ISSN 2313-1527 (PRINT)
ISSN 2313-1535 (ONLINE)

IJEIE

International Journal of Electronics and Information Engineering

Vol. 6, No. 2 (June 2017)

INTERNATIONAL JOURNAL OF ELECTRONICS & INFORMATION ENGINEERING

PUBLISHING OFFICE

Min-Shiang Hwang
Department of Computer Science & Information Engineering, Asia University, Taichung 41354, Taiwan, R.O.C.
Email: mshwang@asia.edu.tw

International Journal of Electronics and Information Engineering is published both in traditional paper form (ISSN 2313-1527) and in Internet (ISSN 2313-1535) at http://ijeie.jalaxy.com.tw

PUBLISHER: Candy C. H. Lin

23-75, P.O. Box, Taichung, Taiwan 40199, R.O.C.

International Journal of Electronics and Information Engineering

Vol. 6, No. 2 (June 1, 2017)

The Encryption Algorithms GOST28147-89-IDEA8-4 and GOST28147-89-RFWKIDEA8-4

Tuychiev Gulom

National University of Uzbekistan, Republic of Uzbekistan, Tashkent
4 Universitet St, Tashkent 100174, Uzbekistan
(Email: blasterjon@gmail.com)
(Received Jan. 13, 2017; revised and accepted Mar. 18, 2017)

Abstract

In the paper created a new encryption algorithms GOST28147-89-IDEA8-4 and GOST28147-89-RFWKIDEA8-4 based on networks IDEA8-4 and RFWKIDEA8-4, with the use the round function of the encryption algorithm GOST 28147-89. The block length of created block encryption algorithm is 256 bits, the number of rounds is 8, 12 and 16.

Keywords: GOST 28147-89; Lai-Massey Scheme; Round Function; Round Keys; Output Transformation

1 Introduction

The encryption algorithm GOST 28147-89 [7] is a standard encryption algorithm of the Russian Federation and based on a Feistel network. This encryption algorithm is suitable for hardware and software implementation, meets the necessary cryptographic requirements for resistance and, therefore, does not impose restrictions on the degree of secrecy of the information being protected. The algorithm implements the encryption of 64-bit blocks of data using the 256 bit key. In round functions used eight S-box of size 4x4 and operation of the cyclic shift by 11 bits. To date GOST 28147-89 is resistant to cryptographic attacks.

On the basis of structure encryption algorithm IDEA [6] and Lai-Massey scheme developed networks IDEA8-4 [8] and RFWKIDEA8-4 [15], consisting from four round function. In the networks IDEA8-4 and RFWKIDEA8-4, similarly as in the Feistel network, in encryption and decryption process using the same algorithm. In the networks used four round function having one input and output blocks and as the round function can use any transformation.

As the round function networks IDEA4-2 [1], RFWKIDEA4-2 [12], PES4-2 [11], RFWKPES4-2 [22], PES8-4 [2], RFWKPES8-4 [10] using the round function of the encryption algorithm GOST 28147-89 created the encryption algorithm GOST28147-89-IDEA4-2 [19], GOST28147-89-RFWKIDEA4-2 [28], GOST28147-89-PES4-2 [27], GOST28147-89-RFWKPES4-2 [29], GOST28147-89-PES8-4 [34] and GOST28147-89-RFWKPES8-4 [34].

In addition, by using SubBytes(), ShiftRows(), MixColumns() and AddRoundKey() transformations of the encryption algorithm AES [5] as round functions of networks IDEA8-1 [15], RFWKIDEA8-1 [15], PES8-1 [9], RFWKPES8-1 [10], IDEA16-1 [13], RFWKIDEA16-1 [17], PES16-1 [21], RFWKPES16-1 [23], IDEA32-1 [14], RFWKIDEA32-1 [39], PES32-1 [16], RFWKPES32-1 [18], IDEA16-2 [13], PES16-2 [21], RFWKIDEA16-2 [17], RFWKPES16-2 [23], IDEA32-4 [14], RFWKIDEA32-4 [39], PES32-4 [16], RFWKPES32-4 [18] created encryption algorithms AES-IDEA8-1 [36], AES-RFWKIDEA8-1 [38], AES-PES8-1 [37], AES-RFWKPES8-1 [20], AES-IDEA16-1 [35], AES-RFWKIDEA16-1 [31], AES-PES16-1 [33], AES-RFWKPES16-1 [33], AES-IDEA32-1 [24], AES-RFWKIDEA32-1 [32], AES-PES32-1 [25], AES-RFWKPES32-1 [25], AES-IDEA16-2 [30], AES-RFWKIDEA16-2 [30], AES-PES16-2 [30], AES-RFWKPES16-2 [30], AES-IDEA32-4 [26], AES-RFWKIDEA32-4 [26], AES-PES32-4 [3], AES-RFWKPES32-4 [40].

In this paper, applying the round function of the encryption algorithm GOST 28147-89 as round functions of the networks IDEA8-4 and RFWKIDEA8-4, developed new encryption algorithms GOST28147-89-IDEA8-4 and GOST28147-89-RFWKIDEA8-4. In encryption algorithms GOST28147-89-IDEA8-4 and GOST28147-89-RFWKIDEA8-4 block length is 256 bits, the key length is changed from 256 bits to 1024 bits in increments of 128 bits and number of rounds equal to 8, 12, 16, allowing the user depending on the degree of secrecy of information and speed of encryption to choose the number of rounds and key length.

2 The Encryption Algorithm GOST28147-89-IDEA8-4

2.1 The Structure of the Encryption Algorithm GOST28147-89-IDEA8-4

In the encryption algorithm GOST28147-89-IDEA8-4 length of subblocks X^0, X^1, ..., X^7, length of round keys $K_{12(i-1)}$, $K_{12(i-1)+1}$, ..., $K_{12(i-1)+7}$, $i = \overline{1...n+1}$, $K_{12(i-1)+8}$, $K_{12(i-1)+9}$, ..., $K_{12(i-1)+11}$, $i = \overline{1...n}$ and K_{12n+8}, K_{12n+9}, ..., K_{12n+23} are equal to 32-bits. In this encryption algorithm the round function GOST 28147-89 is applied four time and in each round function used eight S-boxes, i.e. the total number of S-boxes is 32. The structure of the encryption algorithm GOST28147-89-IDEA8-4 is shown in Figure 1 and the S-boxes shown in Table 1.

Consider the round function of encryption algorithm GOST28147-89-IDEA8-4. First 32-bit subblocks T^0, T^1, T^2, T^3 are summed round keys $K_{12(i-1)+8}$, $K_{12(i-1)+9}$, $K_{12(i-1)+10}$, $K_{12(i-1)+11}$ i.e. $S^0=T^0+K_{12(i-1)+8}$, $S^1=T^1+K_{12(i-1)+9}$, $S^2=T^2+K_{12(i-1)+10}$, $S^3=T^3+K_{12(i-1)+11}$. 32-bit subblocks S^0, S^1, S^2, S^3 divided into eight four-bit subblocks $S^0=s_0^0||s_1^0||...||s_7^0$, $S^1=s_0^1||s_1^1||...||s_7^1$, $S^2=s_0^2||s_1^2||...||s_7^2$, $S^3=s_0^3||s_1^3||...||s_7^3$. Four bit subblocks s_i^0, s_i^1, s_i^2, s_i^3, $i = \overline{0...7}$ transformed into the S-boxes:

$$\begin{aligned} R^0 &= S_0(s_0^0)||S_1(s_1^0)||...||S_7(s_7^0), \\ R^1 &= S_8(s_0^1)||S_9(s_1^1)||...||S_{15}(s_7^1), \\ R^2 &= S_{16}(s_0^2)||S_{17}(s_1^2)||...||S_{23}(s_7^2), \\ R^3 &= S_{24}(s_0^3)||S_{25}(s_1^3)||...||S_{31}(s_7^3). \end{aligned}$$

The resulting 32-bit subblocks R^0, R^1, R^2, R^3 cyclically shifted left by 11 bits and obtain subblocks Y_0, Y_1, Y_2, Y_3: $Y_0=R^0<<11$, $Y_1=R^1<<11$, $Y_2=R^2<<11$, $Y_3=R^3<<11$.

Consider the encryption process of encryption algorithm GOST28147-89-IDEA8-4. Initially the 256-bit plaintext X partitioned into subblocks of 32-bits X_0^0, X_0^1, ..., X_0^7, and performs the following steps:

1) subblocks X_0^0, X_0^1, ..., X_0^7 summed by XOR with round key K_{12n+8}, K_{12n+9}, ..., K_{12n+15}: $X_0^j = X_0^j \oplus K_{12n+8+j}$, $j = \overline{0...7}$.

Table 1: The S-boxes of encryption algorithms

	0x0	**0x1**	**0x2**	**0x3**	**0x4**	**0x5**	**0x6**	**0x7**	**0x8**	**0x9**	**0xA**	**0xB**	**0xC**	**0xD**	**0xE**	**0xF**
S_0	0x4	0x5	0xA	0x8	0xD	0x9	0xE	0x2	0x6	0xF	0xC	0x7	0x0	0x3	0x1	0xB
S_1	0x5	0x4	0xB	0x9	0xC	0x8	0xF	0x3	0x7	0xE	0xD	0x6	0x1	0x2	0x0	0xA
S_2	0x6	0x7	0x8	0xA	0xF	0xB	0xC	0x0	0x4	0xD	0xE	0x5	0x2	0x1	0x3	0x9
S_3	0x7	0x6	0x9	0xB	0xE	0xA	0xD	0x1	0x5	0xC	0xF	0x4	0x3	0x0	0x2	0x8
S_4	0x8	0x9	0x6	0x4	0x1	0x5	0x2	0xE	0xA	0x3	0x0	0xB	0xC	0xF	0xD	0x7
S_5	0x9	0x8	0x7	0x5	0x0	0x4	0x3	0xF	0xB	0x2	0x1	0xA	0xD	0xE	0xC	0x6
S_6	0xA	0xB	0x4	0x6	0x3	0x7	0x0	0xC	0x8	0x1	0x2	0x9	0xE	0xD	0xF	0x5
S_7	0xB	0xA	0x5	0x7	0x2	0x6	0x1	0xD	0x9	0x0	0x3	0x8	0xF	0xC	0xE	0x4
S_8	0xC	0xD	0x2	0x0	0x5	0x1	0x6	0xA	0xE	0x7	0x4	0xF	0x8	0xB	0x9	0x3
S_9	0xE	0xF	0x0	0x2	0x7	0x3	0x4	0x8	0xC	0x5	0x6	0xD	0xA	0x9	0xB	0x1
S_{10}	0xF	0xE	0x1	0x3	0x6	0x2	0x5	0x9	0xD	0x4	0x7	0xC	0xB	0x8	0xA	0x0
S_{11}	0x1	0x8	0x7	0xD	0x0	0x4	0x3	0xF	0xB	0xA	0x9	0x2	0x5	0x6	0xC	0xE
S_{12}	0x2	0xB	0x4	0xE	0x3	0x7	0x0	0xC	0x8	0x9	0xA	0x1	0x6	0x5	0xF	0xD
S_{13}	0x3	0xA	0x5	0xF	0x2	0x6	0x1	0xD	0x9	0x8	0xB	0x0	0x7	0x4	0xE	0xC
S_{14}	0x4	0x5	0xA	0x0	0xD	0x1	0x6	0x2	0xE	0x7	0xC	0xF	0x8	0x3	0x9	0xB
S_{15}	0x5	0x4	0xB	0x1	0xC	0x0	0x7	0x3	0xF	0x6	0xD	0xE	0x9	0x2	0x8	0xA
S_{16}	0x6	0x7	0x8	0x2	0xF	0x3	0x4	0x0	0xC	0x5	0xE	0xD	0xA	0x1	0xB	0x9
S_{17}	0x7	0x6	0x9	0x3	0xE	0x2	0x5	0x1	0xD	0x4	0xF	0xC	0xB	0x0	0xA	0x8
S_{18}	0x8	0x9	0x6	0xC	0x1	0xD	0xA	0xE	0x2	0xB	0x0	0x3	0x4	0xF	0x5	0x7
S_{19}	0x9	0x8	0x7	0xD	0x0	0xC	0xB	0xF	0x3	0xA	0x1	0x2	0x5	0xE	0x4	0x6
S_{20}	0xA	0xB	0x4	0xE	0x3	0xF	0x8	0xC	0x0	0x9	0x2	0x1	0x6	0xD	0x7	0x5
S_{21}	0xB	0xA	0x5	0xF	0x2	0xE	0x9	0xD	0x1	0x8	0x3	0x0	0x7	0xC	0x6	0x4
S_{22}	0xC	0xD	0x2	0x8	0x5	0x9	0xE	0xA	0x6	0xF	0x4	0x7	0x0	0xB	0x1	0x3
S_{23}	0xD	0xC	0x3	0x9	0x4	0x8	0xF	0xB	0x7	0xE	0x5	0x6	0x1	0xA	0x0	0x2
S_{24}	0x1	0x8	0x7	0x5	0x0	0xC	0xB	0xF	0x3	0x2	0x9	0xA	0xD	0x6	0x4	0xE
S_{25}	0x2	0xB	0x4	0x6	0x3	0xF	0x8	0xC	0x0	0x1	0xA	0x9	0xE	0x5	0x7	0xD
S_{26}	0x3	0xA	0x5	0x7	0x2	0xE	0x9	0xD	0x1	0x0	0xB	0x8	0xF	0x4	0x6	0xC
S_{27}	0xF	0xE	0x1	0xB	0x6	0xA	0xD	0x9	0x5	0xC	0x7	0x4	0x3	0x8	0x2	0x0
S_{28}	0xE	0xF	0x0	0xA	0x7	0xB	0xC	0x8	0x4	0xD	0x6	0x5	0x2	0x9	0x3	0x1
S_{29}	0xA	0xB	0xC	0xE	0x3	0xF	0x0	0x4	0x8	0x1	0x2	0x9	0x6	0x5	0x7	0xD
S_{30}	0xB	0xA	0xD	0xF	0x2	0xE	0x1	0x5	0x9	0x0	0x3	0x8	0x7	0x4	0x6	0xC
S_{31}	0xC	0xD	0xA	0x8	0x5	0x9	0x6	0x2	0xE	0x7	0x4	0xF	0x0	0x3	0x1	0xB

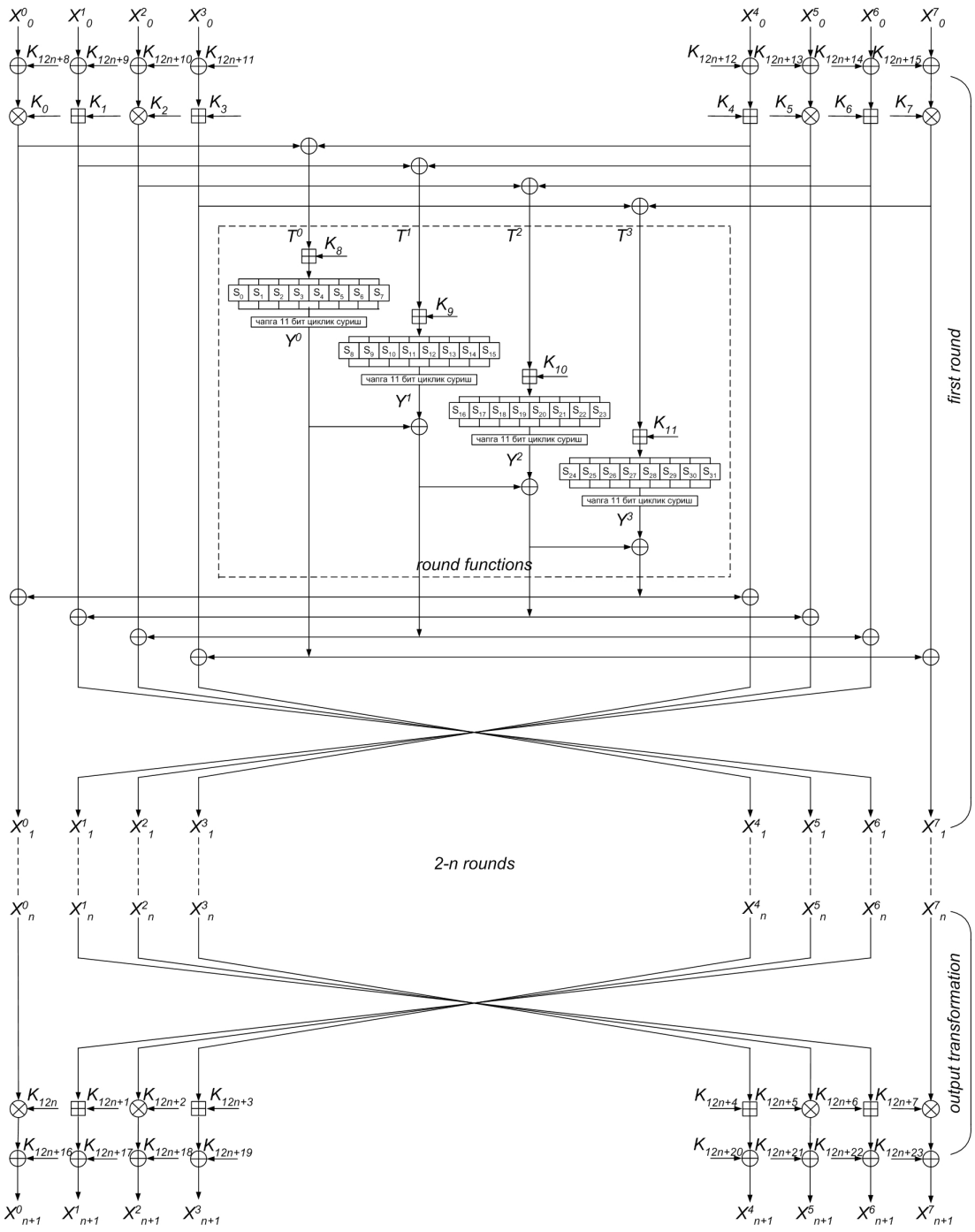


Figure 1: The scheme of encryption algorithm GOST28147-89-IDEA8-4

2) subblocks X_0^0, X_0^1, ..., X_0^7 multiplied and summed with the round keys $K_{12(i-1)}$, $K_{12(i-1)+1}$, ..., $K_{12(i-1)+7}$ and calculated 32-bit subblocks T^0, T^1, T^2, T^3. This step can be represented as follows:

$$\begin{aligned} T^0 &= (X_{i-1}^0 \cdot K_{12(i-1)}) \oplus (X_{i-1}^4 + K_{12(i-1)+4}), \\ T^1 &= (X_{i-1}^1 + K_{12(i-1)+1}) \oplus (X_{i-1}^5 \cdot K_{12(i-1)+5}), \\ T^2 &= (X_{i-1}^2 \cdot K_{12(i-1)+2}) \oplus (X_{i-1}^6 + K_{12(i-1)+6}), \\ T^3 &= (X_{i-1}^3 + K_{12(i-1)+3}) \oplus (X_{i-1}^7 \cdot K_{12(i-1)+7}), i = 1. \end{aligned}$$

3) to subblocks T^0, T^1, T^2, T^3 applying the round function and get the 32-bit subblocks Y^0, Y^1, Y^2, Y^3.

4) subblocks Y^0, Y^1, Y^2, Y^3 are summed to XOR with subblocks X_{i-1}^0, X_{i-1}^1, X_{i-1}^2, X_{i-1}^3, i..

$$\begin{aligned} X_{i-1}^0 &= X_{i-1}^0 \oplus Y^3, \\ X_{i-1}^1 &= X_{i-1}^1 \oplus Y^2, \\ X_{i-1}^2 &= X_{i-1}^2 \oplus Y^1, \\ X_{i-1}^3 &= X_{i-1}^3 \oplus Y^0, \\ X_{i-1}^4 &= X_{i-1}^4 \oplus Y^3, \\ X_{i-1}^5 &= X_{i-1}^5 \oplus Y^2, \\ X_{i-1}^6 &= X_{i-1}^6 \oplus Y^1, \\ X_{i-1}^7 &= X_{i-1}^7 \oplus Y^0, i = 1. \end{aligned}$$

5) at the end of the round subblocks swapped, i.., $X_i^0 = X_{i-1}^0$, $X_i^7 = X_{i-1}^7$, $X_i^1 = X_{i-1}^6$, $X_i^2 = X_{i-1}^5$, $X_i^3 = X_{i-1}^4$, $X_i^4 = X_{i-1}^3$, $X_i^5 = X_{i-1}^2$, $X_i^6 = X_{i-1}^1$, $i = 1$.

6) repeating steps 2-5 n times, i.e., $i = \overline{2...n}$ obtain 32-bit subblocks X_n^0, X_n^1, ..., X_n^7.

