Journal of ELECTRICAL ENGINEERING, VOL. 64, NO. 3, 2013, 196–200

SECURITY EVALUATION AND ENCRYPTION EFFICIENCY ANALYSIS OF RC4 STREAM CIPHER FOR CONVERGED NETWORK APPLICATIONS

Alaa E Din Riad [*] – **Hamdy K. Elminir** [*] — **Alaa R. Shehata** [*]
Taha R. Ibrahim [**]

The trend toward converged networks where voice,IP,video and wireless are combined over the same network infrastructure offers significant - and highly attractive - benefits.But this union introduces new vulnerabilities and new opportunities to break into the network and so we need converged security.In this paper we investigate the possibility of using RC4 as a stream key generator for converged network applications.The RC4 was statistically tested against wide range of tests for inspecting the validity of the encryption,and it's mathematically measured for encryption efficiency.All the consequent proves that RC4with suitable key length is highly secure and efficient enough to be used within the converged network.

K e y w o r d s: converged network, converged security, encryption efficiency, RC4, security evaluation, and stream cipher

1 INTRODUCTION

The term converged network or NGN (Next Generation Networks) relates to the integration of voice (fixed or wireless),data and video services into a single packet network.

The converged network requires converged security that integrates, enhances supplements and expands traditional data security techniques and policies to provide protection for converged communications. As Cryptography is probably the most important aspect of communication security and is becoming increasingly important as a basic building block for computer security in general [1]. Encryption if not implemented correctly, can do the potential to delay voice (or video) packets and adversely affect the performance of VoIP on the converged network, especially if there are multiple encryption points. The overhead of the encryption should have little impact on the performance of converged networks, as those real time packets like VoIP packets are more sensitive to delay than data packets. Hence if we utilized the traditional encryption techniques used within the data network like block ciphers, the voice quality will be degraded because of the resulting delay. From the prospective of converged security, we can minimize the risk to voice quality even more by employing streaming ciphers instead of using block ciphers that is used with data encryption mechanisms, as stream ciphers are almost always faster and use far less code than do block ciphers [2]. The example in this section, RC4, can be implemented in just a few lines of the code that fits our converged network real time traffic applications.

2 RC4 STREAM CIPHER

The RC4 is the most widely used stream cipher. It is part of the SSL/TLS (Secure Sockets Layer/Transport Layer Security) standards that have been defined for communication between web browsers and server's .It is used in the WEP (Wired Equivalent Privacy) protocol that is a part of the IEEE 802.11 wireless LAN standard [2]. RC4 is a variable size stream cipher developed by Ron Rivest for RSA Data Security, Inc. it is used also in other applications as Lotus Notes, Apple computers' AOCE and Oracle secures SQL. The IEEE 820.11i uses the Temporal Key Integrity Protocol (TLIP) and the Advanced Encryption Standards (AES), TKIP uses the RC4 stream cipher as the encryption and decryption algorithm, and therefore all involved parties must share the same secret key [3]. A number of papers have been published analyzing methods of attacking RC4 (*eg* [4–7]). None of these approaches is practical against RC4 with a reasonable key length, such as 128 bits. A more serious problem is reported in [8]. The authors demonstrate that the WEP protocol, intended to provide confidentiality on 802.11 wireless LAN networks, is vulnerable to a particular attack approach. In essence, the problem is not with RC4 itself but the way in which keys are generated for use as input to RC4. This particular problem does not appear to be relevant to other applications using RC4 and can be remedied in WEP by changing the way in which keys are generated. This problem points out the difficulty in designing a secure system that involves both cryptographic functions and protocols that make use of them [3]. In our paper we will use the statistical tests and mathematical measurements to prove that RC4 is secure and random enough to be used within the converged network.

