Statistical Tests For Key Strength Identification In Cryptography

Dr. Addepalli VN Krishna

Abstract: The cryptographic study involves three algorithms, one for Encryption of Plain text to Cipher text, one for Decryption for Cipher text back to Plain text and third for the generation of the Key. Key generation algorithm works on the principle of Randomness. In this work, the randomness of Key is studied by using Statistical methods like Runs Up & Runs Down test, Runs (Above and Below the mean), Chi Square test & Auto correlation test for its usability in Cryptographic study.

Index Terms: Cryptography, Key, Statistical tests, Randomness , Runs Up &Down test, Chi Square test, Auto Correlation test.

1 INTRODUCTION

In cryptographic study, the information to be transferred from the sender to Receiver will be converted to intelligible form so that an unauthorized user even if he intercepts the message canot gets any useful information from it. That is the information which is called as plain text is converted to Cipher text using an algorithm and key which does not have any meaning. At the receiver's side, the received cipher text is converted back to Plain text using the same algorithm and the key. Thus this process involves three algorithms, one for Encryption of Plain text to Cipher text, one for Decryption for Cipher text back to Plain text and third for the generation of the Key. Key generation algorithm works on the principle of Randomness. The better the randomness of the key, the better will be the strength of the key. Thus in this work an attempt is made to check the randomness of some probable keys by using models like Upward and downward test, Runs above and below mean, Chi-square test and Auto correlation random number test.

2 LITERATURE STUDY

2.1 Review Stage

Now a day, every computer is connected to another through internet. In this work authors tried several searching strategies based on differential crypto analysis [1]. They have identified clustering differential paths, the searching algorithm found to provide better results when compared to examining individual paths. They tested this on algorithms like Feistel and SPN structures, whereby the best distinguishers for each of the investigated ciphers were obtained by discovering clusters with thousands of paths. They considered the KATAN block cipher family as a test case, and studied the crypto analytical attacks like boomerang attack and related-key model to obtain the best cryptanalytic results.

- *Dr. Addepalli VN Krishna, is working as Professor, Computer Science & Engineering, Faculty of Engineering, CHRIST (Deemed to be University), Bengaluru. He has around 26 yrs of teaching experience and with around 50 publications in Journals of repute. He is actively involved in Research for Guiding Scholars in their Doctoral studies.*

In this work the authors tried to study the Linear and Differential paths for with high Statistical Bias to perform Linear and Differential crypto analysis on Block ciphers [2]. They considered two known Block ciphers like L Block, TWINE to study the S-Box structure used in them. They provided different means for evaluating the strength of S Box by Differential crypto analytical attacks. In this work the Authors [3] worked on the linear sieve and the cubic sieve methods for computing discrete logarithms over prime fields. They have observed that for some set of prime numbers, cubic sieve method runs about two times faster than the linear sieve method. They also provided a theoretical solution for $X^3 \equiv Y^2 Z \pmod p$ that is of central importance in the cubic sieve method. The authors [4] have studied weakness of the DES algorithm under timing attack. It exploits the engineering aspects involved in the implementation of crypto systems while doing encryption process. The authors [5] considered cryptographic transformations of the Kuznyechik algorithm in relation to differential analysis and the translation of their representations into a more convenient form for cryptanalysis. In the analytical model they have considered 16 cycles of execution of the shift register with linear feedback. They observed the output by algebraic form of a linear transformation. As part of Future work, they identified this concept to be used for differential crypto analysis. In this paper [6] the authors presented two related-key attacks on the full AES. For AES-256 they have shown the first key recovery attack that works for all the keys with time and data complexity, Both attacks are boomerang attacks, which depends on finding local collisions in block ciphers and enhanced with the boomerang switching techniques to gain free rounds in the middle. [7] In this paper the authors applied artificial neural networks for cryptanalyst of identifying weakness in cipher text. The Neural network is trained with cipher text and chosen random key till it tries to match with the original key. The work is analyzed with cipher text only attack. [8] In this work Machine learning algorithms are tried on Cryptographic algorithms, Crypto analysis and Staganegraphy applications. The work abstracts the wok done in these areas and also focuses on future directions of this work. [9] This work focuses on the crypto analytical works done in the last decade. It observes the relevance of Machine learning in crypto analysis of work. This work also focuses on applying Machine learning concepts in Network Analysis. [10] This work presents Neural network analysis on Leight weight algorithm like Simon Cipher. It takes a set of pairs of Plain text and Cipher texts and tries to predict the Keys. The work presents the optimal configuration of Neural Network design for better

predictions of Keys.

