O_002

# Word 2003文書への情報ハイディングシステム
# An Information Hiding System for Word 2003 Documents

北野 宗之† 増田 英孝† 中川 裕志‡
Muneyuki Kitano Hidetaka Masuda Hiroshi Nakagawa

## 1. まえがき

情報ハイディングは，埋め込み媒体が持つ情報の冗長性を利用して別の情報を埋め込む技術であり，文書を埋め込み媒体とした情報ハイディングは，第三者が気づかない作為を文書に施すことによって情報を埋め込み，当事者のみがその文書から秘匿情報を抽出できることを目的とする．類似の技術である暗号化技術が情報の意味を隠蔽するのに対し，情報ハイディング技術は情報の存在自体を隠蔽する技術である．

情報ハイディングの一つの応用として，悪意のある第三者による傍受や検閲等からの脅威を想定したステガノグラフィ（秘密通信路，あるいは秘匿通信）がある．文書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．一方，著作権情報などをデジタルコンテンツに埋め込む電子透かしも，ステガノグラフィとともに情報ハイディングの一応用分野である．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き，不正な二次配布をした場合に流出元を特定できるため，海賊版などの不正コピーの流布に対する抑止効果が期待できる．

これまで文書を対象とした情報ハイディング手法のアプローチの一つとして，行間，文字間隔を調整するなど文書自体を画像として取り扱う手法がある [1]．次に，文書内容そのものはプレインテキストとして扱い，同義語等を利用して文書内容を書き換える方式があり，辞書変換法と呼ばれる [2, 3]．また，テキスト中の改行位置の変更による情報埋め込み手法もある [4]．

そこで我々は，現在広く利用されているワードプロセッサ Word 2003 を対象とし，情報を埋め込んだ文書を Word 2003 上で表示しても情報を埋め込んでいることが発見されにくい情報ハイディング手法を提案する．提案手法は，空白及びタブ文字の色情報が変化しても，Word 上の表示には影響を与えないことを利用して情報を埋め込む．そして，受け取り側で送り手側が指定した保護部分の改竄の有無を検出できる情報ハイディング手法であり，Word 2003 に情報埋め込み検証システムを組み込んで，すべての作業を利用者が Word 2003 上で一括して処理が行える．

## 2. 本提案の目的と応用

本研究では，原著者が指定したある特定の部分のみを保護することを目的とする．最終的な文書の受け手は原著者の保護対象部分に改竄があったかどうかを検証することができる．原著者と最終的な受け手の間に第三者が介入し，文書への追加及び加筆は許すものとする．

†東京電機大学工学部
‡東京大学情報基盤センター

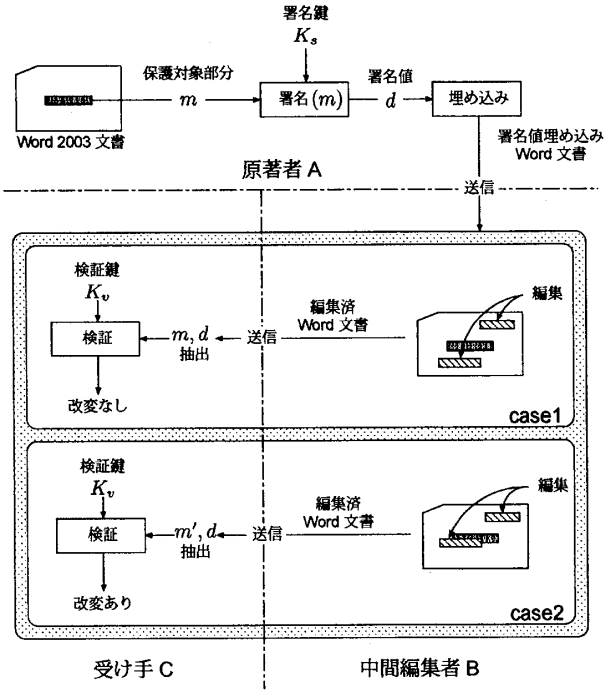


図 1: ビジネスモデルの概念図

本手法の秘匿情報の埋め込み対象は，元の Word 文書の本文に表れる部分についてのみに言及しているが，ファイル名や変更履歴などのドキュメントプロパティを埋め込むことも考えられる．あるいは文書とは全く関係のない通信者間だけが秘密に情報を送受するといった利用方法もある．