7) in output transformation round keys K_{12n}, K_{12n+1}, ..., K_{12n+7} are multiplied and summed into subblocks, i.e.

$$\begin{aligned} X_{n+1}^0 &= X_n^j \cdot K_{12n}, \\ X_{n+1}^1 &= X_n^6 + K_{12n+1}, \\ X_{n+1}^2 &= X_n^5 \cdot K_{12n+2}, \\ X_{n+1}^3 &= X_n^4 + K_{12n+3}, \\ X_{n+1}^4 &= X_n^3 + K_{12n+4}, \\ X_{n+1}^5 &= X_n^2 \cdot K_{12n+5}, \\ X_{n+1}^6 &= X_n^1 + K_{12n+6}, \\ X_{n+1}^7 &= X_n^7 \cdot K_{12n+7}. \end{aligned}$$

8) subblocks X_{n+1}^0, X_{n+1}^1, ..., X_{n+1}^7 are summed to XOR with the round key K_{12n+16}, K_{12n+17}, ..., K_{12n+23}: $X_{n+1}^j = X_{n+1}^j \oplus K_{12n+16+j}$, $j = \overline{0...7}$.

As ciphertext plaintext X receives the combined 32-bit subblocks $X_{n+1}^0 || X_{n+1}^1 || ... || X_{n+1}^7$.

In the encryption algorithm GOST28147-89-IDEA8-4 when encryption and decryption using the same algorithm, only when decryption calculates the inverse of round keys depending on operations and are applied in reverse order. One important goal of encryption is key generation.

2.2 Key Generation of the Encryption Algorithm GOST28147-89-IDEA8-4

In n-round encryption algorithm GOST28147-89-IDEA8-4 in each round used 12 round keys of 32 bits and the output transformation of 8 round keys of 32 bits. In addition, prior to the first round and after the output transformation is applied 8 round keys on 32 bits. The total number of 32-bit round keys is equal to 12n+24. Hence, if n=8 then necessary 120, if n=12 then 168 and if n=16 then 216 to generate round keys. In Figure 1 in encryption used encryption round keys K_i^c instead of K_i, while decryption used decryption round keys K_i^d.

The key encryption algorithm K of length l ($256 \le l \le 1024$) bits is divided into 32-bit round keys K_0^c, K_1^c,..., $K_{Lenght-1}^c$, $Lenght = l/32$, here $K = \{k_0, k_1, ..., k_{l-1}\}$, $K_0^c = \{k_0, k_1, ..., k_{31}\}$, $K_1^c = \{k_{32}, k_{33}, ..., k_{63}\}$,..., $K_{Lenght-1}^c = \{k_{l-32}, k_{l-31}, ..., k_{l-1}\}$ and $K = K_0^c||\, K_1^c||...||K_{Lenght-1}^c$. Then we calculate $K_L = K_0^c \oplus K_1^c \oplus ... \oplus K_{Lenght-1}^c$. If $K_L = 0$ then K_L is chosen as 0xC5C31537, i.e. $K_L =$ 0xC5C31537. Round keys K_i^c, $i = \overline{Lenght...12n+23}$ calculated as follows: $K_i^c = SBox0(K_{i-Lenght}^c) \oplus SBox1(RotWord32(K_{i-Lenght+1}^c)) \oplus K_L$.

After each round key generation the value K_L is cyclic shift to the left by 1 bit. Here *RotWord32()*-cyclic shift to the left of 1 bit of the 32-bit subblock, *SBox()*convert 32-bit subblock in S-box and $SBox0(A) = S_0(a_0)||\, S_1(a_1)||\, ...||\, S_7(a_7)$, $SBox1(A) = S_9(a_0)||\, S_{10}(a_1)||...||\, S_{15}(a_7)$, $A = a_0||\, a_1||...||\, a_7$ and a_i-four-bit subblock, S_i-i-th S-Box.

Decryption round keys are computed on the basis of encryption round keys and decryption round keys of the first round associate with of encryption round keys as follows:

$$\begin{aligned} & (K_0^d, K_1^d, K_2^d, K_3^d, K_4^d, K_5^d, K_6^d, K_7^d) \\ = \; & ((K_{12n}^c)^{-1}, -K_{12n+1}^c, (K_{12n+2}^c)^{-1}, -K_{12n+3}^c, -K_{12n+4}^c, (K_{12n+5}^c)^{-1}, -K_{12n+6}^c, (K_{12n+7}^c)^{-1}, \\ & K_{12(n-1)+8}^c, K_{12(n-1)+9}^c, K_{12(n-1)+10}^c, K_{12(n-1)+11}^c). \end{aligned}$$

Decryption round keys of the second, third and n-round associates with the encryption round keys as follows:

$$\begin{aligned} & (K_{12(i-1)}^d, K_{12(i-1)+1}^d, K_{12(i-1)+2}^d, K_{12(i-1)+3}^d, K_{12(i-1)+4}^d, K_{12(i-1)+5}^d, K_{12(i-1)+6}^d, \\ & K_{12(i-1)+7}^d, K_{12(i-1)+8}^d, K_{12(i-1)+9}^d, K_{12(i-1)+10}^d, K_{12(i-1)+11}^d) \\ = \; & ((K_{12(n-i+1)}^c)^{-1}, -K_{12(n-i+1)+6}^c, (K_{12(n-i+1)+5}^c)^{-1}, -K_{12(n-i+1)+4}^c, -K_{12(n-i+1)+3}^c, \\ & (K_{12(n-i+1)+2}^c)^{-1}, -K_{12(n-i+1)+1}^c, (K_{12(n-i+1)+7}^c)^{-1}, K_{12(n-i)+8}^c, K_{12(n-i)+9}^c, \\ & K_{12(n-i)+10}^c, K_{12(n-i)+11}^c), i = \overline{2...n}. \end{aligned}$$

Decryption keys output transformation associated with the encryption keys as follows:

$$\begin{aligned} & (K_{12n}^d, K_{12n+1}^d, K_{12n+2}^d, K_{12n+3}^d, K_{12n+4}^d, K_{12n+5}^d, K_{12n+6}^d, K_{12n+7}^d) \\ = \; & ((K_0^c)^{-1}, -K_1^c, (K_2^c)^{-1}, -K_3^c, -K_4^c, (K_5^c)^{-1}, -K_6^c, (K_7^c)^{-1}). \end{aligned}$$

Decryption round keys applied to the first round and after the output transformation associated with the encryption round keys as follows: $K_{12n+8+j}^d = K_{12n+16+j}^c$, $K_{12n+16+j}^d = K_{12n+8+j}^c$, $j = \overline{0...7}$.

3 The Encryption Algorithm GOST28147-89-RFWKIDEA8-4

3.1 The Structure of the Encryption Algorithm GOST28147-89-RFWKIDEA8-4

In the encryption algorithm GOST28147-89-RFWKIDEA8-4 the length of subblocks X^0, X^1, ..., X^7, length of round keys $K_{8(i-1)}$, $K_{8(i-1)+1}$, ..., $K_{8(i-1)+7}$, $i = \overline{1...n+1}$, $K_{8(i-1)+8}$, $K_{8(i-1)+9}$, ...,

$K_{8(i-1)+11}$, $i = \overline{1...n}$ and K_{8n+8}, K_{8n+9}, ..., K_{8n+23} are equal to 32-bits. In this encryption algorithm the round function GOST 28147-89 is applied four time and in each round function used eight S-boxes, i.e. the total number of S-boxes is 32. The structure of the encryption algorithm GOST28147-89-IDEA8-4 is shown in Figure 2 and the S-boxes shown in Table 1.

Consider the round function of encryption algorithm GOST28147-89-RFWKIDEA8-4. First 32-bit subblocks T^0, T^1, T^2, T^3 divided into eight four-bit sub-blocks, i.e.

$$\begin{aligned} T^0 &= t_0^0||t_1^0||...||t_7^0, \\ T^1 &= t_0^1||t_1^1||...||t_7^1, \\ T^2 &= t_0^2||t_1^2||...||t_7^2, \\ T^3 &= t_0^3||t_1^3||...||t_7^3. \end{aligned}$$

The four bit subblocks t_i^0, t_i^1, t_i^2, t_i^3, $i = \overline{0...7}$ converted into the S-boxes:

$$\begin{aligned} R^0 &= S_0(t_0^0)||S_1(t_1^0)||...||S_7(t_7^0), \\ R^1 &= S_8(t_0^1)||S_9(t_1^1)||...||S_{15}(t_7^1), \\ R^2 &= S_{16}(t0^2)||S_{17}(t_1^1)||...||S_{23}(t_7^2), \\ R^3 &= S_{24}(t_0^3)||S_{25}(t_1^3)||...||S_{31}(t_7^2). \end{aligned}$$

The resulting 32-bit subblocks R^0, R^1, R^2, R^3 cyclically shifted left by 11 bits and obtain subblocks Y_0, Y_1, Y_2, Y_3: Y_0=R^0<<11, Y_1=R^1<<11, Y_2=R^2<<11, Y_3=R^3<<11 .

Consider the encryption process of encryption algorithm GOST28147-89-RFWKIDEA8-4. Initially the 256-bit plaintext X partitioned into subblocks of 32-bits X_0^0, X_0^1, ..., X_0^7, and performs the following steps:

1) subblocks X_0^0, X_0^1, ..., X_0^7 summed by XOR with round key K_{8n+8}, K_{8n+9}, ..., K_{8n+15}: $X_0^j = X_0^j \oplus K_{8n+8+j}$, $j = \overline{0...7}$.

2) subblocks X_0^0, X_0^1, ..., X_0^7 multiplied and summed with the round keys $K_{8(i-1)}$, $K_{8(i-1)+1}$, ..., $K_{8(i-1)+7}$ and calculated 32-bit subblocks T^0, T^1, T^2, T^3. This step can be represented as follows:

$$\begin{aligned} T^0 &= (X_{i-1}^0 \cdot K_{8(i-1)}) \oplus (X_{i-1}^4 + K_{8(i-1)+4}), \\ T^1 &= (X_{i-1}^1 + K_{8(i-1)+1}) \oplus (X_{i-1}^5 \cdot K_{8(i-1)+5}), \\ T^2 &= (X_{i-1}^2 \cdot K_{8(i-1)+2}) \oplus (X_{i-1}^6 + K_{8(i-1)+6}), \\ T^3 &= (X_{i-1}^3 + K_{8(i-1)+3}) \oplus (X_{i-1}^7 \cdot K_{8(i-1)+7}), i = 1. \end{aligned}$$

3) to subblocks T^0, T^1, T^2, T^3 applying the round function and get the 32-bit subblocks Y^0, Y^1, Y^2, Y^3.

4) subblocks Y^0, Y^1, Y^2, Y^3 are summed to XOR with subblocks X_{i-1}^0, X_{i-1}^1, X_{i-1}^2, X_{i-1}^3, i.. $X_{i-1}^0 = X_{i-1}^0 \oplus Y^3$, $X_{i-1}^1 = X_{i-1}^1 \oplus Y^2$, $X_{i-1}^2 = X_{i-1}^2 \oplus Y^1$, $X_{i-1}^3 = X_{i-1}^3 \oplus Y^0$, $X_{i-1}^4 = X_{i-1}^4 \oplus Y^3$, $X_{i-1}^5 = X_{i-1}^5 \oplus Y^2$, $X_{i-1}^6 = X_{i-1}^6 \oplus Y^1$, $X_{i-1}^7 = X_{i-1}^7 \oplus Y^0$, $i = 1$.

5) at the end of the round subblocks swapped, i.., $X_i^0 = X_{i-1}^0$, $X_i^7 = X_{i-1}^7$, $X_i^1 = X_{i-1}^6$, $X_i^2 = X_{i-1}^5$, $X_i^3 = X_{i-1}^4$, $X_i^4 = X_{i-1}^3$, $X_i^5 = X_{i-1}^2$, $X_i^6 = X_{i-1}^1$, $i = 1$.

6) repeating steps 2-5 n times, i.e., $i = \overline{2...n}$ obtain 32-bit subblocks X_n^0, X_n^1, ..., X_n^7.

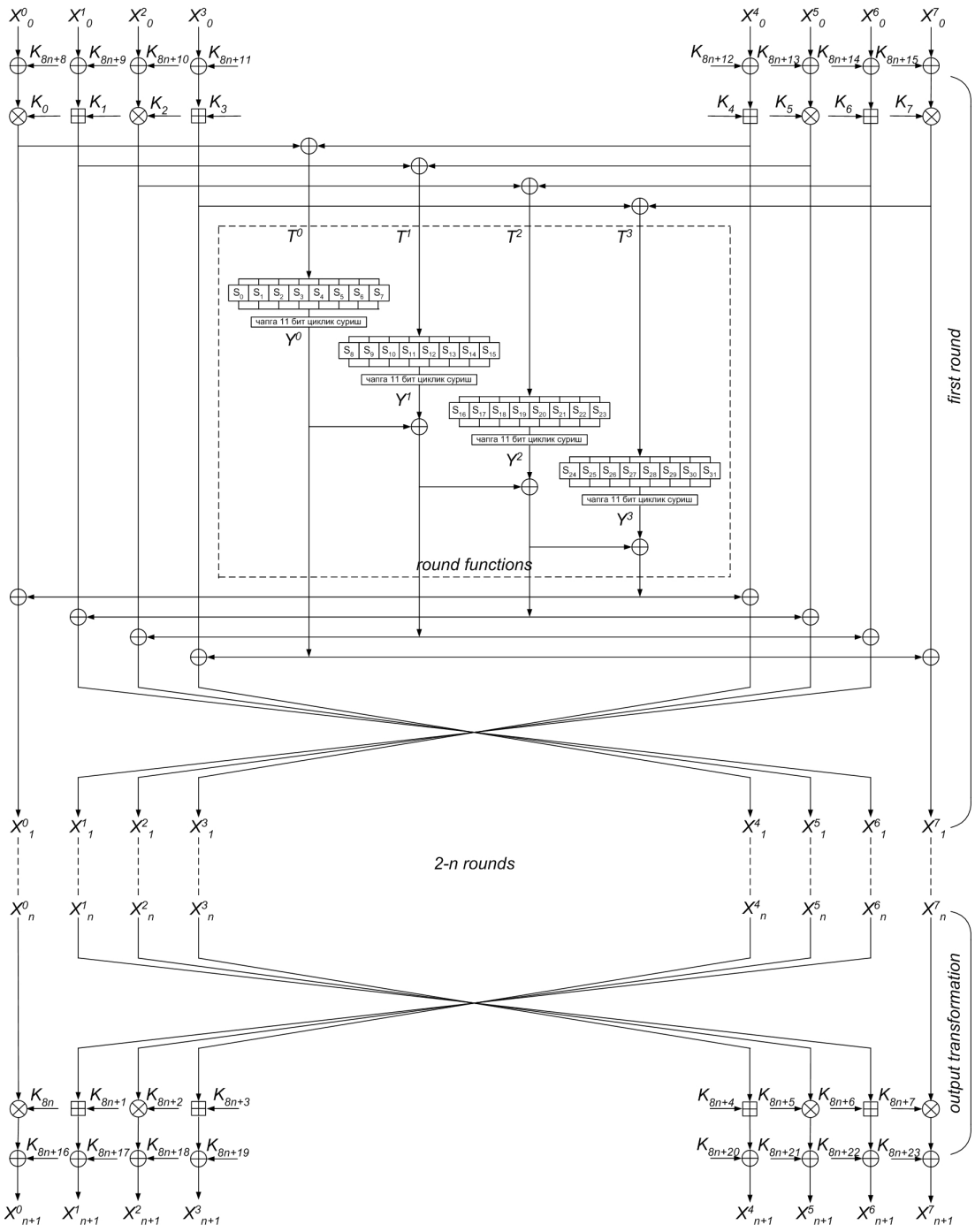


Figure 2: The scheme of encryption algorithm GOST28147-89-RFWKIDEA8-4

7) in output transformation round keys K_{8n}, K_{8n+1}, ..., K_{8n+7} are multiplied and summed into subblocks, i.e. $X^0_{n+1} = X^j_n \cdot K_{8n}$, $X^1_{n+1} = X^6_n + K_{8n+1}$, $X^2_{n+1} = X^5_n \cdot K_{8n+2}$, $X^3_{n+1} = X^4_n + K_{8n+3}$, $X^4_{n+1} = X^3_n + K_{8n+4}$, $X^5_{n+1} = X^2_n \cdot K_{8n+5}$, $X^6_{n+1} = X^1_n + K_{8n+6}$, $X^7_{n+1} = X^7_n \cdot K_{8n+7}$,

8) subblocks X^0_{n+1}, X^1_{n+1}, ..., X^7_{n+1} are summed to XOR with the round key K_{8n+16}, K_{8n+17}, ..., K_{8n+23}: $X^j_{n+1} = X^j_{n+1} \oplus K_{8n+16+j}$, $j = \overline{0...7}$. As ciphertext plaintext X receives the combined 32-bit subblocks $X^0_{n+1}||X^1_{n+1}||...||X^7_{n+1}$.

In the encryption algorithm GOST28147-89-RFWKIDEA8-4 when encryption and decryption using the same algorithm, only when decryption calculates the inverse of round keys depending on operations and are applied in reverse order. One important goal of encryption is key generation.

3.2 Key Generation of the Encryption Algorithm GOST28147-89-RFWKIDEA8-4

In n-round encryption algorithm GOST28147-89-RFWKIDEA8-4 used in each round 8 round keys of 32 bits and the output transformation of 8 round keys of 32 bits. In addition, prior to the first round and after the output transformation is applied 8 round keys on 32 bits. The total number of 32-bit round keys is equal to 8n+24.

The key encryption algorithm K of length l ($256 \le l \le 1024$) bits is divided into 32-bit round keys K^c_0, K^c_1,..., $K^c_{Lenght-1}$, $Lenght = l/32$, here $K = \{k_0, k_1, ..., k_{l-1}\}$, $K^c_0 = \{k_0, k_1, ..., k_{31}\}$, $K^c_1 = \{k_{32}, k_{33}, ..., k_{63}\}$,..., $K^c_{Lenght-1} = \{k_{l-32}, k_{l-31}, ..., k_{l-1}\}$ and $K = K^c_0||\, K^c_1||...||K^c_{Lenght-1}$. Then we calculate $K_L = K^c_0 \oplus K^c_1 \oplus ... \oplus K^c_{Lenght-1}$. If $K_L = 0$ then K_L is chosen as 0xC5C31537, i.e. $K_L =$ 0xC5C31537. Round keys K^c_i, $i = \overline{Lenght...8n+23}$ calculated as follows: $K^c_i = SBox0(K^c_{i-Lenght}) \oplus SBox1(RotWord32(K^c_{i-Lenght+1})) \oplus K_L$. After each round key generation the value K_L is cyclic shift to the left by 1 bit.

Decryption round keys are computed on the basis of encryption round keys and decryption round keys of the first round associate with of encryption round keys as follows:

$$\begin{aligned} & (K^d_0, K^d_1, K^d_2, K^d_3, K^d_4, K^d_5, K^d_6, K^d_7) \\ = \; & ((K^c_{8n)})^{-1}, -K^c_{8n+1}, (K^c_{8n+2)})^{-1}, -K^c_{8n+3}, -K^c_{8n+4}, (K^c_{8n+5)})^{-1}, -K^c_{8n+6}, (K^c_{8n+7)})^{-1}). \end{aligned}$$

Decryption round keys of the second, third and n-round associates with the encryption round keys as follows:

$$\begin{aligned} & (K^d_{8(i-1)}, K^d_{8(i-1)+1}, K^d_{8(i-1)+2}, K^d_{8(i-1)+3}, K^d_{8(i-1)+4}, K^d_{8(i-1)+5}, K^d_{8(i-1)+6}, K^d_{8(i-1)+7}) \\ = \; & ((K^c_{8(n-i+1)})^{-1}, -K^c_{8(n-i+1)+6}, (K^c_{8(n-i+1)+5})^{-1}, -K^c_{8(n-i+1)+4}, -K^c_{8(n-i+1)+3}, \\ & (K^c_{8(n-i+1)+2})^{-1}, -K^c_{8(n-i+1)+1}, (K^c_{8(n-i+1)+7})^{-1}), i = \overline{2...n}. \end{aligned}$$

Decryption keys output transformation associated with the encryption keys as follows:

$$\begin{aligned} & (K^d_{8n}, K^d_{8n+1}, K^d_{8n+2}, K^d_{8n+3}, K^d_{8n+4}, K^d_{8n+5}, K^d_{8n+6}, K^d_{8n+7}) \\ = \; & ((K^c_0)^{-1}, -K^c_1, (K^c_2)^{-1}, -K^c_3, -K^c_4, (K^c_5)^{-1}, -K^c_6, (K^c_7)^{-1}). \end{aligned}$$

Decryption round keys applied to the first round and after the output transformation associated with the encryption round keys as follows: $K^d_{8n+8+j} = K^c_{8n+16+j}$, $K^d_{8n+16+j} = K^c_{8n+8+j}$, $j = \overline{0...7}$.

4 Results

As a result of this study built a new block encryption algorithms called GOST28147-89-IDEA8-4 and GOST28147-89-RFWKIDEA8-4. This algorithm is based on a networks IDEA16-2 and RFWKIDEA16-2 using the round function of GOST 28147-89. Length of block encryption algorithm is 256 bits, the number of rounds and key lengths is variable. Wherein the user depending on the degree of secrecy of the information and speed of encryption can select the number of rounds and key length.