[*] Faculty of Computers and Information Systems, Mansoura University, El Mahalla El Kobra, Mansheyet El Bakry 52 Hosam Shehab Street, 31962 Egypt [**] taha@aou.edu.eg

DOI: 10.2478/jee-2013-0029, ISSN 1335-3632

The RC4, like any other stream cipher, depends on the strength of its key stream. This in turns depends on the degree of randomness of its pseudo random bit generator. The output of such generator needs to meet stronger requirements than other applications. In particular, their output must be unpredictable in the absence of knowledge of the inputs. If the key (seed) is unknown, the next output number in the sequence should be unpredictable in spite of any knowledge of previous random numbers in the sequence. This property is known as forward unpredictability. In addition it should also not be feasible to determine the key (seed) from knowledge of any generated values (*ie* backward unpredictability is also required). No correlation between a seed and any value generated from that seed should be evident; each element of the sequence should appear to be the outcome of an independent random event whose probability is $1/2$.

3 RC4 SECURITY ANALYSIS

Suitable metrics are needed to investigate the degree of randomness and encryption quality for binary sequences produced by RC4, statistical testing and mathematical measurements. These will be used to gather evidence whose output sequences are truly random and have a high encryption quality, and can be used safely in the converged network applications. There are many statistical test suites that can be used in analyzing the output sequence of RC4, we will take a sample output sequence of the RC4's stream key generator (after XOR with plain text) and conduct various statistical tests. Each statistical test determines whether the sequence possesses a certain attribute that a truly random sequence would be likely to exhibit, in our first paper [9]. We evaluated the randomness properties of RC4 resulted sequences using NIST statistical test suite [10] and RC4 passed all the 16 tests with high values, and so this paper will go further for analyzing the encryption quality and the security features for the RC4 encrypted files using mathematical measurement and security evaluation tools.

3.1 Calculating the entropy of RC4 sequence

The entropy of a file is an index of its information contents [11]. It is measured in bits per character, the information content of a message $M[i]$ is defined by

$$M[i] := \log(1/p[i]) = -\log(p[i])$$

where $p[i]$ is the probability, that the message $M[i]$ is transmitted by the message source and log denotes the logarithms to base 2. This means that the information content depends exclusively on the probability distribution with which the source generates the message .The semantic content of the message does not enter into the calculation. The information content of two messages chosen independently of one another equals the sum of the information content of the individual messages, with the aid of the information content of the individual messages, the average amount of information which a source with a specified distribution delivers can be calculated as the individual messages will be weighted according to the probabilities of their occurrence.

$$\text{Entropy}(p[1], p[2], \ldots, p[r]) := \\ -\left[p[1]^* \log(p[1]) + p[2]\log(p[2]) + \cdots + p[r]\log(p[r])\right].$$

The entropy of a source thus indicates its characteristic distribution for documents which contains every character of the character set (0 to 255) the entropy lies between 0bit/char (in a document which consists of only one character) and log(256) bit/char (in a document in which all 256 characters occur equally often) [9].

Test Result

The entropy of the RC4 sequence was founded to be 7.99 out of 8 which means that every character can be founded in our output sequences.In other words,RC4 generator output has a very high information content which indicates high security.

3.2 Floating Frequency Analysis

The floating frequency of a document is a characteristic of its local information content at individual points in the document.The floating frequency specifies how many different characters are to be found in any given 64-character long segment of the document [11].

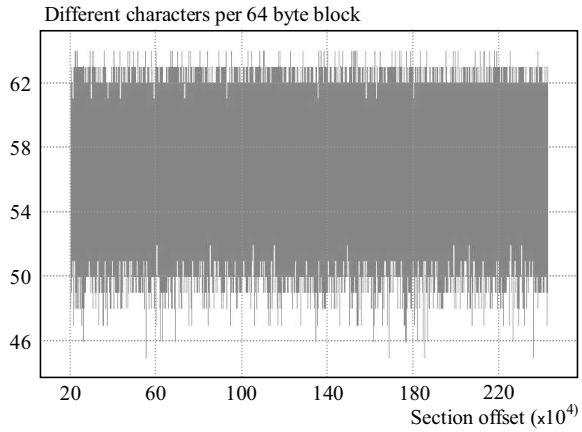

Fig. 1. Floating Frequency Analysis for Rc4 Sequences

Distribution of character frequency

The function considers sequences of text in the active window that are 64 characters long and counts how many different characters are found in this "window".The "window" is then shifted one character to the right and the calculation is repeated.This procedure results in a summary of the document in which it is possible to identify the places with high and low information density. A sequence of length $n > 64$ bytes has n-63 such index numbers in its characteristics.