次に，提案する情報ハイディングシステムを利用したビジネスモデルについて説明する．図 1 に我々が提案する情報ハイディングシステムにおけるビジネスモデルの例を示す．文書を作成する原著者 A は，文書の中の指定した部分（網かけ部）を保護対象 $m$ とし，その部分に署名鍵 $K_s$ を用いてデジタル署名を行い，署名値 $d$ を保護対象部分内に分散させて埋め込む．この署名付き文書を中間編集者 B に送信する．中間編集者 B は，編集する担当箇所においては加筆や訂正などの編集行為が許されている．図 1 では，中間編集者 B が保護対象部分以外の編集を行った場合を case1，保護対象部分に編集を行った場合を case2 として示す．そして，最終的な文書が受け手 C に送信され，C は中間編集者 B によって原著者 A の保護対象部分に変更が加えられていないことを検証鍵 $K_v$ を用いて確かめることができる．更に，この文書全体を $M$ とおくと，中間編集者 B は原著者 A の署名とは独立に $M$ の署名又はハッシュを埋め込み，最

表 1: 日本語 Word 文書の空白数の調査

| | |
|---|---|
| 全文書数 | 114 |
| 平均空白数 | 241.4 |
| 平均ページ数 | 2.4 |
| 1ページ当たりの空白数 | 111.5 |

表 2: チェックマークの仕様

| | |
|---|---|
| R の下位 1 ビット | 保護対象部分の先頭を示す |
| G の下位 1 ビット | 保護対象部分の末尾を示す |
| B の下位 1 ビット | チェックマークであることを示す |

終的な文書として受け手に渡すこともできる.

図 1 のシナリオでは，case1 の編集済文書では保護対象部分 $m$ の改変はない，case2 の編集済文書では保護対象部分 $m'$ は改変されていると判定される.

このモデルの応用として，著作権法における不正な改変と，改変した著作物の不正な引用の阻止が考えられる. これは，社内で取りまとめるマニュアルやドキュメントを一人ではなくグループまたは組織で作り上げようとする場合などの利用，原作者が書いた意見やデータを中間編集者が歪めたことを検出できるシステムである. 例えば，株主総会資料を作成するときに中間編集者が現場から上がってきた会計データを改竄して用いてしまう状況を防ぐことが挙げられる.

## 3. Word 文書の色属性を利用した情報ハイディング

### 3.1 情報埋め込み手法

Word 上の表示には影響を与えないものを選択して情報を埋め込む. 情報の埋め込みにはできるだけ一度に埋め込むことができる情報量が多いことが望まれる. そこで本研究では色属性に着目した. 空白，タブ文字，段落自体を利用すれば，色情報を変更しても Word 上の見かけに影響を与えないことを利用する [5]. この色属性を利用すれば，空白 1 文字当たり 24 ビットの情報を埋め込むことができることになる.

Word では編集記号の表示モードを変更することにより空白文字を "·"，タブ文字を "→" として表示することができる. しかし，空白及びタブ文字の色属性がどんな色であっても "·" 及び "→" の表示は色属性に関係なく一定色であり，色属性の変更が Word 上の表示に影響を及ぼさない§.

空白が多数出現する英語などの言語では、この方法は多数の埋め込み位置を使えるので有力である. これにひきかえ，日本語文書では空白文字の出現頻度が低い. そこで，空白及びタブ文字が実際の Word 文書中にどれくらい出現するのか，筆者以外の第三者が作成した学内外の 114 文書について Word の文字数カウントツールを用いて調査を行った. その結果を表 1 に示す. この表中の空白数は半角の空白，全角の空白，タブ文字の合計数である. これらの空白文字は，主として体裁を整えたり，表示を微調整するために使われていた.

サンプルの 114 文書のうち 14 個以上の空白がある文書は 107 文書であった. 後に 3.3 で述べるようにデジタル署名値の埋め込みには 14 個の空白が必要である. 従って，表 1 の結果からすれば多くの文書では署名値の埋め込みに十分な空白が存在する. また，1 ページ当たりの空白数は 111 文字であり，1 ページ当たりに埋め込むことができる情報量が多いことが挙げられる.

### 3.2 保護対象部分の指定

次に，秘匿情報が埋め込まれた文書を受け取り側で保護対象部分の改竄を検証するためには，保護対象部分の位置を示す情報（以後，チェックマークと呼ぶ）を付加する必要がある. ただし，このチェックマークも Word 上の表示には影響ができるだけ少ないものとする.