It is known, that the S-box encryption algorithm GOST 28147-89 are secret and used as a long-term key. following Table 2 summarizes options openly declared S-box such as: deg -degree of algebraic nonlinearity; NL -nonlinearity; λ -resistance to linear cryptanalysis; δ-resistance to differential cryptanalysis; SAC-strict avalanche criterion; BIC-bit independence criterion.To S-Box was resistant to cryptanalysis it is necessary that the values deg and NL were large, and the values λ, δ, SAC and BIC small. In block cipher algorithms GOST28147-89-IDEA8-4 and GOST28147-89-RFWKIDEA8-4 for all S-boxes, the following equation:deg = 3, $NL = 4$, $\lambda = 0.5$, $\delta = 3/8$, SAC=2, BIC=4, i.e. resistance is not lower than the algorithm GOST28147-89. These S-boxes are created based on Nyberg construction [4].

Table 2: Parameters of the S-boxes encryption algorithm GOST 28147-89

№	Parameters	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
1	deg	2	3	3	2	3	3	2	2
2	NL	4	2	2	2	2	2	2	2
3	λ	0.5	3/4	3/4	3/4	3/4	3/4	3/4	3/4
4	δ	3/8	3/8	3/8	3/8	1/4	3/8	0.5	0.5
5	SAC	2	2	2	4	2	4	2	2
6	BIC	4	2	4	4	4	4	2	4

To the encryption algorithm applied linear cryptanalysis. Attack on 4-round GOST28147-89-IDEA8-4 has a data complexity of 2^{83} chosen plaintexts and on 4-round GOST28147-89-RFWKIDEA8-4 has a data complexity of 2^{75} chosen plaintexts.

5 Conclusions

In this way, built a new block encryption algorithms called GOST28147-89-IDEA8-4 and GOST28147-89-RFWKIDEA8-4 based on networks IDEA8-4 and RFWKIDEA8-4 using the round function of GOST 28147-89. Installed that the resistance offered by the author block cipher algorithm not lower than the resistance of the algorithm GOST 28147-89.

References

[1] M. Aripov and G. Tuychiev, "The network IDEA4-2, consists from two round functions," *Infocommunications: Networks-Technologies-Solutions*, vol. 24, no. 4, pp. 55–59, 2012.

[2] M. Aripov and G. Tuychiev, "The network PES8-4, consists from four round functions," *Materials of the International Scientific Conference on Modern problems of applied mathematics and information technologies-Al-Khorezmiy 2012*, vol. 2, pp. 16–19, 2012.

[3] M. Aripov and G. Tuychiev, "The encryption algorithm AES-PES32-4 based on network PES32-4," *Materials of the International Scientific Conference on Modern Problems of Applied Mathematics and Information Technologies-Al-Khorezmiy 2016*, vol. 2, pp. 28–34, 2016.
[4] U. Bakhtiyorov and G. Tuychiev, "About generation resistance S-Box and boolean function on the basis of Nyberg construction," *Materials Scientific-Technical Conference on Applied Mathematics and Information Security*, pp. 317–324, 2014.
[5] J. Daeman and V. Rijmen, "Aes proposal: Rijndael," *NIST AES Proposal, http://csrc.nist.gov/*, 1998.
[6] X. Lai and J. Massey, *On the Design and Security of Block Cipher*, Diss. ETH Series in Information Processing, No. 9752, 1992. (`http://e-collection.library.ethz.ch/eserv/eth:38650/eth-38650-02.pdf`)
[7] GOST 2814789. National Standard of the USSR. Information Processing Systems. Cryptographic Protection. Algorithm Cryptographic Transformation.
[8] G. Tuychiev, "The network IDEA8-4, consists from four round functions," *Infocommunications: Networks-Technologies-Solutions*, vol. 26, no. 2, pp. 55–59, 2012.
[9] G. Tuychiev, "About networks PES8-2 and PES8-1, developed on the basis of network PES8-4," *Materials of the International Scientific Conference* .
[10] G. Tuychiev, "About networks RFWKPES8-4, RFWKPES8-2, RFWKPES8-1, developed on the basis of network PES8-4," *Materials of the International Scientific Conference.*
[11] G. Tuychiev, "The network PES4-2, consists from two round functions," *Uzbek Journal of the Problems of Informatics and Energetics*, vol. 5-6, pp. 107–111, 2013.
[12] G. Tuychiev, "The networks RFWKIDEA4-2, IDEA4-1 and RFWKIDEA4-1," *Acta of Turin Polytechnic University in Tashkent*, vol. 3, pp. 71–77, 2013.
[13] G. Tuychiev, "About networks IDEA16-4, IDEA16-2, IDEA16-1, created on the basis of network IDEA16-8," *Compilation of Theses and Reports Republican Seminar on Information Security in the Sphere Communication and Information. Problems and Their Solutions*, 2014.
[14] G. Tuychiev, "About networks IDEA32-8, IDEA32-4, IDEA32-2, IDEA32-1, created on the basis of network IDEA32-16," *Infocommunications: Networks-Technologies-Solutions*, vol. 30, no. 2, pp. 45–50, 2014.
[15] G. Tuychiev, "About networks IDEA8-2, IDEA8-1 and RFWKIDEA8-4, RFWKIDEA8-2, RFWKIDEA8-1 developed on the basis of network IDEA8-4," *Uzbek Mathematical Journal*, vol. 3, pp. 104–118, 2014.
[16] G. Tuychiev, "About networks PES32-8, PES32-4, PES32-2 and PES32-1, created on the basis of network PES32-16," *Ukrainian Scientific Journal of Information Security*, vol. 20, pp. 164–168, 2014.
[17] G. Tuychiev, "About networks RFWKIDEA16-8, RFWKIDEA16-4, RFWKIDEA16-2, RFWKIDEA16-1, created on the basis network IDEA16-8," *Ukrainian Scientific Journal of Information Security*, vol. 20, pp. 259–263, 2014.
[18] G. Tuychiev, "About networks RFWKPES32-8, RFWKPES32-4, RFWKPES32-2 and RFWKPES32-1, created on the basis of network PES32-16," *Compilation of Theses and Reports Republican Seminar Information Security in the Sphere Communication and Information. Problems and their Solutions*, 2014.
[19] G. Tuychiev, "Creating a data encryption algorithm based on network IDEA4-2, with the use the round function of the encryption algorithm GOST 28147-89," *Infocommunications: Networks-Technologies-Solutions*, vol. 32, no. 4, pp. 49–54, 2014.
[20] G. Tuychiev, "New encryption algorithm based on network RFWKPES8-1 using of the transformations of the encryption algorithm AES," *International Journal of Computer Networks and Communications Security*, vol. 3, no. 6, pp. 31–34, 2014.

[21] G. Tuychiev, "About networks PES16-4, PES16-2 and PES16-1, created on the basis network PES16-8," *Ukrainian Information Security Research Journal*, vol. 17, no. 1, pp. 53–60, 2015.
[22] G. Tuychiev, "About networks PES4-1 and RFWKPES4-2, RFWKPES4-1 developed on the basis of network PES4-2," *Uzbek Journal of the Problems of Informatics and Energetics*, vol. 1-2, pp. 100–105, 2015.
[23] G. Tuychiev, "About networks RFWKPES16-8, RFWKPES16-4, RFWKPES16-2 and RFWKPES16-1, created on the basis network PES16-8," *Ukrainian Information Security Research Journal*, vol. 17, no. 4, pp. 163–169, 2015.
[24] G. Tuychiev, "Creating a block encryption algorithm based network IDEA32-1 using transformation of the encryption algorithm AES," *Acta NUUz*, vol. 2/1, pp. 136–142, 2015.
[25] G. Tuychiev, "Creating a block encryption algorithm based networks PES32-1 and RFWKPES32-1 using transformation of the encryption algorithm AES," *Compilation Scientific Work Scientific and Practical Conference on Current Issues of Cyber Security and Information Security-CICSIS-2015*, pp. 101–112, 2015.
[26] G. Tuychiev, "Creating a block encryption algorithm on the basis of networks IDEA32-4 and RFWKIDEA32-4 using transformation of the encryption algorithm AES," *Ukrainian Scientific Journal of Information Security*, vol. 21, pp. 148–158, 2015.
[27] G. Tuychiev, "Creating a encryption algorithm based on network PES4-2 with the use the round function of the GOST 28147-89," *TUIT Bulleten*, vol. 34, no. 4, pp. 132–136, 2015.
[28] G. Tuychiev, "Creating a encryption algorithm based on network RFWKIDEA4-2 with the use the round function of the GOST 28147-89," *International Conference on Emerging Trends in Technology, Science and Upcoming Research in Computer Science (ICDAVIM'15), //printed in International Journal of Advanced Technology in Engineering and Science*, 2015.
[29] G. Tuychiev, "Creating a encryption algorithm based on network RFWKPES4-2 with the use the round function of the GOST28147-89," *International Journal of Multidisciplinary in Cryptology and Information Security*, vol. 4, no. 2, pp. 14–17, 2015.
[30] G. Tuychiev, "Development block encryption algorithm based networks IDEA16-2 and RFWKIDEA16-2 using the transformation of encryption algorithm AES," *Information Security in the Light of the Strategy Kazakhstan-2050: Proceedings III International Scientific-Practical Conference (15-16 Oct. 2015, Astana)*, pp. 40–60, 2015.
[31] G. Tuychiev, "The encryption algorithm AES-RFWKIDEA16-1," *Infocommunications: Networks-Technologies-Solutions*, vol. 34, no. 2, pp. 48–54, 2015.
[32] G. Tuychiev, "The encryption algorithm AES-RFWKIDEA32-1 based on network RFWKIDEA32-1," *Global Journal of Computer Science and Technology: E Network, Web, Security*, vol. 15, pp. 33–41, 2015.
[33] G. Tuychiev, "The encryption algorithms AES-PES16-1 and AES-RFWKPES16-1 based on networks PES16-1 and RFWKPES16-1," *International Journal of Electronics and Information Engineering*, vol. 3, no. 2, pp. 53–66, 2015.
[34] G. Tuychiev, "The encryption algorithms GOST28147-89-PES8-4 and GOST28147-89-RFWKPES8-4," *Information Security in the Light of the Strategy Kazakhstan-2050: Proceedings III International Scientific-Practical Conference*, pp. 355–371, 2015.
[35] G. Tuychiev, "New encryption algorithm based on network IDEA16-1 using of the transformation of the encryption algorithm AES," *IPASJ International Journal of Information Technology*, vol. 3, pp. 6–12, 2015.
[36] G. Tuychiev, "New encryption algorithm based on network IDEA8-1 using of the transformation of the encryption algorithm AES," *IPASJ International Journal of Computer Science*, vol. 3, pp. 43–47, 2015.
[37] G. Tuychiev, "New encryption algorithm based on network PES8-1 using of the transformations of the encryption algorithm AES," *International Journal of Computer Networks and Communications Security*, vol. 3, no. 2, pp. 1–5, 2015.

[38] G. Tuychiev, "New encryption algorithm based on network RFWKIDEA8-1 using transformation of AES encryption algorithm," *International Journal of Multidisciplinary in Cryptology and Information Security*, vol. 4, no. 1, pp. 1–5, 2015.
[39] G. Tuychiev, "To the networks RFWKIDEA32-16, RFWKIDEA32-8, RFWKIDEA32-4, RFWKIDEA32-2 and RFWKIDEA32-1, based on the network IDEA32-16," *International Journal on Cryptography and Information Security (IJCIS)*, vol. 5, no. 1, pp. 9–20, 2015.
[40] G. Tuychiev, "The encryption algorithm AES-RFWKPES32-4," *Proceedings International Round Table On the National and Information Security in the Republic of Kazakhstan. The Experience of Foreign Countries*, 2016.

Biography

Tuychiev Gulom is a candidate technical Sciences (Ph.D.), National University of Uzbekistan. He received Pd.D. degree in specialty mathematic from the National University of Uzbekistan. He had published more than 50 scientific articles. He current research interests include block ciphers, Boolean functions.

An Anti-Phishing kit Scheme for Secure Web Transactions

Abdul Abiodun Orunsolu[1], A. S. Sodiya[2], A. T. Akinwale[2], B. I. Olajuwon[3]
(Corresponding author: A. A. Orunsolu)
Department of Computer Science, Moshood Abiola Polytechnic, Abeokuta, Nigeria[1]
(Email: orunsolu.abdul@mapoly.edu.ng)
Department of Computer Science, Federal University of Agriculture, Abeokuta, Nigeria[2]
Department of Mathematical Science, Federal University of Agriculture, Abeokuta, Nigeria[3]
(Received Feb. 11, 2017; revised and accepted Mar. 21, 2017)

Abstract

In this work, an anti-phishing approach was proposed against phishing pages generated by phishing kits. The architecture consists of a Sorter Module (SM) and Signature Detection Module (SDM). The SM is used to separate pages with login attributes and obfuscated scripts from other pages within the system. These sorted pages are fed into the SDM, where the signature of the suspicious page is generated. In SDM, a two-tier classifier is employed to generate phishing label based on signature analysis. Experimental results of the approach indicated a detection accuracy of 100% on specific phishing kit-generated sites and 98% on general phishing/legitimate data. To determine the detection time of the approach, latency analysis of the system was performed. The results indicated a latency 0.3s and standard deviation of 0.367s for the various operations performed by the system during detection. Thus, the approach effectively detects phishing pages by using 'fingerprints' from phishing kits.

Keywords: Electronic Commerce Security; Malicious Sites; Online Scams; Spoofed Pages; Phishing Kits

1 Introduction

Today, there is an enormous growth of digital applications in both private and public domains. Business services and relationships are redefined as people's reliance on the internet technology continues to grow at an unprecedented rate. The proliferation of Internet has created a lot of opportunities in terms of automatic availability of services, global coverage, efficiency, reliability, and zero-delay in service delivery. Despite these noble contributions of the internet technology, the security issues of the online communication have become key concern to the stakeholders.

In the recent times, hackers have continuously managed a host of online black markets which threaten stakeholders' confidence in the usability of internet technology [5, 8]. This range of criminal enterprises includes spam-advertised commerce, botnet attacks, and a vector for propagating malware [9]. The incidence of internet black market climaxed with the advent of phishing, in which both the service providers and online business operators have suffered consequences such as damaged reputation and huge financial losses. Also, unwary users share their own negative experiences at these malicious sites [3].

Despite the existence of various anti-phishing measures, the frequency of phishing incidences continues to increase [1]. For instance, RSA's online fraud report showed estimated losses of over $5.9 billion by global organizations in nearly 450,000 attacks in 2013 [17]. Figure 1 presents the ominous illustration of unique phishing sites detected from January to September 2015 [17]. Generally, cyber criminals use phishing predominantly as a technique for obtaining identity related information employing both the social engineering and technical subterfuge [18].

In the past, setting up a fake website with similar feel like the benign site could be achieved by copying HTML document of a website and modifying them. However, the rise in the number of phishing sites may be unconnected with HTML approach as phishers now prefer exploit kits/phishing kits to the former [17]. These exploit kits simplify the creation of fraudulent websites by stealing the source code of legitimate web pages [7]. According to RSA online report in 2014, abundant tools and offerings have flooded the underground markets, thereby making the lives of phishers and would-be-phishers easier. While the creation of these malicious sites continues to give phishers opportunity to create fraudulent services, majority of internet users remain ignorant, unconscious, or negligent of the adverse effects of phishing [1]. Although a vast number of literatures identified the unpopular consequences of phishing toolkits, there is paucity of literature that concentrates on how to counter phishing from exploit kits perspective.

Figure 1: Phishing volume from JAN-SEPT 2016 [17]

To this end, we report an anti-phishing technique based on the features of phishing kits as reported in some extant literature [13, 15]. The rationale behind the design is simple. The proliferation of phishing campaign has been attributed to the availability of phishing kit to hackers who employ the tool with minimal efforts [8]. The proposed technique creates an arm race between the hackers who employ the tool to create phishing pages and the approach which disrupts the efficacy of the tool by flagging such page as fake. In addition, the approach will reduce the degree of trust associated with phishing toolkits since there is now a defense model that targets their architecture. This will make hackers spend time checking the efficiency of kit before their deployment. Moreover, on the part of phishing kits authors, our research presents an attacking model for which they must consider defense when deploying their kits. We affirm that this is an advantage since attack is the best form of defense.

The rest of the paper is organized as follows: Section 2 presents related works, where our approach is compared with other existing anti-phishing techniques. The overall architecture and design details of our

proposed methodology are discussed in Section 3. In Section 4, the implementation and the evaluation of the proposed method are presented. Conclusions and future works are presented in Section 5.

2 Related Works

Anti-phishing research has attracted a lot of interests from security experts from both academics and IT industries. Researchers have developed a plethora of countermeasures such as list-based approaches (i.e. Whitelist and Blacklist), heuristics approaches, hybrid approaches or multifaceted mechanisms [6]. Table 1 presents a summary of related works in comparison with our proposed scheme with respect to the following parameters:

1) Kit defamation: ability of the system to detect phishing sites using toolkit-related "fingerprints", thereby discouraging its motivation to would-be-phishers.

2) Drop email discovery: ability of the system to detect obfuscated e-mail addresses within a phishing page.

3) Client independent: ability of presenting the anti-phishing components as a non-browser hand-on or plugin.

4) Search engine independence: ability of the system to make detection without any input from search engine

Table 1: Comparison of related works with our approach

Works	Kit Defamation	Drop e-mail Discovery	Client Independence	Search Engine Independence
Aparna et al.	No	No	Yes	Yes
Shahriar et al.	No	No	No	No
Han et al.	No	No	Yes	No
Olivo et al.	No	No	No	Yes
Gowtham et al.	No	No	Yes	No
Our work	Yes	Yes	Yes	Yes

A number of studies have examined the reasons that people fall for phishing attacks. For instance, Dhamija et al. [12] identified lack of computer system knowledge, lack of knowledge of security and security indicators, visual deception and bounded attention. The authors further showed that a large number of people cannot differentiate between legitimate and phishing web sites, even when they are made aware that their ability to identify phishing attacks are being tested. In a similar vein, Sheng et al. [18] investigated the demographic analysis of phishing susceptibility. Their works showed that women were more susceptible to phishing than men [4].

In a more recent study, Mohammed et al. [2] conducted user study with the use of eye tracker to obtain objective quantitative data on user judgment of phishing sites. Their results indicated that users detected 53% of phishing sites even when primed to identify them with little attention on security indicators.

A new method based on hybrid approach which depends on profiling phishing attacks has offered significant progress in the quest against phishing. Islam et al. [9] investigated a three-tier classification

approach to detect phishing emails where accuracy of detection is up to 97%. However, this technique suffered from lengthy training time and complex analysis. In a similar vein, Gowtham et al. [6] studied the characteristics of legitimate and phishing pages by proposing heuristics, which are characteristics that are found to exist in phishing attacks. The authors extracted 15 features as heuristics for evaluating the phishiness of a webpage. However, this approach suffered from client-side exploits such as Java exploit attacks and high categorization time.

Cova et al. [13] presented an interesting research on analysis of phishing kits. The authors demonstrated that there is no such thing as a free phishing kit in the underground economy. This was based on the analysis of a large number of kits. The authors found that the kit authors developed backdoors in their kits using obfuscated code. These backdoors were used to send a copy of information collected by the inexperienced kit users to third parties. In the same vein, McCalley et al. [15] analyzed a "backdoored" phishing kits distributed by the infamous Mr-Brain hacking group. The authors showed a number of obfuscated codes used by the kit creator that allowed a third party to access the credentials of internet victims. In our work, we used the analysis of these works to present a defence framework against phishing kits.

Han et al. [18] developed an Automated Individual White-List which described trusted Login User Interface where the users submit his credentials. Every LUI would case a warning except if trusted. Once a LUI was trusted, its features would be stored locally in a whitelist. The approach can effectively defend dynamic pharming attacks as well as protecting web digital identities. However, the approach is limited by new login problem, in which users are warned when submitting credentials on a website for the first time.

Antonio et al. [11] presented an anti-phishing approach in which multi-factor authentication model was used to protect users' identities as well as increasing customers' trusts in e-commerce.

Our approach offers the following contributions to anti-phishing research:

1) We propose an interesting angle to detect phishing websites based on the analysis of phishing kits and some other generalized features of a typical phishing attacks.

2) We evaluate the proposed approach using phishing pages created by a numbers of individuals.

3) We evaluate the system for both accuracy and performance.

3 The Proposed Scheme

The research reported in this paper is focused on phishing detection using the analysis of phishing kits. Phishing kits are usually distributed as .zip, .gz, .tar, or .rar archives that contain two types of files. The first file displays a copy of the targeted web site and the second file contains the scripts used to save the phished information and send it to phishers. Phishing kits are distributed on websites with instructions for the phishers to insert their email address in order to receive the stolen information. One of the distinguishing features of kit-weaved sites is drop email address usually passed to the PHP "action" file. In certain cases, the drop email address is obfuscated into a hexadecimal code [15, 13]. In general, majority of phishing kits contains:

1) Resources to replicate target site using HTML pages, JavaScript and CSS files, images and other media files [12, 17].

