テクニカルノート

改行位置を利用したテキストステガノグラフィ

滝澤 修[†1] 松本 勉[†2] 中川 裕志[†3]
村瀬 一郎[†4] 牧野 京子[†4]

プライバシ保護などに利用できるステガノグラフィ（秘匿通信）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提案されてきた．本論文では，ディ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テキストステガノグラフィを提案する．提案手法はドキュメントのレイアウト情報を利用しないため，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を保つ手段として利用できる．

Steganography on Digital Documents by Adjustment of New-line Positions

Osamu Takizawa,[†1] Tsutomu Matsumoto,[†2]
Hiroshi Nakagawa,[†3] Ichiro Murase[†4] and Kyoko Makino[†4]

In the usual steganography 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 proposes 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.

1. まえがき

ステガノグラフィは，埋め込み媒体（以下，カバーデータ）が持つ情報の冗長性を利用して秘匿情報（以下，エンベデッドデータ）を埋め込むのが基本であり，画像や音響信号など冗長度の高い媒体に対しては，人間に識別できるような劣化をきたすことなく比較的実現しやすい．それに対して，文字コードには冗長性がまったくないため，ディジタルドキュメントをカバーデータとする場合は，行間や語間を調整するなどレイアウトを操作して埋め込む手法が主に提案されている[1)]．しかし，ディジタルドキュメントとしては，電子メールのように，レイアウト情報を有さないテキスト（プレーンテキスト）が多くを占めており，プライバシが保たれた文字通信の手段として，プレーンテキストをカバーデータとするステガノグラフィの需要は多いと考えられる．

ハイフネーションが必要な英語とは異なり，膠着言語である日本語の場合，改行の位置は，禁則処理などの例外を除けば，単語の途中であっても比較的自由である．その性質に着目し，本論文では，日本語のディジタルドキュメントを対象として，1 行あたりの文字数が秘匿情報を表すように改行コードを文書内に挿入することで，ステガノグラフィを実現する手法を提案する．

2. 提案するテキストステガノグラフィ

2.1 カバーデータへのエンベデッドデータの埋め込み

提案手法では，ワープロ文書のようにパラグラフの末尾のみに改行コードが入っていて，画面上もしくは

†1 独立行政法人通信総合研究所（現在，独立行政法人情報通信研究機構）
Communications Research Laboratory (Presently National Institute of Information and Communications Technology)
†2 横浜国立大学大学院環境情報研究院
Graduate School of Environment and Information Sciences, Yokohama National University
†3 東京大学情報基盤センター
Information Technology Center, The University of Tokyo
†4 株式会社三菱総合研究所
Mitsubishi Research Institute, Inc.

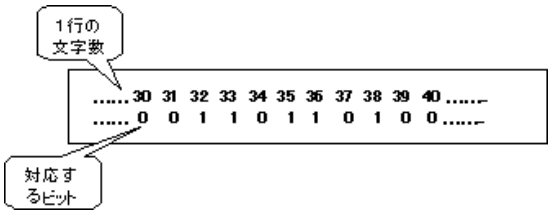

図 1 テーブルの例
Fig. 1 An example of the table.

印字上の行の折り返し部分には改行コードが入っていないディジタルドキュメントをカバーデータとする（以下，カバーテキスト）．埋め込むべきエンベデッドデータは 0 もしくは 1 のビット列とする．ビットと，1 行あたりの文字数との対応をテーブルとしてあらかじめ定めておく．**図 1** に，テーブルの例を示す．エンベデッドデータの埋め込み処理に際しては，そのテーブルを参照し，エンベデッドデータの各ビットに応じた 1 行文字数になるように，カバーテキストの冒頭から改行コードを挿入していく．ただし，テーブルから 1 行あたりの文字数を選ぶ際に，一般的な表示系における行幅をおおむね均一に保つことを重視する☆．そのため，埋め込み処理にあたって，基準とする 1 行の行幅を定めておく．ここで行幅は，当該行における全文字の文字幅の総和と定義する．文字幅は，いわゆる半角文字については 1 文字につき 1，いわゆる全角文字については 1 文字につき 2 と定義する．したがって，おおむね均一な行幅にするためには，半角文字が多い行の場合は文字数が多くなり，全角文字が多い行の場合は少ない文字数になる．

このようにして，エンベデッドデータに応じた改行コードが挿入されてできた文書（以下，ステゴテキスト）は，一般的な表示系において行幅がほぼ均一に保たれ，見た目の自然性が保たれる．

なお埋め込み処理に際して，各行の行幅が一定範囲以内の振れ幅に収まるように制限する．そのため，パラグラフ末の短い行のように，その振れ幅に収まらない行に対しては，埋め込み処理を行わずスキップする．この振れ幅の制限値は，エンベデッドデータの抽出に際しても必要となる．

2.2 ステゴテキストからのエンベデッドデータの抽出

埋め込み処理時に使用したものと同じテーブルを用い（すなわちテーブルが鍵となる），ステゴテキストの各行の文字数をテーブルと照合し，得られる各ビットをエンベデッドデータとして抽出する．ただし，振れ幅の制限値以下の行には，エンベデッドデータが埋め込まれていないと解釈する．

