Cryptanalysis of Linear Congruence Generators

Abstract

Multiplicative congruential generators have been first suggested by D.H.Lehmer as an arithmetic procedure to generate pseudo random numbers. A mild variation of it is the linear congruence generator. Over many years both these generators were widely used in simulations and reported to have good statistical properties and favorable cycle length. Cryptanalysts have come up with numerous complex methods to cryptanalyze the generators mentioned above. We discuss a simple method to cryptanalyze both multiplicative and linear congruence generators, which make them unsuitable as raw input to simulations and various cryptosystem.

If we take n truly randomly numbers between 0 and 1 and truncate them to a finite accuracy, so that each is an integer multiple of $1/v$ for some given value v, then the n dimensional points generated will have an extremely regular structure. Multiplicative and linear congruence generators happen to have such a regular structure. The n-tuples $P_1=(u_1,u_2,\ldots,u_n)$, $P_2=(u_2,u_3,\ldots,u_{n+1}),\ldots$ of uniform variations produced by the generator are viewed as points in the unit cube of n dimensions that has a perfectly regular structure. Furthermore, we show that all the points are found to lie on a relatively small number of parallel hyperplanes.

We further discuss the theory to parallel hyperplanes defined by the equations

$c_1x_1+c_2x_2+\ldots+c_nx_n=0,(+ \text{ or } -)1, (+ \text{ or } -)2,\ldots$

whose solutions are $P_i=(u_i, u_{i+1},\ldots, u_{i+n-1})$, formed by successive tuples of the pseudo random generator. We make use of this property that the points lie mainly in a few hyperplanes to cryptanalyze linear congruence generator. We then illustrate the theory discussed using a practical example.

Introduction

The basic idea is to analyze the set of points generated by

$r_{(i+1)}=k*r_{(i)}+c \bmod m. \rightarrow(1)$

Such a generator (1) is called a linear congruence generator. If c=0, it is a multiplicative congruence generator as shown below (2)

I.e. $r_{(i+1)}=k*r_{(i)} \bmod m. \rightarrow(2).$

Let $r_1,r_2,r_3. . . 0<r_i<m$ be the sequence of residues modulo *m* generated by (2) and let u1,u2,u3, . . . be the sequence viewed as a fraction of *m.*

$\{ 1/m\ (r_1,r_2,. . . r_n) \mid 0<r_i<m)\}.$

Let $P_1=(u_1,. . . ,u_n)$, $P_2=(u_2,. . . . u_{n+1})$, $P_3=(u_3,. . . ,u_{n+2}),. . .$ be points of the unit n-cube formed by taking n successive *u*'s. The arrangement of points in 2-dimensions is as shown below.

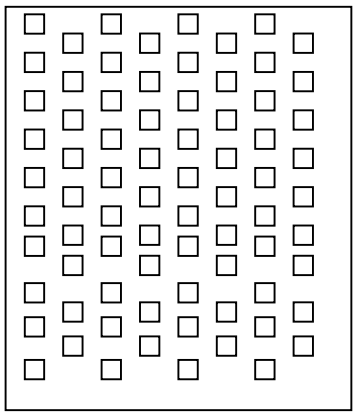


Fig 1. Lattice Structure of points generated by the multiplicative generator

We see that all of the points can be covered by a relatively small number of parallel hyperplanes (Hyperplanes are lines in 2-dimension, since hyperplanes are vectors on a subspace of co-

dimension 1). The points may be covered by a family of vertical lines or diagonal lines or, lines at a given inclination with the co-ordinate axes. In each case, there is a different count on the number of hyperplanes required to cover all the points.

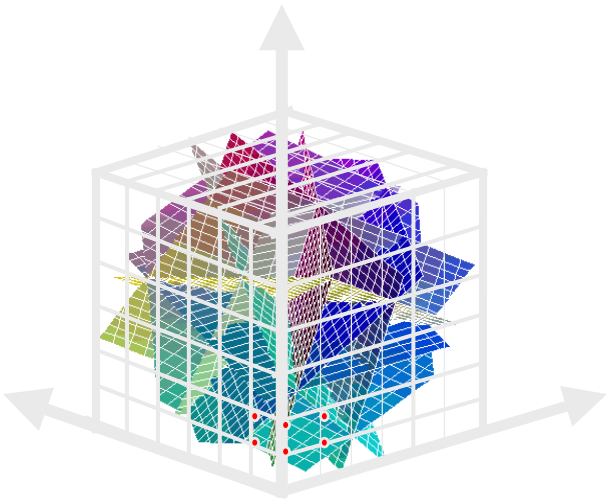


Fig 2. Intersection of hyperplanes in 3-dimension.