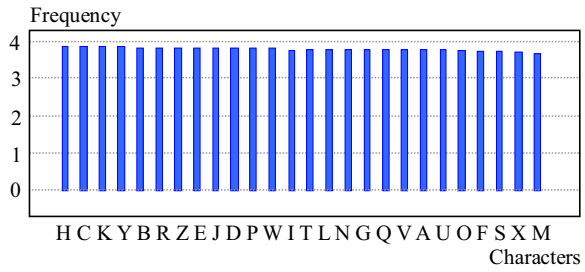

Fig. 2. Histogram Analysis of the RC4

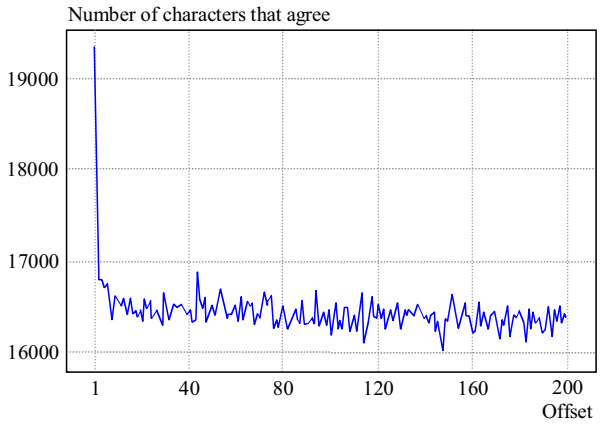

Fig. 3. Original File Autocorrelation Analysis

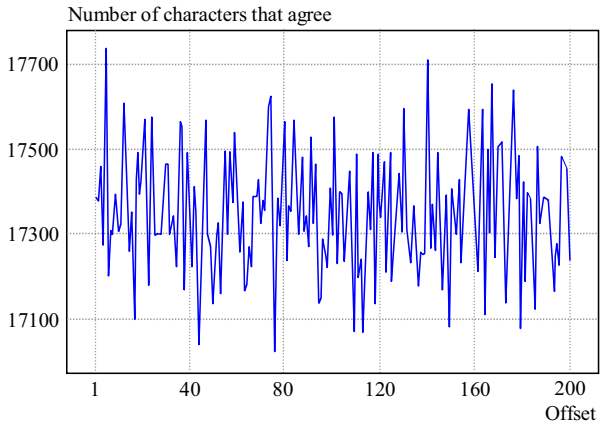

Fig. 4. The RC4 Encrypted File Autocorrelation Analysis

Depending on the structure and content of the data in the active window,different profiles are produced [9]:

- In images (bmp files) values obtained usually lie between 5 and 20.
- With text one can expect the values to fluctuate between 18 and 30.
- Executable programs have values between 30 and 40 in their actual program sections,(in the data section they fluctuate around 20).
- Compressed files,encrypted data and keys themselves have values just below the maximum value of 64.

As can be shown from Fig. 1, the structure of the RC4 sequences follows the profile expected from a truly random and secured sequence, and the location of the key cannot be specified away from the whole sequence.

3.3 Histogram Analysis

In statistics, a histogram is a graphical display of tabulated frequencies shown as bars. It shows what proportion of cases fall into each of several categories.It is a form of data binning.The categories are usually specified as non-overlapping intervals of some variable. The categories (bars) must be adjacent. The intervals (or bands, or bins) are generally of the same size, and are most easily interpreted if they are [12].

The histogram of a document expresses the frequency distribution of the characters in a document by graphical form in a corresponding window.The x-axis of the histogram contains all the characters in the character set: In a text window the character set contains the letters of the alphabet selected in Text Options,while in a window for hexadecimal inputs and outputs the character set contains the numbers 0 to 255 (see ASCII Table). The frequency of each character is shown (as a percentage) on the vertical axis.

From Fig. 2 we can see that the histogram of the RC4 encrypted file is fairly uniform and is significantly different from the graph expected from unencrypted file. Consequently no information can be deduced from it.

