令和７（２０２５）年度　卒業研究報告書概要

課程，学籍番号，氏名	課程：電気・電子情報工学課程，学籍番号：B243277，氏名：三原　颯東
工学分野名：情報通信システム	指導教員名：市川　周一

題　　目：和　置換暗号による難読化命令列の解読

（英　）Decryption of obfuscated instruction sequence by substitution cipher

Abstract

This study aims to examine the capabilities and limitations of statistical cryptanalysis on simple substitution ciphers by evaluating n-gram frequency analysis for both natural-language text and instruction sequences. Although substitution ciphers preserve statistical structure, naive 1- to 3-gram frequency analysis achieved only limited decoding accuracy for text (36–49%) and instruction sequences (44–50%). In contrast, combining frequency information with optimization methods significantly improved the decryption accuracy for text, achieving up to 99% using greedy search and simulated annealing. However, the same approach was ineffective for instruction sequences, with match rates remaining below 34%, and extended scoring functions using combined n-gram log-probabilities showed no further improvement. These results indicate that while modern computational resources enable partial recovery of instruction-sequence ciphertext, achieving high-accuracy decryption remains an open challenge.

概　　要

本研究では，単一置換暗号に対する統計的解読手法の特徴整理と限界の明確化を目的とし，文字列と命令列を対象とした n-gram 頻度分析法の有効性を検証した．単一置換暗号は置換写像により平文と暗号文の同型性が保たれるため，暗号文中の 1,2,3 -gram の統計的偏りが平文の統計的構造を強く反映する．市川ら(2008) は，セキュアプロセッサの多様化を低コストで実現する方法として，命令コード部に単一置換暗号を利用するというアプローチを提案した．市川らは，命令 mnemonic の種類は自然言語文字よりも多く，分布特性も異なることから，復号手法が単純であっても実用時間内で解読することは困難と主張した．しかし，これらの仮説は約 18 年前の技術環境に基づくものであり，現代の計算環境，統計的手法を前提とした再評価が必要であるという点が，本研究の問題設定である．

本研究では初めに，文字列を対象としたときの単一置換暗号に対する頻度分析法の特徴を評価した．頻度分析に基づく 1 から 3 -gram 頻度表を用い，単一置換暗号化テキストを復号したところ，1-gram の一致率は 46 % , 2 -gram は 36 % , 3-gram は 49 % を示した．これにより，単純頻度分析では適切に復号できないことが明らかになった．

次に，命令列を対象としたときの単一置換暗号に対する頻度分析法の特徴を評価した．実験には，複数の C 言語のコードを用い，mnemonic の 1 から 3 -gram 頻度を収集した．その結果，命令列において高頻度に現れるパターンが存在し，mov や cmp などの命令列の出現率が全体の 40 から 70 % を占めた．そこで，文字列と同様に 1 から 3 -gram 頻度表を用い，単一置換暗号化テキストを復号したところ，1-gram の一致率は 50 % , 2 -gram は 44 % , 3-gram は 45 % を示した．これにより，単純頻度分析では適切に復号できないことが明らかになった．

文字列と命令列を対象としたときの単一置換暗号に対する頻度分析法と，最適化手法を組み合わせた手法の復号精度を評価した．最適化には，n-gram の log 確率をスコア関数として用いた．文字列では最適化手法として，貪欲法と焼きなまし法を採用し，それぞれ一致率は 97.9～ 98.1 % , 98.1 ～ 99.3 % と高い値を示したことから，単純頻度分析と最適化手法を組み合わせることで，文字列の置換暗号は高精度に復号可能であることを確認した．また，命令列では，最適化手法として貪欲法を採用し，評価を行った．貪欲法では一致率が 13.5 % ～ 33.3 %を示し，文字列の時とは異なり高い精度での復号が確認できなかった．そこで，スコア関数を複数の n-gram の log 確率の合成値とした時の一致率の評価を行った．一致率は 3.7 % ～ 6.4 % となり復号精度の向上は確認できなかったが，実用時間は 21.8～ 29.6 秒と実行不可能な時間はかからなかった．

本研究では，現代の計算能力と統計モデルを用いることで，部分的には命令列を対象とした暗号文の復号が可能であることが示された．しかし，高い精度の復号を実現するには至っていないため，命令列でより有効な復号手法を見出すことが今後の課題である．

発表する際の課程を記入
（学籍が他課程所属の学生も発表する課程を記入すること）

電気・電子情報工学　課程

発表番号 91