2) Automated email address field to obtain and validate victim's information e.g. PHP getmxrr ( ) function [17].

3) Obfuscation techniques for hiding drops and backdoor within the phishing kit.

From the foregoing, we present the overall system design of the anti-phishing kit architecture. In order to design an effective architecture, a two-stage anti-phishing service is created into a single workflow. These services consist of: (1) a Sorter Module (SM) and (2) a Signature Detection Module (SDM). The SM retrieves the suspicious loading sites and checks for presence of login form. This is because; the primary motivation of phishing attacks is to fraudulently obtained users' credentials using spoofed sites. The SM performs the second check for obfuscation detection. Obfuscation is a technique used to hide attacks from static detection by causing the appearance of malicious string to change. In this way, the code evades detection tools. If no login form or obfuscated code is observed in the source code by SM, the loading page is allowed to continue its session. However, if these traits are observed on the site, the SM transferred the page to the SDM. In SDM, the suspected page is processed further. Here, the features of the pages are extracted and their signatures are obtained. In this work, 18 heuristic features are considered.

From the viewpoint of phishing kits, heuristics such as drops, hexadecimal scripts etc. are referred to as Third party heuristics. To improve the performance of signatures on general phishing datasets, the URL characteristics and keyword identity are considered. Then, all these signatures are used by hybrid classifier to generate appropriate label for the loading sites. The hybrid classifier is optimized using machine learning algorithms consisting of Na?ve Bayes and Support Vector Machine [16].

4 Components Of Anti-Phishing Kit System

The proposed Anti-Phishing kit (APK) methodology consists of the Sorter Module (SM) and Signature Detection Module (SDM) as depicted in Figure 2. These two components are used to analyze each webpage visited and determine the status of the page with phishing kit. To detect kit generated sites, the APK loads the webpage, w, a user wants to visit, and scan through the source file of w for various inherent phishing kit signatures and the other relevant anti-phishing signatures.

4.1 The Sorter Module

For a given webpage w, the SM component of APK first identifies the presence of login fields. This is to reduce superfluous computation on webpages without login fields as most phishing websites are meant to steal user credentials. Generally, the presence of login form on a page is usually characterized by presence of Form tags, Input fields, and login keywords (e.g. password, PIN, ID, username, social security number, account number etc.). The Input fields usually hold user input and login attributes which distinguish a login form from other types of forms.

The SM employs Latent Dirichlet Allocation (LDA) to manage the three login form properties. This is due to the sensitivity of LDA to changes in keywords usage which make it good for handling synonyms. For instance, a phisher may replace the word 'password' with 'secret word or phrase' to circumvent login detection process. LDA posits that one way of sorting the content of w is to look at the set of words it uses. Because words carry very strong semantic information, webpages that contain similar content i.e. set of login words, will most likely use a similar set of words. As such, mining an entire corpus of webpages can expose sets of words that frequently co-occur within webpages. These sets of word may be interpreted as topics. The modeling process of LDA can be described as finding a mixture of topics z for each website w i.e. $P(z|w)$, with each topic described by terms t following another probability distribution, i.e., $P(t|z)$. This formalization is given in Equation (1).

$$P(t_i|w) = \sum_{j=1}^{Z} P(t_i|z_i = j)P((z_i = j)|w). \quad (1)$$

Where $P(t_i|w)$ is the probability of the i^{th} term for a given w and z_i is the latent topic. $P(t_i|z_i = j)$ is the probability of t_i within topic j. $P(z_i = j|w)$ is the probability of picking a term from topic j in the webpage. The number of latent topics, in the case of login attributes, has to be defined in advance. In this way, LDA estimates the topic-term distribution $P(t|z)$ and the webpage-topic distribution $P(z|w)$ from an unlabeled corpus of documents using Dirichlet priors for the distributions and a fixed number of topics [14].

The functionality of SM extends to detecting presence of obfuscated code in the loading page. The work of Cova et al. and McCalley showed that most phishing kits employed obfuscation to hide their malicious activities. In the light of this, we adopt the work of Xu et al. [10] to perform the detection of obfuscation by SM subsystem due to its lightweight attribute.

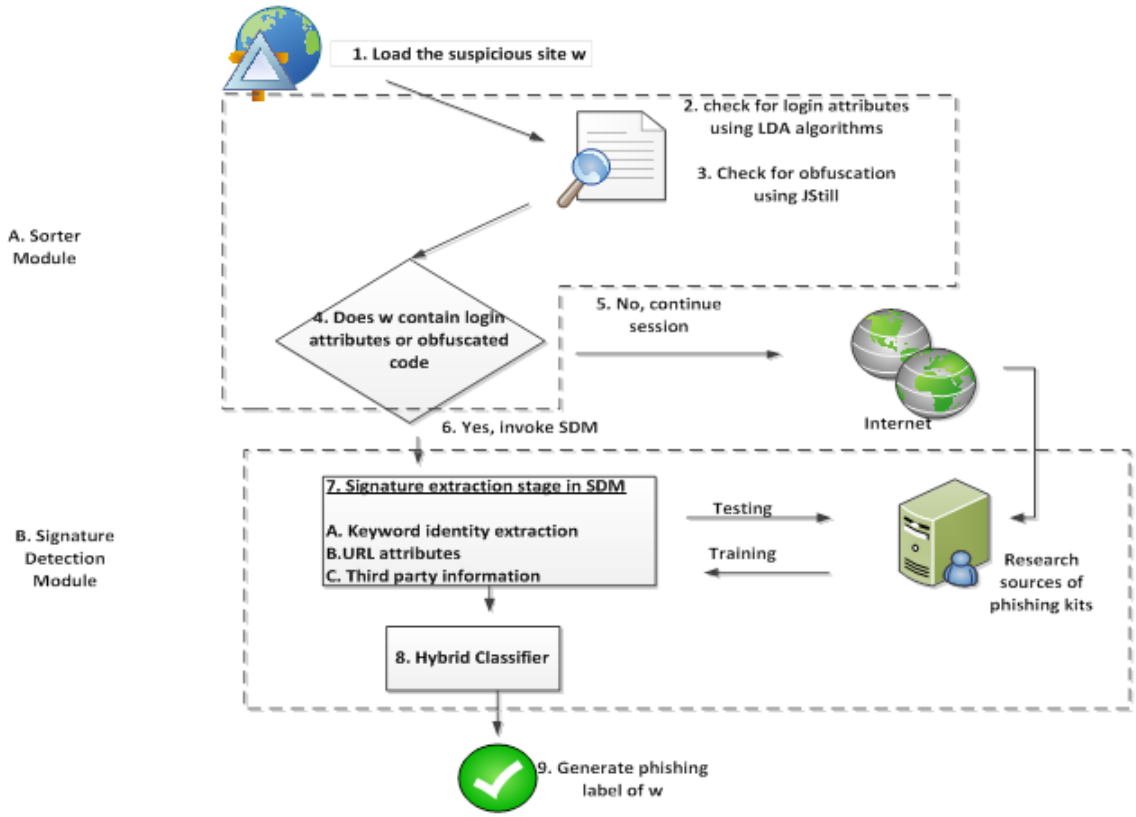


Figure 2: Anti-phishing kit defense architecture

4.2 The Signature Detection Module

To determine whether a sorted page is toolkit generated or not, the Signature Detection Module is invoked. The main heuristics used by SDM in classifying a website are the signatures extracted from

the analysis of phishing kit code (called Third party heuristics), URL characteristics and keyword identity.

Third party heuristics are extracted features known of a typical phishing kit. For instance, drops and hexadecimal code are used to hide detection of kits features (Figure 3). These scripts are inserted by phishers to obtain undue advantage from the users of the kit. In addition, these features retain some information about the kit used. The following section provides discussion on these heuristics.

1) Drop attributes/Bad forms: Phishing kits creators usually include backdoors in their scripts. These backdoors consist of hidden drops, which covertly send victims' information to addresses different from that of kit users. Therefore, the presence of hidden drops is an indication that a page is kit-related.

2) Hexadecimal character codes/Address obfuscation: One distinguishing feature of phishing kits is to obfuscate email addresses within a message or website. Hexadecimal codes are used to obfuscate the planted backdoors in the kits (Cova, et al., 2008). If a website contains obfuscated address which can be recovered with instrumentation or set of command, then the page is phishing kit-related.

3) Name of toolkit: Some toolkits usually add the name of the tool used to fake a site in the source code of the website as comment e.g. ¡!– created with –¿. The name can be found at the top of the HTML document or the copyright section at the footer of the webpage as comment which is not visible to users. Therefore, if a page contain a name not related to its content, then it is kit-based.

4) Blank Redirection page: Some phishing toolkits create a blank page as the first page which will redirects to the fake page. The blank page contains some information belonging to the kit used in the source code of the page. If a site contains a blank redirection page with information unrelated to its content, then it is kit-based.

5) URL of Toolkit: Apart from the toolkit name that can be found in faked sites, the URL of the toolkit used can also be found as a comment. This is to link new users to the toolkit in order for them to download it. The URL is usually commented out so that it will not be visible to web users. If a page contains out-of- comment URL not related to the page, then it is kit-based.

Keyword extraction heuristics are particularly important in identifying target organization of a phishing kit. The keyword identity set of a webpage is usually extracted by considering Document Object Model tree. The keyword extraction in SDM adopts the Term Frequency-Inverse Document Frequency (TF-IDF). The TF-IDF is a numerical statistic which reflects how important a feature is to a document in a corpus. It is often used as a weighting factor in information retrieval and text mining. The TF-IDF value increases proportionally to the number of times a feature appears in the document, but is offset by the frequency of the feature in the corpus.

It is interesting to note that the base of the log function does not matter and constitutes a constant multiplicative factor toward the overall result. Thus, a term t has a high TF-IDF weight by having a high term frequency in a given document D (i.e. a feature is common in a document) and a low document frequency in the whole collection of documents (i.e. is relatively uncommon in other documents).

The Keyword extraction heuristics used in our proposed system are discussed as follows:

1) Domain name credibility: The domain name credibility feature determines the genuineness of the target organization by phishing kit creators using Google's PageRank system. If kit contains file for one or more target organization, then the rank of the hosting domain is compared with a threshold value (usually 5 on 0 to 10 scale) indicating the legitimacy of the site.

2) Domain name identity: Most of the website domain names have relationship to their contents. The keywords in this domain name are usually part of the base domain URL. If the keyword identity set of a page is not related to its contents, then it is phishing. Otherwise, it is legitimate and non-kit related.

3) Out-of-position brand name: Legitimate sites often put their brand name into their domain name. On the other hand, phishing sites are always hosted on compromised or newly registered domains. If the domain keywords are not related to its brand, then the page is suspicious.

4) Age of domain: This feature checks the age of the domain name. Many phishing pages claim the identity of known brand which has relatively long history. If the age of domain does not correspond to its WHOIS lookups, then it is likely to be deceptive.

The URL identity of a webpage is determined by analyzing the patterns from its hyperlinks structure. In a legitimate website most of the links points to its own domain or associated domain, but in phishing sites (including the kit-based phish sites) most of the links point to foreign domain to imitate the behavior of a legitimate page. For URL identity extraction, the SDM consider the "href" and "src' attributes of the anchor links, particularly ¡a¿, ¡area¿, ¡link¿, ¡img¿, and ¡script¿ tags from the DOM tree of a webpage. For each anchor, the SDM extracts the base domain portion from the URL, and then calculate the number of times each base domain appears. The base domain that has the highest frequency will be the URL identity of the webpage. This step is necessary in determining the behavior of the URL embedded in a suspicious webpage. In the end, the following features are considered from URL identity to generalize the detection accuracy of the proposed system:

1) URL of original site: Most phishing sites put the URL of the original site faked as comment at the top of the html page. This is to show where the website was copied from. If such feature exists on a page, then it is phishing and possibly kit-based.

2) Presence of user-info in the domain name: In this feature, the presence of @ or dash (-) is checked for within the URL. If such feature is found, then the page is a phish site.

3) IP address behavior (either irreversible or reversible): In this feature, the system checks whether the URL address of a website is a permanent IP address which does not have DNS entries. In most phishing site, the practice is usually an IP address-based URL because of its low cost. Therefore, if such feature exists, then the page is a phish site.

4) Number of dots in the URL: This feature counts the number of dots in the URL as most phishing pages tend to use more dots in their URLs. If this feature exists on a page, then it is a phish site.

5) Domain name in the path of the URL: This feature checks for the presence of dot separated domain or host name in the path part of the URL. If this feature exists in a page, then it is a phish site.

6) Presence of Foreign anchors: This feature examines of foreign anchors in a webpage. If a page contains too many foreign anchors, then it is likely to be deceptive.

7) Cookie domain: This feature checks the transmission of text data by a web server to a web client. This text data is called HTTP cookies which are used for maintaining information about client users. If a website has a domain cookie which is in a foreign domain, then it may be deceptive as most legitimate websites have their own domain cookies or no cookies.

8) Port Number behavior: This feature compares the port number part of a domain name with the stated protocol part of a URL. If the protocol does not match the port number, then the page is a phish site.

9) SSL protected: Secure login pages of benign sites often have an SSL certificate while most phishing sites do not. This feature examines the certificate of a webpage and whether it is issued by trusted certificate authority or not. If the page's claimed identity does not appear in the attached certificate, then the page is likely to be phishing and kit-related.

```
$hostname = gethostbyaddr($ip);
$message = "Chase Bank Spam ReZulT\n"; ...
$message .= "User ID : $user\n";
 $messege .= "hostip" $message .= "Full Name :
$fullname\n"; ...
 $message .= "City : $city\n";
 $messege .= "port";
 $message .= "State : $state\n"; ...
 $message .= "Mother Maiden Name : $mmn\n";
 $messege .= "@"; ...
mail($to,$subject,$message,$headers);
        mail($messege,$subject,$message,$headers);
```

Figure 3: Sample of drop email code (Cova et al. 2008)

The algorithm in Algorithm 1 presents the main structure of the proposed system.

Algorithm 1 Anti-Phishing Kit Defense Algorithm

1: Input: Web page W, Anti-phishing signatures S
2: Otput: Phishiness level of W
3: Begin
4: Load and parse page W
5: Generate the DOM from HTML of W
6: Check if $W \ni FORM$ input
7: Check if $W \ni ObfuscatedJavaScriptcode$
8: If (6) .OR. (7) $\in W$, invoke SM for preprocessing
9: else Exit
10: If (8) is .True. extract signature of $W \ni$ preprocessed DOM (W)
11: Extract URL heuristics, Keyword extraction heuristics and Third-party heuristics from W
12: Send the signature (W) to NB-SVM classifier trained with dataset S
13: Display the status of W
14: End

To detect the status of the analyzed webpage, the SDM uses a hybrid classifier consisting of Na?ve Bayes and Support Vector Machine based on the extracted features. The SDM uses Na?ve Bayes (NB) as vectorizer and Support Vector Machine as classifier. The main problem associated with using SVM as classifier is the effort needed to transform text data into numerical data which is sometimes termed as "vectorization". It is natural to use the NB as the vectorizer for classifier based on the vector space model, such as SVM, which typically requires preprocessing to vectorize the raw text documents into numerical values. In this way, NB is used as a pre-processor for selected features in the front end of the SVM to vectorize corpus before the actual training and classification are carried out. The main procedure of the proposed hybrid classifier is described in Algorithm 2. Algorithm 2: A hybrid classifier algorithm - create an ensemble of classifiers using NB and SVM.

Algorithm 2 Two-Tier Classifier of SDM

1: Input: 18-dimension feature vector space, Training data of labeled examples S consisting of the signature
2: Otput: Label (1: (Phished) PT related; 0: (Benign) non-PT related)
3: For all signature i extracted from W
4: Compute the conditional probability of signatures of analyzed phishing kit given the signature of W
5: Construct the probability as input into SVM
6: Find the optimal hyper plane for signature (W) and signature S
7: Classify W
8: End

5 Implementation And Evaluation

Our Anti-Phishing defense system is implemented using VB.NET Framework on a machine with Windows 7 OS. The machine runs on an Intel Core i5 processor with 4 GB RAM and 450 GB Hard drive. We trained the algorithm using a set of web pages consisting of toolkit-generated pages and genuine pages. A preliminary test showed that this implementation can accurately detect the absence of login form on most tested sites. It loads the webpage a user wants to visit, and scans through the source file using the attributes of its keyword extraction, URL behaviour and Third-party heuristics.

In order to evaluate the effectiveness of this system, we recruited the service of ethical hackers consisting of 100 students of a Computer Security class and 4 external research collaborators to create phishing pages using toolkits of their choice. In addition, phishing and legitimate data were obtained from openly available research database sources such as PhishTank, Millersmiles, Alexa ranking and Cybercrime Archive to evaluate the performance of the system on general phishing data corpus.

We conducted three experiments to evaluate the performance of the proposed system. The first experiment and the second experiment were used to evaluate the accuracy of the proposed approach in terms of True Positive, False Positive, False Negative and True Negative. The True Positive means the actual data and predicted categories are true. The True Negative means actual and predicted categories are negative. The False Positives means the predicted should have been negative instead classified as positive. The False Negatives means predicted should have positive instead classified as negative. Accuracy is a measure of how accurate the learned system makes prediction on unseen test instances.

In the first experiment, 258 kits-generated sites created by ethical hackers were subjected to the prediction of APK. Exactly 208 of these cloned pages were generated by students of Computer Security class during a Security Lab assignment in March 2015. The remaining 50 were created by ethical

hackers collaborated into this study in December 2014. Table 2 presents the number of sites created by each phishing toolkit used by the students during the cloning process. These kits perfectly faked the original sites with similar look and feel that can deceive even an experienced web user (Figure 4). One of the phishing toolkits that are used by the ethical hackers is the HTTRACK which is an easy to use offline browser utility. This toolkit enables a phisher to download a webpage from the internet to a local directory and thereafter build recursively all directories (e.g. html files, images, link structure and other files) from the server to the phisher's computer. It is important to state that these two groups of collaborators (i.e. the computer security students and ethical hackers) did not have the knowledge of our defense system.

Table 2: Distribution of kit-created pages

Name of Kit	Number of Webpages
Cyotek	20
A1	38
Fresh Web	30
HTTRACK	80
Webclone Maker	20
Web2Disk	20
Total	208

In general, these toolkits always obfuscate the links of the login-page. These toolkit-generated websites were built offline and were later put online. Then, the APK was used to download the source code of these toolkit-generated sites and checked for phishing signatures. In the end, the approach correctly labeled the 258 websites as phishing. APK was able to detect all the 258 websites due to the common weakness of redirection to the real web page these phishing toolkits mimicked.

Figure 4: PT-generated first merchants bank home page

In the second experiment, the performance of the system was tested on general phishing dataset corpus. A total number of 200 phishing pages and legitimate pages were compiled over the period of 4 months from September 2014 to December 2014. Specifically, our data consists of 100 phishing pages from PhishTank (2014), Millersmiles and CyberCrime Archive (http:// cybercrime-tracker.net). On the other hand, the 100 legitimate pages were obtained from Alexa Ranking Top List which contains well-known websites with high ranking (http://www. alexa.com). These legitimate sites are popular sites which are usually target by phishers because of their large numbers of subscribers. This is easy as all the registered users of a website may not all be security conscious and sometimes, changes to operational issues of such websites may not get to all these users on time. Figure 5 showed the experimental results of the system accuracy using confusion matrix for the 200 dataset.

From the results, the system accuracy is 98% with low false negatives of 0.04%. All the correctly detected phishing sites exhibited features captured in the approach while the incorrectly labeled sites were developed with advanced features not available in phishing toolkit technology. A closer examination of these undetected sites revealed that phishers built those sites from scratch to escape possible detection by anti-phishing techniques.

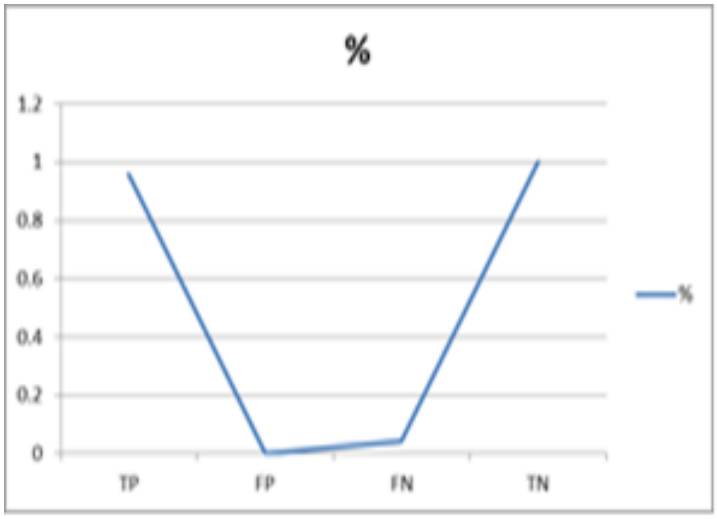


Figure 5: Experimental results on general data corpus

In the third experiment, the runtime analysis of the proposed system was evaluated to determine the usability issue of delay that users may experience during the detection process of APK. Using standardized timing procedure; we run experiments for different testing sites. In the process, we calculated the time between the initiations of a transaction (e.g. loading of suspicious page) to the time when the system completes the detection process. The average latency for the various operations in APK is presented in Figure 6. In our experimental analysis, we found that it took 0.3s for APK to download and check the login status of a website. The signature extraction and detection takes 0.4s and the system used 0.4s to check clean sites. The standard deviation of the system is 0.367s.