そこで，画面に表示される文字の色属性を 1 ビット 程度変更しても，一般的には利用者が認識できないことを利用して，保護対象部分の先頭と末尾の位置にチェックマーク情報を付加する. まずここでは，保護対象部分及び前後の文字が文書全体に反映されているデフォルト色属性で記述されているものとして説明する. 受け取り側では，検証時にデフォルト色属性からの文字色の 1 ビットの差を見つけて，チェックマークに囲まれた部分が保護対象部分であることを知ることができる.

Word で指定できる色属性の値は Red，Green，Blue がそれぞれ 8 ビット からなる 16 進の "RRGGBB" の 24 ビット値または "auto"（自動）である. "auto" は黒色と見なすことができるため，"000000" として取り扱う.

この保護対象部分を指定するには，利用者の立場からすると専用の GUI を用意することが望ましい. 今回は，Word のプラグインとして実現することを試みた. GUI は保護したい部分を選択して反転表示させ，メニューバーから保護対象部分指定を選択するだけで，保護対象部分の指定を行うことができる.

### 3.3 情報埋め込み

ここでは文書の送り手が秘匿情報を文書内に埋め込む手法を示す. 図 2 にその概要を示し，その流れは以下の通りである. 実装には C#を用い，Word 2003 のプラグインとして組み込んだ [6]. また，埋め込むデジタル署名値の生成及び署名値の検証については，一般的に利用されているデジタル署名の方式として，デジタル署名アルゴリズムの DSA を用いることにした. 生成されるデジタル署名値は 320 ビットとなる. 色属性値は，空白文字の色属性 24 ビットすべてを埋め込みに利用する.

1. Word 2003 上で文書作成者であるユーザ（原著者）は，保護対象部分を指定した Word 文書を DOC 文書として出力する
2. チェックマークを基に，指定された保護対象部分のテキスト $m$ を抜き出す
3. 署名アルゴリズム $S$ より，ユーザの秘密鍵 $S_K$ を使って $m$ のデジタル署名 $d$ を生成する
4. 署名値 $d$ を保護対象部分に含まれる空白文字の色属性に分散させて埋め込む

§Word 2003 の互換ソフトウェアである OpenOffice では空白及びタブ文字の編集記号に色が反映されてしまう.

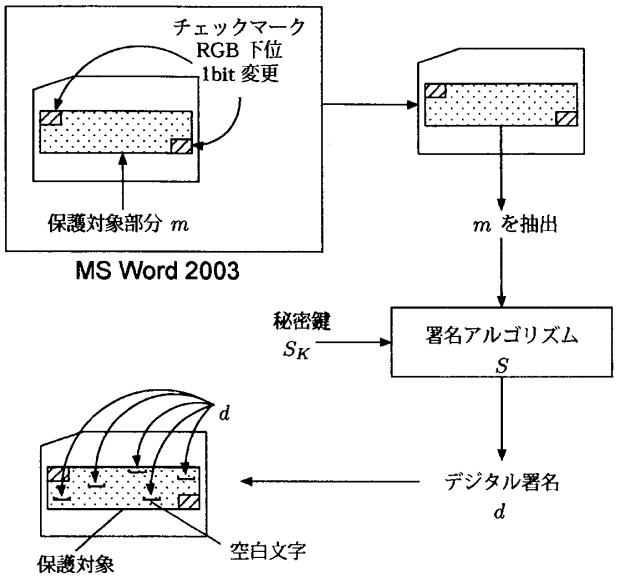


図 2: 情報埋め込み方法

ここで，デジタル署名をすべて埋め込むためには 14 個の空白文字が必要である．従って，3.1 で述べたように保護対象部分はあらかじめ広く取っているものとし，デジタル署名の埋め込みに必要な空白が保護対象部分に含まれていることを想定している．しかし，保護対象部分が狭い場合まで考えるとなるとデジタル署名を埋め込むことは困難である．その場合にはデジタル署名ではなく，ハッシュ値またはハッシュ値の一部を直接埋め込む方法が考えられる．ハッシュ値を用いれば，埋め込みに必要な空白は数個程度となる．このような場合には，ハッシュ関数の 1 つである SHA-1 より生成された 160 ビットのハッシュ値を適用する．

