様式 C -19

科学研究費補助金研究成果報告書

平成 21 年　3 月 31 日現在

研究種目：基盤研究(B)
研究期間：2005～2008
課題番号：１７３６０１７４
研究課題名（和文）符号木の特性解析とそのデータ圧縮および暗号への応用

研究課題名（英文）Analysis of Coding Tree Characteristics and its Applications to Data Compression and Cryptosystems

研究代表者
山本 博資（Yamamoto Hirosuke）
東京大学・大学院新領域創成科学研究科・教授
研究者番号：30136212

研究成果の概要：
さまざまな用途の符号化で利用されている符号木(Coding Tree)に関する未解決問題やその応用に関して多くの研究成果を得た．具体的には，データ圧縮用の符号である固定長-可変長符号(FV 符号)の同期系列の特性評価，FV 符号の一般情報源に対する性能評価，符号木を動的に更新するスプレイ符号の性能評価，動的再起的に符号木を成長させる T-code の性能評価およびその乱数検定への応用，符号木を用いた放送型暗号方式のグループ鍵更新法，などに関して理論的な解析を行なうと共に，新しい応用法を明らかにした．

交付額

（金額単位：円）

	直接経費	間接経費	合　計
２００５年度	2,100,000	0	2,100,000
２００６年度	2,600,000	0	2,600,000
２００７年度	2,600,000	780,000	3,380,000
２００８年度	2,500,000	750,000	3,250,000
年度			
総　計	9,800,000	1,530,000	11,330,000

研究分野：工学
科研費の分科・細目：電気電子工学
キーワード：符号木，FV 符号，一般情報源，同期系列，データ圧縮，暗号システム

１．研究開始当初の背景

符号木(Coding tree)は，データ圧縮やデータ検索などで使用される基本的なデータ構造であり，1950 年代から数多くの研究がなされて来ている．しかし，符号木を動的に更新しながら符号化を行なうスプレイ符号や T-code などに関しては理論的解析が困難であるため，多くの未解決問題が残されていた．また，符号木を暗号システムや乱数検定に応用するという新しい分野に関しても，まだ，十分に研究がなされていない状態であった．

２．研究の目的

本研究では，符号木の未解決問題，特に符号木の動的な特性に関する幾つかの重要な未解決問題を解析すると共に，符号木の新しい応用分野である暗号システムへの応用に関して，理論的性能評価を行なっている．具体的には下記の項目を明らかにすることを目的として研究を行なった．

(1) FV 符号(固定長可変長符号)の同期系列の特性評価を，符号木の形状に基づく分類を利用して行なう．
(2) FV 符号の一般情報源に対する性能評価を符号

木の性質を用いて行なう.
(3) 動的に大きく符号木が更新されるスプレイ符号の性能評価を, 理論およびシミュレーションにより行なう.
(4) 動的かつ再起的に符号木を成長させる T-code の特性を理論的に評価する. さらにその特性を用いて, 乱数検定への応用法を示す.
(5) 符号木の放送型暗号システムへの応用法およびその性能評価を理論的に行なう.

３．研究の方法

本研究は, 情報理論的なさまざまな手法を用いて, 上記研究目的欄の(1)-(5)の項目に対して研究を行なったが, その手順は主に下記の通りである.
(1) 問題点の整理
(2) 関連研究の調査
(3) 定理の予想／新しい応用や符号化法の提案
(4) 定理の証明／提案方式の性能評価
(5) 証明の検証／評価の理論的な検証
(6) シミュレーションによる性能評価
(7) 学会での発表および他の研究者との討論により, 理論の正しさの検証を行なう. 問題点や新しい発想・着眼点が得られれば上記(3)に戻って, 理論の拡張を行なうという繰り返しを行い, 理論を洗練化している.

なお, 上記のうち, (1)-(4)は研究代表者が研究協力者の協力を得て行なった. また, (5)は連携研究者が主に担当した. (6)は連携研究者と研究協力者が分担して行なった.

