自然言語文を用いた秘密分散の提案

滝澤修　　山村明弘

独立行政法人通信総合研究所

視覚復号型秘密分散の考え方を自然言語文に適用した新しい秘密分散法を提案する。日本語文を対象とし、複数枚の分散テキスト(share text)を重ね合わせて、1 文字目を上層から下層、2 文字目を上層から下層、…に順次並べていくと、その文字列の中に秘密テキスト(secret text)が現れるようにするものである。積み重ねる順序を変えることによって、別の秘密テキストを浮かび上がらせることも可能である。分散テキストを重ね合わせることは簡単な機械処理によって実現できる。重ね合わせて得られた文字列の中から秘密テキストを復号することは、人間による視認でも可能だが、意味のあるフレーズは 2 文字以上の形態素の連なりになっている確率が高い性質を利用して、形態素解析器を援用して切り出す方法を提案する。分散テキストは、文データベースを用いて合成することで、一見自然な文に見せかけることができる。

A Proposal of Secret Sharing Using Natural Language Text

Osamu TAKIZAWA　　Akihiro YAMAURA

Communications Research Laboratory

Modifying the idea of the visual cryptography, we propose a method of sharing a secret key using natural language texts. Our target here is restricted to Japanese texts. Each participant obtains a share, which is a Japanese text in our scheme. When a certain number of participants retrieve the secret key, they supply their shares and pile up these natural language texts. The sequence of the first, second (and so on) letters occurred in the pile shows the secret text. The order of the pile is significant, and changing the order may yield the distinct secret text. It is easy to pile the shared natural language texts by computer operation. Human eyes can recognize the secret text from the piled texts, however, we aim to construct a natural language text secret sharing scheme employing a morphological analyzer because a meaningful phrase is a chain of morphemes consisting of more than two words with a high probability. We can make a shared natural text look like a natural text without any secret meaning by synthesizing using a text database.

1. はじめに

複数のメンバーが分散して保有する情報を合わせた場合にのみ秘密情報を復号できる秘密分散法（secret sharing）[1][2]の一つの実現形態として、Naorら[3]によって提案された視覚復号型秘密分散法（Visual Cryptography または Visual Secret Sharing，以下VSS）は、計算機を使わず人間の視認によって復号可能な新しい暗号として、研究や実用化が進められている[4][5][6]。

ところで有史以来の古典的な暗号は、自然言語テキストを対象とするものがほとんどを占めていた。画像におけるピクセルに対応するのは、自然言語テキストの場合は文字である。しかし自然言語テキストを情報隠蔽媒体とするには、文字コードには冗長性が全く無く[7]、また言葉は意味をもつことから、情報を隠蔽するために文字を改変すれば僅かな改変でも露見してしまう難点がある。そのため、画像におけるピクセル単位の操作のような方法は文字には単純には適用できない。また自然言語テキストは、画像等の他の媒体と比べて情報量が少ないため、隠蔽できる情報量も少なく、実用的に不十分な場合が多い。そのため、自然言語テキストを情報ハイディングや秘密分散などの情報隠蔽媒体とする研究は、一部の例外[8][9][10]を除き、あまり多く見られない。しかし、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自然言語テキストでの情報交換は主流の位置を占めており、情報伝達手段としての自然言語テキストの重要性は今後も変わらないと考えられる。従って自然言語テキストを情報隠蔽媒体として扱う暗号法には、多くの応用が期待できる。

そこで本稿では、自然言語テキストを情報隠蔽媒体とする秘密分散法（Text Secret Sharing，以下TSS）の実現の可能性について検討し、一つの簡単な方法を提案する。

2. テキスト秘密分散法の考え方

TSSは、秘密テキスト(secret text)を複数の分散テキスト(share text)に分散して隠蔽し、テキストを“重ね合わせて”秘密テキストを復号するもの、と定義できる。VSSでは、分散画像はノイズのような無意味画像が使われることが多い[11]。従ってTSSの場合も分散テキストが無意味な文字列であっても構わないが、そうすると秘密分散を使っていることを見抜かれる懸念があり、それは既に脅威となりえることなので、自然言語処理的な工夫により、できるだけ自然な文にすることが望ましいといえる。

ところで、TSSにおいて、VSSにおける“重ね合わせる”ことに対応するのはどのような処理であろうか？例として以下のような方法が考えられる。

（方法1）

分散テキストを複数枚重ね合わせて、一致している文字を拾い読みすると、秘密テキストになっているようにする方法が考えられる。図1にイメージを示す。

図1　方法1のイメージ
分散テキストを重ね合わせて上から眺めた場合
（丸印を付した個所が秘密テキスト「盛岡駅前が会場です」）

この方法は、文字をイメージとして扱っている点で、VSSの一種に位置づけることができ、VSSのように視認で復号できる特長がある。重ね合わせる順序は鍵にならない。また、分散テキストが完全に揃っていない場合でも、秘密テキストは完全ではない状態（過剰な文字が混入している）であるものの、過剰文字を間引いて読むことで大まかに解読できてしまう懸念がある。

