International Journal of Information Technology and Knowledge Management
July-December 2008, Volume 1, No. 2, pp. 475-483

INFORMATION THEORY TESTS BASED PERFORMANCE EVALUATION OF CRYPTOGRAPHIC TECHNIQUES

Yudhvir Singh & Yogesh Chaba

Abstract

Cryptography is an emerging technology, which is important for network security. Research on cryptography is still in its developing stages and considerable research effort is required. This paper is devoted to the security and attack aspects of cryptographic techniques. We have discussed the dominant issues of security and various information theory characteristics of various cipher texts. The simulation based information theory tests such as Entropy, Floating Frequency, Histogram, N-Gram, Autocorrelation and Periodicity on cipher text are done. The simulation based randomness tests such as Frequency Test, Poker Test, Runs Test and Serial Test on cipher text are done. Finally, we have benchmarked some well-known cryptographic algorithms in search for the best compromise in security.

Keywords: Cryptography, Ciphers, Secret Key Cryptography, Security, Attack, Attacks Analysis and Performance.

1. Introduction

Common design criterion for every cryptosystem is that the algorithms used for encryption and decryption are publicly known and that the security of the scheme only relies on the secrecy of a short secret called the key [1][2]. Figure 1 shows the brief overview of all parts of an encryption system. It considers classical symmetric encryption only, *i.e.*, both the sender and the receiver of a secret message share a common secret key K. An efficient encryption algorithm E, takes the key and a plaintext and outputs a ciphertext, and an efficient decryption algorithm D, takes the key and a ciphertext and outputs a plaintext. It requires that the decryption of a ciphertext yield the original message again. This property is called correctness of the encryption scheme. Usually it denotes symmetric keys by K; K_1; K_2; … K_n, messages by m; m_1; m_2; … m_n, and cipher texts by c; c_1; c_2; … c_n. The substitution cipher is the oldest and one of the most famous ciphers in history and literature [2]. Then a key K of the substitution cipher is a permutation of the set $\{a; b; \ldots ; z\}$. A message $m = m_1||m_2||: : ||m_n$ is encrypted by computing $c = K(m_1)||K(m_2)|| \ldots ||K(m_n)$ where the m_i are single letters and || denotes concatenation of strings. As K is a permutation of set $\{a, b, \ldots.z\}$, there exists the unique inverse permutation K^{-1}.

A small example based on the key K is shown in figure 2, if the plaintext ***abc*** is encrypted under this key, the ciphertext becomes ***ECH***. The simplest attack on any

