数理解析研究所講究録
1504巻 2006年 123-135

線形合同法による暗号化用擬似乱数生成法

木村大生*，齋藤誠慈*
*:大阪大学大学院情報科学研究科情報数理学専攻

Generator of Pseudo-random Number for Encryption by Linear Congruential Method

Motoki Kimura*, Seiji Saito*
*Graduate School of Information Science and Technology, Osaka University

アブストラクト

本論文では，線形合同法と閾値関数によってデジタル計算の丸め誤差を時系列の生成に影響を与えない擬似乱数の生成法を提案し，生成された二値の擬似乱数列に対して，暗号化への適用を考慮した強度評価を行う．

1． 暗号化手法

現代暗号では，暗号アルゴリズムを公開することで，暗号自体の仕組みが分かり，暗号化における安心感が得られる．そして暗号鍵のみを秘密にすることにより暗号文から平文を得ることができないようにしている．このことにより，暗号鍵をもたない第三者には，暗号文から平文を解読することが困難になる． 従って，暗号鍵の作成と管理の方法が重要となる．図1は暗号通信の様子を示したものである．

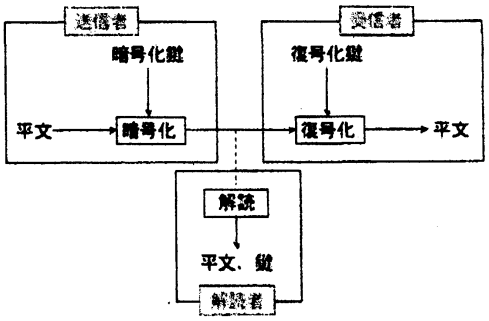

図1 暗号通信の様子

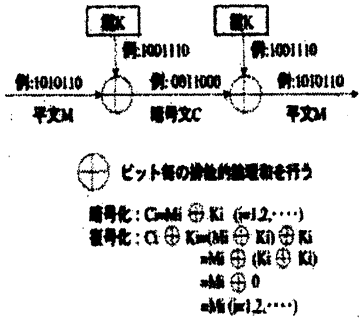

図2 ストリーム暗号

送信者は，伝送したい平文を暗号化鍵により暗号化して送信する．受信者は，受け取った暗号文を復号化鍵により復号して平文情報を得る．送信路において，送信情報を得ようとする第三者を解読者と呼び，解読者は送信情報を不正に取得し，様々な攻撃法により平文情報を得ようとする．暗号化における頑強性が低いと，解読される可能性が高

くなる. 図1において暗号化鍵と復号化鍵が同一である暗号アルゴリズムを共通鍵暗号方式と呼び, 異なるものを公開鍵暗号方式と呼ぶ. これらのアルゴリズムにはそれぞれ長所と短所があり, 前者は, 暗号化と復号化で使用する鍵が同じことから, 暗号化してやり取りするごとに個別の鍵を必要とするため, 鍵の管理が煩雑になる危険性がある. また, 暗号化した鍵で復号できてしまうため, 鍵を相手に送る場合は, 第三者に秘密裏に送らなければならず, 鍵の送り方にも注意を必要とする.

これまでのデジタル暗号法は非常に精度が高く, 誤差の影響が同様に働く同機種間での通信にしか使うことができず, 非常に高価で機種依存性の強いものであった. この丸め誤差増大の問題を解決しないことには, デジタル暗号法を実用化するのは極めて困難である. 丸め誤差の問題を解消することで, デジタル暗号方式は擬似乱数の高速生成と, 初期値に大きく依存する性質という頑強な鍵を持った暗号方式となり得る.

ストリーム暗号(図2)の頑強性は鍵系列として用いる乱数の性質に依存する. 本節ではデジタル暗号法, 特にデジタルカオス暗号法に用いる乱数系列の性能評価項目について述べる. 実際に各項目の満たすべき基準値等は評価を行う際に記載する. 暗号用乱数には一般的乱数が重視する統計的性質に加えさらに予測不可能性が求められる. 予測不可能性とは, 乱数の一部から他のビットが予測できないという性質である. 実際には予測不可能性を示すのは容易ではない. そのため, 以下の性質を以て代用することが多い.

長周期性　統計的乱数性　線形複雑度

ただしこれらを満たすことは予測不可能性の必要条件に過ぎない. さらに暗号用擬似乱数に必要な条件には生成する擬似乱数同士の無相関性や, 出力結果から入力に関する情報を得がたいといった条件が求められる. これらに関しては

相関関係　相互情報量

に関してそれぞれの性質を考察する.