Figure 6: Average latency analysis of APK

6 Conclusions And Future Work

In this paper, the concept of phishing detection based on the analysis of phishing kit was presented and discussed. This is achieved through the Sorter Module and Signature Detection Module components of the APK architecture. The Sorter Module detects the presence of login fields and obfuscated code on a suspicious page to prevent superfluous computation in the Signature Detection Module. This is necessary because most phishing pages are set up to have access, and subsequently, steal users' data. The sorted pages with login fields and obfuscated code are sent to Signature Detection Module. In SDM, signatures are extracted from the sorted pages and its features are subsequently generated. A hybrid classifier is used to correctly label the extracted signatures. The work is implemented and evaluated using dataset from standard dataset from openly available research data sources such as PhishTank. Three experiments were conducted during the evaluation process. The first experiment, in which the performance metrics were evaluated, indicated that the accuracy of the proposed system is 100% with no false positives for specifically kit-generated sites. Whereas in the second experiment, which determines the accuracy of APK on general phishing dataset corpus from openly available data sources indicated 98% accuracy with low false positives. In the third experiment, the associated latency with the proposed system was evaluated. The evaluation results indicated a very low latency with insignificant bandwidth overhead. Future works will determine the accuracy of the system on large datasets from mainly phishing toolkit-generated sites and open phishing dataset corpus with adequate consideration for evasion technique such as randomization in toolkits. We hope to devise a method to have access to online black market to obtain data for phishing toolkit-generated sites from professional phishers to see if the submissions of ethical hackers have any resemblance with their swindles.

References

[1] N. Ajaya, M. Luthfor, S. Nitesh, H. Leane, "A multi-modal neuro-physiological study of phishing detection and malware warning," in *Proceedings of CCS*, USA, 2015.

[2] M. Alsharnouby, F. Alaca, S. Chiasson, "Why phishing still works: User strategies for combating phishing attacks", *International Journal of Human Computer Studies*, vol. 82, pp. 69–82, 2015.

[3] S. Antonio and P. Xavier, "Phishing secrets: History, effects, and countermeasures," *International Journal of Network Security*, vol. 11, no. 3, pp. 163–171, 2010.

[4] S. Aparna, K. Muniasamy, "Phish indicator: An indication for phishing sites," in *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, pp. 481–487, 2015.

[5] APWG, *Phishing Activity Trends Report*, 3rd Quarter, 2016. (`https://docs.apwg.org/reports/apwg_trends_report_q3_2016.pdf`)

[6] Y. Cao, W. Han, Y. Le, "Anti-phishing based on automated individual white-list," in *Proceedings of the 4th ACM workshop on Digital Identity Management*, pp. 51–60, 2008.

[7] B. Cheng, J. Fu, "Social Bots Detection on Mobile Social Networks," *International Journal of Network Security*, vol. 19, no. 1, pp. 163–166, Jan. 2017.

[8] M. Cova, C. Kruegel, G. Vigna, "There is no free phish: An analysis of "Free" and live phishing kits," in *USENIX Workshop on Offensive Technologies*, 2008.

[9] R. Gowtham, I. Krishnamurthi, "A Comprehensive and efficacious architecture for detecting phishing pages," *Computers and Security*, vol. 40, pp. 23–37, 2014.

[10] W. Han, Y. Cao, E. Bertino, J. Yong, "Using automated individual white-list to protect web digital identities," *Expert Systems with Applications*, vol. 39, no. 15, pp. 11861–11869, 2012.

[11] R. Islam, J. Abawajy, "Multi-tier phishing detection and filtering approach," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 324–335, 2013.

[12] P. Kathryn, et al., "The design of phishing studies: Challenges for researchers," *Computers and Security*, vol. 52, pp. 194–206, 2015.

[13] R. Krestel, P. Fankhauser, W. Nejdl, "Latent dirichlet allocation for tag recommendation," in *Proceedings of the Third ACM Conference on Recommender Systems*, pp. 61–68, 2009.

[14] H. McCalley, B. Wardman, G. Warner, *Analyis of Backdoored Phishing Kits*, Advances in Digital Forensics 7, pp. 155–168, 2011.

[15] E. Medvet, E. Kirda, C. Kruegel, "Visual-Similarity based phishing detection," in *Proceedings of 4th conference on Security and Privacy in Communication Networks*, 2008.

[16] A. A. Orunsolu and A. S. Sodiya, "An anti-phishing kit scheme for secure web transactions", in *Proceedings of the 3rd International Conference on Information Systems Security and Privacy*, pp. 15–24, 2017.

[17] RSA, *The RSA Anti-Fraud Command Center*, 2014. (`https://www.emc.com/collateral/solution-overview/10580-afcc-sb.pdf`)

[18] S. Sheng, et al., "Who falls for phish? A demographic analysis of phishing susceptibility and effectivenessfor interventions," in *Conference on Human factors in Computing Systems*, 2010.

[19] W. Xu, F. Zhang, S. Zhu, "JStill: Most static detection of obfuscated malicious Javascript code," in *Proceedings of the Third ACM Conference on Data and Application Security and Privacy*, pp. 117–128, 2013.

Biography

Orunsolu A. Abiodun is a Lecturer in the Department of Computer Science, Moshood Abiola Polytechnic, Abeokuta, South West, Nigeria. He is currently on his Ph.D degree at the Federal University of Agriculture, Abeokuta with research focus on web-related attacks/online fraud detection. He has a number of scholarly journal articles and conference proceedings on research areas such as Network

Security, Cryptographic protocols, Information retrieval and Intelligent systems.

A. S. Sodiya is a Professor of CS from the Federal University of Agriculture, Abeokuta. He is the President of Information Technology Systems and Security Professionals of the Nigeria Computer Society. He has a number of scholarly journal articles and conference proceedings to his credit.

A. T. Akinwale is a Professor of CS from the Federal University of Agriculture, Abeokuta. He has successfully supervised a number of Ph.D students and MSc students. He has a number of journal articles and conference proceedings to his credit. His area of research interests are similarity measures, databases, network protocols and computer security.

B. I. Olajuwon is a Professor of Mathematical Science from the Federal University of Agriculture, Abeokuta. He has a number of journal articles and conference proceedings to his credit.

A Hybrid Digital Signature Scheme on Dependable and Secure Data

Addepalli V. N. Krishna[1], Addepalli Hari Narayana[2], Somanchi Krishna Murthy[3]
(Corresponding author: Addepalli V. N. Krishna)

Department of Computer Science and Engineering, Christ University, Bangaluru, India[1]
(Email: adapalli.krishna@christuniversity.in)
Electrical Engineering, IIT-Indore, India[2]
Applied Mathematics Department, DIAT, Pune, India[3]
(Received Jan. 27, 2017; revised and accepted Mar. 1 & Apr. 4, 2017)

Abstract

In Digital Signature Standard Algorithm, a Discrete Logarithm Problem is used to calculate signature. The minimum Key length to be used is 1024 bit length. The work involves a Discrete Logarithm Problem computation, an Inverse and Modular operations at sender's side. At Receiver's side it involves two Discrete Logarithm Problem computations and modular operations which involves more computing resources as the key length needed is 1024 bit length In the present work, a Cubc Spline Curve based Public Key algorithm (CSCPKC) is used for Discrete Logarithm computation and the key length used is approximately 120 bit which needs less computing resources. It involves a secret matrix key to be shared among the participants which generates Random number sequence to be used in Signature generating algorithm. This model works well for a relatively small team of participants in having an Authentication process for Secured and Smooth data transfer with limited computing resource utilization.

Keywords: Cubic Spline Curve Public Key Cryptography; Digital Signature; Private Matrix Key; Random Number; Side Channel Attacks

1 Introduction

Digital Signature Algorithm consists of two numbers that contains values which are computed as per specified algorithm within parameters. This mechanism helps for authentication of users and also it verifies the integrity of the message. Digital signatures are generated through DSA and verified. Signatures are generated in association with Public and Private Keys. Thus each signatory has their own set of Public and Private keys which are used for Authentication process.

Any symmetric encryption scheme uses a private key for secure data transfer [20]. In their work on "A new Mathematical model on encryption scheme for secure data transfer [12]", the authors considered not only key but also time stamp and nonce values to increase the strength of sub key generated. In addition the nonce value can also be used for acknowledgement support between participating parties. The model can be further improved by considering a non linear model where the key values vary with the data generated [11].

Most of the products and standards that use public-key cryptography for encryption and digital signatures use RSA or D-H algorithm today [5]. Recently, Elliptic Curve Cryptography has begun to

challenge RSA. The principal attraction of ECC, compared to RSA, is that it appears to offer better security for a smaller key size, thereby reducing processing overhead.

Some recent works on application of ECC are cited here. [2, 3] Explains the engineering of ECC as a complex interdisciplinary research field encompassing such fields as mathematics, computer science and electrical engineering. The work [18] deals with adoption of Knapsack algorithm on ECC and its added advantages lie more security in real time applications. [7] Specifies the standard specifications for public key cryptography. Encryption to data supports the very important features like security, Confidentiality to data & Authentication of users. [15, 17] discussed the features of Numerical data analysis which helps in building a mathematical model. In works [9, 10], the authors discussed a new public key algorithm which is based on Cubic Spline curve Public Key Cryptography (CSCPKC). They made a comparative study of Cubic Spline curve based cryptography with ECC algorithm in terms of Key length and computing resources. They also worked on different scalar operations on CSCPKC which forms the security of the proposed algorithm. In [8] the author discussed a new & simple algorithm which generates a random number sequence. The sequence is not a time bound algorithm but it depends on the vector being used in the algorithm. Thus in this algorithm it generates a random sequence which can be used in Digital signatures and which consumes low computing resources when compared to standard random number generator algorithms. [14] discusses the Standard Digital Signature Scheme approved by Information Technology Laboratory National Institute of Standards and Technology Gaithersburg, USA. In work [6], the author deals with application of Cubic Spline Interpolation in cryptosystems using Choatic Mapping Concept and discussed the strengths against crypto analysis. In work [22], the authors represented the Cubic spline curve in terms of Symmetric encryption mechanism and its crypto analytical strength. The works [1, 13, 19, 21, 24] deals with Survey, Relevance and importance of DSA in authentication process. The works deals with application of Block ciphers or application of ECC on DSA and its improvements in authentication process.

2 Modelling of The Problem

The work may be divided to two parts. The first part deals with the generation of random number sequence. The second part deals with generation of Cubic Spline curve Public Key algorithm to be used in Digital signature for Authentication purpose.

Algorithm to Generate Random Sequence:

1) A random matrix is being used as a key. Let it be A.
2) Generate a Ternary vector for N values, i.e from 0 to $N-1$. Let this be B.
3) Multiply $A \times B$.
4) Consider a modulus function on the product of Step 3 by some prime number.
5) Convert the output of Step 4 to decimal which forms a random number generated sequence.

Modeling of Cubic Spline Curve Problem (CSCPKC Algorithm):

Global Parameters:

- T_1, T_N: The first and the last data points (Considering the problem as natural Spline);
- n: number of nodal points on the curve considered (Δx being defined by number of points considered on the cubic spline curve);
- G: Base Sequence considered;
- t: Number of iterations considered (which specifies Δt considered in the algorithm);

- K: Random number considered from Random number sequence generator algorithm;
- P: Field considered.

For the 2 point on the curve:

$$\begin{aligned} B(2) &= -\alpha\frac{\Delta t}{\Delta x^2} \bmod P \\ A(2) &= 1+2\alpha\frac{\Delta t}{\Delta x^2} \bmod P \\ D(2) &= Y(2)+\alpha\frac{\Delta t}{\Delta x^2}\times D(1) \bmod P. \end{aligned}$$

For the points 3 to $n-2$:

$$\begin{aligned} B(I) &= -\alpha\frac{\Delta t}{\Delta x^2} \bmod P \\ C(I) &= B(I) \\ A(I) &= 1+2\alpha\frac{\Delta t}{\Delta x^2} \bmod P \\ D(I) &= Y(I). \end{aligned}$$

For the $n-1$ point:

$$\begin{aligned} C(N-1) &= -\alpha\frac{\Delta t}{\Delta x^2} \bmod P \\ A(N-1) &= 1+2\alpha\frac{\Delta t}{\Delta x^2} \bmod P \\ D(N-1) &= Y(N-1)+(\alpha\frac{\Delta t}{\Delta x^2})\times D_N. \end{aligned}$$

These conditions imply that T_1 is known in terms of T_2. Thus the point 2 is a relation between T_1, T_2 & T_3. But since T_1 is known, this relation reduces to a relation between T_2 and T_3. This process of substitution can be continued until T_{n-1} can be formally expressed as T_n. But since T_n is known we can obtain T_{n-1}. This enables us to begin back substitution process in which T_{n-2}, T_{n-3}, $\cdots$, T_3, T_2 can be obtained for one iteration. Thus the problem is solved by Tridiagonal matrix algorithm and the process is repeated for 'i' iterations.

A New Hybrid Digital Signature Algorithm:

Consider a CBSPKC algorithm, with global parameters like G, P, Public key being PB, X be the Private Key, K be the random number considered from the Random sequence generator algorithm.

Sender's Signature:

1) Calculate $V = G^K \bmod P$.
2) Calculate $V_1 = V^x \bmod P$.

Receiver's Verification:

1) Calculate $R = PB^K \bmod P$, if $R = V_1$, then verified.
2) If Integrity of the message is also to be considered, G can be replaced with $G + H(m)$ where $H(m)$ represents the hash value of the message sent.

Example:

Random Number Generator:

$$\begin{aligned}
\text{For } n &= 0:80\\
n_1 &= floor(n/3);\\
r_1 &= n - n_1 \times 3;\\
n_2 &= floor(n_1/3);\\
r_2 &= n_1 - n_2 \times 3;\\
n_3 &= floor(n_2/3);\\
r_3 &= n_2 - n_3 \times 3;\\
r_4 &= n_3;\\
r &= [r_4\ r_3\ r_2\ r_1];\\
r &= r';\\
a &= [3\ 4\ 2\ -6;\ 4\ -5\ 2\ 6;\ 3\ -2\ 6\ 8;\ 6\ -3\ 2\ 8];\\
r &= modulo(a \times r, 3);\\
r &= r(4,1) + r(3,1) \times 3 + r(2,1) \times 9 + r(1,1) \times 27\\
\text{end}
\end{aligned}$$

Sequence generated is 0, 8, 4, 74,79, 75, 80, 78, 77, 74, 53, 49, 45, 7, 3, $\cdots$ which is random in nature.

Depending on the session participation, random number can be considered. For the given problem the session considered is 14, so the random number considered $(K) = 7$.

CSCPKC:

Boundary Conditions: Both sides maintained at known data values, i.e $T_1 = 4$, $T_N = 7$;
Global Parameters: G, T_1, T_N, α, Δx, Δt;
N**:** Ternary Vector of 81 values considered;
n**:** 11 points considered on the cubic spline curve;
K**:** Random number;
t**:** Private key;
P**:** Field;
PB**:** Public key, (G^t);
Global Parameters:

$$\begin{aligned}
\alpha &= 4, \Delta t = 3, \Delta x = 3, t = 5, T_1 = 4, T_n = 7;\\
G &= 4\ 6\ 23\ 8\ 8\ 25\ 8\ 6\ 6\ 11\ 7.
\end{aligned}$$

Generating Public Key from Private key:

- Private key: $t = 5$;
- Public key: $(G)^5 = PB = G_2$;
- $PB = G_2 = 4\ 32\ 9\ 33\ 16\ 1\ 17\ 6\ 19\ 32\ 7$.

Sender's Signature:

1) Calculate $V = G^K \bmod P$;

$$V = 4\ 26\ 10\ 35\ 13\ 39\ 18\ 24\ 13\ 3\ 7\ = G_1$$

.

2) Calculate $V_1 = V^x \bmod P$;

$$V_1 = 4\ 11\ 29\ 0\ 3\ 11\ 13\ 11\ 14\ 15\ 7 = G_3.$$

Receiver's Verification:

1) Calculate $R = PB^K \bmod P$, if $R = V_1$, then verified.
2) $R = 4\ 11\ 29\ 0\ 3\ 11\ 13\ 11\ 14\ 15\ 7 = V_1 = G_3$ (Hence proved).

Complexity:

Consider the equation,

$$A_B = G^x \bmod P$$

where g is the generator; P be the field. Thus if we go by the complexity of the discrete logarithm problem, it is of the order of $e^{((\ln P)^{1/3} \ln(\ln P))^{2/3}} \times O(n) \times O(N)$ where n refers to number of nodal point considered on the curve and N refers to size of Ternary vector considered.

3 Conclusion

The work deals with development of new digital signature algorithm which can be used in a limited environment. The main advantage with this mechanism is it consumes very less computing resources for authentication purpose. It provides a combination of Random number generator algorithm which needs a Matrix Private Key to be shared among the participants and CSCPKC algorithm for generating the sequence which can be mapped for Digital signature. The work may be extended to Digital Signature Standard algorithm which needs more complex construction and computing resources for authentication and verification.

References

[1] N. Barik, K. Sunil, "A study on efficient digital signature scheme for E-governance security", *Global Journal of Science & Technology*, vol. 2, no. 3, 2012.
[2] R. C. C. Cheng, N. J. Baptiste, W. Luk, P. Y. K. Cheung, "Customizable elliptic curve cryptosystems", *IEEE Transactions on VLSI Systems*, vol. 13, no. 9, pp. 1048–1059, 2005.
[3] A. Cilardo, L. Coppolino, N. Mazzocca, L. Romano, "Elliptic curve cryptography engineering", *Proceedings of the IEEE*, vol. 94, no. 2, pp. 395–406, 2006.
[4] Cryptography Stack Exchange, *Time Complexity to Solve Discrete Log Problem*, Apr. 29, 2017. (`http://crypto.stackexchange.com/questions/12893/time-complexity-to-solve-discrete-log-problem`)
[5] W. Diffie, "The first ten years of public key cryptography", *Proceedings of IEEE*, vol. 76, no. 5, pp. 560–577, 1988.
[6] F. Hwu, C. Y. Ho, *The Interpolating Random Spline Cryptosystem and the Chaotic-map Public-key Cryptosystem*, University of Missouri - Rolla, CSc-93-09, May 1993. (`http://cs.mst.edu/media/academic/cs/documents/technicalreports/93-09.pdf`)
[7] IEEE, *Standard Specifications for Public Key Cryptography*, IEEE Standard 1363, 2000.
[8] A. V. N. Krishna, "A new algorithm for Random number generation in network security", *Journal for Scientific & Industrial Research*, vol. 64, pp. 791–793, 2005.
[9] A. V. N. Krishna, H. Narayana, "A cubic spline curve public key cryptography", *accepted with Journal for Discrete Mathematical Sciences and Cryptography.*

[10] A. V. N. Krishna, H. Narayana, V. K. Madhura, "Window method based cubic spline public key cryptography", *International Journal of Electronics and Information Engineering*, vol. 4, no. 2, pp. 94–102, 2016.
[11] A. V. N. Krishna, "A new non-linear model based encryption scheme with time stamp & acknowledgement support", *International Journal of Network Security*, vol. 13, no. 3, pp. 202–207, 2007.
[12] A. V. N. Krishna, A. V. Babu, "A new model based encryption scheme with time stamp & acknowledgement support", *International Journal of Network Security*, vol. 11, no. 3, pp. 172–176, 2010.
[13] P. Kuppuswamy, P. M. Appa, S. Q. Y. Al-Khalidi, "A new efficient digital signature scheme algorithm based on block cipher", *IOSR Journal of Computer Engineering*, vol. 7, no. 1, pp. 47–52, 2012.
[14] NIST, *Digital Signature Standard (DSS)*, Information Technology Laboratory National Institute of Standards and Technology Gaithersburg, MD 20899-8900, July 2013.
[15] Numerical Methods for Scientific and Engineering Computation, M.K.Jain, SRK Iyengar and RK Jain, New Age International Publishers.
[16] S. V. Patankar, *Numerical Heat Transfer and Fluid Flow*, McGraw Hill, 1980.
[17] R. Ramanna, *Numerical methods*, pp. 78–85, 1990.
[18] R. Ramasamy, R. A. Prabakar, M. I. Devi, M. Suguna, "Knapsack based ECC encryption and decryption", *International Journal of Network Security*, vol. 9, no. 3, pp. 218–226, 2009.
[19] S. Singh et al., "Survey on techniques developed using digital signature: Public key cryptography", *International Journal of Computer Applications*, vol. 117, no. 16, May 2015.
[20] W. Stallings, *Cryptography and Network Security*, Prentice Hall, 4th Edition, 2006.
[21] S. R. Subramanya, K. Byung, "Digital signatures", *IEEE POTENTIALS*, pp. 5–8, 2006.
[22] N. Sun, T. Ayabe, K. Okumura, "An animation engine with cubic spline interpolation", in *Proceedings of International Conference on Intelligent Information Hiding & Multimedia Signal Processing*, pp. 109–112, 2008.
[23] S. Yakoubov, V. Gadepally, N. Schear, E. Shen, A. Yerukhimovich, "A survey of cryptographic approaches to securing big-data analytics in the cloud", in *IEEE High Performance Extreme Computing Conference (HPEC'14)*, pp. 1–6, 2014.
[24] Q. Zhang, Z. Li, C. Song, "The improvement of digital signature algorithm based on elliptic curve cryptography", in *2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC'11)*, pp. 1689–1691, 2011.