3 METHODOLOGY OF WORK

3.1 Randomness testing of Keys by Statistical Tests

Statistical Tests	*Model-Mean*	*Model-standard deviation (σ)*	*Acceptable level*	*Acceptable value for a significance level 0.05*
Runs test (Up's And Down)	*μ = (2*N-1)/3*	*(√(16*N-29)/90).*	*(a – μ)/ σ.*	*-1.96-1.96*
Runs Test (Above and Below the mean)	*(2*n1*n2/ N) +0.5)*	*√ ((2*n1*n2(2*n1*n2-n1-n2))/((n1+n2) 2 * (n1+n2-1)))*	*(a – μ)/ σ.*	*-1.96-1.96*

Runs test (Up's and Down)
Let N be the number in a sequence of values
If 'a' be the number of runs in a sequence, the mean is calculated as μ = (2*N-1)/3 and
Standard Deviation σ 2 = (√(16*N-29)/90).
Acceptable level = (a – μ)/ σ.
If the value lies between -1.96-1.96, the sequence accepts the hypothesis of Randomness.

Runs Test (Above and Below the mean)
Let N be the number in a sequence of values,
If 'a' be the number of runs in a sequence, n1 be the number of upward runs and n2 be the number of downward runs
The mean is calculated as μ = (2*n1*n2/N) +0.5) and
Standard Deviation σ 2 = ((2*n1*n2(2*n1*n2-n1-n2))/((n1+n2) 2 * (n1+n2-1)))

Chi Square test:
If N be the number of values of the sequence
By considering E = N/n >= 5, n(No. of levels) can be considered for dividing the sequence in multiple levels of values. If O I = Number of values of sequence at each level, E I = Considered value of sequence based on the number of levels, $X 2 = (O_I - E I)^2 / E$ will be calculated. If the value is < 16.9 for a confidence level of 0.5, the hypothesis of Independence of values of sequence is accepted.

Auto correlation test:
In this test, for the given sequence, the values are considered at predefined intervals. A Value say 'M' will be calculated as i+(M+1)*m <=N, where I is the initial value of the sequence considered, m is the interval being taken and N be the number of values of sequence. The the mean of sequence is calculated as
$\rho = (1/ (M+1)[\Sigma R_{i+km} * R_{(k+1)m}])-0.25.$
σ = (√ (13*M +7))/ 12*(M+1)
Significance level = ρ / σ
If the value lies between -1.96-1.96, the sequence accepts the hypothesis of Randomness.

4 EXAMPLE

Let the sequence considered as Key to be tested for randomness is
4 8 5 2 4 6 3

Runs Up & Runs Down test
Runs Up= 2, Runs Down =2
Total number of runs = 4, N= 7
Mean μ = (2*N-1)/3 = 4.33
Standard Deviation σ 2 = (√(16*N-29)/90).= 0.92
Acceptable level = (a – μ)/ σ.= (4- 4.33)/ 0.95 = -0.347
Since the value lies between -1.96-1.96, the sequence accepts the hypothesis of Randomness.

Runs Test (Above and Below the mean)
The number of values in a sequence, N= 7
The number of runs in a sequence, a = 4, n1= the number of upward runs=2,
n2 be the number of downward runs=2
Mean μ = (2*n1*n2/(n1+n2)) +0.5) = 2.5
Standard Deviation σ 2 = ((2*n1*n2(2*n1*n2-n1-n2))/((n1+n2) 2 * (n1+n2-1))) = 0.66
Acceptable level = (a – μ)/ σ = 1.85
Since the value lies between -1.96-1.96, the sequence accepts the hypothesis of Randomness

Chi Square test
N=7, to satisfy the condition, E=7/n>= 5
n considered is 1;

Interval	*O I*	*E I*	*$(O_I - E I)^2 / E_I$*
2-3	*1*	*7*	*5*
3-4	*1*	*7*	*5*
4-5	*2*	*7*	*4*
5-6	*1*	*7*	*5*
6-7	*1*	*7*	*5*
7-8	*0*	*7*	*5*
8-9	*1*	*7*	*5*

$X^2 = (O_I - E I)^2 / E_I = 4.8$
Since the value obtained is 4.8 < 16.9 for a confidence level of 0.5, the hypothesis of Independence of values of sequence is accepted.