2.　線形合同法・閾値関数による擬似乱数発生法

さまざまなシミュレーション結果から, FIPS140-2 の基準を満たし, 擬似乱数を生成するときに適した関数は,「指数項を持ち, 軌道の変化が激しいこと」「滑らかで, 使用区間において微分可能であること」「二値系列生成に関して偏りがないこと」等の特徴を有することである[1]. 上記条件を満たし, 計算速度が速いと考えられる擬似乱数生成の関数として次の線形合同を考える.

$$x_{n+1} = \alpha x_n \quad (mod \quad M)$$
$$\alpha = 2, \quad k = 940087, \quad m = 7427466391 \tag{1}$$
$$M = 742766391$$

k,m は十分大の素数で，x_n は計算過程では整数型により誤差を防ぎ，k < m として，値は1以下の小数値で，初期値は k/m である．α = 2 以外では，良好な結果は得られない[1].

線形合同法で生成される実数値系列 { x_n }から二値系列 { X_n }への変換方法がいくつかある[2]．本研究では，閾値関数法を用いる．生成される実数値系列{x_n}に対し二値系列{ X_n }を次のように定める：

(2) $$X_{n+1}=\begin{cases}0 & (x_n\in[d,u)) \\ 1 & (x_n\in[u,e])\end{cases}$$

ただしu は閾値で d < u < e である．一般的には区間の中点 u = (d+e)/2 を用いる．法 M の値により，擬似乱数列{ X_n }の周期は影響を受ける．

式(1),(2)により生成した二値の擬似乱数系列{ X_n }の統計的乱数性の評価には米国の NIST (National Institute of Standards and Technology,商務省技術標準局) 認定の擬似乱数検定法である FIPS140-2 (Federal Information Processing Standards,商務省連邦情報処理規格) を用いた[2],[3]．FIPS140-2 は暗号モジュールのセキュリティ要件となっており，暗号方式と実装方法の双方を保証するものである．乱数性を保証するには以下の4つのテストに有意水準1%で基準値を満たす必要がある．いずれのテストも生成した{0,1}の二値の擬似乱数 20000 ビットを対象としている．

1) **モノビットテスト**　生成した 20000 ビット(101 ビット-20100 ビット，20101 ビット-40000 ビット) 中の 0，1 の各ビットの個数に関する評価である．0 と 1 での出現個数がほぼ等しいことが望まれる．

2) **ポーカーテスト**　20000 ビット(101 ビット-20100 ビット，20101 ビット-40000 ビット)を4ビットずつ 5000 個の組に分け，区切られた4ビットの出現頻度に関する評価である．区切られた4ビットの値を 10 進法で表現すると 0-15 に分類されるが，0-15 各々の出現回数を，f(i)，I =0,1,…,15,とした時に，$\frac{16}{5000}\sum_{i=0}^{15}f(i)^2-5000$ の値を算出する．f(i)，i =0,1,…,15,のそれぞれの値に偏りがありすぎるとその式の値は大きくなり，基準値を満たさなくなる．

3) **ランテスト**　0 または 1 のビットが連続する個数に関する評価である．同じビットの続く塊をラン(連)とし，ビット値0(または1)が2つ続けば，その塊を0(または1)に関する長さ2のランと呼ぶ．生成した擬似乱数列の中に 0，1 それぞれに関する長さが1ー5 のランと 6 以上のランの出現個数を評価するものである．

4) **長ランテスト**　0 または 1 に関するランの最大値を制限する．

本研究におけるシミュレーションでは初期値を変化させ，各々の初期値について擬似乱数を 25 メガバイト発生させ検定を 1 万回ずつ行った．擬似乱数生成の際，有理数で取った初期値は，法 M の大きさによって，最初の数十ビット以上に 0 が続く可能性がある．また，初期値をわずかに(10^{-10} 程度)変更した場合に，ビット値の変化は 40-50 ビット目あたりから出現し始めた．従って検定の際には生成した擬似乱数の最初の t ビットは検定範囲に含めないようにした．先述のシミュレーション結果では，t=100 として検定範囲を 101 ビット目からとしている．実際に生成した擬似乱数を暗号化に用いる場合には t の値も暗号化鍵として利用することが可能である．式(1)，(2)による擬似乱数列 $\{X_n\}$の FIPS140-2 の検定シミュレーションを試みた結果を以下に示す．閾値 u = 0.5 である

表 1 101-20100bit の結果

	基準値 (NIST)	0の回数	1の回数	判定
モノビットテスト	9725～10275	10061	9939	○
ポーカーテスト	2.16～46.17	5.01		○
ランテスト-1	2315～2685	2427	2449	○
ランテスト-2	1114～1386	1275	1245	○
ランテスト-3	527～723	906	652	○
ランテスト-4	240～384	300	305	○
ランテスト-5	103～209	170	139	○
ランテスト-6 以上	103～209	171	158	○
長ランテスト	25 以下	13	12	○