保護対象部分のテキストのデジタル署名値またはハッシュ値を計算した後，保護対象部分に出現する 1 番目の空白文字から順にこれらの値を埋め込んでいく．

図 3 に埋め込み前の文書，図 4 に埋め込み後の文書を示す．この 2 つの文書に Word 表示上の違いは見られないため表面的に見ただけでは情報が埋め込まれていることが発見されにくい．

### 3.4 情報抽出と検証

ここでは情報埋め込みが行われた DOC 文書に対して，受け取り側で秘匿情報を抽出して改竄されているかどうかを検証する．図 5 にその概要を示す．

1. 空白文字に埋め込まれた色属性から署名 $d$ を復元する
2. チェックマークを基に，保護対象部分のテキスト $m'$ を抜き出す
3. 検証アルゴリズム $V$ より，公開鍵 $P_K$ を使って署名 $d$ とテキスト $m$ が正しいことを検証する

以上の工程により，保護対象部分 $m$ について改竄の有無を検出することができる．

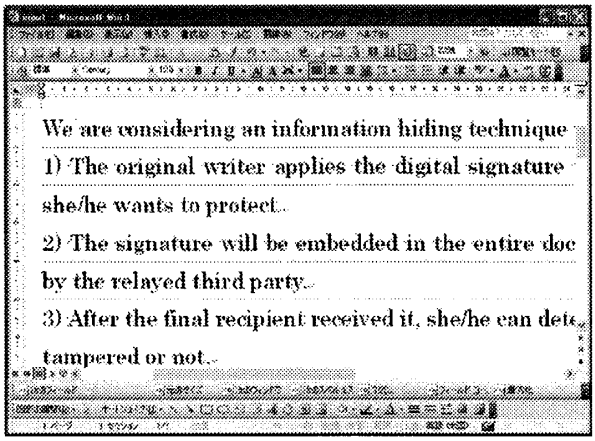


図 3: 埋め込み前の Word 画面表示例

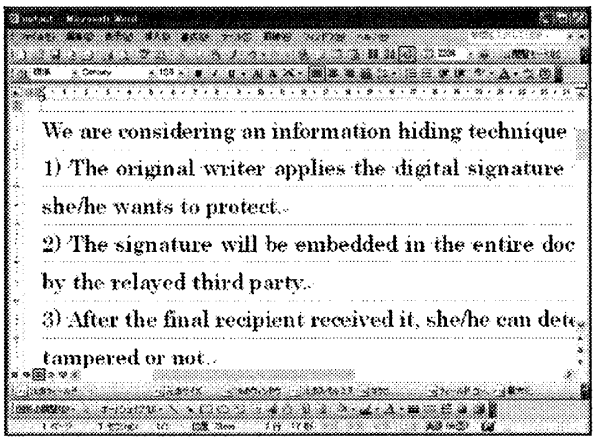


図 4: 埋め込み後の Word 画面表示例

2 の保護対象部分のテキストを抜き出すために，3.2 のチェックマークの色情報を利用する．また，空白文字の色属性に埋め込まれた署名値の復元は，1 番目の空白文字から順に抽出していく．

図 4 の DOC 文書に対して，以下の項目について Word 上での編集を行った場合について，検証を行った．

1. 保護対象部分のテキストを変更する
2. 保護対象部分の空白文字に埋め込んだ色情報の一部を変える
3. 情報を埋め込んである保護対象部分の空白文字を削除する
4. 保護対象部分に任意の色属性を指定した空白文字を挿入する
5. 保護対象部分の先頭または末尾の色情報を変更
6. チェックマークと同じ色情報をランダムな位置に挿入
7. 保護対象部分とは関係のない部分を変更する

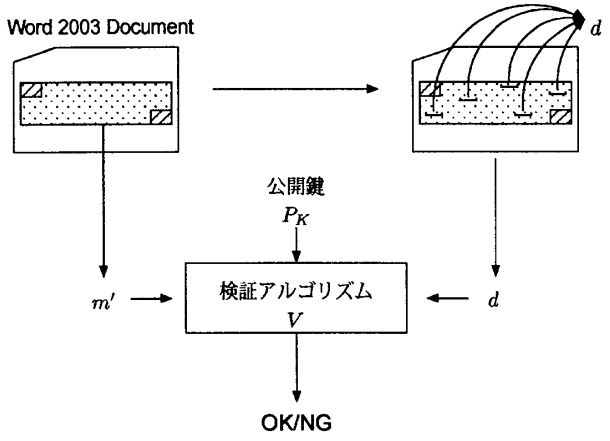


図 5: 情報抽出と検証方法

表 3: 文書編集による検証結果

| 評価項目 | Word 編集 |
|---|---|
| 保護対象部分のテキストを変更 | × |
| 空白文字の色属性を変更 | × |
| 情報を埋め込んだ空白文字を削除 | × |
| 別の色つき空白文字を挿入 | × |
| チェックマークの色変更 | × |
| チェックマークと同じ色を他の挿入 | × |
| 関係のない部分を変更 | ○ |