Biography

Addepalli V. N. Krishna, working as Professor, Computer Science & Engineering, Faculty of Engineering, Christ University, having 22 years of teaching Experience. He was involved with WASE Program (WIPRO academic Software Engineering) a Collaborative Program between Wipro Technologies & BITS, Pilani between 2004-2011.He has 28 papers published In Journals of national and international repute. His Areas of Interest are Cryptography, Data Modeling, Data Mining and presently guiding 3 Students for their Doctoral studies.

Addepalli Hari Narayana is studying Electrical Engineering, IIT, Indore. His areas of Interest are Digital Signal Processing, Microprocessors and cryptography.

Somanchi Krishna Murthy is working as Head, Associate Professor, Department of Applied Mathematics, DIAT, Pune. He has experience of around 18 yrs and published work in Journals and conferences

of National and International repute. His areas of Interest are Mathematical Modeling, Cryptography and Advanced Reinforced Polymers.

Evaluating the Segmentation Methods of Image Logs to Identify Natural Fractures in Hydrocarbon Wells

Milad Karami[1], Ahmad Keshavarz[2], Hamed Jelodar[1]

(Corresponding author: Milad Karami)

Department of Computer, Science and Research, Islamic Azad University, Bushehr, Iran[1]

(Email: karamimilad1@gmail.com)

Electrical Engineering Department, Persian Gulf University, Bushehr 75168, Iran[2]

(Received Feb. 13, 2017; revised and accepted Apr. 23, 2017)

Abstract

Image logs are one of the strong tools used to identify and interpret information of carbonate wells. With the aim of this virtual images, different phenomena such as fractures, faults and layers that have important role in carbonate wells can be identified and interpreted. Any physical rupture or separation of stone that is due to increasing tensions on stone resistance which may be formed by excavation factors in the stone is called fracture. Because of some factors such as excavation dust, heating and errors due to machine, grabbed images due to noise existence are not qualified. Today by using image processing techniques in the different areas such geology and oil, image quality can be improved. The use of intelligent ways to detect fractures in these images can be useful as other areas. One of the basic steps in fractures separation in image logs, is images segmentation. The purpose of segmentation is to isolate the fractures routes from background pixels in images. In this paper, first to remove noise from image logs, different filters were implemented and studied on the images. The results of both qualitative and quantitative criteria were evaluated. The results showed that Gaussian low-cross filtering method is the most optimized method to improve the images and remove the noise. Then some of the classification different methods on sample images of boreholes was operated and evaluated qualitatively and quantitatively. The results show that the color information-based methods and the use of clustering concept has the optimal results for the grabbed images classification from boreholes wall to identify fractures.

Keywords: Detection; Fractures; Image Logs; Image Processing; Segmentation

1 Introduction

Due to the existence of huge reservoirs of oil and gas in Iran, for maximum progression and utilization of the available resources, business and industry in this area are of particular interest. Therefore, identification and recognition of carbonate reservoirs are critically essential in this industry [6]. With the development of science and new technologies, utilizing new tools and methods to enhance the pace of work and reducing the costs and risks can be used in this area. For this reason, smart methods and techniques are being applied with particular interest in this field instead of traditional methods of this industry.

Accurate identification of existing phenomena in reservoirs using smart techniques in order to utilize the carbonate reservoirs with higher efficiency is very important. Fractures are one of the most effective parameters in exploitation and exploration of reservoirs and in other fields such as rock mechanics and water resources; since it plays influential roles in fluid motion and wellbore stability and modeling boreholes [9].

Any type of failure or physical separation of rock which is caused by excess stress on rock strength, resulting from natural mechanisms, drilling factors, etc. in the rock, is called fracture [6].

Fracture identification by image logs, for reasons including cost and less time, and also due to reasons such as not being able to identify phenomena larger than the diameter of the core by core test and the impossibility of multiple impressions on a regional core, has been taken into particular consideration and use toward other techniques of reservoir identification such as core test. However, because of the complexity of the reservoir, accurate and manual identification of fractures and other phenomena in image logs is difficult and even impossible and time-consuming in some cases. Hence, using smart and automatic methods can help to reduce these problems.

At the moment, there are different methods for identification of fractures in boreholes, but fracture identification by image logs, for reasons including cost and less time, and also due to reasons such as not being able to identify phenomena larger than the diameter of the core by core test and the impossibility of multiple impressions on a regional core, has been taken into particular considerationand use toward other techniques of reservoir identification such as core test. However, because of the complexity of the reservoir, accurate and manual identification of fractures and other phenomena in image logs is difficult and even impossible and time-consuming in some cases and also, due to the complexity and high density of fractures, commentary by human experts is not the same in many cases.

Because of the importance of small-scale fractures in naturally fractured reservoirs, as well as issues such as wellbore stability and fluid motion of wells through the fractures network, a new technology emerged in the oil industry to detect fractures in oil industry called image logs. In photography of the wells after measuring the physical properties of rocks of reservoirs' wall with different methods, in order to evaluate the existing features and phenomena in the walls, these signals are transduced into virtual images which depict these phenomena and features. The virtual images are called image logs.

Conducted research show that studies in the field of fractures identification have been carried out intelligently by image logs which possess both advantages and disadvantages. In 2004, Wu et al. worked on enhancing the images of carbonate reservoirs using a combination of different types of image logs of a well. As a result of enhancing the obtained images, fracture in these images could be identified easier by experts [15]. In 2005, Wang studied the detection of edges to find the fractures in the rocks on CCD images taken from the rocks, using valley-edge algorithm without the application of threshold values. He prevented the occurrence of noise in the images using multi-scale technique for detecting the edges [11]. In 2007, Wang implemented a new algorithm to identify the fractures in these images through images taken from rock with advanced CCD cameras. After removing the noise by middle-consuming filter, he segmented the edge-based image and separated the fracture from other phenomena in images by extracting eleven features using support vector machine clustering [12]. In 2005, Liu segmented the images of carbonated reservoirs using two-dimensional wavelet transform. He separated sub-images of fracture and porosity from the image of the well [15]. In 2008, Kherroubi separated the traces of fracture by morphological operations. Then, with basic information about the layers of wells, made an effort to find the main direction of fracture and afterwards, separated the fractures from other phenomena and classified footprints belonging to a fracture using clustering algorithm [5]. In 2010, Wang detected the edges using the ultraviolet image taken from the rocks alongside the optical images. He detected the edges by canny edge detector and thresholding. He then found connection points of the fractures by removing the noise and narrowing the peaks. Moreover, he connected the pieces related to a fracture and filled the cuts [13]. In 2010, Wang used a particular type of neural network called PCNN to identify

the fractures through a new method for detecting the edges [4].

Liu in 2005, segmented the images of carbonated reservoirs by two-dimensional wavelet transform in an article titled segmentation of micro-resistance images using two-dimensional dynamic wavelet transform. He separated sub-images of fracture and porosity from the image of the well [7]. To identify the fractures, in 2010, he used a particular type of neural network called PCNN in an article titled PCNN (Pulse coupled neural network) edge detector for images of rock fracture through a new method for detecting the edges [4]. In 2013, Assous et al. represented a new algorithm for detection of plane features on micro-resistance images of the well in a report titled automatic detection of plane features in image logs. To separate the sinusoidal plane features, they had also relied on information of the edge in the main part of their work [2].

Due to the need to identify natural fractures including normal open fractures, which are the path of fluid oil transfer under the earth and the presence of complexities that were mentioned above, the use of smart and image processing methods in order to help oil industry experts to identify geological phenomena have attracted attentions. Using analysis techniques and image processing can result in achieving acceptable outcomes in this area, but despite the importance of fractures, perhaps because of the complexities involved in this field such as poor quality of images and very similar phenomena in images, etc. this issue has been of less attention and effort [13].

Given that today in oil industry, visual diagrams are used as a key tool for identification and interpretation of geological phenomena such as fracture, exact interpretation and achieving accurate and precise results in wells' phenomena such as fracture, is heavily dependent on the appropriate and fine quality of the images [11]. Poor pictures can lead to poor interpretation as well as wrong results and false reservoir modeling for geologists. Thus, enhancing image quality and image interpretation quality control are two key parameters in identification and detection of fractures [11].

In this article, given the importance of enhancing image logs in order to identify geological phenomena and interpreting them, various filters have firstly been implemented on the images to improve their quality and accuracy of different methods has been evaluated with an appropriate standard. After qualitative and quantitative examination of the implemented methods, the most efficient method to improve the quality of studied images is introduced. The next major step in identifying fractures, is the segmentation of images taken from the well. Studying previous research shows that currently there is no smart and automatic method for segmentation of image logs because of the complexities involved in identifying fractures and no comprehensive work has been done in this area. Therefore, because of the importance of fracture, this paper aims to provide an algorithm and method for segmenting image logs in order to identify fractures intelligently. First, the required theories are introduced in the methodology chapter, then the proposed algorithm is explained in the implementation chapter and is implemented on sample images and the results are evaluated and discussed. The final chapter concludes and summarizes the paper.

2 Methodologies

In image processing, image enhancement is one of the essential steps in achieving optimized results in further procedures such as image segmentation. Segmentation is one of the necessary and vital steps. Precise segmentation determines the probable success or failure during computer analysis process. Segmentation is often the most difficult requirement for image analysis, especially in rock mass image processing in which the complications and surface properties of the rock add to the complexity of these issues. For this reason, certain research has been conducted in the field of image processing [5]. The process of dividing a digital image into several areas is called segmentation. Resulting sections are different objects that possess similar features including texture, color, etc. The outcome of segmentation is a series of areas which form the whole image together. All pixels of an individual area have similar

properties such as color, intensity, texture, etc. [18].

2.1 Image Enhancement

Image enhancement refers to manipulating an image in order to achieve a more appropriate one in comparison with the main image for a particular application [2]. The most optimized method for different contexts is noise removal and enhancement of image quality, thus separate studies and experiments should be conducted for various applied fields. There is no general theory for enhancing image quality but in general, the best index to investigate the results of image quality enhancement are human experts and human visual system [3].

In this chapter, we briefly introduce the theories of various filters for image processing, and then we propose an appropriate quantitative criterion to analyze the accuracy of filters when there is no noiseless reference image for investigating the performance of filters. This criterion accommodates with human interpretation of the enhanced images and is the most proper quantity for practical applications of image processing in which there is no noiseless reference image such as images of boreholes.

One type of filters are local filters. These filters enhance the images relying on local information, amount of color, and bright surfaces of pixels of an image. These filters possess greater application range and freedom of action in comparison with frequency filters. For instance, unlike frequency filters, these filters can also carry out non-linear operations [3]. Therefore, in this paper, we focus on some of the most applicable local filters for enhancing the images of wells.

General mechanism of local filters is based on the fact that for each pixel, a neighborhood is considered that includes the neighboring pixels of the studied one and a mathematical operation is conducted on these pixels. Implementing this mathematical operation on every pixel within the noisy image results in a new enhanced image [3]. Studied filters in this paper are briefly described below.

2.1.1 Average Filter (Mean)

Mean filter is used to remove unnecessary details and noise in images. After implementing on noisy images, this filter makes them opaque and reduces the noise. The general relationship of this filter for one pixel of the image and the defined neighborhood of the studied pixel is described as follows [3].

$$g(x,y) = \frac{\sum_{s=-a}^{a}\sum_{t=-b}^{b} w(s,t) f(x+s,y+t)}{\sum_{s=-a}^{a}\sum_{t=-b}^{b} w(s,t)} \tag{1}$$

$f(x,y)$ is the pixel brightness of the input image, $g(x,y)$ is the pixel brightness after applying the filter, and $w(s,t)$ weights applied to the input image pixels for implementation weighted average filter.

2.1.2 Laplacian Filter

This filter for each pixel of the noisy image, calculate the second derivative in the vertical and horizontal direction and leads to improve the noisy images. This filter relationship is as follows [3].

$$\nabla_f^2 = \frac{\partial_f^2}{\partial_x^2} + \frac{\partial_f^2}{\partial_y^2} \tag{2}$$

2.1.3 Unsharping Filter

These filters reduce the parts and sudden changes in brightness of image pixels which are caused by factors such as noise and lead to improve images. The implementation of this filter is as follows [3]:

Step 1: It amazes the original image.

Step 2: The original image is subtracted from amazed image. The resulting image is called mask.

Step 3: Masks produced in Step 2 is added to the original image. The resulting image will be an improved image.

2.1.4 Sobel Gradient Filter

This filter uses the first derivative property to the vertical and horizontal direction which is called gradients to improve Pictures [3]. Filter is taken fromfirst derivative image by masking on the image vertically and horizontally and can cause to improve Pictures.

2.1.5 Gaussian Low-pass Filter

Gaussian low-pass filter (GLPF) is defined as the following equation. In this respect, D is the gap in the frequency domain to the center and m is Gaussian filter variance. This filter also reduces noise, and can have the ability not to remove the useful thumbnail image information instead of noise and just removes noise [3].

$$H(u,v) = e^{\frac{-D^2(u,v)}{2\sigma^2}} \tag{3}$$

2.1.6 Optimal Evaluation Criteria

Digital images are subject to various types of distortion. The distortions in various stages of processing, namely the production, processing, compression, storage, transmission and resetting the image may be applied to the image. Each of these distortions may cause loss of visual quality (Subjective) of the image. Thus, there must be some methods for evaluating images, that by using them we can score for visual quality of images and image processing algorithms [17].

There are two basic method for evaluating images. In the first method the images evaluated by using human observers. This method is called tendentiously. In applications where there are the images finally are used by the human user, this method is only the right approach. But the shortcoming of this method is hard work, time-consuming and high cost. Especially in ways that we are going to get to optimal response by repeat, applying this method is inappropriate [17].

The second method for evaluating images which is called unbiased method, uses quantitative criteria for evaluating the image. Whatever these quantitative criteria selected lower, the result of the evaluation with this method is closer to the first approach. Although, recently some criteria such as PSNR, MSE are used in articles that have no direct relationship with the visual quality of image, but the purpose of most present researches is to provide a criteria more appropriate to evaluate images so that visual quality is considered [17].

There are one of the major issues in the impartial methods or absence of the original image (the image without distortion) is distorted to compare with the image. Impartial procedures with regard to this issue are divided into three categories. In the first approach which is called Full-reference is called, the original image exists totally. In contrast, the approach without reference there is no original image. In the third approach that refers to the reference reduced, some features of the original image is available for side information.

Given that for image logs,there is the reference image without noise, a suitable quantitative measure without reference that can dealt with the quantitatively assess the methods of omitting noise and improve the quality of images. The results of the quantitative criteria should be adjusted with qualitative evaluation of human observers and human visual system. Studies showed that one of most optimal

quantitativecriteria according to our willingness is a criteria named Q.This criteria concerned with quantitative evaluation according to our willingness through decomposition of the eigenvalues of image gradient matrix [17]. The general algorithm of quantitative parameter calculation Q is as follows:

Step 1: The input image is divided into a number of smaller images with the specific dimensions which is called mask.

Step 2: For divided image above, we find the masks that are not anisotropic.

Step 3: The parameter value is calculated with the following formula for thefound non-anisotropic masks in the previous step.

$$Q_k = S_1 \frac{S_1 - S_2}{S_1 + S_2} \tag{4}$$

S_1 and S_2 is the gradient eigen value of any masks from divided image.

Step 4: Quantitative parameter Q for whole image is calculated from the following equation.

$$Q = \frac{1}{M} \sum_{k=1}^{m} Q_k \tag{5}$$

M is the total number of evaluated image and m is the masks due to being non anisotropic the small amount was calculated for them in Step 3.

2.2 Segmentation of Images

It is a vital and essential step in image processing. Most done segmentation methods in various applications, are based on monochrome images. For process the colorful images processing there needs to more calculation than gray image processing. But today, with increased speed and reduced cost of computing, processing of colorful images is also taken into consideration in the last decade. Colorful images could enhance the quality of segmentation images [18]. Colorful images have more information than monochrome images. Each pixel in an image containing information about brightness and saturation levels. It provides the total color of an image that can lead to better and more trusted classification of images. There are different models to show the colors such as: RGB, CMY, and HSV. Usually, the colorful model RGB is selected to segmentation images because of its simplicity and faster processing. In a colorful image each pixel is displayed by a combination of red, green and blue. They must have a steady ratio for colorful images for the areas belonging to a sector. Necessities of a good classification are includes: pixels of a section should not contain different colors, and each section should just have a label. All certain pixels must belong to a region labeled. The luminance severity of a segment should be acceptably uniform. In this section the segmentation methods that have been used and implemented in this article are explained.

2.2.1 Edge Detection

Edges detection means the process of detecting and locating brightness sharp discontinuities in an image brightness. Discontinuities are sudden changes in intensity of pixels brightness that determine the objects boundary in the image. Classical edge detection method requires a convolution of the image with an operator (a two-dimensional filter). These operators are designed in such way to be sensitive huge gradient in picture, which show zero in uniform areas. There is lots of edge detector operator that each are sensitive to a certain type of edges [1]. One of the Gaussian edge detectorsis Canny

edge detector. Canny edge detector despite the complexity of the calculations much better answer than other methods commonly used to detect edges in noisy images. This is an algorithm known as best edge detector [1, 8]. This algorithm first lights the image and eliminates noise. Then the image gradient to highlight regions calculated by high spatial derivatives. Then the algorithm dealt with searching in these highlighted pixels and each pixel cannot be removed as the maximum. They are determined according to the two threshold values in a hysteresis.

2.2.2 Thresholding Method

In thresholding method, the image directly divides to different areas based on kevel of brightness or properties. In images that background and the object dived into two mode categories in brightness histogram, to isolate object from in an image a threshold value T can be determined that isolates these two mode. Then all points that brightness is more in them than threshold value, belong to a class that are lower than this threshold, belong to another class [14]. Equation (6) shows this.

$$g(x,y) = \left\{ \begin{array}{l} 1 \text{ if } f(x,y) > T \\ 0 \text{ if } f(x,y) < T \end{array} \right\} \tag{6}$$

2.2.3 Morphological Operators

The word morphology refers to a branch of biology science that explains about the form and structure of animals and plants. In mathematics, morphology is as a tool for new and descriptive performance of areas such as borders, construct, and the Canucks Hall [14]. Mathematic language is the morphology of sets theory. The morphology provides a unique and powerful approach of image processing issues [14].

2.2.4 C-means Clustering

This algorithm is a clustering algorithm that by using the distance measurement of samples from each other, divides them into several distinct classes. This algorithm can be used for images segmentation assuming the number of classes is specified. This algorithm is an unsupervised classification that divides the input data based on the vital distance to several classes defined [18]. The algorithm can be defined as follows:

1) First the clusters number is determined by the user.
2) As same as number of clusters defined values as the average initial value of each cluster are chosen randomly.
3) The data distance is calculated to each average.
4) Due to the distance value for each data to the average have the smallest distance; it has belonged to the closest cluster.
5) After the data clustering for clusters obtained, the mean of each data recalculated.
6) Using criteria such as the rate of the means changes in two successive stages or the number of algorithm iteration, the end of clustering algorithm determined. Otherwise Steps 3 to 5 are repeated again.

2.2.5 Self-Organized Map Neural Network (SOM)

Among the neural network, SOM is an unsupervised networks usually it is used for segmentation of images [10]. This network is a competitive type that is used for clustering data. In this network neurons with respect to the input data compete with each other to be activated. The arrival of a new data to the Ethernet interface, all neurons of output layer according to a distance criterion, calculate their distance to this data.Neurons with closest distance criteria conquers. Then weights of this neuron and neighboring neurons proportional to their distance to the winner neuron updated.

2.2.6 Clustering Algorithms Evaluation

There is many methods have been proposed yet for images segmentation, but still there is no a very compelling valuation parameter to assess segmentation methods [16]. This makes the comparison of different algorithms and determining most optimal parameters for a clustering method hard [16]. Each person has its own standards for a good segmentation and considers the different applications of different criteria for optimal segmentation [16]. So in general, usually optimal segmentations dependent on the images application and determining the exact parameters of assessment is difficult. Most of evaluation methods are thematic segmentation (based on understanding and human interpretation). There are approaches objective and based on mathematical parameters, but the major of these approaches are used to assess the segmentation results for algorithms have reference image which prepared by human to compare [16]. With regard to this there is no reference for evaluating segmentation of image logs and no possibility to isolate all segmentation pixels manually to create reference images due to complexity of these images. To evaluate segmentation results, subjective assessment based on human perception can be used. But to evaluate quantitatively the segmentation algorithms can be generally considered parameters that assess different methods with reference and non-reference image by using them [16].

