様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２５年　６月５日現在

機関番号：12601
研究種目：挑戦的萌芽研究
研究期間：2010～2011
課題番号：22656085
研究課題名（和文）　T-complexity を用いた暗号用乱数検定法の開発

研究課題名（英文）　Development of a randomness test method based on T-complexity for cryptography

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（和文）：
T-複雑度(T-complexity)に基づく T-乱数検定法を提案し，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-複雑度に基づく LZ-乱数検定法における大きな欠点である複雑度の分布が離散的になるという欠点が T-複雑度にはなく，理想的な正規分布を持ち，性能のよい乱数検定が行えることを示した．さらに， T-複雑度と LZ-複雑度の中間的な性質を持つ RP-複雑度を定義し，その RP-複雑度に基づく RP-乱数検定法についても性能評価を行った．

研究成果の概要（英文）：
T-randomness test is proposed based on T-complexity and the performance of the test is evaluated by theoretical analyses and simulation. LZ-randomness test based on LZ-complexity has a defect such that its probability is discrete. But, it is shown that T-complexity has the ideal normal distribution and T-randomness test works well. RP-complexity, which has an intermediate characteristic between LZ-complexity and T-complexity, is also defined and evaluated.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2010 年度	1,700,000	0	1,700,000
2011 年度	1,400,000	420,000	1,820,000
年度			
年度			
年度			
総　計	3,100,000	420,000	3,520,000

研究分野：工学
科研費の分科・細目：電気電子工学，通信・ネットワーク工学
キーワード：T-複雑度，LZ-複雑度，乱数検定，ユニバーサルデータ圧縮符号

１．研究開始当初の背景

暗号・情報セキュリティシステムの安全性は，秘密鍵などで使用される乱数の安全性に大きく依存している．しかし，真性乱数の生成はコストがかかるために，多くの場合擬似乱数が用いられている．そのため，擬似乱数が真性乱数と区別

が付かない良い乱数であるか否かを判定する乱数検定は，非常に重要な役目を担っている．NIST（National Institute of Standards and Technology）の乱数検定ツールは，世界的に最も広く利用されている．しかし，それに含まれていたLZ-乱数検定法が，第一種の誤り確率が大きいため削除された．その結果，NISTにはユニバーサルデータ圧縮理論に基づく検定法が一つも組み込まれていない．この大きな欠点を改善するために，LZ-乱数検定法に代わる有用な検定法の開発が急務となっていた．

２．研究の目的

T-符号(T-code)で使用されていたT-分解(T-decomposition)で，系列を分解したときの分解数として定義されるT-複雑度(T-complexity)を用いた乱数検定法(T-乱数検定法)を与えると共に，その性能評価を理論およびシミュレーションにより評価することを目的とした．具体的な目的は下記の項目であった．

(1) 与えられた系列のT-複雑度を効率よく求めるアルゴリズムを開発する．

(2) T-複雑度に基づく乱数検定を，NISTの乱数検定ツールに合わせた仕様で作る．

(3) T-複雑度に関する理論解析を行なう．

(4) 幾つかの検出が困難な偏りのある乱数系列を用いて，NIST乱数検定ツールとT-乱数検定法との性能を比較する．

(5) 他の複雑度に関しても検討を行う．

３．研究の方法

上記「研究目的」の(1)-(3)に関しては，理論的な研究のため，研究代表者が連携研究者や研究協力者等と議論を重ねることにより，実施した．また，(4)の性能比較は，パソコンを用いてシミュレーションを行うことにより，性能評価を行った．

4．研究成果

(1)従来のT-分解は，系列の最後から前向きに分解するため，全ての系列を読み込んでから分解を始めるオフラインアルゴリズムであった．それを，系列の前方から後方に向けてオンラインで分解できるアルゴリズムを開発した．さらに，分解に使用する分解木のサイズが大きくならない効率のよいアルゴリズムを与えた．