（方法2）

複数枚の分散テキストを重ね合わせて、冒頭文字の位置を合わせ、1文字目を上層から下層、2文字目を上層から下層、…に順次並べていくと、その文字列の中に秘密テキストが現れるようにする方法が考えられる。図２にイメージを示す。

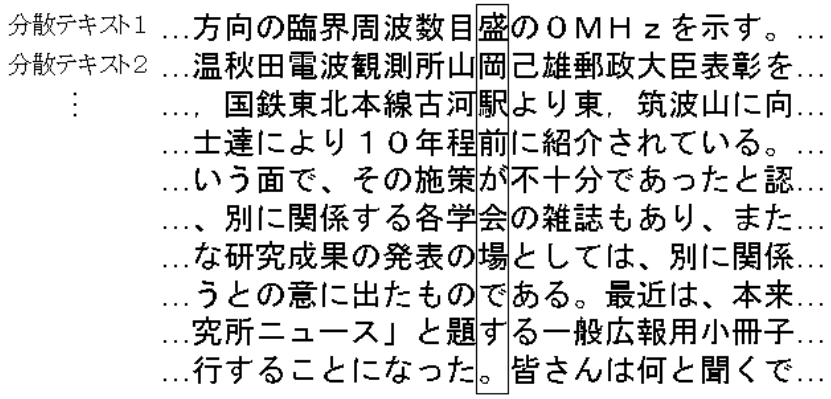


図2　方法2のイメージ
分散テキストを重ね合わせて横から眺めた場合
（四角で囲った範囲が秘密テキスト「盛岡駅前が会場です」）

重ね合わせて得られる文字列は、無意味な文字列の一部に、意味のある秘密テキストが混じっている形になる。図２の場合、重ね合わせて得られる文字列は、縦に左から右へ読んでいくことで得られ、「**…波測線0のる発た」な数所古年施各表もとっ目山河程策学のの題た盛岡駅前が会場です。の己よに不のとある0雄り紹十雑しるーM郵東介分…**」（一部）となる（下線部が秘密テキスト）。この方法は、重ね合わせる順序も鍵になっているので、順序を入れ替えることで別の秘密テキストも隠蔽することが原理的には可能である。但し分散テキストが完全に揃っていない場合でも、秘密テキストは完全ではない状態（一部の文字が歯抜け）であるものの、大まかに解読できてしまう懸念がある。

この方法は、重ね合わせる行為自体に計算機の補助が必要になる。また秘密テキストが隠蔽されている個所は視認で抽出できなくはないが、結構面倒なため、やはり計算機の補助が必要となる。従って、物理的な重ね合わせ操作および視認によって直ちに復号できるVSSのような簡便さに欠けている難点がある。

以下の節では、方法2の可能性について更に詳しく検討する。

3. テキスト秘密分散の一方法

本節では、前節の方法2を実現した簡単な処理プログラムを実装した結果について述べる。

3．1　分散テキストの生成方法

第2節で述べたように、分散テキストは自然な文章であることが望ましい。単語を組み合わせることによって意味的に自然な文を合成することは、計算機では膨大な知識と複雑なアルゴリズムを必要とする。そこで、ある程度の長さの文を大量にデータベース化しておき、その文を組み合わせることによって分散テキストとする方法を考える。ここで使うデータベースは、2つ以上の文を続けても文間の文脈が極端には不自然にならないように、同じ分野の文によって構成されていることが重要と考えられる。

分散テキストを重ねて秘密テキストがうまく現れるようにする方法を考える。例えば、「滝澤より、文京グリーンコートセンターオフィス16階でお待ちしています。」という秘密テキストを10枚の分散テキストに1箇所に隠蔽させることを想定すると、それぞれの分散テキストの中に、

分散テキスト１　「…滝ンィて…」
分散テキスト２　「…澤コスい…」
分散テキスト３　「…よー１ます…」
分散テキスト４　「…り トセ階。…」
分散テキスト５　「…、６で…」
分散テキスト６　「…文ンタお…」
分散テキスト７　「…京グー待…」
分散テキスト８　「…リオち…」
分散テキスト９　「…ーフし…」
分散テキスト10

という、それぞれ3～4文字のフレーズを同じ位置に入れなければならないことになる。例えば分散テキスト1の場合、「滝ンィて」という無意味なフレーズを内包させて、どのように分散テキストを作るかが問題となる。以下の対処方法が考えられる。