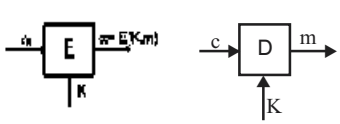

Figure 1: Basic Encryption System

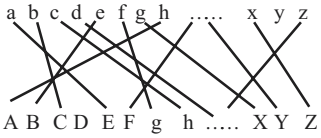

Figure 2: Substitution Cipher

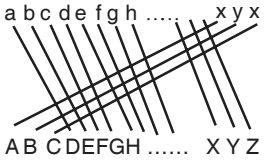

Figure 3: Shift-Clipher

encryption scheme is brute-force attack, where each possible key is used to decrypt a ciphertext and the resulting message is investigated [4][6]. Thus the practicability of this attack relies precisely on the size of the key-space. For the substitution cipher the size of the key-space is $26! \cong 2^{86}$, so a brute-force attack on a substitution cipher is clearly impractical. However, using statistical attacks, the substitution cipher can be broken quite easily. A special form of the substitution cipher is the shift-cipher as shown in figure 3. Permutation K is chosen from a restricted set, by shifting the letters by a fixed number of positions in the alphabet. Given a number A_K {0; … ; 25}, the permutation K maps each letter to its A_Kth successor, wrapping around after the Z and starting with A again if necessary. If one identifies $\{a; \ldots ; z\}$ with {0; … ; 25} then $K(B) = B + A_K$ mod 26. A variant is the so-called ROT-13 scheme, where $K_{KK} = 13$ [3][4]. The Vigenere Cipher is a generalization of the Shift-Cipher [10]. While the latter relies on a single permutation K to shift every letter by A_K positions, the Vigenere cipher uses n such permutations K_0; … ; K_{n-1} to shift different letters by a different number A_{K0}; … ; A_K of positions. Vermal cipher was the first scheme, in which a security proof is given for cryptographic algorithm [10]. It may be defined over a variety of message spaces, for example $\{a; \ldots ; z\}$ or {0; 1}. Here it is defined over bit strings. Let $M = C = K = \{0; 1\}^n$ be bit strings of a fixed length n. Both encryption and decryption are defined by the *XOR* of the key with the plaintext or ciphertext, respectively. The Vernam has the disadvantage that the key is as long as the message and may not be reused. Thus deploying this scheme necessarily requires securely transmitting a large amount of keys. While this is done in very sensitive environments, it is impractical for essentially any application. Hill encryption algorithm the plaintext and key are arranged in a block/ matrix fashion of two rows and two columns [8][9]. Then certain calculations (generally matrix multiplication and modulo) are performed to get resultant as cipher text and

same process to get in reverse i.e. plain text. Encryption using binary addition technique a text document is encrypted through binary addition [4]. Encryption using binary Exclusive-OR provides an encryption using the binary Exclusive-OR operation key with plaintext [5][6]. It is shown that with many encrypted documents it is easy to work out the key and hence the plaintext as well. The stages involved in an encryption using binary addition are similar to those performed here, except that the document that is encrypted is a text document. In Playfair technique a 5×5 matrix is filled with the plaintext characters [2]. For example, the different letters of a keyword are inserted first, followed by the remaining letters. The plaintext is divided into pairs; these digraphs are encrypted using the following rules: If both letters can be found in the same column, they are replaced by the letters underneath. If both letters can be found in the same row, take the letters to their right. If both letters of the digraph are in different columns and rows, the replacement letters are obtained by scanning along the row of the first letter up to the column where the other letter occurs and vice versa. Double letters are treated by special rules, if they appear in one diagraph. They can be separated by filler. In Solitaire algorithm the key is created by means of a pack of cards and rules that are agreed upon in advance [7][8]. A pack of cards is unsuspicious to outsiders, shuffling the deck provides a certain amount of coincidence, cards can be transformed into numbers easily and a transposition cipher can be carried out without any further aid. Sender and receiver have to own a deck of cards shuffled in the same manner. A key stream is generated that consist of as many characters as the message to be encrypted.

2. SIMULATION ENVIRONMENT

The cryptanalyst's main goal is to break the cryptographic system in every possible way, with his existing knowledge and available infrastructure. Apart from this an adversary pursues, there are different possibilities a cryptanalyst might exploit for different information from an encryption scheme, and *e.g.* information theory tests and randomness analysis test. The simulator CrypTool is used to analyze information theory tests and randomness analysis test. This tool contains some in-built mechanisms for these test and analysis.

Information Theory Tests: The information theory tests mechanisms for cryptographic techniques are:

Entropy	Calculate the entropy of a document.
Floating Frequency	Calculate the floating frequency of a document.
Histogram	Calculate the character frequency of a document.
N-Gram	Analyze the frequency of N-Grams of a document.
Autocorrelation	Perform autocorrelation of characters in a document.
Periodicity	Analyze the periodicity of a document.

Randomness Analyses Tests: The Randomness analysis contains the following tests:

Frequency Test	Checks the random quality of the active document with the Frequency test.
Poker Test	Checks the random quality of the active document with the Poker test.
Runs Test	Checks the random quality of the active document with the Run and Long-Run test.
Serial Test	Checks the random quality of the active document with the Serial Test.