(2) LZ-複雑度は，その累積確率分布が図1のような不連続な分布となるが，T-複雑度は図2に示すように，ほぼ理想的な正規分布を持つことを明らかにした．

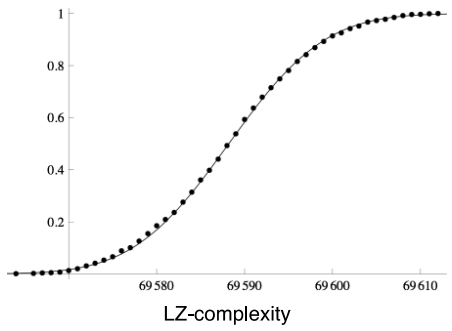

図1 LZ-複雑度の累積分布

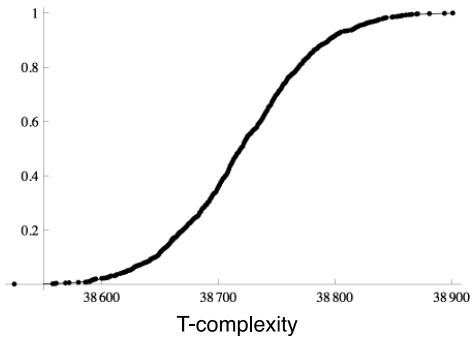

図2 T-複雑度の累積分布

(3) T-複雑度に基づくT-乱数検定法を，NISTの乱数検定法の仕様に合わせて，次のような手続きとして構成した．

Test Procedure

C1 Set α and I to a given significance level and a given trial number, respectively, e.g. $\alpha = 0.01$ and $I = 10^3$.

C2 Generate a sequence of length N. Compute the T-complexity t of the sequence.

C3 Compute $z = \frac{t-\mu}{\sigma}$.

C4 Compute P-value= $\mathrm{erf}\left(\frac{|z|}{\sqrt{2}}\right)$.

C5 If the number of trials is less than I, go to C2.

C6 Compute $r_\alpha = \frac{\#\{\text{P-value} : (\text{P-value}) \geq \alpha\}}{I}$.

C7 Compute $\xi = \frac{r_\alpha - (1-\alpha)}{\sqrt{\frac{\alpha(1-\alpha)}{I}}}$.

C8 Test the null hypothesis $H_0 : \xi \sim \mathrm{N}(0, 1)$.

C9 If H_0 is rejected, conclude that sequences are non-random.

(4) 乗算合同法で生成される擬似乱数系列は，3バイトごとの数値を3次元空間上にプロットすると一様分布にはならず，格子状に偏った分布となることが知られている．NISTの乱数検定ツールなどではこの偏りをうまく検出できないが，T-乱数検定を用いると，この擬似乱数の偏りを検出できることを明らかにした．また，$Y_1Y_2Y_3\cdots$の各Y_iを8ビットとし，Y_{3j}, Y_{3j+1}は真の乱数で生成し，Y_{3j+2}を$Y_{3j}+Y_{3j+1}$の下位8ビットとしたとき，NISTの乱数検定ツールではうまく検出できない．しかし，T-乱数検定を用いると，この偏りをうまく検出できることを明らかにした．

(5) T-複雑度は，LZ-複雑度のような欠点（分布が離散的になるという欠点）が無く，T-複雑度に基づくT-乱数検定はよい性能を有している．一方，図3に示すように乱数系列のLZ-複雑度はLZ複雑度の最大値にほぼ等しいが，乱数系列のT-複雑度は，図4に示すようにT-複雑度の最大値をよりかなり小さい．

乱数系列は圧縮不可能な系列であり，乱数の複雑度が全ての系列の中で最も大きな複雑度を持つと，人間の直観によく一致する．しかし，T-複雑度はその条件を満たしていない．