（対処1）分散テキスト中に多少おかしなフレーズがあっても許容する

秘密テキストの正確な復号を実現するために、分散テキストの自然さを犠牲にする方法である。例えば上掲の例における分散テキスト1の「滝ンィて」を内包した文を無理やり合成する。この方法では、おかしなフレーズがある場所を手がかりとして隠蔽情報の存在を見破られる危険性がある。

（対処2）復号された秘密テキストが原文とは多少異なっていても許容する

分散テキストの自然さを優先して、秘密テキストの正確な復号を犠牲にする方法である。例えば上掲の例における分散テキスト1の「滝ンィて」を「滝シィテ」程度に調整することを許容する。この方法では、秘密テキストが形態素的におかしな文になるので、３．2項で述べる形態素解析を用いた復号が困難になることが問題となる。またこの調整の自動化は難しいと思われる。

（対処3）秘密テキストをちぎる

秘密テキストを複数の部分秘密テキストにちぎって、「部分秘密テキストの最大文字数≦分散テキストの枚数」とし、分散テキスト上で置く位置を散らせる。但しちぎる際に形態素を分断するような変なちぎり方をすると、３．2項で述べる形態素解析を用いた復号が困難になるので、工夫が必要である。また、置く位置を散らせた状態でうまく分散テキストを合成することは難しい。

（対処4）秘密テキストの文字数≦分散テキストの枚数とする

最も簡単に実現できる方法であるが、十分な長さの秘密テキストを使えないか、もしくは大量の分散テキストを使わなければならない等の点で、実用上の大きな制約がある。

本稿では、とりあえず対処4を採用することにする。

３．２　秘密テキストの復号支援方法

第2節で述べた通り、方法2は秘密テキストが隠蔽されている個所を抽出することが視認では比較的難しいため、対策として例えば秘密テキストの開始と終了の個所にフラグシーケンスを付加する方法が考えられる。しかし、第１節で述べたように自然言語テキストを情報隠蔽の媒体として用いる場合、隠蔽できる情報量が少ない制約があるため、秘密テキストもできるだけ短いことが望ましい。そこで、意味のあるフレーズとそうでないフレーズとは視認で見分けがある程度可能という自然言語の性質を利用することで、フラグシーケンスを使わなくても抽出できるような手立てを考える。

自然言語処理における基本的な処理の一つとして、文を形態素（語を構成する最小単位）に分解する形態素解析がある。形態素解析をすると、意味の無いフレーズは、1文字の形態素の並びになる場合が圧倒的に多い。そのため、2文字以上の形態素が多く現れる個所は意味のある秘密テキストの部分である可能性が高いことになる。この性質を復号に援用する。

3．3　具体的な処理の例

3．1項の対処4に基づく分散テキストの生成機能、および3．2項の秘密テキストの復号支援機能をperlで実装した。分散テキストの生成のための文データベースとしては、通信総合研究所の広報紙「CRLニュース」の20年分の全記事[12]を使用した。このデータベースを採用したのは、同じ分野の文によって構成されているという条件を考慮したためである。データベースのサイズは約5MBである。また、秘密テキストの復号支援については、形態素解析器「茶筌」[13]を用いた。

「盛岡駅前が会場です。」を秘密テキストとした場合に生成した分散テキスト（10枚）のうち、例として2枚を以下に示す。いずれも、自然言語テキストとして不自然ではないと思われる。

「最近は、本来の電離層を介する伝搬よりも、むしろ宇宙通信に対して電離層が与える影響に関する研究の方が活発になっている傾向がある。もっとも内側の太線の円は衛星軌道を地球上に投影したものを表わすと同時に半径方向の臨界周波数目盛の０ＭＨｚを示す。」

「この現象は雷放電による電波が電離層上部の多種類のイオンと作用し、特に重水素イオンと共鳴作用をすることによって生じたものと考え、これを重水素ホイッスラと呼ぶことにした。卒直に言って、当所は一般へのＰＲという面で、その施策が不十分であったと認めざるを得ない現況である。」

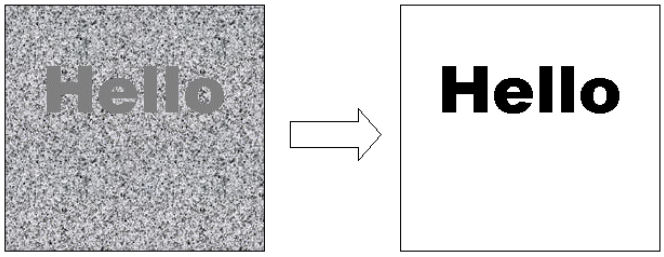
な	助動詞
数	名詞-一般
所	名詞-接尾-一般
古	接頭詞-名詞接続
年	名詞-一般
施	未知語
各	接頭詞-名詞接続
表	名詞-一般
もと	助詞-係助詞
目	動詞-自立
山	名詞-接尾-一般
程	名詞-固有名詞-地域-一般
策	助詞-副助詞
学	名詞-一般
の	名詞-接尾-一般
の	助詞-連体化
題	名詞-非自立-一般
た	名詞-一般
盛	助動詞
岡	名詞-固有名詞-地域-一般
駅	名詞-一般
が	助詞-格助詞-一般
会	名詞-一般
で	助動詞
。	記号-句点
の	助詞-連体化
己	名詞-一般
よ	副詞-一般
不	接頭詞-名詞接続
の	名詞-非自立-一般
と	助詞-格助詞-一般
ある	動詞-自立
０	名詞-数
雄	名詞-一般
り	助動詞
紹	未知語
十	名詞-数
雑	名詞-一般
し	動詞-自立
る	名詞-数
一	記号-アルファベット
Ｍ	