４．研究成果

(1) FV 符号では, 伝送中の雑音などで符号語にビット誤りが生じると, 符号語の区切り位置がずれ(同期がずれ), その後の復号が連続して誤るという「復号誤り伝播」が生じる欠点がある. しかし, 符号木に「同期系列」が存在すると, 同期がずれていてもその同期系列により自動的に同期が回復する. 今までは, どのような符号木が同期系列を持ち, どのような符号木が同期系列を持たないかが知られていなかったが, 本研究では, 同期系列に関して下記の事柄を明らかにした.
① 符号（あるいは符号木）が与えられたとき, その符号が同期系列を持つか否かを判定するアルゴリズム, および全ての同期系列を見つけることができるアルゴリズムを与えた.
② 同期系列を持たない符号木の特徴を, 整理分類した.
③ 符号木の葉の数 n が与えられたとき, 符号木の種類の総数は n の Catalan 数で与えられるが, 同期系列を持たない符号木の総数を②の分類を利用して, 理論的に導出し, n の Fuss-Catalan 数を用いて表現できることを示した.
④ １つの系列が２つ以上の符号木の同期系列になっているとき, その系列をそれらの符号木の共通同期系列というが, 与えられた複数の符号木の共通同期系列を求めるアルゴリズムを与えた.
⑤ ２つの符号木の共通同期系列を利用することにより, 与えられた符号語の系列から, その符号化に使用されている符号木を同定する手法を与えた. このことにより, データ圧縮用の木符号を利用した簡易暗号は, 共通同期系列を利用することにより暗号解析ができる可能性があることが明らかになった.

(2) 確率的な解析を理論的に行なう場合に, 一般的に情報源の定常性やエルゴード性を仮定することが多いが, 定常性やエルゴード性などを仮定しない理論解析手法として「情報スペクトル手法」がある. 本研究では, 定常性やエルゴード性などを仮定しない「一般情報源」に対する FV 符号の性能に関して, 情報スペクトル手法を適用することにより, その性能を評価した. 詳しくは, 一般情報源に対する平均最適な FV 符号がどのような性質を満たさなければならないかを解析し, 最適な符号の場合に, 正規化符号長と正規化自己情報量の差が, 情報源出力データの系列長が長くなるにつれて, 確率的にゼロに収束することを明らかにした.

(3) FV 符号など符号木をデータ圧縮に利用する場合, 固定した符号木を用いるより, 符号木をデータに依存して動的に変化させながら利用する方が, データの特性変化への追随性がよくなる. 他方, データが定常な場合は, 動的に大きく符号木を変化させると逆に性能が悪化する. 動的に符号木を大きく変化させる符号として, スプレイ符号が知られているが, 静的な符号木を用いる場合に比べて, 動的な符号木を利用した符号は理論的な性能評価が非常に困難となる. そのため, スプレイ符号の性能評価は, 従来十分に研究されていなかったが, 本研究では, 下記の事柄を明らかにした.
① ２分枝のスプレイ符号を, 一般のα分枝の場合に拡張すると共に, その理論的な性能評価を理論およびシミュレーションにより行ない, 下記のことを明らかにした.
② α分枝のスプレイ符号を, エントロピー $\mathrm{H}(\mathrm{X})$ を持つ長さ N の定常無記憶データに適用したとき, その符号語長が $2\mathrm{H}(\mathrm{X})+3+c/N$ 以下となる(c は定数). つまり, スプレイ符号は動的な追随が非常に速いだけでなく, データが定常なときでも, その悪化の程度は最適な場合の２倍以下で収まる. その結果, トータルの性能が

比較的よく，その傾向はシミュレーション結果とも一致している．

③ α分枝のスプレイ符号の場合，兄弟節点が増えるため，節点の入れ換えを行なうときに，親節点のどの兄弟節点と入れ替えるのが最適になるかが分からない．本研究では，
- 右隣選択（親の右隣の兄弟節点と入れかえる）
- 右端選択（右端の兄弟節点と入れ替える）
- 最大順位選択（順位の一番大きな兄弟と入れ替える）
- 最小順位選択（順位の一番小さい兄弟と入れ替える）