Quantitative criteria that can be used to evaluate algorithms are includes [16]:

- Coherence and uniformity of classification areas with respect to certain properties, such as color.
- Areas where are in the different classification groupssuch as color, they must be as possible different according to the hypothesis, such as distance and neighborhood.
- Inner regions of sectors should be simple as much as possible without holes (holes).
- Boundaries of different sectors as possible are simple and sharp.
- There be no extra sectors because classification in images.

3 The Proposed Algorithm Implementation

Any filters presented in the previous was implemented on a set of image logs. The results both qualitatively and quantitatively were evaluated by assessment criteria Q which was introduced in the previous section.

Figure 1 shows a number of examined images containing noise. There are different geological phenomena such as natural and artificial fractures and layers. The quality of these images is reduced because of the noise caused by factors such as drilling mud, heat and etc.

Figure 1 shows the results obtained after implementation discussed filters in this article.

Examining the results above qualitatively shows that the average filters and Gaussian low-pass have better results compared to other filters. Between these two filters, although mean filter has removed the small details and more noise, but leads to more blurring the image too. This can lead to reduce the

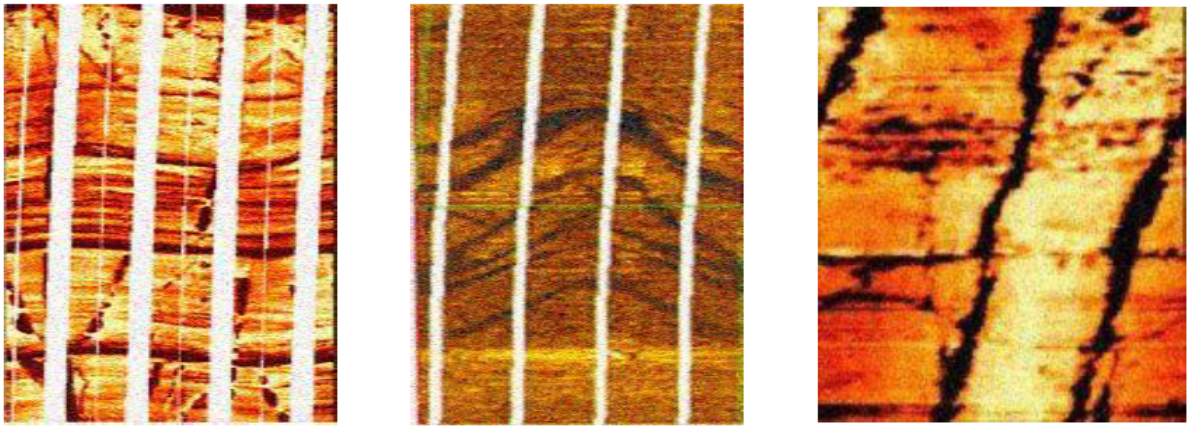


Figure 1: Noisy visual well diagrams

quality of image processing in other steps on these images such as segmentation. Therefore, Gaussian filter method due to rely on both the principle of noise reduction and maintaining the main information of image is quantitatively a more optimal method to improve the image logs. Table 1 shows the results of quantitative filters to improve image with evaluation criterion Q.

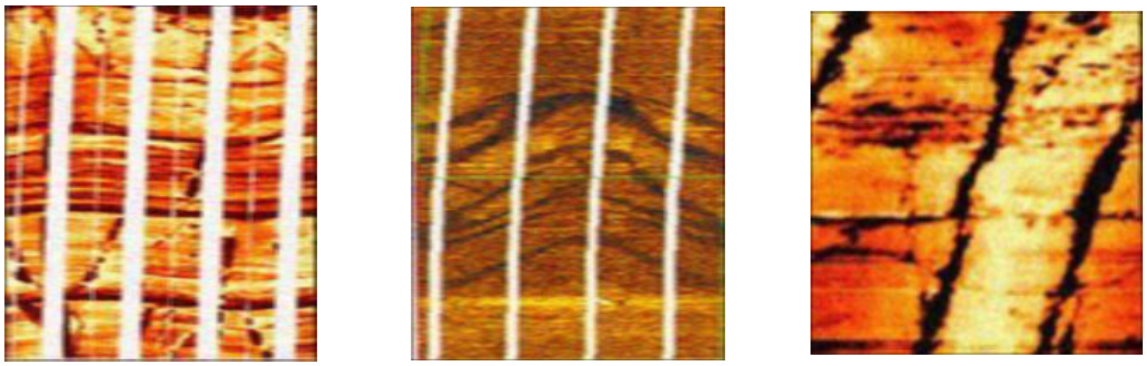
Table 1: The results of quantitative evaluation with criterion Q for different filters implemented on sample image

Figures	Mean filter	Laplacian filter	Unsharping filter	Sobel filter	Gaussian low-pass filter
Figure 1	2/7	0/049	0/014	1/59	4/4
Figure 2	5	0/009	0/008	0/3	7/8
Figure 3	2/7	0/59	0/15	0/6	3/4

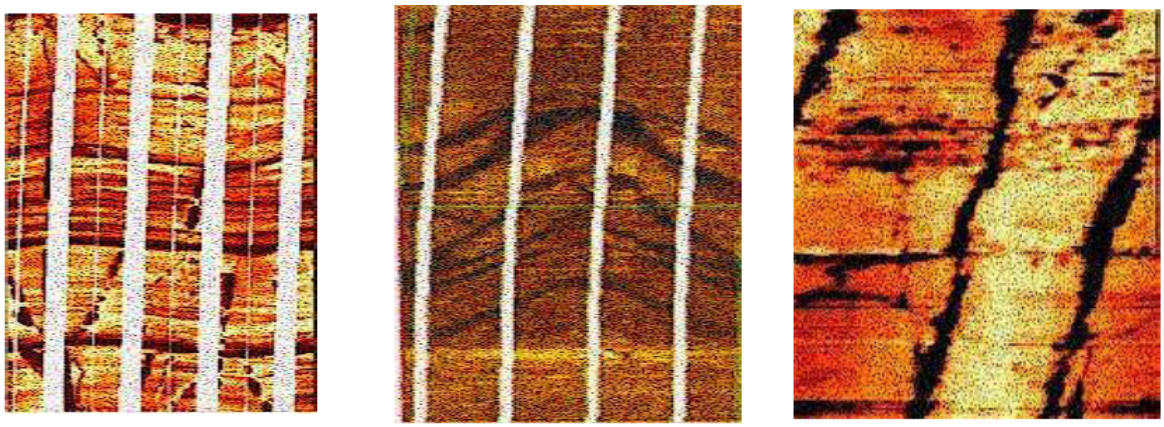
Examining the quantitative results with the criterion Q in Table 1 show that Gaussian low-pass filter has the best accuracy to remove noise and improve the carbonate wells images. Next stages after removing noise are images segmentation. After the implementation of each approach, the results are qualitatively and quantitatively evaluated. Then the final results of all methods and their performance compared to each other and the best way to segmentation the image logs determined. In the following section the implementation and evaluation of each of the methods will be discussed and investigated.

Figure 3 shows some of image logs that Gaussian low-pass filter is applied on them and the proposed algorithm to segmentation optimally in order to detect natural fractures implemented on them. In these images various phenomena such as artificial fractures, natural fractures, layout of ground is seen in the boreholes. This phenomena is seen with a different color than the other image pixels. The artificial fractures in the images that result from drilling are seen as vertical dark bands. Natural fractures and ground layouts are seen as the dark sinus.

The subjective examination of segmentation method results based on edge detector in Figure 4 shows that the method indicates many details in segmentation image which is not related to fracture

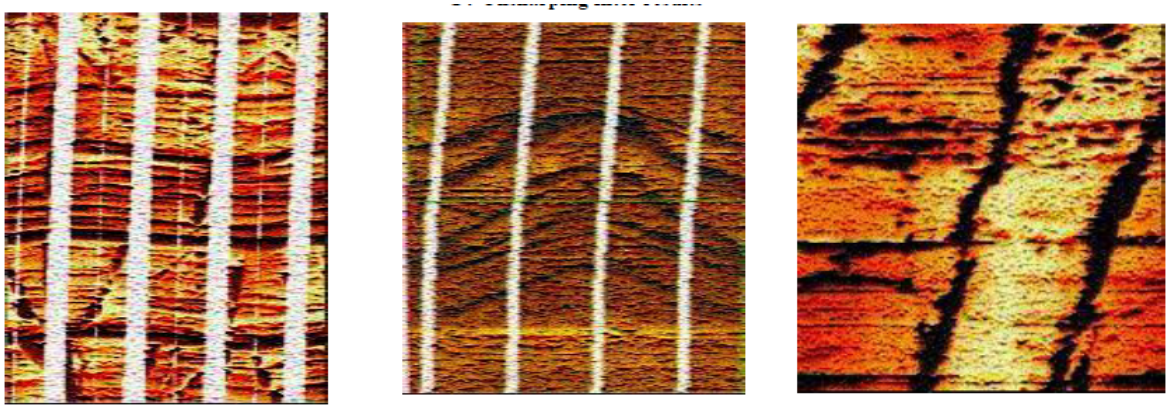


(a) A. Average filter results (mean)

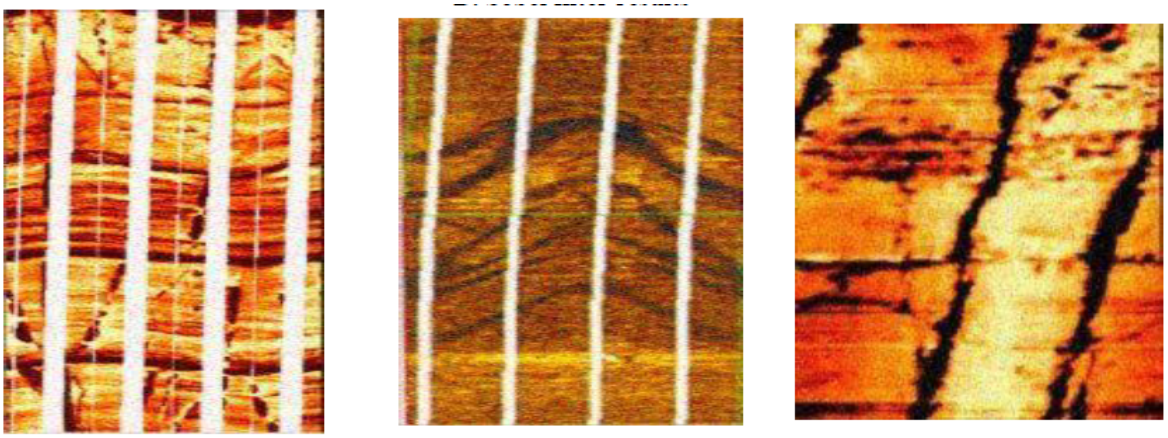


(b) B. Laplacian filter results

(c) C. Unsharping filter results



(d) D. Sobel filter results



(e) E. Gaussian filter results

Figure 2: The results obtained from implementing different filters on noisy well diagrams

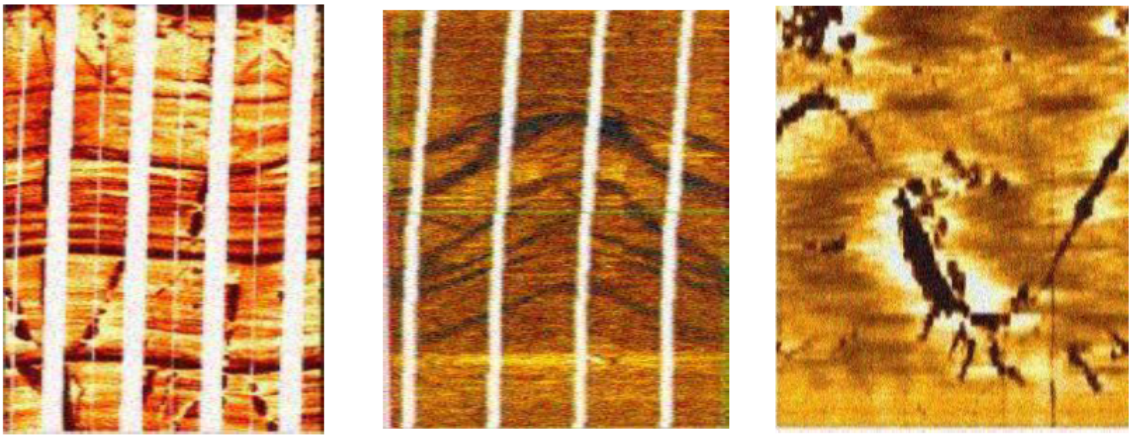


Figure 3: Sample visual well diagrams

Figure 4: Segmentation results with Canny edge detection method

and phenomena like that. Therefore, from subjective and visual interpretation point of view this is not an appropriate method to classify the image logs. Figure 5 shows the morphology method segmentation results.

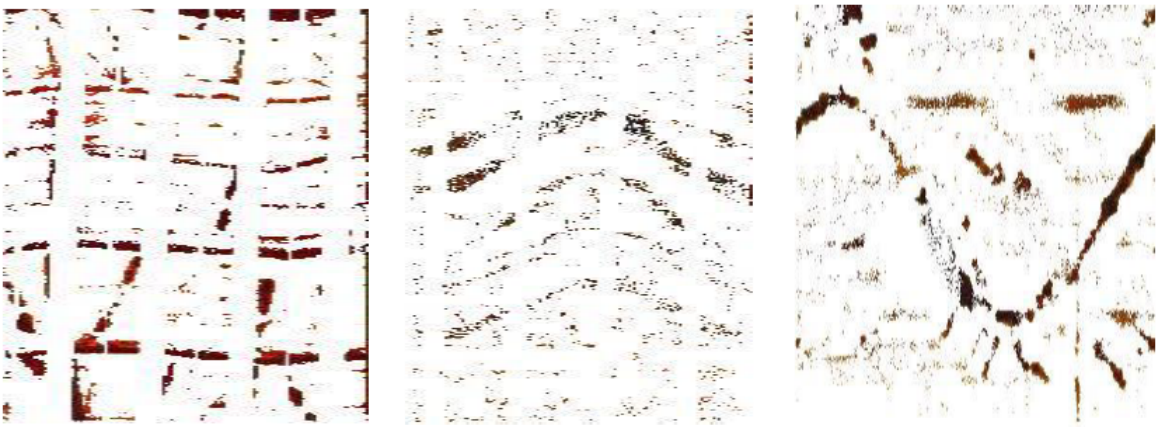


Figure 5: Segmentation results with morphology method

Examining the fig.5 shows that morphology method has better results for segmentation than Canny edge detector method, but still there is many small and additional sectors in image and the sectors belong to the fractures are damaged. Figure 6 shows the segmentation method Results by using thresholding.

Figure 6 shows the better results compared to the two previous segmentation method subjectively and visually, but by investigating more accurately the images of isolated sectors have notch and holes inside each component that indicates the accuracy of its classification and error is insufficient. Figure 7 shows the results of K-means approach to classify sample image logs.

Examining the results of segmentation method using K-means algorithm shows that this method indicates too better results compared to previous methods. Because the resulting images have very little

Figure 6: Segmentation results with thresholding method

Figure 7: Segmentation results with K-mean method

cuts and holes in the resulting images sections and additional and small images have been removed to a large extent. Figure 8 shows the results of segmentation method using the SOM algorithm.

Examining the Figure 8 shows subjectively and visually that this method doesn't have the issues of segmentation of first three investigated method and its quality is very good. The visual examination shows that SOM and K-means algorithm have no close and similar results. These two methods are the most optimal segmentation method for image logs. To determine the quantitative evaluation of studied segmentation methods the holes extent criteria and cuts existed in images obtained and the lack of additional small components in images as two quantitative parameter was examined. Table 2 shows the results sample images according to the holes and cuts extent in the isolated sections. Whatever holes and cuts extent are lower in the isolated sectors and coherence of isolated components are more the accuracy of method is better and also the images that contains no additional and small components their accuracy is more. Whatever the value of the quantity be more the accuracy of the studied approach is more.

Figure 8: Segmentation results with SOM method

Table 2: Classification methods evaluation

Method	Figure 1	Figure 2	Figure 3	Figure 4
Canny edge detection method	3537/5	4611/3	4668/3	4962/3
Morphology method	1739/3	148/16	5793/15	9087/7
Thresholding method	2391/10	7459/13	7485/17	1948/16
K-means method	3046/13	1016/21	623/28	3566/18
SOM method	3068/18	1845/24	975/33	1529/24

Examining the Table 2 shows that K-means and SOM methods from the point of view of integration sectors isolated and lack of cut and holes in them and also lack of additional fine components their evaluation criteria is better compared to other methods, therefore, the best methods for aspect of quantitative to segmentation of image logs are K-means and SOM and K-means.

4 Conclusion

In this article the quality of image logs as a powerful and important tool in the field of oil to identify and interpret the Geological phenomena using a variety of filters were tested. To evaluate the results of noise removal both qualitative and visual and by a quantitative measure that doesn't rely on the reference image without reference was evaluated. The results of this accuracy parameter Q were consistent with the visual results and Gaussian low-pass filter was most optimal to remove noise in image logs (wells wall virtual image). Then in the next section segmentation of image logs as one of the most important steps in identifying fractures in the oil wells was evaluated. Several algorithms were implemented and a proper criteria to evaluate the results in this area was introduced. The results showed that the color-based method and clustering concept such as SOM and K-means are the best methods for segmentation the images.

References

[1] S. S. Al-Amri, N. V. Kalyankar, S. Khamitkar, "Image segmentation by using threshold techniques", *Journal of Computing*, vol. 2, no. 5, pp. 83–86, 2010.

[2] S. Assous, P. Elkington, S. Clark, J. Whetton, "Automated detection of planar geological features in borehole images", *Geophysics*, vol. 79, no. 1, pp. 11–19, 2014.

[3] R. Gonzalez, R. E. Wood, *Digital Image Processing*, Third Edition, Prentice Hall, 2008.

[4] C. He, W. Wang, "A PCNN-based edge detection algorithm for rock fracture images", *IEEE Symposium on Photonics and Optoelectronic (SOPO'10)*, pp. 1–4, 2010.

[5] J. Kherroubi, "Automatic extraction of natural fracture traces from borehole images", *19th IEEE International Conference on Pattern Recognition (ICPR'08)*, pp. 1–4, 2008.

[6] F. Khoshbakht, *Application of Borehole Image Logs in Fracture Study in One of Oil Field of South West Iran*, Thesis of Msc, Tehran University, Tehran, 2006.

[7] R. L. Liu, Y. Q. Wu, J. H. Liu, Y. Ma, "The segmentation of FMI image based on 2-D dyadic wavelet transform", *Applied Geophysics*, vol. 2, no. 2, pp. 89–93, 2005.

[8] R. Maini, H. Aggarwal, "Study and comparison of various image edge detection techniques", *International Journal of Image Processing*, vol. 3, no. 1, pp. 1–12, 2011.

[9] S. Prensky, *Advances in Borehole Imaging Technology and Applications*, London: Geological Society, Special Publications, 1999.

[10] J. S. Sengar, N. Sharma, "Review: Competitive learning algorithm of neural network", *IJCTA*, vol. 2, no. 1, pp. 1480–1485, 2011.

[11] W. Wang, "An edge based segmentation algorithm for rock fracture tracing", in *IEEE International Conference on Computer Graphics, Imaging and Vision*, pp. 43–48, 2005.

[12] W. Wang, H. Liao, Y. Huang, "Rock fracture tracing based on image processing and SVM", in *Third IEEE International Conference on Natural Computation (ICNC'07)*, pp. 632–635, 2007.

[13] W. Wang, X. Wang, "Micro rock fracture image acquisition and processing", in *2nd IEEE International Workshop on Intelligent Systems and Applications (ISA'10)*, pp. 1–4, 2010.

[14] Z. Wang, A. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures", *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[15] H. Wu, C. Texaco, "Image enhancement and 3D visualization of borehole image logs in characterization of fracture in Kuwait", in *11th Abu Dhabi International Petroleum Exhibitation and Conference*, pp. 1–5, 2004.

[16] H. Zhang, J. E. Fritts, S. A. Goldman, "Image segmentation evaluation:a survey of unsupervised methods", *Computer Vision and Image Understanding*, vol. 110, no. 1, pp. 260–280, 2008.

[17] X. Zhu, P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content", *IEEE Transactions on Image Processing*, vol. 19, no. 12. pp. 3116–3132, 2010.

[18] T. Zuva, O. O. Olugbara, S. O. Ojo, S. M. Ngwira, "Image segmentation, available techniques, developments and open issues", *Canadian Journal on Image Processing and Computer Vision*, vol. 3, no. 4, pp. 20–29, 2011.

Biography

Milad Karami is received the Master. Degree in computer software engineering from Islamic Azad University, Bushehr, Iran. He currently is a lecture in SAMA College and his research interests include Image Processing, computer vision and Pattern Recognition.