Auto Correlation Test
A Value say 'M' will be calculated as i+(M+1)*m <=N, where I is the initial value of the sequence considered = 2, m is the interval being taken = 2,
N be the number of values of sequence=7
So M = 1;
Then the mean of sequence is calculated as
ρ =(1/ (M+1)[R 2 * R2 + R 4 * R4])-0.25 = 34 .
σ = (√ (13*M +7))/ 12*(M+1)= 0.186
Significance level = ρ / σ = 183

If the value lies between -1.96-1.96, the sequence accepts the hypothesis of Randomness. Thus for the given model, it may not accept the hypothesis. But since the sequence considered is very small in number, this model may not be suitable for testing randomness in this sequence.

5 RESULTS ANALYSIS

The results of different statistical tests considered for evaluating the strength of Key are shown in table 1:

Tests	*Mean*	*Standard Deviation*	*Significance level*	*Acceptable range*	*Remarks*
Runs Up & Runs Down test, , and	*4.33*	*0.92*	*-0.347*	*-1.96 – 1.96*	*Accepted for hypothesis of independence of numbers*
Runs (Above and Below the mean)	*2.5*	*0.66*	*1.85*	*-1.96 – 1.96*	*Acceptable*
Chi Square test	*NA*	*NA*	*4.8*	*<=16.9*	*Acceptable*
Auto Correlation test	*34*	*0.186*	*183*	*-1.96 – 1.96*	*Since the sequence considered is very small, this test may not be suitable for testing Independence of numbers.*

In this work an attempt is made to evaluate the strength of the Key by using Statistical methods like. It is observed that the considered sequence accepts the hypothesis of Independence for the first three tests and Not suitable to apply for Auto correlation test.

6 CONCLUSION & FUTURE WORK

The considered sequence accepts the hypothesis of Independence for the Runs Up & Runs Down test, Runs (Above and Below the mean), Chi Square test and may not be suitable for Auto correlation test. The work may be extended by using Machine Learning algorithms to evaluate the functionality of randomness being considered for Key used in encryption process.

7 REFERENCES

[1] Jiageng Chen et al., Towards Accurate Statistical Analysis of Security Margins: New Searching Strategies for Differential Attacks, IEEE Transactions on Computers (Volume: 66 , Issue: 10 , Oct. 1 2017)

[2] Jiageng Chen et al., Improved Differential Characteristic Searching Methods, 2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing,

[3] Abhijit Das &C. E. Veni Madhavan, On the cubic sieve method for computing discrete logarithms over prime fields , International Journal of Computer Mathematics Volume 82, 2005 - Issue 12

[4] F.Pereira Ribeiro, ACM Transactions on Information and System Security, 1999

[5] Gayrat & Avazjon (2019), Representation of the Block Data Encryption algorithm in Analytical form for Differential crypto analysis, IJIRIS, International Journal of Innovative Research in Information Security, Volume VI, 38-42.

[6] A Biryukov, Advances in Cryptology - ASIACRYPT 2009, 15th International Conference on the Theory and Application of Cryptology and Information Security, Tokyo, Japan, December 6-10, 2009

[7] Riccardo Focardi and Flaminia L. Luccio, Neural Cryptanalysis of Classical Ciphers, CUER-WS.org, Vol 2243/paper10

[8] Mohammed M. Alani, Applications of Machine Learning in Cryptography: A Survey, arXiv.1902.04109v1[cs.CR], 11 Feb., 2019

[9] SambasivaRao Baragada, 2P Satyanrayana Reddy, A Survey on Machine Learning approaches to Cryptanalysis, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 2, Issue 4, July – August 2013

[10] Kowsic Jayachandiran, A Machine Learning Approach for Cryptanalysis, RIT Computer Science _ Capstone Report _ 20175.