の 4 通りを考え，親節点の順位付けについても
- 変化優先順位方式（生起した節点の更新後の順位を一番高くする）
- 順位保存順位方式（生起した節点の更新後の順位を一番低くする）

の２通りを，子節点の順位付けについては
- 最小順位方式（入れ替えられた節点の順位をゼロにする）
- 最大順位方式（入れ替えられた節点の順位を最大にする）
- 順位保存方式（入れ替えられた節点を含めて順位を保存する）

の３通りの方式を考えた．これらの全ての組み合わせに対して，カルガリーコーパスを用いて性能評価をした．その結果，親の兄弟節点の選択に対しては「最大順位選択」が，親節点の順位付け規則については「順序保存方式」が，子節点の順位付け規則については「最小順位方式」の性能が一番よいことが分かった．

(4) T-code は，1984 年に Mark Titchener が考案した動的かつ再起的に符号木を成長させながら符号化する符号である．T-code の符号木は再起的な T-augmentation アルゴリズムにより生成されるが，その逆操作である T-decompression を用いて，与えられたデータ系列を T-code の符号木の最長符号語系列に対応させることができる．その T-decompression によるデータ系列の分解数を，T-complexity と呼ぶが，T-complexity は，データ系列の複雑度を表す１つの指標となる．この T-complexity に関して，本研究では下記の成果を得た．

① T-complexity が最大になるデータ系列に対して，T-complexity がデータ系列長 n のどのような関数になるかを理論的に導出するための手法（微分方程式近似手法）を提案した．その結果，従来シミュレーション結果のグラフマッチングで予想されていた結果が理論的に導出できることを明らかにした．さらに，データ系列が T-complexity 最大系列でなく，完全な乱数系列である場合に対しても同様の手法が適用できることを示し，乱数系列の場合の T-complexity の大きさが系列長 n のどのような関数になるかを理論的に導出した．

② T-complexity を用いた新しい乱数検定手法を提案した．
米国商務省標準技術局 NIST(National Institute of Standards and Technology) は暗号用の乱数検定ツールを定めている．その中に，LZ(Lempel-Ziv)符号に基づく LZ-complexity を利用した LZ 乱数検定法が含まれていたが，良好な乱数が棄却される割合が期待値より大きくなる欠点があり，NIST 乱数検定ツール Version 1.7 からは, LZ 乱数検定法は除外されている．T-complexity は LZ-complexity の拡張と見なすことができるため，T-complexity を乱数検定に利用することが考えられるが，本研究では具体的に T-complexity を用いた乱数検定の詳しい手順（T 乱数検定法）とその性能を理論およびシミュレーションにより評価した．その結果，LZ 検定で生じるような問題点が生じないことを明らかにすると共に，偏った擬似乱数系列で，NIST の他の検定法では棄却されないが，T 検定で棄却できる擬似乱数系列が存在することを示した．

(5) 放送型暗号(Broadcast encryption)では，契約した正規の受信者のみに情報を配信する必要があり，そのためにグループ鍵を用いて暗号通信がなされる．しかし，正規受信者のグループから脱退する者や新たに加入する者が生じた場合に，そのグループ鍵を安全にかつ効率よく更新する必要がある．そのグループ鍵更新法に関して下記の成果を得た．

① 正規の受信者を論理的な木構造を用いて分類する LKH(Logical Key Hierarch)方式において，グループ鍵およびサブグループ鍵を非常に単純なアルゴリズムで更新する Slcuk-Shidu 方式を取り上げ，その脱退時の鍵更新コストが漸近的に最適であることを理論的に明らかにした．Slcuk-Shidu 方式では脱退時の鍵更新コストだけを考えているが，加入時と脱退時の両方の鍵更新コストを考えた方式に簡単に拡張することができ，その場合でも脱退時の鍵更新コストが漸近的に最適になることを示した．

② 情報コンテンツごとで契約グループが異なり，それらが束(lattice)構造的な階層構造をなしているとき，その束構造に離散

対数問題をうまく組み込むことにより，脱退／加入時に，単に素数を公開するだけで安全にコンテンツごとに新しいグループ鍵を更新できる新しい鍵更新方式を提案した.