○=基準内

表 2 20101-40100bit の結果

	基準値 (NIST)	0の回数	1の回数	判定
モノビットテスト	9725～10275	10008	9992	○
ポーカーテスト	2.16～46.17	21.74		○
ランテスト-1	2315～2685	2510	2579	○
ランテスト-2	1114～1386	1239	1167	○
ランテスト-3	527～723	647	628	○
ランテスト-4	240～384	304	314	○
ランテスト-5	103～209	162	165	○
ランテスト-6 以上	103～209	149	160	○
長ランテスト	25 以下	14	12	○

○=基準内

表 3 40101-60100bit の結果

	基準値 (NIST)	0の回数	1の回数	判定
モノビットテスト	9725～10275	9749	10251	○
ポーカーテスト	2.16～46.17	27.70		○
ランテスト-1	2315～2685	2579	2397	○
ランテスト-2	1114～1386	1222	1292	○
ランテスト-3	527～723	625	634	○
ランテスト-4	240～384	291	349	○
ランテスト-5	103～209	147	152	○
ランテスト-6 以上	103～209	136	175	○
長ランテスト	25 以下	12	10	○

○=基準内

式(1)によって生成した擬似乱数 $\{ X_n \}$における初期値の依存性を求める．異なる初期値 x_0，$y_0 = x_0 + \Delta$，$\Delta > 0$，の各々の軌道を $\{ x_n \}$，$\{ y_n \}$として，$L_n = |x_n - y_n|$ とおき，$2^i x_n < M$，$i=0,1,\cdots,k-1$ とすると，次のようになる：

(3) $$L_{k+n} = |2^{k+n}\Delta - iM| \quad (i = 0,1,...,2^n - 1)$$

Δは微小であっても，k が大のとき，L_{k+n} も大きくなる．

式(1)における初期値 k, m をわずかに変更させた場合に生成された 5 通りの擬似乱数 $\{X_n\}$について，同一ビットでの比較を行った．表 4 のように，相違率は 50± 1 程度となった．ただし比較には生成開始 301bit-20300bit までの 20000bit を用いた．

表 4 式(1)の初期値 k, m を変化させたときの 0 と 1 の出現相違の比率

k	940087	940087	940097	940097	940097
m	7427466419	7427466439	7427466391	7427466419	7427466439
相違(%)	49.70	50.64	49.98	49.79	50.38

3. 予測不可能性

本手法(1), (2)で生成する擬似乱数における周期長は, 別の初期値 k, m に換えても初期値に依存せず, 線形合同法における法 M に依存することが実験的に判明した(表 5 を参照).

表 5 初期値 k, m によらず, 法 M の 2 種(7427466391, 7427466419) によって, 式(1)における 1, 2, …, M-1 の出現する周期長は 2 通り (3713733195, 7427466418) となる.

初期値 k	初期値 m	法 M(A, B の 2 種)	周期長 T
940087	7427466391	(A) 7427466391	(a) 3713733195
		(B) 7427466419	(b) 7427466418
	7427466419	(A) に同じ	(a)に同じ
		(B) に同じ	(b)に同じ
	7427466439	(A) に同じ	(a)に同じ
		(B) に同じ	(b)に同じ
940089	7427466391	(A) に同じ	(a)に同じ
		(B) に同じ	(b)に同じ

表 5 における (A), (B) の場合は, 次のように考察できる[7].

(A) $M = 2^n - 1$ のとき, 次のように整数系列は変化して, 周期長は n となる.

$1 \to 2^1 \to 2^2 \to 2^3 \to \cdots \to 2^n = M + 1 \equiv 1 \pmod M$

(B)) $M = 2^n + 1$ のとき, 次のように整数系列は変化して, 周期長は 2n となる.

$1 \to 2^1 \to 2^2 \to 2^3 \to \cdots \to 2^n \to 2^{n+1} = M + (2^n - 1) \equiv 2^n - 1 \pmod M$

ここまで, n+1 回の変化し, さらに n-1 回だけ変化するから, 周期長 2n である.

$2^n - 1 \to 2(2^n - 1) \equiv (2^n + 1) - 2^2 \to (2^n + 1) - 2^3 \to \cdots \to (2^n + 1) - 2^n \equiv 1 \pmod M$

以上から, タイプ(B)のとき, 周期長が長くなり, 擬似乱数生成にはこちらが相応しい.