また、分散テキストを重ね合わせて得られる文字列を形態素解析した結果の一部を右に示す。秘密テキストである「盛岡駅前が会場です。」の一節（括弧の部分）が、2文字形態素の連なりになっており、適切な閾値を設けることにより、高い精度で機械的に切り出せることが示唆されている。

4．　考察

提案した手法では、形態素解析を援用して、秘密テキストをより視認しやすくする操作を施している。これはVSSにおいて、分散画像の重ね合わせによりコントラストが低下した秘密原画像を、右下図のように、輪郭線処理などを援用してより視認しやすくすることに相当すると言える。

秘密テキスト自体が無意味な文字列（例えばパスワードなど）の場合には、形態素解析による復号支援処理は機能しない。その場合は、無意味な文字列の前後を意味ある文字列ではさむ等の対応をとる必要がある。VSSの場合でも、無意味な画像を秘密原画像とした場合には、視認によって抽出することは困難である。



提案した手法は、分散テキストが欠けていても大まかに復号できてしまうという、秘密分散方式としてはかなり大きな欠点を抱えている。これは人間が自然言語テキストを読解する知的能力に関係することであるため、自然言語テキストを秘密分散法の情報隠蔽媒体とする場合につきまとう本質的な問題といえる。

重ね合わせの順序を入れ替えることで別の秘密テキストも隠蔽する方法については、今回は検討しなかったが、原理的には可能であり、今後の課題としたい。また3．1項の対処3で述べた、秘密テキストをち

ぎって散らして配置する方法についても検討したい。

【謝辞】
本研究のきっかけを与えて下さった、PuKyong National University の Prof.Ji-Hwan Park に感謝する。また、自然言語テキストを情報隠蔽媒体として適用する方法に関して、横浜国立大学の松本勉教授、東京大学の中川裕志教授、三菱総合研究所の村瀬一郎、井上信吾、牧野京子の各氏から有益な助言を賜っていることに感謝する。

【参考文献】
[1] A.Shamir, “How to share a secret”, Communications of the ACM, 612-613, 1979.
[2] G.Blakley, ″Safeguading cryptographic keys″, Proceedings of AFIPS National Computer Conference, 313–317, 1979.
[3] M.Naor and A.Shamir, “Visual Cryptography”, Advances in Cryptology-Eurocrypt’94 , 1-12, 1994.
[4] 加藤拓，今井秀樹：“視覚復号型秘密分散法の拡張構成方式”，電子情報通信学会論文誌，Vol.J79-A, No.8, 1344-1351, 1996年8月.
[5] 有井幸太，盛拓生，坂井一雄，今井秀樹：“積み重ね順序を鍵とする視覚暗号方式”，SCIS2000, 2000年1月.
[6] 視覚復号型暗号製品「あわすとでーる」凸版印刷株式会社，http://www.toppan.co.jp/aboutus/release/article463.html, 2001年4月.
[7] 松井甲子雄，“電子透かしの基礎”，森北出版，1998年.
[8] 中川裕志,木村浩康,三瓶光司,松本勉,“辞書変換法に基づく日本語テキストへの情報ハイディング”，情処論，Vol.41, No.8, 2272- 2279, 2000年.
[9] 松本勉，中川裕志，村瀬一郎，“ネットワーク向けインフォメーションハイディング技術開発 テキスト用フィンガープリンティング方式 FinPri.txt の開発”，情報処理振興事業協会 次世代デジタル応用基盤技術開発事業 先端的情報化推進基盤整備事業 論文集, 97-104, 2000 年 6 月.
[10] 滝澤修，“情報埋込・抽出方法及びその装置並びに記録媒体”，特願 2001-67597.
[11] Moon-Soo Kim, Seong-Han Shin, Ji-Hwan Park, “New Construction for Multiple Visual Secret Sharing”, SCIS2000, 2000 年 1 月.
[12] 通信総合研究所，“CRL ニュース”，創刊号～第 238 号, 1976 年～1995 年.
[13] 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.