3.4 Autocorrelation Analysis

The autocorrelation of a random process describes the correlation between values of the process at different points in time, as a function of the two times or of the time difference [14]. The autocorrelation of a sequence is an index of the similarity of different sections of the sequence. It is sometimes possible to work out the key length of an encrypted file from its auto correlation.The purpose of this empirical test of independence is to check correlation between succeeding outcomes of the pseudorandom number generator and/or between the binary sequence S and version of S that has been displaced by t positions [13].In other words the autocorrelation function $C(t)$ measures the similarity of sequence $(S[i]) = S[1]\,S[2]\ldots$ to sequence $(S[i+t]) = S[t+1]\,S[t+2]\ldots$ which is shifted by t places.A sequence of length n is examined and the following parameters are defined

$A(t) :=$ the number of sequences $(S[i])$ and $(S[i+t])$ in the segment under consideration which agree.

$D(t) :=$ the number of sequences $(S[i])$ and $(S[i+t])$ in the segment under consideration which do not agree.

The autocorrelation function $C(t) = (A(t) - D(t))/n$.

Figure 3 illustrates the autocorrelation analysis of a video file called “Nasheed El Arkam” with 18 MB size, from which we can see that there are a high correlation in the sequence at some shift value, while Fig. 4 illustrates that after encrypting this file using RC4 algorithm, there is no correlation at all, which in turns means good statistical properties for the RC4 resulted sequence.

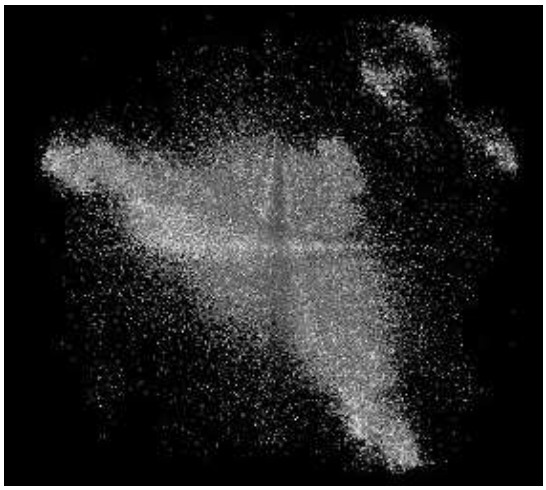

Fig. 5. Plain image space visualization

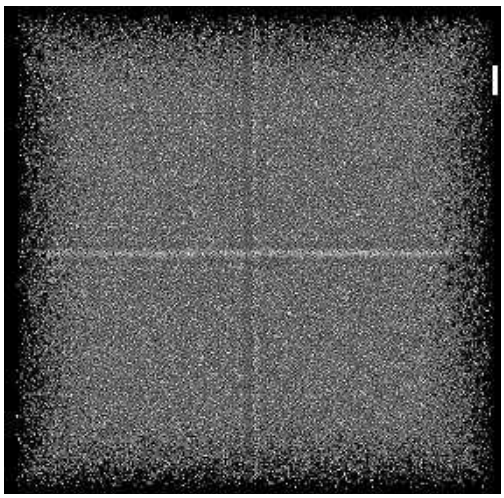

Fig. 6. Encrypted Image Space Visualization

Table 1. Complexity analysis for RC4 algorithm

Sequence Length	Number of LFSRs
531 k	271875
1.410 M	721811
2.417 M	1237503

Table 2. Encryption quality (efficiency) of RC4 stream cipher

File type			
Image (.jpg)	Document (.xls)	Audio (.mp3)	Video (.flv)
742.7	784	4934	60780.6445

Table 3. Encryption quality of RC4 ,RC5 and RC6 for Lena image

secret key length	algorithm type		
128 bit	RC4	RC5	RC6
	7197.476	724.469	722.141

3.5 Random number visualization analysis

Structures in random number sequences can also be visualized graphically.CrypTool [13] implements an algorithm that is called phase space visualization which was first implemented by Dan Kaminsky of DoxPara in his program *Phentropy* (Part of the Paketto Keiretsu Toolkit).