In this paper CrypTool is used as a simulator to conduct the experiments and to get the result. Only alphanumeric and special characters are used for analysis of cryptographic techniques. These specifications are selected in option menu of the CrypTool and visual results are set in window option of the CrypTool. For the input plaintext, around 50-sample text are taken and encrypted with various algorithms. The output of above plaintext is cipher text, analyzed with analysis option in CrypTool. Some of the cryptographic algorithms are implemented in *C*, and their output is taken as cipher text, which is then copied in some text file and that text file is used for the analysis with CrypTool. In analysis option, attack analysis is selected and analyzed with options information tests and randomness test techniques. The parameters of above automatic analysis techniques are assigned or selected manually as per the requirements of techniques. In this paper results are interpreted on the basis of various parameters such as key dimensions generated, derived keys, superposition analysis, correlation of distribution of the cipher texts and language texts, comparing cipher text with English text and plaintext etc. Based on the above parameters for a single test on one cryptographic technique the major interpretations are made as: ZS – Zigzag pattern of output cipher text, correlating similar with input plaintext, ZD – Zigzag pattern of output cipher text, different/not correlating with input plaintext**,** N_1–N_2 – Range of corresponding characters/symbols, N – Value of the corresponding parameter or the character position/ value, *F* – Tests Failed, *P* – Test Passed.

3. SIMULATION RESULTS

Results obtained through simulation are shown in table 2, 3 and 4. The table column denotes the tests method and table rows indicating the cipher text method implemented with encryption algorithm. Values of information theory test parameters are used to interpretation of test, e.g. the value of entropy should be as high as possible, floating frequency has wide range of values, the autocorrelation has none correlating or wide range with different zigzag patterns, and no periodicity or no offset or no cycles are preferred. Values of histogram test parameters are used to interpretation of test, e.g. the value of histogram parameters should be as wide range, but patterns should be least repeating frequencies. Values of randomness test parameters are used to interpretation of test, e.g. the value of such tests should be pass and not more than the MTV of that

test. The details descriptions of various results are shown in figure 4. Table 2 indicates the results obtained by applying information theory tests on different algorithms. When the entropy of cipher text is analyzed Hill & Vigenere are better then other encryption algorithms. Based on frequency analysis Hill cipher is better and ROT-13, Substitution

Table 2
Results of Information Theory Tests

Parameters Encryption Algorithm	*Key & Key Length*	*Entropy/Max. Entropy*	*Floating Frequency (avg± m)*	*Auto correlation*	*Periodicity Offset, length, no of cycle, value*
Plain Text	—	4.51/6.14	25, 4	1-9, Z	176, 2, 2 0D0A
Caesar	9-1	4.51/6.14	25, 4	1-9, Z, S	176, 2, 2 0D0A
Rot-13	13-1	4.03/4.70	24, 3	1-9, Z, S	176, 2, 2 0D0A
Vigenere	Networks -8	5.33/6.14	35, 4	0-6,10, Z, D	176, 2, 2 0D0A
Hill	Networks -8	5.78/6.14	42, 6	0-7, Z, D	NO
Substitution	Networks -8	4.33/6.14	24, 2.5	1-9, Z, D	189, 3, 2, 2, 0D0A
Byte Addition	96B3A05	5.77/8.00	41, 4	None	NO
Binary Ex-OR	96B3A05	5.92/8.00	44, 3	None	NO
Vernam	Text doc- 100	5.81/8.00	41, 4	0-4, Z, D	NO
Playfair	6*6 Matrix 8	4.47/6.14	23, 3	1-10, Z, D	NO
Solitare	Networks8 / deck52 asce	4.47/6.14	24, 2	0-26, Z, D	NO