Figure 2. can be viewed as the intersection of hyperplanes in 3-dimension.with a unit cube enclosing it. Let $1/v$ be the maximum distance between lines, taken over all the families of parallel hyperplanes in n-dimension, v is called the accuracy of the pseudo random number generator in n-dimensions. If there are m points in n-dimensions, the maximum dimensional accuracy obtained is $v_n<m^{1/n}$. Therefore, the maximum number of hyperplanes required to cover all points is $(n!.m)^{1/n}$. Table below gives the upper bound for the number of hyperplanes containing all n tuples.

	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10
$m=2^{16}$	73	35	23	19	16	15	14	13
$m=2^{32}$	2953	566	220	120	80	60	48	41
$m=2^{48}$	119086	9065	2021	766	391	240	167	126

Table 1

On a binary computer with 32-bit words, $m=2^{32}$, fewer than 41 hyperplanes will contain all 10-tuples and fewer than 566 hyperplanes will consist of all 4-tuples.

We now turn our attention to the equation of hyperplanes. Given below are two hyperplanes of dimension 2.

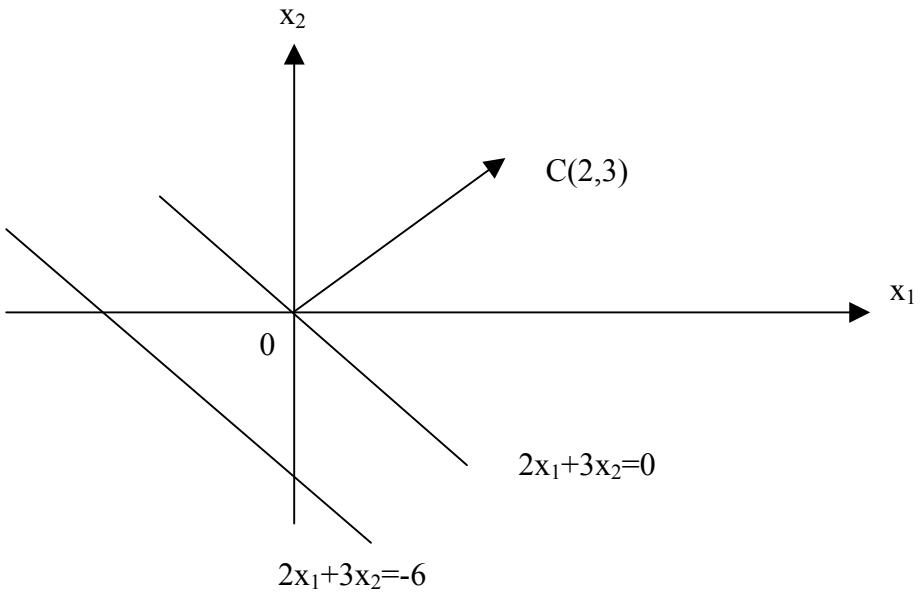


Fig 2. Parallel hyperplanes in 2-dimension

Clearly different families of parallel hyperplanes have different values for *c,* the point normal to the hyperplane. We now try to formulate a mathematical structure for the above discussions. Our goal is to show that all points $P_1, P_2, P_3, \ldots$, of the multiplicative congruence generator will lie on a set of parallel hyperplanes.

Theorem: If $c_1,c_2, \ldots,c_n$ are any choice of integers such that

$c_1+c_2k+c_3k^2+ \ldots + c_nk^{n-1}==0$ modulo m, then all the points $P_1, P_2, P_3, \ldots$ will lie on a set of parallel hyperplanes defined by the equation

$c_1x_1+c_2x_2+\ldots+c_nx_n=0,(+ \text{ or } -)1, (+ \text{ or } -)2, \ldots$ such that all the points fall in fewer than $(n!.m)^{1/n}$ hyperplanes.