次に初期値と使用開始ビット数を以下のように一定として, 法を上記のような形に限

らず，一般の数として様々に変えてシミュレーションを行って周期長を算出した結果，その周期長 T にはいくつかのパターンが見られた．k=940087, m=7427466391, t=300$とし，法 M には，基準として用いている M=7427466391 より大きい素数を順に取った．その結果を表 6 に示す．

表 6　法 M は，$2^n \pm 1$ とは異なる 4 つを選び，式(1)，(2)による擬似乱数を生成し 301-20300bit までの結果を検討した．0 と 1 の出現結果は FIPS140-2 の検定法の基準内である．周期長 T は，左から M-1，(M-1)/2，(M-1)/3，(M-1)/2 である．

法		7427466419		7427466439		7427466451		7427466463	
		出現回数		出現回数		出現回数		出現回数	
	基準値（NIST）	0	1	0	1	0	1	0	1
モノビット	9725～10275	9904	10096	10138	9862	9928	10072	9912	10088
ポーカー	2.16～46.17	3.88		12.20		9.02		9.77	
ラン-1	2315～2685	2464	2499	2529	2502	2557	2541	2554	2518
ラン-2	1114～1386	1257	1240	1212	1230	1265	1234	1243	1243
ラン-3	527～723	626	596	646	624	613	614	627	628
ラン-4	240～384	329	311	339	293	304	344	319	340
ラン-5	103～209	161	172	141	151	148	148	153	135
ラン-6以上	103～209	150	160	174	142	155	160	137	162
長ラン	25以下	13	14	11	14	12	13	17	13
周期長		7427466418		3713733219		2175822150		3713733231	

全てのテストにおいて基準内となっている．

法 M が素数のとき，周期長はどのように分布するかを，表 7 に示した．法 M には素数を用いることで，40%近い割合で，理論上の最大周期長である M-1 周期となることがわかった．

表 7　整数 200000 までを 5 区間に分け，各区間の素数を法 M としたとき，式(1)による系列の周期長 T の分布を示した．

区間	～50000	～100000	～150000	～180000	～200000
素数の個数	668	1228	5132	9591	17983
$T = M - 1$	38.2	38.3	37.5	37.6	37.8
$T = (M-1)/2$	28.0	28.3	28.2	28.4	28.2
$T = (M-1)/3$	6.6	7.2	6.9	6.7	6.7
$T = (M-1)/4$	5.1	5.0	4.8	4.8	4.7
$T = (M-1)/5$	2.2	2.0	1.9	1.7	1.8
$(M-1)/10 \leq T \leq (M-1)/6$	10.9	10.8	11.6	11.6	11.6
$T < (M-1)/10$	8.1	8.3	9.1	9.2	9.7
合計(%)	100	99.9	100	100	100

式(1),(2)による擬似乱数の系列を，NISTのFIPS140-2検定によって暗号化用擬似乱数の統計的評価を表7,8に示す．FIPS140-2検定において基準外となった結果に関しても，モノビットテストやランテストの基準値からわずかにずれる程度であるため，式(1),(2)の手法は統計的乱数性の高い擬似乱数を生成すると判断できる(表8)．

表 8　表の左上(301bit-)，右上(20000301bit-)，左下(51860301bit-)，右下(10764031bit-)のFIPS140-2検定結果

	基準値(NIST)	0の回数	1の回数	判定
モノビットテスト	9725～10275	10060	9940	○
ポーカーテスト	2.16～46.17	1.29		○
ランテスト-1	2315～2685	2324	2344	○
ランテスト-2	1114～1386	1282	1241	○
ランテスト-3	527～723	649	650	○
ランテスト-4	240～384	305	317	○
ランテスト-5	103～209	165	138	○
ランテスト-6以上	103～209	172	160	○
長ランテスト	25以下	10	12	○

○＝基準内

	基準値(NIST)	0の回数	1の回数	判定
モノビットテスト	9725～10275	9878	10122	○
ポーカーテスト	2.16～46.17	16.62		○
ランテスト-1	2315～2685	2573	2567	○
ランテスト-2	1114～1386	1252	1262	○
ランテスト-3	527～723	636	642	○
ランテスト-4	240～384	275	293	○
ランテスト-5	103～209	156	172	○
ランテスト-6以上	103～209	172	150	○
長ランテスト	25以下	10	11	○

○＝基準内

	基準値(NIST)	0の回数	1の回数	判定
モノビットテスト	9725～10275	10291	9709	×
ポーカーテスト	2.16～46.17	28.50		○
ランテスト-1	2315～2685	2426	2525	○
ランテスト-2	1114～1386	1249	1245	○
ランテスト-3	527～723	651	649	○
ランテスト-4	240～384	306	296	○
ランテスト-5	103～209	171	147	○
ランテスト-6以上	103～209	182	175	○
長ランテスト	25以下	10	12	○

○＝基準内　×＝基準外

	基準値(NIST)	0の回数	1の回数	判定
モノビットテスト	9725～10275	9927	10073	○
ポーカーテスト	2.16～46.17	32.45		○
ランテスト-1	2315～2685	2585	2442	○
ランテスト-2	1114～1386	1112	1220	×
ランテスト-3	527～723	635	652	○
ランテスト-4	240～384	330	303	○
ランテスト-5	103～209	131	161	○
ランテスト-6以上	103～209	153	156	○
長ランテスト	25以下	16	15	○

