Department of Electrical Engineering, IIT Bombay
EE 793 - Topics in Cryptology

Home Paper Assignment Report

LC distribution of the exponential map in prime fields

Prepared by:
Prasann Viswanathan Iyer (190070047) (Leader)
Harshit Shrivastava (18D070011)
Durgesh Ahire (18D070004)
Kaustubh Bhargao (18D070014)
Rishabh Sureka (18D070023)

Instructor: Prof. VR Sule
Date: April 23, 2023

Contents

1 Brief description of the problem

1.1 Introduction

This document contains the solution for the assignment problem for EE 793.

The objective of this problem is to gather data on the distribution of Linear Complexity (LC) of recursively generated sequences by the exponential map in the group of units of prime fields. This is a number theoretic question that has not been deeply studied, and the gathered data is expected to provide insights into properties of the exponential map over prime fields. In this problem, we choose a prime number of bit length n between 16 to 32 bits, fix a primitive element of Fp, and compute the LC of a sample of 1000 randomly chosen x in Fp.

1.2 Overview of Problem

Given a prime number p and a primitive element ζ of $\mathbb{F}_p$, we define the map $F(x) : \mathbb{F}_p \rightarrow \mathbb{F}_p$ as $y = F(x) = \zeta^x \bmod p$. We then generate a sample of recursive sequences $S(\zeta, x) = y, F(y), F^2(y), \ldots, F^{M-1}(y)$ in $\mathbb{F}_p$ for $M = O(n^2)$ using 1000 randomly chosen x in $\mathbb{F}_p$. We then compute the LC of $S(\zeta, x)$ using the Berlekamp-Massey (BM) algorithm over $\mathbb{F}_p$ and draw the histogram of LC. Finally, we repeat this process for five different primes p.

1.3 Theoretical Understanding

The LC of a sequence is the length of the shortest linear feedback shift register (LFSR) that can generate the sequence. The BM algorithm is a method for finding the minimal LFSR that generates a given sequence. The exponential map in the group of units of a prime field is a widely studied object in number theory and has many interesting properties. The distribution of the LC of recursively generated sequences by the exponential map is a largely unexplored area, and this problem aims to provide insights into this distribution.

2 Computation and Analysis

Five different primes p were chosen as follows:

$$p = [957097, 890501, 417293, 612649, 496291]$$

We decided to use 6-digit primes, considering the time required for simulations. ζs were chosen randomly from the range $(2, p)$, for each prime p.

$$\zeta = [733474, 130002, 319358, 128675, 137979]$$

We then compute the linear complexity values using the Berlekamp-Massey algorithm. The simulation code is attached below.

```python
from sage.misc.prandom import randrange
from sage.matrix.berlekamp_massey import berlekamp_massey
from sage.plot.histogram import Histogram, histogram
import math, random, time
from matplotlib import pyplot as plt
import numpy as np

primes = [957097, 890501, 417293, 612649, 496291]
zetas = [randrange(2, p) for p in primes]
Ms = [randrange(1, 10) * ((int(math.log2(p))+1)**2) for p in primes]
     # O(n^2) where n is number of bits (18) here

# zetas
# [733474, 130002, 319358, 128675, 137979]
# Ms
# [2000, 1200, 722, 3200, 2888]

lc_hist = []
for z,p,M in zip(zetas, primes, Ms):
    print(z,p,M)
    input_vec = np.random.randint(2,p,size=1000)
    lc_values = []
    for x in input_vec:
        y = Mod(x,p)
        S = []
        for i in range(math.ceil(M/2)*2):
            y = Mod(z,p) ** y
            S.append(y)
        lc_values.append(berlekamp_massey(S).degree())
    lc_hist.append(lc_values)

for lc in lc_hist:
    print(min(lc), max(lc))

for lc in lc_hist:
    plt.style.use('ggplot')
    plt.hist(lc, bins=50)
    plt.show()
```

Listing 1: Python example

3 Results

Below are the histogram plots for values of Linear Complexity on 1000 sequences for each pair of (p, ζ). We ran the simulations on multiple sets of 5 primes and obtained similar results for each case.