Table 3
Histogram Analysis of Encryption Techniques

Parameters Encryption Algorithm	*ASCII Histogram*	*Diagram*	*Tri-gram*	*4-gram*	*N-gram*
Plain Text	1, 3-6, 11, 13	1-20	3, 2,1	3,2,1	21
Caesar	2-6, 11, 13	1-20	3, 2, 1	3,2,1	21
Rot-13	2-6, 11, 13	1-20	3, 2, 1	3,2,1	21
Vigenere	1-6, 1, 3	2, 1	2, 1	1	1-
Hill	1, 5, 35	2, 1	2, 1	1	1-
Substitution	1-7, 10, 21, 35	4, 3, 2, 1	3, 2, 1	1	1-
Byte Addition	1, 4, 8	2,1	1	1	1-
Binary EX-OR	1-4	4, 2, 1	1	1	1-
Vernam	1-4, 6	5, 4, 3, 2,1	1	1	1-
Playfair	1-6, 10, 11	3, 2, 1	1	1	1-
Solitare	1-6, 16	4, 3, 2, 1	3, 2, 1	1	1-

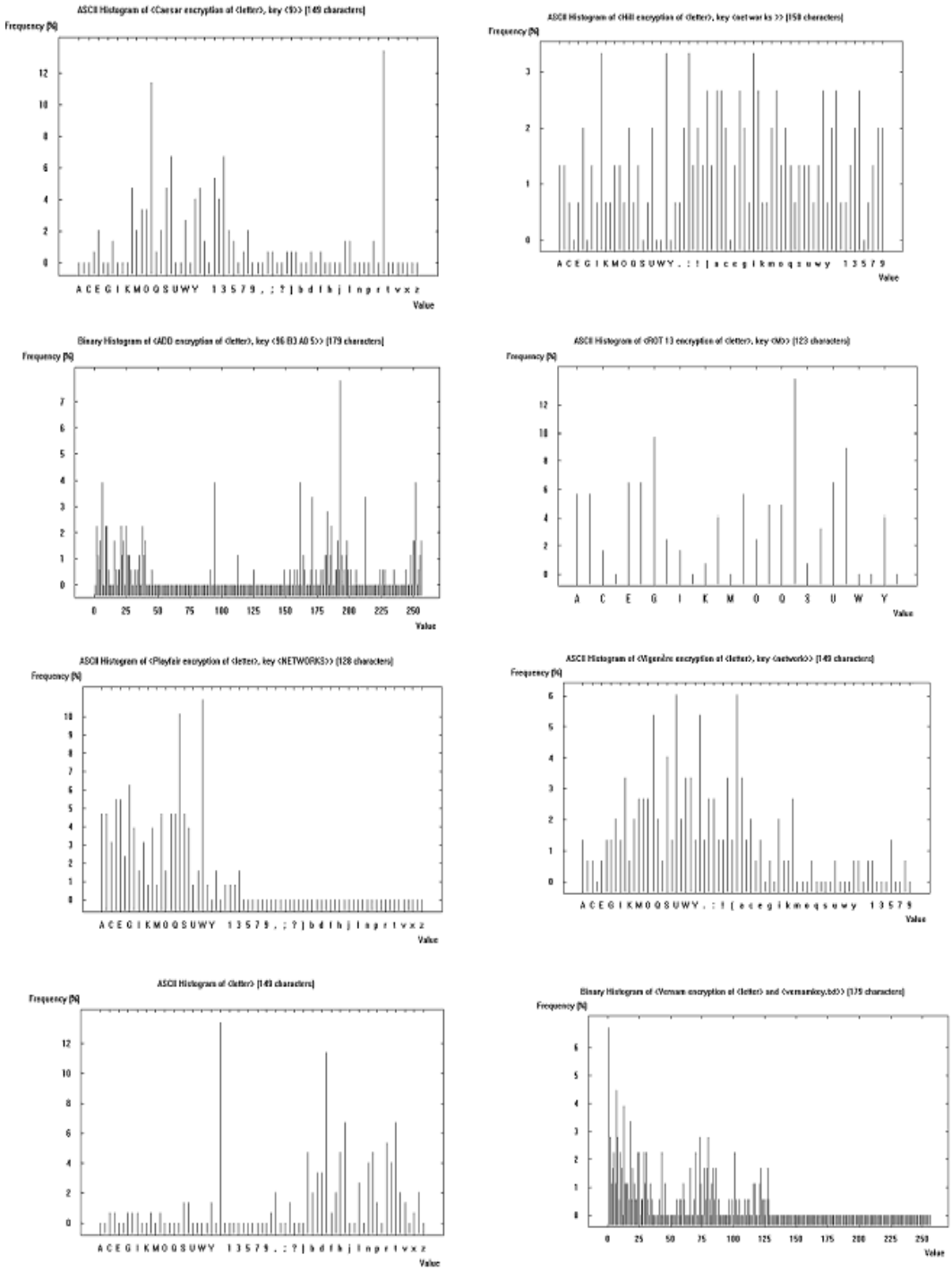

Figure 4: Analysis of Various Parameters for Ciphers

and Solitare can be easily analyzed, because these have least variation factor. When Correlation is analyzed Byte Addition and Binary EX-OR are more effective and Caesar and ROT-13 are the worst ciphers. Based on periodicity the Byte Addition, Binary EX-OR, Vernam, Playfair and Solitare are better then other ciphers. Further simulation is repeated for Histogram and Randomness Tests are shown in table 3 and 4.

Table 4
Randomness Analysis of Encryption Techniques

Parameters Encryption Algorithm	*Frequency Test A = 0.05, MTV = 3.841*	*Poker Test MTV = 14.07*	*Run Test MTV = 9.488*	*Serial Test MTV = 5.99*	*Long Run Test MTV = 34*
Plain Text	32.58, F	37.63, F	14.04, F	33.98, F	9, P
Caesar	32.58, F	45.11, F	40.23, F	53.32, F	7, P
Rot-13	26.28, F	32.70, F	27.66, F	25.79, F	7, P
Vigenere	31.38, F	78.28, F	118.58, F	88.99, F	6, P