○＝基準内　×＝基準外

閾値関数(2)の妥当性を示すために，線形合同法(1)による実数値系列の状態での分布状態を調べる．生成された実数値系列$\{x_n : 0 \le x_n \le 1\}$ の総数 n = 20000 を p 個の均等な区間に分割し， i 番目の部分区間 $[(i-1)/p, i/p]$, $1 \le i \le p$, の出現回数 f(i) に関して，式(4)で統計量 χ^2 を算出し，自由度 p-1 の χ^2 検定を行った([2])．

$$\chi^2 = \sum_{i=1}^{p} \frac{\{f(i)-(n/p)\}^2}{n/p} \tag{4}$$

区間数 p =10 の時の 301 回目から 6 回検定を行った各区間の出現回数 f(i) と統計量 χ^2 は表 9 の通りである．

自由度 9 の χ^2 分布の有意水準 5%は 16.92，1%の限界点は 21.67 である．1000 回分の統計量を算出したところ 1 %の限界点を約 10%の割合で超えてしまう．次に分割区間数を減らし，区間数 p = 2 とし，同様に統計量を算出した．総数 n=1000000 について

6 回検定し，自由度 1 のχ^2分布の有意水準 5%，1%は各々3.84，6.63 である(表 10 を参照)．

表 9　i 番目の部分区間における出現回数 f(i) と統計量 χ^2　(6 回の検定)

f(1)	2054	1991	1805	1928	1932	2067
f(2)	2006	2021	1899	1958	2004	2089
f(3)	2055	2031	1945	1980	2020	2021
f(4)	1919	2026	1953	1979	2090	2046
f(5)	2010	1952	2128	2074	1965	1937
f(6)	2006	2021	1899	1958	2004	2089
f(7)	1968	2036	2000	2006	2112	1978
f(8)	1967	1942	2082	2046	1943	2006
f(9)	2016	1952	2129	2073	1966	1937
f(10)	1993	2028	2160	1992	1958	1830
統計量χ^2	7.024	6.320	64.505	11.185	16.685	30.128
判定(5%)	○	○	×	○	○	×
判定(1%)	○	○	×	○	○	×

これらの結果より，本手法(1)，(2)を用いる際には，閾値を区間の中点 0.5 に設定し，初期値 k と m は，互いに素であればよく，法 M には素数を用いることで，予測不可能性の高い擬似乱数列が得られることを，実例を挙げて確認した．

表 10　○は仮説 H_0[母集団は平均 0.5 の 2 項分に従う]は棄却できないことを意味する．

	1	2	3	4	5	6
f(1)	50000	49743	50211	50075	49993	50004
f(2)	50000	50257	49789	49925	50007	49996
統計量χ^2	0	2.642	1.781	0.225	0.002	0.00064
5%	○	○	○	○	○	○
1%	○	○	○	○	○	○

最後に，式(1)，(2)による擬似乱数の系列に関する線形複雑度を評価する([4])．線形複雑度とは，与えられた系列を生成することの出来る最短の LFSR(線形シフトレジスタ linear feedback shift register)の段数をいう．LFSR と LFSR による擬似乱数生成の

様子を図3に示す．LFSRは式で表すと次式のように再帰的になる．$s_i=c_1s_{i-1}+c_2s_{i-2}+\cdots+c_ns_{i-n}\ (\mathrm{mod}\ 2)$．このとき，次の特性多項式が定義できる．$C(D)=1+c_1D+c_2D^2+\cdots+c_nD^n$ ただし，係数 c_i は法2における値を用いる．線形複雑度を次に示すBerlekamp-Masseyのアルゴリズムを用いて算出した(図4)．