First we show that if $c_1+c_2k+c_3k^2+ \ldots + c_nk^{n-1}==0$ modulo m, then $c_1u_i+c_2u_{i+1}+\ldots+c_nu_{i+n-1}$ is an integer for every *i*. Let [ ] denote the greatest integer function. The sequence

$r_1,r_2,\ldots r_n$ can be rewritten as

r_1, kr_1-m[(kr_1)/m], k^2r_1-m[(k^2r_1)/m], . . . and the sequence u_1, u_2, . . . ,u_n can be written as
r_1/m-[r_1/m], (kr_1)/m-[(kr_1)/m], (k^2r_1)/m-[(k^2r_1)/m], . . .
Clearly if $c_1+c_2k+c_3k^2+$ . . .+ c_nk^{n-1} is a multiple of *m,* then $c_1u_i+c_2u_{i+1}+$. . .+c_nu_{i+n-1} is an integer. Now it remains to be shown that there are integers c_1,c_2, . . .,c_n not all zero such that
$c_1+c_2k+c_3k^2+$ . . .+ c_nk^{n-1}==0 modulo m.
We have seen that $c_1u_i+c_2u_{i+1}+$. . .+c_nu_{i+n-1}=0,(+ or -)1, (+ or -)2,. . .We now show that there exist non-zero c_1,c_2, . . .,c_n using a general theorem on linear forms by Minkowski, using the basic result that a symmetric convex set of volume 2^n in *n* space must contain a point (other than the origin) with integer co-ordinates.

Minkowski's Lemma

Let C be a bounded, symmetric, convex domain in R^n . Let a_1, a_2,. . . ,a_n be linearly independent vectors in R^n. Let A be the n*n matrix whose rows are the a_i^s . If vol(C)>2^n|det A|, there exist integers c_1,c_2, . . .,c_n (not all zero) such that $c_1a_1+c_2a_2+$. . .+c_na_n (element of)*C*

If we consider set D of all (c_1,c_2, . . .,c_n) (element of) R^n such that
$c_1a_1+c_2a_2+$. . .+c_na_n (element of)*C.* It is easily seen that D is bounded, symmetric and convex because *C* is. Moreover, D=A^{-1}C so that by linear algebra,
Vol(D)= Vol(C) (|det A|)$^{-1}$
Thus, if Vol(D)>2^n , then D contains the lattice points (c_1,c_2, . . .,c_n) not equal to zero such that
$c_1a_1+c_2a_2+$. . .+c_na_n (element of)*C*. But Vol(D)>2^n is equivalent to
Vol(C)> 2^n |det A| , as desired.

We now look at a live cryptanalysis of linear congruence generators in 2-dimesnion. Before that, as a corollary of Minkowski's lemma, we show that-
A lattice L in R^2 contains a non-zero vector *alpha,* such that
|*alpha*|^2<=(4*(Area of parallelogram spanned by a lattice basis for *L*.))/ pi, pi=3.1415…
Let dl= Area of parallelogram spanned by a lattice basis for *L*.
Let the convex *S* be a circle of radius *r* about the origin. Minkowski's lemma guarantees the existence of a non-zero lattice point in *S,* provided

Area(*S*)>4.dl

i.e. pi*r^2 >4.dl or

r^2>(4.dl)/pi

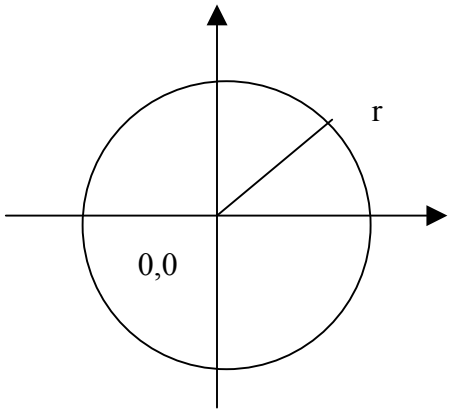


Figure 3. Circle of radius *r* about the origin

So for any positive number *e,* there is a lattice point alpha with
$|alpha|^2$ <((4.dl)/pi)+*e*. Since there are only finitely many lattice points in a bounded region and since *e* can be arbitrarily small, there is a lattice point satisfying the above inequality.

Illustration of Cryptanalysis of Linear Congruence Generator

Since-"Random numbers fall mainly in the planes"