Hill	38.89, F	34.06, F	16.29, F	33.63, F	7, P
Substitution	35.88, F	73.53, F	31.96, F	40.20, F	9, P
Byte Addition	8.14, F	25.62, F	13.79, F	23.67, F	10, P
Binary Ex-OR	7.84, F	13.78, P	47.68, F	14.43, F	17, P
Vernam	120.84, F	150.80, F	60.58, F	157.14, F	21, P
Playfair	37.51, F	37.51, F	101.12, F	101.12, F	5, P
Solitare	34.63, F	150.89, F	112.8, F	57.87, F	6, P

Table 3 indicates the results obtained by applying histogram test on different algorithms. When ciphers are analyzed Byte addition is difficult to analyzed, followed by Binary EX-OR, Playfair and Vernam while Caesar, ROT-13 can be easily analyzed. Table 4 indicates the results obtained by applying randomness tests on different algorithms. When ciphers are analyzed, Byte addition and Binary EX-OR are difficult to analyzed, followed by Solitare and Hill ciphers.

As shown in figure 5, it is clear from textual analysis, that Caesar and ROT-13 following the same pattern as plaintext, so it can easily deduced the data of cipher text with the help of correlation, histogram and frequency analysis test. These are the results of implementing various tests on different cipher texts generated with various cryptographic algorithms and it has been found that traditional cryptographic algorithm are more prone to information theory analysis. While others bit manipulation based such as Byte addition and Binary EX-OR are difficult to analyze and others such Hill, Vernam, Play-fair and Solitaire seems to be moderate to analyze with these tests.