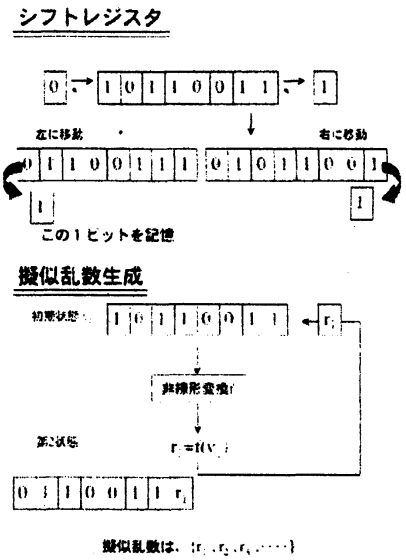

図3　LFSRとその擬似乱数生成

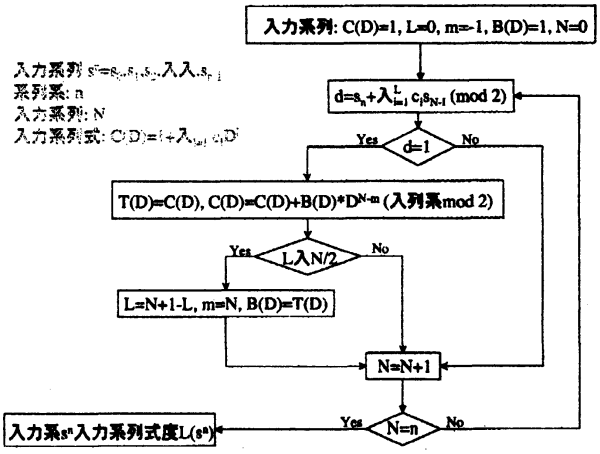

図4　Berlekamp-Masseyのアルゴリズム

系列 $s^n=s_0s_1s_2\cdots s_{n-1}$ の線形複雑度 $L(s^n)$ を算出するには，このアルゴリズムを n 回ループすることになる．初めに初期状態として $C(D)=1, L=0, m=-1, B(D)=0$ を与え，反復変数を N とする．d の値はLFSRにおけるフィードバック値に相当し，各時点における特性多項式 $C(D)$ の係数 c_i , $i=0,1,\cdots,n-1$; $c_i=0,1$, と，次に入力される擬似乱数値で決められる．d の値が1の場合に限り，線形複雑度が高くなる可能性を含むことになる．長さ n の系列に対し，n 回のループを行なうが，k 回目 $(k=1,2,\cdots,n)$ のループの際に算出されている線形複雑度 $L(s^k)$ が $k/2\}$ で増加していることが望ましいとされる．

単純な系列の場合を図5に示す．単純な系列における線形複雑度は，系列長 n に対し，複雑度は $k/2$ で増加せずに，急激に増加するか(レジスタの段数が多数必要)，周期長に依存した値以降は一定になる(擬似乱数の系列長に比して少ない段数で表現可能である)．$L(s^k)$ が極端に小さな値であるか，また極端な増加の仕方をする系列は線形複雑度が低いことになる．

実際に線形合同法で生成した擬似乱数列の系列長1000までの線形複雑度を図6に示す．さらに線形複雑度を算出し，擬似乱数の系列長 n に対し，線形複雑度は $n/2$ に近い形で増加していくことを確認した．

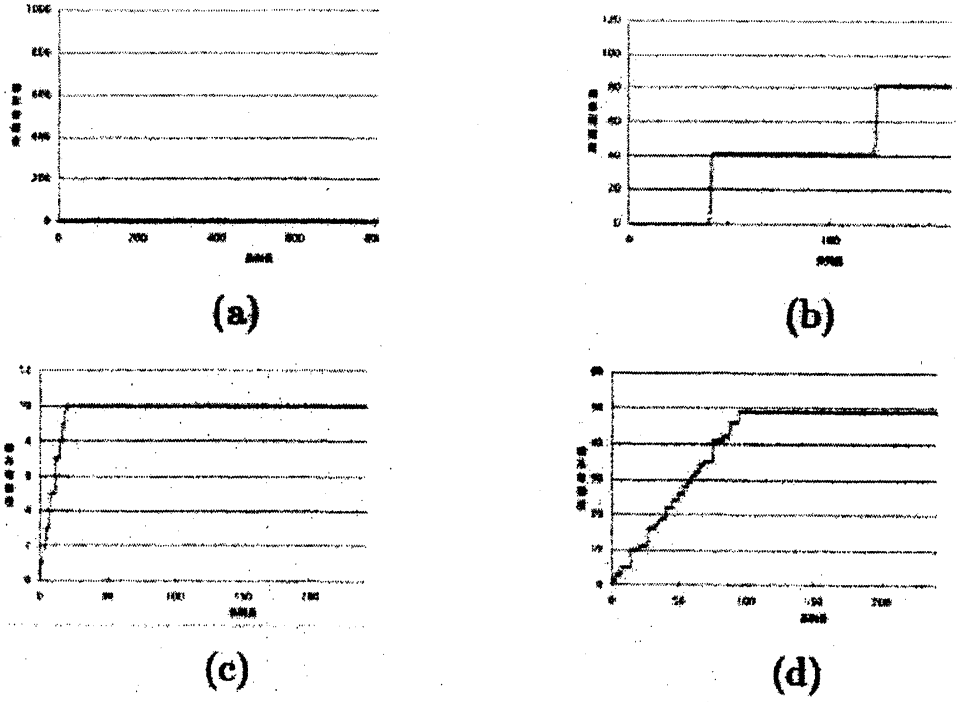

図 5　単純な系列の場合：(a)系列長 1000, 周期 1000　(b)：系列長 200, 周期 200
(c)系列長 300, 周期 10　　(d)系列長 300, 周期 50