3. 提案手法の評価

3.1 解読攻撃への耐性評価

提案手法では，テーブルが鍵となるため，テーブルにおけるビットと 1 行あたりの文字数との対応をランダムに定義しておけば，テーブルを持っていない者による解読は困難である．したがって，暗号文攻撃に対する耐性はあるといえる．しかし，手法が既知で，エンベデッドデータとステゴテキストのペアが得られれば，テーブルを再現できるので，既知平文攻撃に対する耐性は弱いといえる．したがって，エンベデッドデータ自体を暗号化しておくなどの対抗策を講じる必要がある．

3.2 無効化攻撃への耐性評価

提案手法は，改行する位置を人為的に付け替えることによる無効化攻撃に対しては無力である．ただし，同一のエンベデッドデータを繰り返し埋め込むことにより，局所的な無効化に対する耐性を持たせることは可能である．

4. 考　　察

4.1 提案手法の特徴

提案手法は，エンベデッドデータが各行の文字数として埋め込まれているため，ハードコピーにもデータが残り，ハードコピーの複写を繰り返しても劣化消失しにくい特徴がある．ドキュメントのレイアウト情報にエンベデッドデータを埋め込む従来の手法では，複写を繰り返すことによってデータが劣化消失する問題があった．また挿入されるのは改行コードだけであるので，エンベデッドデータを埋め込むことによってカバーテキストの意味内容には変質をきたさない．したがって，著作物に ID 情報などを埋め込む，電子透かしとしての応用も可能である．さらに，膠着言語であれば，日本語に限らず中国語など他言語にも適用可能である．

4.2 ステゴテキストの自然性に関する考察

解読攻撃や無効化攻撃を防ぐためには，改行の位置情報にエンベデッドデータが埋め込まれていることを攻撃者に気づかせない工夫を講じることが有効である．そのためには，ステゴテキストの自然性を保つことが重要になる．改行の位置に関するテキストの自然性は，行幅の振れ幅が小さいという「見た目の自然性」と，

☆ ここでいう一般的な表示系は，行幅を均等割付しないことを想定している．均等割付をした場合，行幅は均一になる一方，文字間のスペーシングが行ごとに不均一になる．

文節や形態素の境界など読解しやすい位置に改行があるという「言語としての自然性」の 2 つに依存する．言語としての自然性を保つためには，改行可能な位置に制約を課することになるため，必然的に振れ幅は大きくなり，したがって「見た目の自然性」とはトレードオフの関係になる．言語としての自然性としては，句読点などの禁則処理以外は行わないという，最も緩やかな制約から，同一文字種（漢字，ひらがな，カタカナ，数字，アルファベットなど）の文字列の途中での改行を制限するという，強い制約まで考えられる☆．

5. ま と め

本論文では，日本語のディジタルドキュメントを対象として，各行の文字数がエンベデッドデータを表すように改行コードを挿入することで，ステガノグラフィを実現する手法を提案した．提案手法はおおむね 1 行につき 1 ビットのエンベデッドデータしか埋め込めないため，埋め込めるデータ量を増やすための改良が今後の課題である．

謝辞 日頃ご議論くださる横浜国立大学松本研究室の吉岡克成氏，鈴木雅貴氏，三菱総合研究所の村野正泰氏，井上信吾氏，赤井健一郎氏，通信総合研究所（現情報通信研究機構）の山村明弘氏に感謝する．

参 考 文 献

1) 松井甲子雄：電子透かしの基礎，森北出版 (1998).

(平成 15 年 12 月 5 日受付)
(平成 16 年 3 月 5 日採録)

滝澤 修（正会員）

1987 年京都大学大学院工学研究科電気工学専攻修士課程修了．1997 年大阪大学博士（工学）．現在，独立行政法人情報通信研究機構（旧通信総合研究所）セキュリティ高度化グループ主任研究員．自然言語処理の情報セキュリティへの応用や，非常時防災通信の研究等に従事．著書に「実践情報科教育法」（東京電機大学出版局，分担執筆）等．1990 年度電子情報通信学会学術奨励賞ほか．言語処理学会，人工知能学会，計量国語学会，日本災害情報学会等の会員．

松本 勉（正会員）

1986 年東京大学大学院博士課程修了．工学博士．同年横浜国立大学工学部専任講師．同助教授，教授を経て，2001 年より同大学大学院環境情報研究院教授．1981 年より暗号や情報セキュリティの研究に従事．「明るい暗号研究会」創設メンバ．現在，暗号アルゴリズム，情報利用管理，ディジタル証拠性，情報ハイディング，バイオメトリクス，人工物メトリクス，耐タンパーソフトウェア等に広く関心を持つ．国際暗号学会 IACR 理事．暗号技術検討会構成員．ASIACRYPT'96 プログラム委員長．ASIACRYPT2000 実行委員長．電子情報通信学会より「情報セキュリティの基礎理論」への貢献に関して業績賞を受賞．

中川 裕志（正会員）

1975 年東京大学卒業．1980 年同大学院博士課程修了．工学博士．同年より横浜国立大学勤務．1999 年より東京大学情報基盤センター教授．言語情報処理の研究に従事．2004 年より言語処理学会会長．

村瀬 一郎（正会員）

1986 年名古屋大学卒業．同年より株式会社三菱総合研究所勤務．情報セキュリティの調査研究に従事．

牧野 京子

1982 年東京学芸大学卒業．同年より株式会社三菱総合研究所勤務．ソフトウェア工学および情報セキュリティの研究に従事．

☆ 異なる文字種の境界は形態素の境界である場合が多いという経験則に基づく．