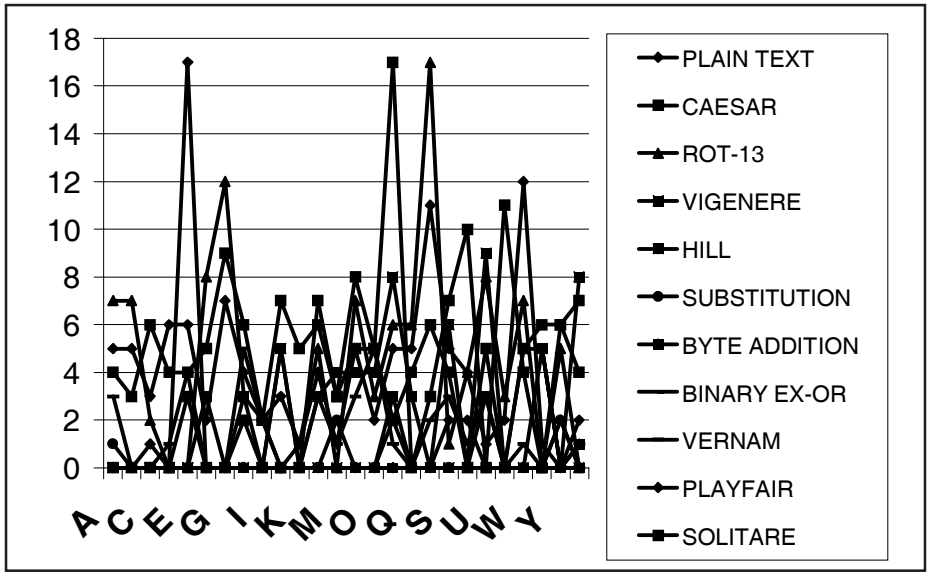

Figure 5: Textual Analysis of Ciphers

4. CONCLUSION

Encryption techniques were analyzed using CrypTool simulator with various tests such as Information Theory Tests, Histogram Analysis Tests and Randomness Tests. With these automatic test techniques it was found that the traditional cryptographic techniques are weaker and bit manipulation based are stronger and other cryptographic techniques are moderate to analyze. The Caesar and ROT-13 cipher was found to be weakest with the simulation based tests analysis and Byte Addition and Binary EX-OR cipher are the strongest among these ciphers.

REFERENCES

[1] Oded Goldreich, "Cryptography And Cryptographic Protocols", *Distributed Computing*, Springer-Verlag 2003 (2003) 16: 177–199

[2] Bart Prenel, Vincent Rijmen, Antoon Bosselaers, "Recent Developments in the Design of Conventional Cryptographic Algorithms", 18 September 1998.

[3] Weis Lucks, Bogk, "On the 2ROT13 Encryption Algorithm", *Sicherheit Von 1024 Bit RSA Schlüsseln Gefährdet*, (April 1, 2005).

[4] Daniel J. Bernstein, "Comparison of 256-Bit Stream Ciphers at the Beginning of 2006".

[5] Zhijie Shi, Ruby B. Lee, "Bit Permutation Instructions for Accelerating Software Cryptography", *Proceedings of the IEEE International Conference on Application-Specific Systems*, *Architectures and Processors*, July 10-12, 2000, Boston, Massachusetts, USA, 138–148.

[6] Coppersmith *et al*., "Cryptanalysis of Stream Ciphers With Linear Masking", *Proc. Crypto 2002, LNCS 2442*, (Springer 2002).

[7] Bernhard Esslinger, The CrypTool Script: Cryptography, Mathematics and More, (8th edition – distributed with CrypTool version 1.4.10), Germany, (July 25, 2007).

[8] Boris Pogorelov, Marina Pudovkina, "Properties of the Transformation Semi Group of the Solitaire Stream Cipher".

[9] Murray Eisenberg, "Hill Ciphers and Modular Linear Algebra", Copyright C1998 By Murray Eisenberg, (November 3, 1999).

[10] B. Tuckerman, "A Study of the Vigenere-Vernam Single and Multiple Loop Enciphering Systems," *Ibm Research Report Rc2879*, (14 May 1970), Yorktown, Heights NY.

Yudhvir Singh[1] **& Yogesh Chaba**[2]
Deptt of CSE
GJUST
Hisar
E-mail: *yudhvirsingh@rediffmail.com*[1]
yogeshchaba@yahoo.com[2]