The lattice structure of points produced by the congruential RNG

$x_{(n)}$=a*$x_{(n-1)}$+c mod m makes it easy to find a, c and m if one is given a few pairs $x_{(i)}$,$x_{(i+1)}$ produced by the RNG. The idea is this: points (x_1, x_2), (x_3, x_4), (x_5, x_6).....with Unit-cell volume the modulus of the generator, the 'm' of $x_{(n)}$ =a*$x_{(n-1)}$ +c mod m. This means that the parallelepiped formed by any three points of the lattice has volume a multiple of the unit cell's volume, which is the (unknown) modulus m. So we need only choose any three points whose coordinates are successive integers from the RNG, find the absolute value of the determinant of the matrix. This is the volume of the parallelepiped and it must be a multiple of the unknown

modulus. It is clear that we have chosen *n*=2, i.e. over 2-dimensional space. This is because a parallelepiped is a parallelogram in two dimension and thus more convenient to find the volume of a parallelepiped in two dimension when compared to higher dimensions. Usually five or six such determinant values (i.e. the area of the parallelogram in determinant form) will have gcd that determine m, this is the volume of a parallelepiped determined by three points in the plane with coordinates successive integers from the generator. Then solving $a*(x_3-x_1) =(x_4-x_2) \bmod m$ for 'a' determines the multiplier.

Example: Consider the congruential sequence

5,13,10,4,11,6,15,14,12,8,…

generated by $x_{(n)} =2*x_{(n-1)} +3 \bmod 19$.

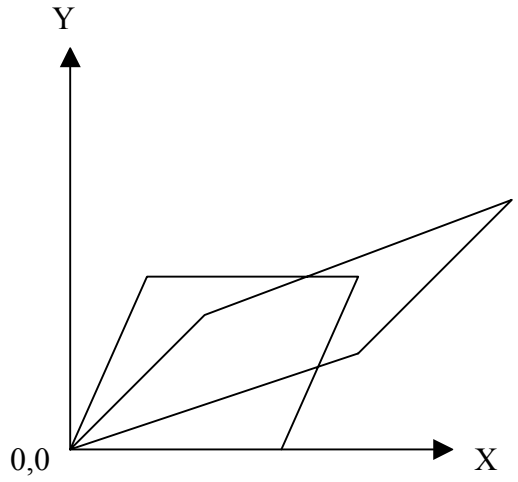
Form points from successive x's, for example

(5,13), (10,4), (11,6), (15,14),(12,8),...

Subtract the first point from the others. If (a,b) is the first point and (p,q) is the second point then we find (p-a, q-b)

(Translate the lattice to include (0, 0)):

(5,-9),(6,-7),(10,1),(7,-5),...



Form a few determinants:

Abs ((5*-7)-(-9*6)) =19=m

Abs ((6*1)-(-7*10)) =76=4*m

Abs ((10*-5)-(7*1)) =57=3*m

All determinants will be multiples of m.

We needn't use successive points; for example:

Abs ((5*-5)-(7*-9)) =38=2*m

We will usually not be lucky enough to get m from the first two determinants, but successive gcd's quickly become a constant sequence

....m, m, m, m, m,

Let = = indicate congruence. If for ax = =b (mod m) and a is relatively prime to m, the solution (unique mod m) of the linear congruence is $X=b*a^{(\text{phi }(m)-1)}$ (mod m) where phi (m) is the Euler's Totient Function or simply by the Extended Euclid's algorithm

From the first (or any other) point of the translated lattice, say (10,1), solve

a*10=1 mod 19, to get a=2,

From the original points, a*5+c=13 mod 19,gives

2*5+c==13 mod 19,c=3.

We get c=3 and thus the generator

$x_n = a*x_{n-1} + c \bmod m$, has a=2, c=3 and m=19.

Acknowledgement

The entire cryptanalysis of linear and multiplicative congruence generator and the statistics in table 1, presented here is due to [1]. The discussion of Minkowski's lemma and its corollary, presented are due to [4],[5],[6]. The intersecting hyper-planes in Fig 2. is derived from John Stembridge's page, http://www.math.lsa.umich.edu/~jrs/

Bibliography

1. George Marsaglia, *Random Numbers Fall Mainly In The Planes*, Mathematics Research Lab, Boeing Scientific Research Laboratories, Seattle, Washington.

2. Donald.E.Knuth, *Semi Numerical Algorithms*, Volume 2.

3. Shaum's, *Linear Algebra*, MGH.

4. G.Hadley, *Linear Algebra*, Addison-Wesley

5. Jody Esmonde and M.Ram Murthy, *Problems in Algebraic Number Theory*, Springer Edition.

6. Artin, *Algebra,* PHI-EEE

7. Tom Apostol, *Introduction to Analytical Number Theory*, Springer International, Student edition, 1989.

8. Neal Koblitz, *A Course In Number Theory and Cryptography*, Springer, Second edition, 1994.

9. Bruce Schneier, *Applied Cryptography*, Wiley Publications, Second edition, 2001.

10. George Marsaglia, Newsgroups: sci.math.symboli, Archives.

-Compiled by Sarad A.V