項目の 1〜4 は保護対象部分に関わる改竄で 5，6 は保護対象部分の位置に関わる無効化であるため，検証結果は不一致 (×) を返すことを想定し，7 では一致 (○) を返すことを想定している．本システムでの検証結果は表 3 のようになり，我々が想定していたものと一致している．このうち 2 については，文書の空白に余裕があれば同一の署名値を繰り返し埋め込むことで，局所的な色属性変更による耐性は持たせることができる．6 に関しては保護対象部分のチェックマークの開始位置以前に開始マークが，開始位置以降終了位置以前に終了マークが出現する場合に不一致となる．このように攻撃者に情報ハイディングの存在を知られてしまった場合には無効化攻撃が可能となってしまうが，何らかの攻撃によってデータの信頼性がなくなってしまったことを検知することができる．データの改竄については，使用するデジタル署名の強さに依存してなりすましを防ぐことができる．

今回は検証システムを受け取り側で持っている場合を想定しており，その場合に情報埋め込みが行われた DOC 文書を受け取ることにより，受け取り側で保護対象部分の改竄の有無を検証できる．

ところで，本手法では空白文字の色属性に署名値を埋め込むため，当然ながら元文書に比べ，ファイルサイズの増加が生じる．そのため，ファイルサイズが異常に大きくなっていることによって見破られるということも考えられる．そこで，文書の平均的なページ数を持つサンプルとしてファイルサイズ 47,104 バイト の DOC 形式の文書を用意し，情報を埋め込む際のファイルサイズを調べた結果，署名値を埋め込む元文書と比べ，DSA を用いたときの差分は 0 バイトであった．これはバイナリ形式固有の最適化処理によるものと考えられる．従って，実際に情報埋め込みをした文書のサイズを見ると，大きくなっているとはいえないので，ファイルサイズによって見破ることは難しいと考えられる．

## 4. むすび

Word 2003 でファイル入出力可能な DOC 文書に対して，保護対象部分を指定してそのデジタル署名を秘匿情報として埋め込む手法を提案した．提案手法では，空白及びタブ文字の色属性を利用して情報を埋め込み，受け取り側で情報を抽出して改竄の有無を検証できる．

今回は Word 2003 で作成された DOC 文書について，提案手法を用いたシステムの試作を行った．その結果，保護対象部分の埋め込み，抽出後の検証が行えることが分かった．また，保護対象部分の改竄の有無の検出実験を行った結果，あらかじめ期待された結果が得られた．

## 謝辞

本研究を進めるにあたって有益なアドバイスをして頂いた 横浜国立大学 松本勉教授，三菱総合研究所 村瀬一郎氏，井上信吾氏，赤井健一郎氏，牧野京子氏，独立行政法人情報通信研究機構 滝澤修氏，吉岡克成氏 に感謝致します．

## 参考文献

[1] J. Brassil et al., “Electronic Marking and Identification Techniques to Discourage Document Copying, ” IEEE J. Selected Areas in Communications, vol.13, no.8, pp.1495–1504, 1995.

[2] M. Chapman, G. Davida, “Hiding the Hidden: A Software System for Concealing Ciphertext as Innocuous Text, ” Proc. First International Conference on Information and Communication Security, pp.335–345, Beijing, China, Nov.1997.

[3] 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–，” 情処学論，vol.42, no.9，pp.2339–2350，Sept.2001.

[4] 滝澤修，中川裕志，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，” 情処学論，vol.45，no.8，pp.1977–1979，2004.

[5] 北野宗之，増田英孝，中川裕志，“Word 2003 XML 文書への情報ハイディングシステム”，情処研報，2005-CSEC-30，pp.205–212，2005.

[6] E. Carter, E. Lippert, Visual Studio Tools for Office: Using C# With Excel, Word, Outlook, and InfoPath, Microsoft Net Development Series, Addison Wesley Professional, Boston, MA, 2005.