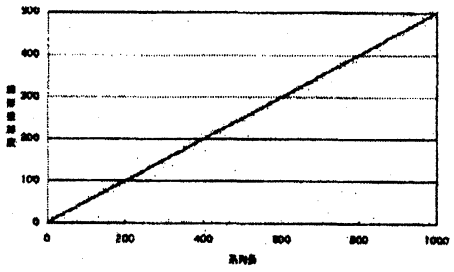

図 6　式(1), (2)によって生成された擬似乱数の線形複雑度

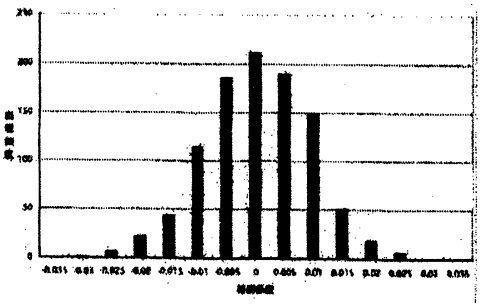

図 7　鍵値 1, 2 における相関係数に対するヒストグラム

4.　相関関係

式(1), (2)により生成する擬似乱数列は使用する鍵値によって決まる．鍵値を変更して生成した異なる二種の実数値擬似乱数系列間の相互相関や，同一の鍵値を用いて生成した一種の実数値擬似乱数列における自己相関値を算出する([2])．相関値の算出には

二値変換前の実数計算(1)を用いていることに留意する.

異なる鍵1,2で生成された各々の系列 $\{x_n\}$, $\{y_n\}$ 間の相互相関係数を20000bit毎に算出した. 相互相関係数は99%以上が相関係数の絶対値が0.025未満となり, 異なる鍵で生成される擬似乱数列間の相関はほとんど見られないことがわかった. 異なる鍵値で生成した擬似乱数列に, 高い相関があれば別々の鍵で生成した擬似乱数列からの類推が懸念されることになる. 相互相関係数 ρ は次式で算出した.

$$\rho = \frac{S_{xx'}}{S_x S_y}, \quad S_{xx'} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_{i+k} - \bar{x}'), S_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}, S_{x'} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{i+k} - \bar{x}')^2}$$

$\bar{x}, \bar{y}$は, それぞれ$\{x_n\}$, $\{y_n\}$の平均値で, N=20000である. 算出された相関係数 ρ に関する度数分布図を区間幅0.005として求めた. 図7において鍵値は, わずかに異なる鍵値1と鍵値2で生成した擬似乱数列の相関係数を1000回算出した場合である.

鍵値1 : k=940087, m=7427466391, M=7427466391, t=300

鍵値2 : k=940089, m=7427466391¥, M=7427466391, t=300

相互相関係数の絶対値の最大は0.028である. 鍵値をわずかに変更するだけで, 生成される擬似乱数系列間の相関はほとんどないと言える.

また同じ鍵で生成された1つの系列に対し, 1bitずつずれを持たせたときの20000bit毎の自己相関係数も算出した. また自己相関係数も数ビットのずれがあるだけで, 相関係数は極端に小さくなるため, 1つの系列内での相関もほとんど見られないことを確認した. 鍵値 k, m, t 等で生成した実数値擬似乱数 $\{x_i\}$と, ビットずれを持った実数値擬似乱数列 $\{x_{i+k}\}$, k = 0, 1, 2, …, 1000, との相関係数 ρ を次式によりによって算出した.

$$\rho = \frac{S_{xx'}}{S_x S_y}, \quad S_{xx'} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_{i+k} - \bar{x}'), S_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}, S_{x'} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{i+k} - \bar{x}')^2}$$

なお, N=20000, $\bar{x}'$は$\{ x_{i+k} : i \geq 0\}$の平均である. 同一の鍵値で生成した擬似乱数列であっても, 数ビットのずれがあると, 系列間の自己相関係数は急激に小さくなり, それらの系列間にはほとんど相関がないと考えられる(図8).

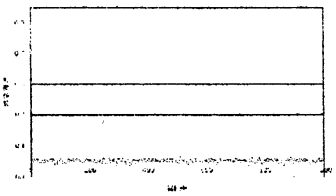

図8 遅れbit k に対する自己相関係数の度数分布

5. 相互情報量

相互情報量を算出することによって，暗号文からの平文の類推のしやすさの程度がわかる．相互情報量を算出する際に用いる平文に様々な偏りを持たせることで，平文が特徴的である時の暗号文への影響を調べることが出来き，擬似乱数そのものの性質評価に繋がる．さらに，算出された相互情報量の値と，用いた平文の分布の関連性を導き，相互情報量から，擬似乱数の統計的な分布の様子を考察する．

