報ハイディング－日本語文書への応用－”，情報処理学会論文誌，Vol.42，No.9，pp.2339 - 2350，2001.
6 情報処理振興事業協会，“インフォメーションハイディングの技術調査報告書”，http://www.ipa.go.jp/security/fy10/contents/crypto/report/Information-Hiding.htm，1998.
7 松本勉，糸山大志，“Lawful Access の無効化を狙う暗号通信の検出は容易か？”，信学技報 ISEC96-79，pp.159-164，1997.
8 井上信吾，村瀬一郎，滝澤修，松本勉，中川裕志，“XML におけるステガノグラフィ手法の提案”，電子情報通信学会 暗号と情報セキュリティシンポジウム(SCIS2002)，pp.301-306，2002.
9 滝澤修，松本勉，中川裕志，村瀬一郎，牧野京子，“改行位置を利用したテキストステガノグラフィ”，情報処理学会論文誌，Vol.45，No.8，pp.1977-1979，2004.
10 奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座(松本研究室)，“日本語形態素解析システム茶筌 version 2.0 for Windows”，1999.
11 松本勉，吉岡克成，鈴木雅貴，赤井健一郎，滝澤修，牧野京子，中川裕志，“改行位置によるテキスト文書の改ざん検出”，電子情報通信学会 暗号と情報セキュリティシンポジウム (SCIS2004)，pp.983-988，2004.

滝澤　修
情報通信部門セキュリティ高度化グループ主任研究員　博士(工学)
コンテンツセキュリティ、非常時防災通信

中川裕志
東京大学教授　工学博士
自然言語処理

牧野京子
株式会社三菱総合研究所
ソフトウェア工学

松本　勉
横浜国立大学教授　工学博士
情報セキュリティ

村瀬一郎
株式会社三菱総合研究所
情報セキュリティ