Figure 5 shows the space visualization of plain image from which we can recognize characteristic patterns which can indicate the inner structure of the input data.While Fig. 6 which represents the space visualizing of the RC4 stream for the same image,thus we can see its visualization as a uniform fog-like filling which indicates true randomness.

3.6 Complexity Analysis

In this section,we will use the Berlekamp-Massey algorithm to find the shortest linear feedback shift register (LFSR) for a given RC4 sequence. Equivalently, it is an algorithm for finding the minimal polynomial of a linearly recurrent sequence [16].

The algorithm

1. Let $s_0, s_1, s_2, \ldots, s_n$ be the bits of the stream.
2. Initialize three arrays b, c and t each of length n to be zeroes, except $b_0 \leftarrow 1$, $c_0 \leftarrow 1$; assign $N \leftarrow 0$, $L \leftarrow 0$, $m \leftarrow 1$.
3. While N is less than n:
   - Let d be $s_N + c_1 s_{N-1} + c_2 s_{N-2} \cdots + c_L s_{N-L}$
   - If d is zero, then c is already a polynomial which annihilates the portion of the stream from $N - L$ to N; increases N by 1 and continue.
   - If d is 1, then
     - Let t be a copy of c.
     - Set $C_{N-m} \leftarrow C_{N-m}$ XOR b_0, $C_{N-m+1} \leftarrow C_{N-m+1}$ XOR b_1 ... up to $C_{n-1} \leftarrow C_{n-1}$ XOR $b_{n-N+m-1}$
     - If $L \leq \frac{N}{2}$,set $L \leftarrow N + 1 - L$, set $m \leftarrow N$, and let $b \leftarrow t$; otherwise leave L, m and b alone.
     - Increment N and continue.

At the end of the algorithm, L is the length of the minimal LFSR for the stream,and we have

$C_L s_a + c_{L-1} s_{a+1} + c_{L-2} s_{a+2} + \cdots = 0$ for all a [14].

The results obtaining after implementing this algorithm are in Tab. 1.

4 ENCRYPTION QUALITY MEASUREMENT OF RC4 STREAM CIPHER

There is a need to develop mathematical measure to evaluate the degree of encryption quantity which we call the encryption quality [17]. We will evaluate RC4 encryption quality as a function of secret key length. A measure

of encryption quality may be expressed as the deviation between the original and encrypted files, for example the quality of image encryption [18] may be determined as follows.

Let F and F' denote the original image(plain image) and the encrypted image (cipher image) respectively (each of size $M * N$ pixels with L grey levels), and $F(x,y), F'(x,y) \in \{0,\ldots,L-1\}$ are the grey levels of the images F and F' at position (x,y) $(0 \le x \le M-1,$ $0 \le y \le N-1)$. Let $H_L(F)$ denote the number of occurrences of each grey level L in the original image (plain image) F. Similarly, $H_L(F')$ denotes the number of occurrences of each grey level L in the encrypted image (cipher image) F'. The encryption quality represents the average number of changes to each grey level L and is expressed mathematically as

$$\text{Encryption Quality } = \frac{1}{256}\sum_{L=0}^{255} \left|H_L(F') - H_L(F)\right|.$$

With the application of encryption to an image,a change takes place in pixels values as compared to those values before encryption. The higher the change in pixels values,the more effective will be the image encryption and hence the encryption quality. Table 2 illustrates the high values resulted from encryption quality analysis for RC4 which indicated its high quality.

To evaluate the encryption quality of the RC4 by comparing it to other know block ciphers, we take the image of lena.bmp size 512×512 and compared the encryption quality obtained from RC4 with those given from the RC5 and RC6 block ciphers[18]. The results listed on Tab. 3. This illustrates the high quality of encryption obtained from RC4 compared with other ciphers.

5 CONCLUSION

This paper introduces a successful efficient implementation of RC4 stream cipher for converged networks real time communications; providing its testing, verification, security evaluation and encryption efficiency. After evaluating and testing processes for RC4, it can be concluded that RC4 is a very secure and efficient stream cipher that can be used to secure the converged networks.