相互情報量算出手順は以下の通り 1)-7)である．

1)平文を任意のパターンで準備する．

2)平文の持つ情報量を算出する．実際に『０』と『１』の出現確率を求め，それぞれ p_0，p_1とする．平文の持つ情報量は，得られる情報量の期待値より次式で表される．

$$H = p_0(-\log_2 p_0) + p_1(-\log_2 p_1)$$

3)鍵値を定め，擬似乱数を生成させ，暗号化を施す．

4)暗号文における出現頻度を求める．生成された暗号文に関して『０』,『１』の出現頻度を求め，それぞれ c_0，c_1とする．

5)各ビット毎に平文ー暗号文の組み合わせについて観察し，出現頻度を求める．平文が『０』かつ暗号文が『１』である確率を pc_{01}などとする．

6)暗号文が分かったというの条件付きでの平文の情報量を算出する．暗号文の状態が決定した時の，平文の情報量は次のような式で表すことができる．

(H)

$$H' = c_0\{pc_{00}(-\log_2 pc_{00}) + pc_{10}(-\log_2 pc_{10})\} + c_1\{pc_{01}(-\log_2 pc_{01}) + pc_{11}(-\log_2 pc_{11})\}$$

7)相互情報量 I= H - H' を算出する．

偏りのある平文パターンに対して，擬似乱数の乱数性が十分でなければ，最も単純な暗号文生成法のひとつである，平文と擬似乱数の排他的論理和を用いて生成した暗号文には偏りが現れることになる．その場合，暗号文からの平文推定が容易になり，相互情報量が大きくなる．すなわち，相互情報量の大きさは，擬似乱数そのものの乱数性を評価する基準となり得る．

擬似乱数の乱数性に関連して，擬似乱数の分布と相互情報量の関係を調べる．擬似乱数の一様分布に関する統計量 χ^2 を次式により算出する．

$$(5) \quad \chi^2 = \sum_{i=1}^{p} \frac{\{f_i - (n/p)\}^2}{n/p}$$

本研究では，閾値を 0.5 とした閾値関数を用いて擬似乱数系列を生成しているため，p=1 とした．また相互情報量を算出する際には 80000bit の系列を用いたので，n=80000 と

している．f_iはそれぞれ擬似乱数がi=0,1となる80000bit中の出現頻度である．式(5)で算出された統計量χ^2と式(H)で算出したlbit毎でのそれぞれの鍵値での相互情報量の平均値との関係を次の図9にプロットした．

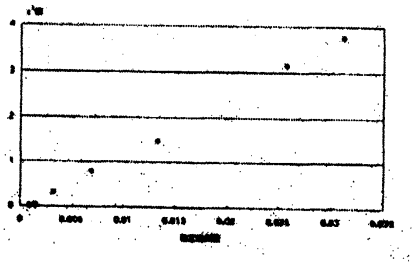

表11　鍵値(k, m, M)

鍵1	0.0597	7427406301	7427406301
鍵2	0.0659	7427406301	7427406301
鍵3	0.0687	7427406419	7427406419
鍵4	0.0681	7427406509	7427406509
鍵5	0.0661	7427406501	7427406501
鍵6	0.0672	7427406439	7427406439
鍵7	0.0667	7427406451	7427406451
鍵8	0.0727	7427406257	7427406257
鍵9	0.0659	7427406499	7427406499
鍵10	384000	68427805317	68427805317

図9　擬似乱数の一様分布性に関する値χ^2と相互情報量との相関関係

鍵値は全10種とし，それぞれの値は表11に示す．擬似乱数として用いない最初のbit数tは300で固定した．図11からわかるように、相互情報量の値と，擬似乱数の一様分布性における平均からのずれを示すχ^2値には強い正の相関関係があることがわかる．相互情報量とχ^2値の間の相関係数は0.9996であった．

6．　今後の課題

本研究では，線形合同法・二値化閾値関数を用いた擬似乱数の生成法を提案しその評価を行った．その擬似乱数は，予測不可能性(長周期性，統計的乱数性)，自己・相互相関関係，相互情報量と統計的性質の関連などの観点からみて良好な結果を得た．

今後は本研究で用いた検定方法や，本手法で算出された検定結果を基準に他の関数や，またカオス関数を用いた際の擬似乱数の生成法の考案が望める．カオス関数の持つ初期値依存性を有効に利用し，暗号化用擬似乱数の生成法の新提案である．カオス関数をそのまま利用するには，丸め誤差の蓄積をはじめとする様々な課題も未解決である．

参考文献

[1] 木村大生：線形合同法関数を用いた暗号用擬似乱数の生成とその強度評価，
大阪大学大学院情報科学研究科修士論文，2007.

[2] 香田徹：離散力学系のカオス，コロナ社，1998.

[3] 米国商務省技術標準局，http://www.nist.gov/

[4] A. J. Menezes et al.:Handbook of Cryptography, CRC Press, 1996.