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 1 件）

① H.Koga and H.Yamamoto, “ Asymptotic properties on the codeword lengths of optimal FV cods for general sources, ” IEICE Transactions on Fundamentals, vol.51, no.4, pp.1546 -1555, 2005（査読有）

〔学会発表〕（計 11 件）
【国際シンポジウム】
① K.Hamano and H.Yamamoto, “ A randomness test based on T -codes, ” ISITA20008 (2008 International Symposium on Information and its Applications), pp.1095 -1100, Dec.7 -10, 2008, Auckland New Zealand (査読有)
② T.Honda, Y.Tashiro, H.Yamamoto, “ FV code trees with no self -synchronizing string, ” ITA workshop Paper -378, Jan. 29 -Feb.3 2007, San Diego, USA
③ Y.Tashiro and H. Yamamoto, "FV code trees with no synchronizing string," AEW5, Oct.25 -27, 2006, Jeju, Korea
④ H.Sakai and H.Yamamoto, “ Asymptotically optimality of tree -based group key management schemes, ” IEEE -ISIT2005 (International Symposium on Information Theory), pp.2275 -2279, Sep.4 -9, 2005, Adelaide Australia （査読有）,

【国内シンポジウム】
⑤ 濱野健二, 山本博資, ”T -complexity 最大系列の特性について, ”第 31 回情報理論とその応用シンポジウム, pp.1011 -1016, 2008, 鬼怒川
⑥ 濱野健二, 山本博資, ”T -code を用いた乱数検定法, ” 電子情報通信学会技術報告, IT2007 -55, pp.43 -50, 2008
⑦ E.Taniguchi and H.Yamamoto, “ Broadcast encryption over multiple hierarchical service set, ” SCIS2008 (Symposium on Cryptography and Information Security, 4A2 -1, 2008
⑧ 水門善之, 山本博資, ”多分木 Jones 型スプレイ符号の符号木更新アルゴリズム, ”電子情報通信学会技術報告, IT2006 -1118, pp.197 -202, 2007
⑨ 水門善之, 山本博資, ”多分木スプレイ符号の圧縮率に対する理論的性能評価 ”,29 回情報理論とその応用シンポジウム, pp.521 -524, 2006,函館
⑩ 田代雄介, 山本博資, ” 同期系列を持たない FV 符号木 ”,電子情報通信学会技術報告, IT2006 -28, pp.19 -24, 2006
⑪ 大川徳之, 原田邦彦, 山本博資, "反辞書法に基づくモデル選択と算術符号を用いたデータ圧縮", 信学技報, IT2005 -49, pp.41 -46, 2005

６．研究組織
(1)研究代表者
山本博資 (Yamamoto Hirosuke)
東京大学・大学院新領域創成科学研究科・教授
研究者番号：30136212

(2)研究分担者（平成 17 年度～平成 19 年度）
- 古賀 弘樹 (Koga Hiroki)
  筑波大学・大学院システム情報工学研究科・准教授
  研究者番号：20272388
- 有村 光晴 (Arimura Mitsuharu)
  湘南工科大学・工学部・講師
  研究者番号：80313427

(3)連携研究者（平成 20 年度）
- 古賀 弘樹 (Koga Hiroki)
  筑波大学・大学院システム情報工学研究科・准教授
  研究者番号：20272388
- 有村 光晴 (Arimura Mitsuharu)
  湘南工科大学・工学部・講師
  研究者番号：80313427

(4)研究協力者
- 濱野 健二 (Hamano Kenji)
  東京大学・大学院新領域創成科学研究科・大学院生(博士課程)
- タニグチ エリオット (Tanguchi Elliot)
  東京大学・大学院新領域創成科学研究科・大学院生(博士課程)
- 水門善之（Suimon, Yoshiyuki）
  東京大学・大学院新領域創成科学研究科・大学院生（修士課程, 2005 -2006 当時）
- 田代雄介（Tashiro, Yusuke）
  東京大学・工学部 4 年生（2005 当時）