References

[1] SHINDER, D.: Deploying VoIP: Weigh the pros and cons of convergence http://articles.techrepublic.com.com/ 5100-10878_11-6187594.htm, May 31, 2007, last visit 7/8/2010.

[2] Computers at Risk: Safe computing in the information age, USA National Research Council,1991.

[3] STALLING, W.: Cryptography and Network Security, Prentice Hall, 2005.

[4] KNUDSEN, L.*et al*: Analysis Method for Alleged RC4, Proceedings ASIACRYPT '98, 1998.

[5] MISTER, S., ANDTAVARES, S.: Cryptanalysis of RC4-Like Ciphers, Proceedings Workshop in Selected Areas of Cryptography SAC '98, 1998.

[6] FLUHRER, S.—McGREW, D.: Statistical Analysis of the Alleged RC4 Key Stream Generator, Proceedings Fast Software Encryption, 2000.

[7] MANTIN, I.—HAMIR, A.: A Practical Attack on Broadcast RC4, Proceedings Fast Software Encryption, 2001.

[8] FLUHRER, S.—MANTIN, I.—SHAMIR, A.: Weakness in the Key Scheduling Algorithm of RC4., Proceedings Workshop in Selected Areas of Cryptography, 2001.

[9] RIAD, A. M.—SHEHATA, A. R.—HAMDY, E. K.—ABOU-ALSOUAD, M. H.—IBRAHIM, T. R.: Evaluation of the RC4 Algorithm as a Solution for Converged Networks, J. Electrical Engineering **60** No. 3 (June 2009).

[10] A Statiistiical Test Suiite for Random and Pseudorandom Number Generators for Cryptographic Applications, NIST Special Publication 800-22, 2001.

[11] ESSLINGER, B.: The CrypTool Script: Cryptography, Mathematics and More, www.cryptool.org,2008.

[12] HOWITT, D. ANDCRAMER, D.: Statistics in Psychology, Prentice Hall, 2008.

[13] ESSLINGER, B.: Cryptology with cryptool v 1.4.20, cryptool v1.4.21, www.cryptool.com, 2009.

[14] Wikipedia.

[15] SCHIESTL, C.: Pseudozufallszahlen in der Kryptographie, Klagenfurt, 1999.

[16] http://www.bookrags.com/wiki/Berlekamp-Massey_algorithm, Last visit 8/8/2010.

[17] HOSAM, El-dinH.: Encryption Efficiency Analysis and Security Evaluation Of R C6 Block Cipher for Digital Images, International Journal of Computer, Information, System Science and Engineering **1** No. 1 (winter 2007).

[18] HOSSAM, El-din H. AHMED—KALASH, H. M.—FARAG ALLAH, O. S.: Encryption Quality Analysis of RC5 Block Cipher Algorithm for Digital Images, Journal of Optical Engineering **45** (2006).

Received 28 March 2011

Alaa E Din Reyad (Prof) graduated in Mansoura University from electrical engineering department in 1982. Obtained Master degree in 1988, and Doctoral degree in 1992. He is the Dean of Faculty of Computers and Information Systems Mansoura University - Egypt, and a member of Egyptian Universities Promotion Committees. His main research activity is currently in intelligent information systems and e-Learning.

Hamdy K. Elminir was born in EI-Mahala, Egypt in 1968. He received the BSc in Engineering from Monofia University, in 1991 and completed his master degree in automatic control system in 1996. He obtained his PhD degree from the Czech Technical University in Prague in 2001. Currently he is head of communication engineering departement, Misr Highier Institute for Engineering and Technology, Egypt.

Taha R. Ibrahim was born in El-Mahalla, Egypt in 1974. He received the BSc in electronic engineering from Faculty of Electronic Engineering in Menouf, Egypt, and his MSc degree in converged network security in El Mansoura University-Egypt. Now he is working as an IT specialist in the Arab Open University-Egypt.

Alaa Eldin Rohiem Received the MSc degree in Electrical Engineering from MTC in 1994 and PhD degree in Electrical Engineering from University of Kent, UK in 2000. He has also worked as a lecturer of Communication Engineering at MTC. Associative Professor in Communication Engineering at MTC. His research interests are in cryptographic algorithms, protocols for cryptography and computer network security.