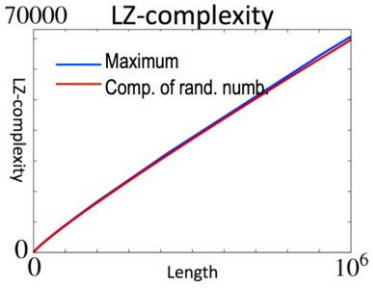

図3 乱数系列の LZ-複雑度

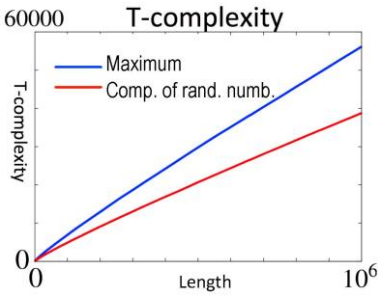

図 4 乱数系列の T-複雑度

そこで，LZ-複雑度の欠点が生じず，かつ乱数系列の複雑度がほぼ複雑度の最大値に等しくなる複雑度を新たに定義する．

LZ-複雑度では，増分分解(incremental parsing)を行う分解木を用いて部分系列pが切り出される．そのとき，図5のように分解木は一つの葉から2つの枝だけを伸ばして成長する．これに対して，T-分解では，図6のように葉に分解木そのものを接続して成長させる．そのため，LZ-複雑度を求めるための分解木は成長が非常に遅く，T-分解の分解木は成長が非常に速い．LZ-複雑度およびT-複雑度のそれぞれの欠点は，これらの分解木の成長の速度に由来している．そこで，これら両者の中間的な成長を実現する畳語分解を図7のように定義する．これは，切り出された系列pのみを分解木に付加することで分解木を成長させるものである．

この分解木を用いると，切り出される系列pは以前に切り出された系列qを用いて$p=qq'u$と表せる．ここで，q'はqの語頭，uは0または1である．つまり，qの繰返しを利用して切り出される．そのため

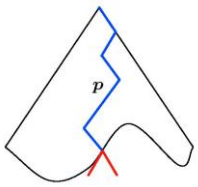

図 5 増分分解木

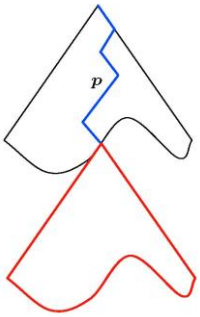

図 6 T-分解木

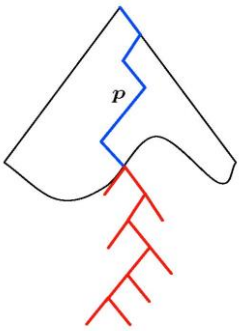

図 7 畳語分解木

この分解法を畳語分解(reduplication parsing)と名付けている．また，系列を畳語分解で分解したときの分解数をRP-複雑度と定義する．

このRP-複雑度を用いて，T-乱数検定と同様に，RP-乱数検定法を構成できる．このRP-乱数検定法は，T-乱数検定法より若干検出能力は劣るものの，(4)で述べたような検出が困難な偏った乱数系列を検出することができる．一方，乱数系列のRP-複雑度は，図8に示すように，最大RP-複雑度に近い値を持つ．

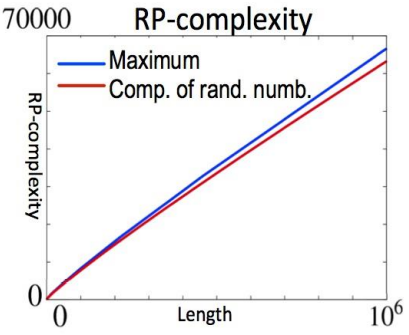

図 8 乱数系列の RP-複雑度

(6) LZ-複雑度，T-複雑度，RP-複雑度の特徴をまとめると次のようになる．

A:真の乱数系列の複雑度の分布が正規分布をする．
B: 真の乱数系列は，全ての系列の中で最も大きな複雑度を持つ．

上記 A と B の指標に対する各複雑度の特徴を表 1 に示す．

表 1　各複雑度の比較

	A	B
LZ-複雑度	bad	very good
T-複雑度	very good	bad
RP-複雑度	good	good

T-複雑度の基づくT-乱数検定法およびRP-複雑度に基づくRP-乱数検定法は，LZ-複雑度に基づくLZ-乱数検定法のような欠点がなく，LZ-乱数検定法が削除されてしまったNIST乱数検定ツールを補完する検定法として使用できる．

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. Kenji Hamano and Hirosuke Yamamoto, A Randomness Test based on T-Complexity, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 査読有り, vol.E93-A, No.7, pp. 1346-1354, July 2010

〔学会発表〕（計１件）
1. 真矢滋，山本博資，畳語分解に基づく文字列の複雑度の提案とその乱数検定への応用，電子情報通信学会情報理論研究会，技術報告 IT2013-8, pp.35-40, 2013 年 5 月 24 日，福井県あわら市

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
（なし）

(3)連携研究者
國廣　昇（KUNIHIRO NOBORU）
東京大学・大学院新領域創成科学研究科・准教授
研究者番号：60345436

岩本　貢（IWAMOTO MITSUGU）
電気通信大学・先端領域研究センター・特任准教授
研究者番号：50377016

(4)研究協力者
元大学院生：濱野　健二（HAMANO KENJI）
元学部学生：真矢　滋（MAYA SIGERU）