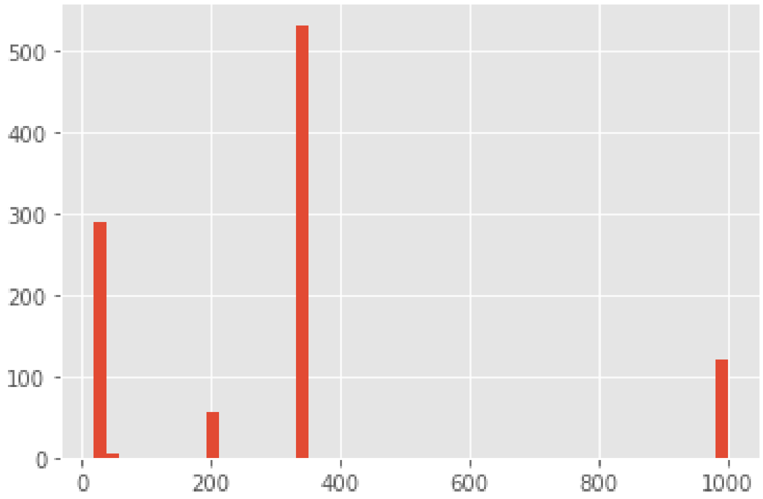

Figure 3.1: LC of $S(\zeta, x)$ for x in $\mathbb{F}_p$, $p = 957097$

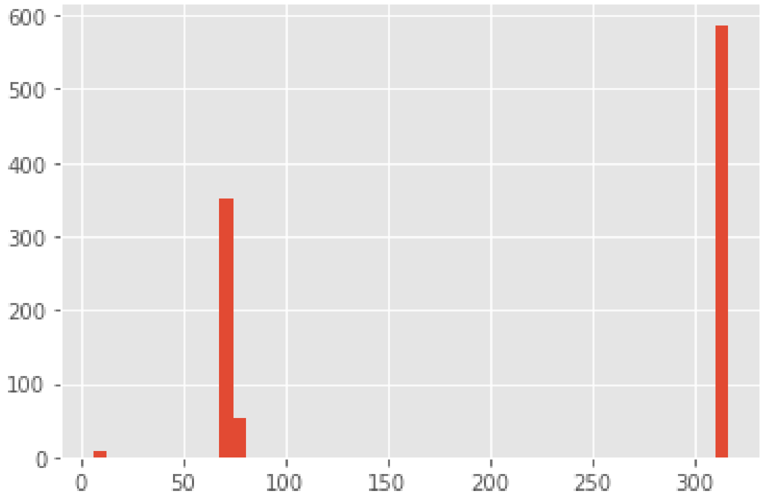

Figure 3.2: LC of $S(\zeta, x)$ for x in $\mathbb{F}_p$, $p = 890501$

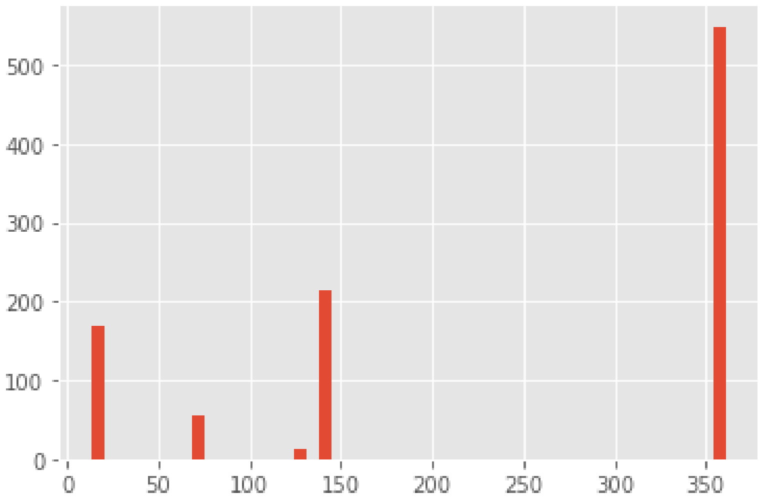

Figure 3.3: LC of $S(\zeta, x)$ for x in $\mathbb{F}_p$, $p = 417293$

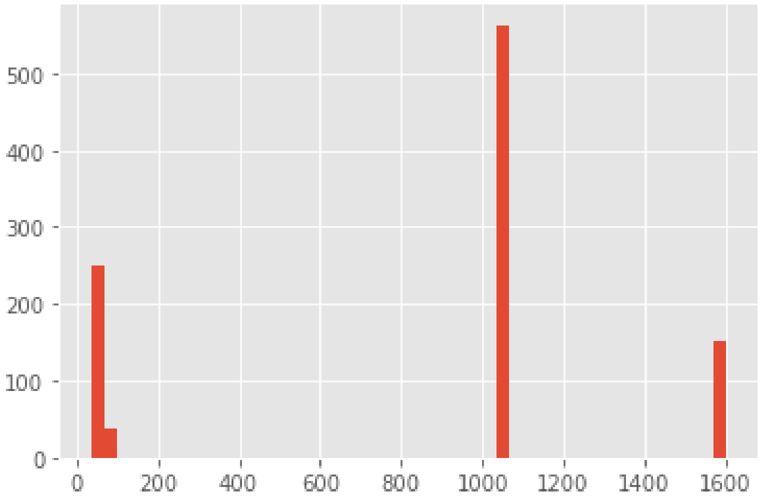

Figure 3.4: LC of $S(\zeta, x)$ for x in $\mathbb{F}_p$, $p = 612649$

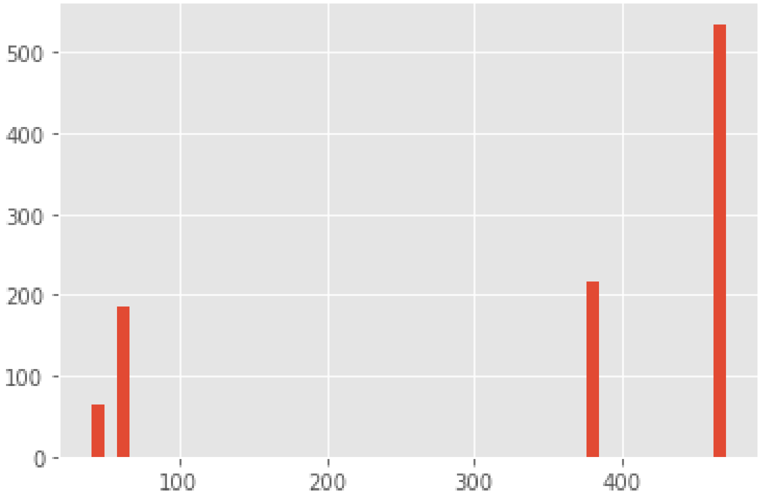

Figure 3.5: LC of $S(\zeta, x)$ for x in $\mathbb{F}_p$, $p = 496291$

4 Observation and Conclusion

We observe that there are clusters of linear complexities at the smaller end and then a large chunk of sequences that have high linear complexity for each ζ but the claim that most LCs are expected to be $> \lfloor M/2 \rfloor$ wasn't observed.

Our reasoning for this is that the primes and zetas chosen were very large and in comparison when we take M to be of the order $O(n^2)$ where n is the number of bits, the generated sequences are not long enough for LCs to mode at values $> \lfloor M/2 \rfloor$.

In simulations involving smaller primes and zetas with M of the order of $O(2^n)$ we observe clusters at $> \lfloor M/2 \rfloor$.

Because the sequences are not long enough for the Linear Complexities to be be more than $> \lfloor M/2 \rfloor$, obtaining an inverse from them is rarely possible.

5 References

1. Programming in Sage
2. Berlekamp-Massey algorithm to find the minimal polynomial of a linear recurrence sequence
3. Measuring execution time in Sage