Ahmad Keshavarz received the B.S. degree from Shiraz University in 2001 at shiraz, Iran. He received M.S. and Ph.D. degree from Tarbiat Modares University (TMU) in 2004 and 2008 at Tehran,

Iran, respectively. He joined the electrical engineering faculty of the Persian Gulf University (PGU) at Bushehr, Iran in 2008. His research interests include Remote Sensing Image Processing, Medical Image Processing, computer vision and statistical pattern recognition.

Hamed Jelodar is received the Master Degree in computer software engineering from Islamic Azad University, Bushehr,Iran .He currently is a PHD Candidate in the School of Computer Science , Nanjing University of Science and technology, Nanjing, China.

Analysis of One Certificateless Encryption for Secure Data Sharing in Public Clouds

Lihua Liu[1], Wenping Kong[1], Zhengjun Cao[2], Jinbo Wang[3]
(Corresponding author: Zhengjun Cao)

Department of Mathematics, Shanghai Maritime University, Shanghai 201306, China[1]
Department of Mathematics, Shanghai University, Shanghai 200444, China[2]
(Email: caozhj@shu.edu.cn)
Science and Technology on Communication Security Laboratory, Chengdu 610041, China[3]
(Received Mar. 10, 2017; revised and accepted Apr. 28, 2017)

Abstract

Certificateless public-key encryption (CL-PKE) is a useful primitive which removes the need for authenticating the invoked public key for the intended receiver. So far, most CL-PKE schemes were based on bilinear pairings. In order to reduce the heavy pairing computations, Seo et al. [IEEE TKDE, 2014, 2107-2119] proposed a CL-PKE scheme without pairing operations. In this note, we show that Seo et al.'s certificateless encryption is insecure because it binds only a user's partial public key with the user's identity. An adversary can replace the remaining public key of the user to launch man-in-middle attacks. We also present an improvement using the technique developed by Schnorr.

Keywords: Certificateless Cryptography; Cloud Computing; Man-in-middle Attack

1 Introduction

In a traditional public-key encryption, the sender has to authenticate that the invoked public key is the legitimate public key for the intended receiver. The verification relies on a public key infrastructure (PKI) which vouches for the connection between the intended receiver's identity and a particular public key. Identity-based encryption removes the need for a PKI, replacing it with the need for a key generation centre (KGC) that computes all users' private keys. In 2003, Al-Riyami and Paterson [1] proposed a new primitive, certificateless public-key encryption (CL-PKE), which avoids the drawbacks of both traditional public-key encryption and identity-based encryption.

Most CL-PKE schemes were based on pairing computations [2, 3, 4]. In 2010, Chevallier-Mames et al. [9] explored the problem of secure delegation of elliptic-curve pairing. In 2013, Liao and Hsiao [11], proposed a novel multi-server remote user authentication scheme using self-certified public keys for mobile clients. Other algorithms for outsourcing of bilinear pairings were discussed in [5, 6, 8, 14]. In 2016, Hsien et al. [10] presented a survey of public auditing for secure data storage in cloud computing. Liu et al. [12] discussed the problem of public auditing for shared data storage with user revocation. The works [7, 13] have pointed out some problems of user authentication and access control in cloud computing scenario.

In 2014, Seo et al. [16] proposed a CL-PKE scheme without pairing operations. The scheme considers two types of adversaries: Type I adversary is a third party attacker who does not know the master key

but can replace public keys of users; Type II adversary is a malicious KGC who has the master key but is unable to replace public keys of users. In this note, we show that Seo et al.'s encryption is insecure because it binds only a user's partial public key with the user's ID. An adversary (Type I) can replace the remaining public key of the user to launch man-in-middle attacks. We also present an improvement using the technique developed by Schnorr [15].

2 Review of Seo et al.'s Encryption

The scheme [16] involves five entities: the data owner, users, the security mediator (SEM), the key generation center (KGC) and the storage service. The SEM, KGC, and the storage service are assumed to be semi-trusted. Namely, they are trusted for executing the protocols correctly but are not trusted for the confidentiality of the data and the keys. Neither the KGC nor the SEM can decrypt the encrypted data for specific users. The scheme can be briefly described as follows.

Setup. KGC generates the system parameters $(p, q, n, k_0, g, y, H_1, H_2, H_3, H_4, H_5, H_6)$ and sets the master key as x. We refer to the original [16] for details.

SetPrivateKey. The user A picks $z_A \in \mathbb{Z}_q^*$ and sets it as the private key.

SetPublicKey. The user A computes $U_A = g^{z_A}$.

SEM-KeyExtract. For the identity ID_A, KGC picks $s_0, s_1 \in \mathbb{Z}_q^*$ and computes

$$w_0 = g^{s_0}, w_1 = g^{s_1}, d_0 = s_0 + xH_1(ID_A, w_0), d_1 = s_1 + xH_2(ID_A, w_0, w_1).$$

KGC sets d_0 as the SEM-key for A. After A proves the knowledge of the secret value z_A such that $U_A = g^{z_A}$, KGC sets (U_A, w_0, w_1, d_1) as the A's public keys.

Encrypt. To encrypt M for ID_A and (U_A, w_0, w_1, d_1), it proceeds as follows:

1) Check whether $g^{d_1} = w_1 \cdot y^{H_2(ID_A, w_0, w_1)}$.
2) Compute $r = H_3(M, \sigma, ID_A, U_A)$ for $\sigma \in \{0,1\}^{k_0}$.
3) Compute C_1, C_2 and C_3 as follows:

$$\begin{aligned} C_1 &= g^r, \\ C_2 &= (M \| \sigma) \oplus H_4(U_A^r) \oplus H_5(w_0^r \cdot y^{H_1(ID_A, w_0) \cdot r}), \\ C_3 &= H_6(U_A, (M \| \sigma) \oplus H_4(U_A^r), C_1, C_2). \end{aligned}$$

Output the ciphertext (C_1, C_2, C_3).

SEM-Decrypt. Given (C_1, C_2, C_3), (U_A, w_0, w_1, d_1) and ID_A, SEM proceeds as follows:

1) Check that ID_A is a legitimate user.
2) Check whether $C_3 = H_6(U_A, C_2 \oplus H_5(C_1^{d_0}), C_1, C_2)$.
3) Send C_1 and $C_2' = C_2 \oplus H_5(C_1^{d_0})$ to A.

User-Decrypt. Given (C_1, C_2'), A proceeds as follows:

1) Compute $M \| \sigma = H_4(C_1^{z_A}) \oplus C_2', \quad r = H_3(M, \sigma, ID_A, U_A)$.
2) Check whether $g^r = C_1$.
3) Return the decrypted message M.

3 Cryptanalysis of Seo et al.'s Encryption

In Seo et al.'s certificateless encryption scheme the A's public key is set as (U_A, w_0, w_1, d_1). Clearly, the encrypter cannot authenticate that (U_A, w_0, w_1, d_1) is the legitimate public key for the intended receiver A. Thus, one has to specify a mechanism to bind these parameters to ID_A. Since

$$d_1 = s_1 + xH_2(ID_A, w_0, w_1), \tag{1}$$

both w_0, w_1 are bound to ID_A. However, we find *the parameter U_A is not bound to ID_A at all.* Therefore, an adversary (Type I) can launch the man-in-middle attack by replacing U_A with $\widehat{U} := g^{\widehat{z}}$ where $\widehat{z}$ is selected by the adversary. See the following description for the attack.

Encrypt (invoking the forged parameter). To encrypt M for ID_A and $(\widehat{U}, w_0, w_1, d_1)$, it proceeds as follows:

1) Check whether $g^{d_1} = w_1 \cdot y^{H_2(ID_A, w_0, w_1)}$.
   ♢ *It passes because U_A is not involved in the process.*
2) Compute $r = H_3(M, \sigma, ID_A, \widehat{U})$ for $\sigma \in \{0,1\}^{k_0}$.
3) Compute C_1, C_2 and C_3 as follows:

$$\begin{aligned} C_1 &= g^r, \\ C_2 &= (M\|\sigma) \oplus H_4(\widehat{U}^r) \oplus H_5(w_0^r \cdot y^{H_1(ID_A, w_0)\cdot r}), \\ C_3 &= H_6(\widehat{U}, (M\|\sigma) \oplus H_4(\widehat{U}^r), C_1, C_2). \end{aligned}$$

   Output the ciphertext (C_1, C_2, C_3).

SEM-Decrypt. (invoking the forged parameter). Given ID_A, (C_1, C_2, C_3) and $(\widehat{U}, w_0, w_1, d_1)$, SEM proceeds as follows:

1) Check that ID_A is a legitimate user.
2) Check whether $C_3 = H_6(\widehat{U}, C_2 \oplus H_5(C_1^{d_0}), C_1, C_2)$.
   ♢ *It passes because*

$$\begin{aligned} & C_2 \oplus H_5(C_1^{d_0}) \\ = \; & C_2 \oplus H_5(g^{rd_0}) \\ = \; & C_2 \oplus H_5(g^{r(s_0 + xH_1(ID_A, w_0))}) \\ = \; & C_2 \oplus H_5(g^{rs_0} g^{rxH_1(ID_A, w_0)}) \\ = \; & C_2 \oplus H_5(w_0^r \cdot y^{r \cdot H_1(ID_A, w_0)}) \\ = \; & (M\|\sigma) \oplus H_4(\widehat{U}^r) \end{aligned}$$

3) Send C_1 and $C_2' = C_2 \oplus H_5(C_1^{d_0})$ to A.

Adversary-Decrypt. Intercepting (C_1, C_2') from the SEM, the adversary computes

$$\begin{aligned} M\|\sigma &= H_4(C_1^{\widehat{z}}) \oplus C_2', \\ r &= H_3(M, \sigma, ID_A, \widehat{U}). \end{aligned}$$

If $g^r = C_1$, then output M, σ.

♢ *It succeeds because* $H_4(C_1^{\widehat{z}}) \oplus C_2' = H_4(C_1^{\widehat{z}}) \oplus C_2 \oplus H_5(C_1^{d_0}) = H_4(C_1^{\widehat{z}}) \oplus (M\|\sigma) \oplus H_4(\widehat{U}^r) = M\|\sigma$.

Adversary-Encrypt. Given M, σ, ID_A and the legitimate parameters (U_A, w_0, w_1, d_1), the adversary computes

$$\begin{aligned} \widehat{r} &= H_3(M, \sigma, ID_A, U_A), \\ \widehat{C}_1 &= g^{\widehat{r}}, \\ \widehat{C}_2 &= (M\|\sigma) \oplus H_4(U_A^{\widehat{r}}). \end{aligned}$$

Send $(\widehat{C}_1, \widehat{C}_2)$ to the intended receiver A.

User-Decrypt. Receiving $(\widehat{C}_1, \widehat{C}_2)$, A proceeds as follows:

1) Compute $M\|\sigma = H_4(\widehat{C}_1^{z_A}) \oplus \widehat{C}_2$ and $\widehat{r} = H_3(M, \sigma, ID_A, U_A)$.
2) Check whether $g^{\widehat{r}} = \widehat{C}_1$.
3) Return the decrypted message M.

   ♢ *It succeeds because* $H_4(\widehat{C}_1^{z_A}) \oplus \widehat{C}_2 = H_4(g^{\widehat{r}z_A}) \oplus (M\|\sigma) \oplus H_4(U_A^{\widehat{r}}) = M\|\sigma$.

4 A Possible Revising Method

As mentioned before, the parameter U_A is not bound to the identity ID_A at all. In order to bind them together, it only needs to set

$$d_1 = s_1 + xH_2(ID_A, U_A, w_0, w_1) \tag{2}$$

instead of the original $s_1 + xH_2(ID_A, w_0, w_1)$. In such case, when a sender wants to encrypt a message for the identity ID_A, the sender first checks whether

$$g^{d_1} = w_1 \cdot y^{H_2(ID_A, U_A, w_0, w_1)} \tag{3}$$

With this procedure, the sender can verify that the invoked parameters U_A, w_0, w_1, d_1 are truly bound to ID_A.

Note that the intractability in Equation (3) is based on that the discrete logarithm of y with respect to g is intractable and that the output of the hash function H_2 is always assumed to be random. The technique was developed by Schnorr [15] and broadly adopted by many cryptographic schemes.

5 Conclusion

We show that Seo et al.'s certificateless encryption is flawed, and put forth a revising method by using a public hash function to bind together all public keys of the intended receiver. We would like to stress that man-in-middle attacks should be carefully evaluated when one designs a certificateless encryption scheme.

Acknowledgements

We thank the National Natural Science Foundation of China (61303200, 61411146001). We are grateful to the reviewers for their valuable suggestions.

References

[1] S. Al-Riyami and K. Paterson, "Certificateless public key cryptography," in *Proceedings of Advances in Cryptology - ASIACRYPT 2003*, pp. 452–473, Taipei, Taiwan, Dec. 2003.

[2] D. Boneh and X. Boyen, "Short signatures without random oracles," in *Proceedings of Advances in Cryptology - EUROCRYPT 2004*, pp. 56–73, Interlaken, Switzerland, May 2004.

[3] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in *Proceedings of Advances in Cryptology - CRYPTO 2001*, pp. 213–229, Santa Barbara, California, USA, Aug. 2001.

[4] D. Boneh, G. Lynn, and H. Shacham, "Short signature from the weil pairing," in *Proceedings of Advances in Cryptology - ASIACRYPT 2001*, pp. 514–532, Gold Coast, Australia, Dec. 2001.

[5] S. Canard, J. Devigne, and O. Sanders, "Delegating a pairing can be both secure and efficient," in *Proceedings of Applied Cryptography and Network Security - ACNS 2014*, pp. 549–565, Lausanne, Switzerland, June 2014.

[6] Z. J. Cao, L. H. Liu, and O. Markowitch, "On two kinds of flaws in some server-aided verification schemes," *International Journal of Network Security*, vol. 18, no. 6, pp. 1054–1059, 2016.

[7] T. Y. Chen, C. C. Lee, M. S. Hwang, and J. K. Jan, "Towards secure and efficient user authentication scheme using smart card for multi-server environments," *Journal of Supercomputing*, vol. 66, no. 2, pp. 1008–1032, 2013.

[8] X. F. Chen et al., "Efficient algorithms for secure outsourcing of bilinear pairings," *Theoretical Computer Science*, no. 562, pp. 112–121, 2015.

[9] B. Chevallier-Mames et al., "Secure delegation of elliptic-curve pairing," in *Proceedings of Smart Card Research and Advanced Application, 9th IFIP WG 8.8/11.2 International Conference - CARDIS 2010*, pp. 24–35, Passau, Germany, April 2010.

[10] W. F. Hsien, C. C. Yang, and M. S. Hwang, "A survey of public auditing for secure data storage in cloud computing," *International Journal of Network Security*, vol. 18, no. 1, pp. 133–142, 2016.

[11] Y. P. Liao and C. M. Hsiao, "A novel multi-server remote user authentication scheme using self-certified public keys for mobile clients," *Future Generation Computer Systems*, vol. 29, pp. 886–900, 2013.

[12] C. W. Liu, W. F. Hsien, C. C. Yang, and M. S. Hwang, "A survey of public auditing for shared data storage with user revocation in cloud computing," *International Journal of Network Security*, vol. 18, no. 4, pp. 650–666, 2016.

[13] C. W. Liu, W. F. Hsien, C. C. Yang, and M. S. Hwang, "A survey of attribute-based access control with user revocation in cloud data storage," *International Journal of Network Security*, vol. 18, no. 5, pp. 900–916, 2016.

[14] L. H. Liu and Z. J. Cao, "A note on 'efficient algorithms for secure outsourcing of bilinear pairings'," *International Journal of of Electronics and Information Engineering*, vol. 5, no. 1, pp. 30–36, 2016.

[15] C. Schnorr, "Efficient signature generation for smart cards," in *Proceedings of Advances in Cryptology - CRYPTO 1989*, pp. 239–252, Santa Barbara, California, USA, Aug. 1989.

[16] S. H. Seo, M. Nabeel, X. Y. Ding, and E. Bertino, "An efficient certificateless encryption for secure data sharing in public clouds," *IEEE Transactions on Knowledge Data Engineerings*, vol. 26, no. 9, pp. 2107–2119, 2014.

Biography

Lihua Liu is an associate professor with Department of Mathematics at Shanghai Maritime University. She received her Ph.D. degree in applied mathematics from Shanghai Jiao Tong University. Her research

interests include combinatorics, cryptography and information security.

Wenping Kong is currently pursuing her M.S. degree from Department of Mathematics, Shanghai Maritime university. Her research interests include combinatorics and cryptography.

Zhengjun Cao is an associate professor with Department of Mathematics at Shanghai University. He received his Ph.D. degree in applied mathematics from Academy of Mathematics and Systems Science, Chinese Academy of Sciences. He served as a post-doctor in Computer Sciences Department, Université Libre de Bruxelles, from 2008 to 2010. His research interests include cryptography, discrete logarithms and quantum computation.

Jinbo Wang received his Ph.D. degree in applied mathematics from Shanghai University. His research interests include applied cryptography and network security.

Mobile Malware and Defending Systems: Comparison Study

Abdullah A. Al-khatib, Waleed A. Hammood
(Corresponding author: Abdullah A. Al-khatib)

Research Centre for Software Technology and Management (SOFTAM)
Faculty of Information Science & Technology (FTSM), University Kebangsaan Malaysia
43600 UKM, Bangi Selangor, Malaysia
(Email: khteb2003@gmail.com)

(Received Apr. 12, 2016; revised and accepted Aug. 10 & Nov. 11, 2016)

Abstract

Due to the enormous growth of the new mobile technology, the smart phones provides news functionality to users. The most impressive features that the user have the ability to access the online store and install various kind of apps. However, there are some application are malicious used by malware writes to lunch different kind of attacks to users information. Therefore, this study focus on summarizing different kinds of mobile malware by highlight the strength and weakness of different category of malware. In addition, summarize the recent proposed frameworks to defend smart phone from malware attacks.

Keywords: Malicious; Mobile Malware; Smart Phone

1 Introduction

The massive growth and development of technologies, especially in networking, have totally changed the way of communication. Moreover, creating, storing, sharing and managing information has become easy and fast. Like the other side of the coin, these advancements have unfortunately opened the doors of information vulnerability.

Nowadays, enemies have started employing network oriented warfare in mobile malware applications. It is evident that information infrastructures have pathetically become targets of malware writers. The hacker attacks have become more and more aggressive, and mobile users have been victimized and destructed [5].

Nowadays, the mobile technology is an interconnected set of information sources that enables the user to interact with the market place to access the online market, banking and transactions. With the existing of the malware writers that are strive to enrich themselves to spy on the user information there is a high demands to protect user information and protect the phone systems [13]

2 Related Research

Smartphones witness a massive growth currently in our society. Due to increase number of smart phones users they became a target and victim to malware threads. This study have been categorized

into two main parts first started with the history of mobile malware evolution and list the malware categorizes. In addition, list the effect of these malware and the possible damages caused on the smart phones. The second part of this study, aim to widely investigate the behavior of smart phone malware propagation [9].

Basically, the smart phones integrate the communications capability of cell phones with PDAs (personal digital Assistant) [8]. These kind of technology provide the ability to the user to access different type of ubiquitous services which is including using emails, share online information, and download online apps. In spite of the usage of such service increases the vulnerability to malware attacks, the smart phones still suffers from a proper way to protect the information from the malware writers [12].

2.1 Mobile Malware

Mobile malware have a variety types that can be able to affect and propagate to harm the victims, these malware normally it's attached with files or programs. In addition, some of malware can be installed on the devices with the user action by clicking on some multimedia messages or downloading not trusted application [14].

- Malware Classes: the purpose behind using the malware is to case damaging to the system using the virus, worm, Trojan, spyware, backdoor etc.
- Infection Vectors: there are plenty ways for infection vectors to deliver the malicious to smart phones or to the targeted system such as Sms/mms, Internet Access, Bluetooth, and USB files.
- Risk of Malware: After the system compromised by the malware, the damages might be on the services to the users, affecting the system performance, economic loss, spy on information.
- Malware on Mobile Platforms: The majority of the modern mobile devices uses different kind of OSs, such as iOS, Android, J2ME, Windows Mobile, and Symbian [10]. Figure 1 shows the mobile risks by type from January to June in 2012 [10].

2.2 Problems of Current Models and Future Trends

In this section, the limitation of the smart phone malware propagation model is highlighted, and mentions the possible future works in this filed [4].

1) Diversity of propagation models. The current propagation of malware models were designed respectively to each malware. Such as WPM model is based on a Bluetooth worm.
2) Difficulty comparing performance among different propagation models. Generally, the comparing between two propagation models consider a hard task due to the complexity of the simulation process which is based on mathematical models.
3) Modelling based on partial information: it is obvious that current model have some limitation to consider all the possible inputs for the model.

Future Trends: from the previous issues, investigation and exploring more solution will provide a promising result in this area.

According to [7] the analysis of prevent these kind of attacks and identify the detection techniques. There are several reasons behind making the smart phones are targeted and vulnerable to security attack. Some of the reasons are:

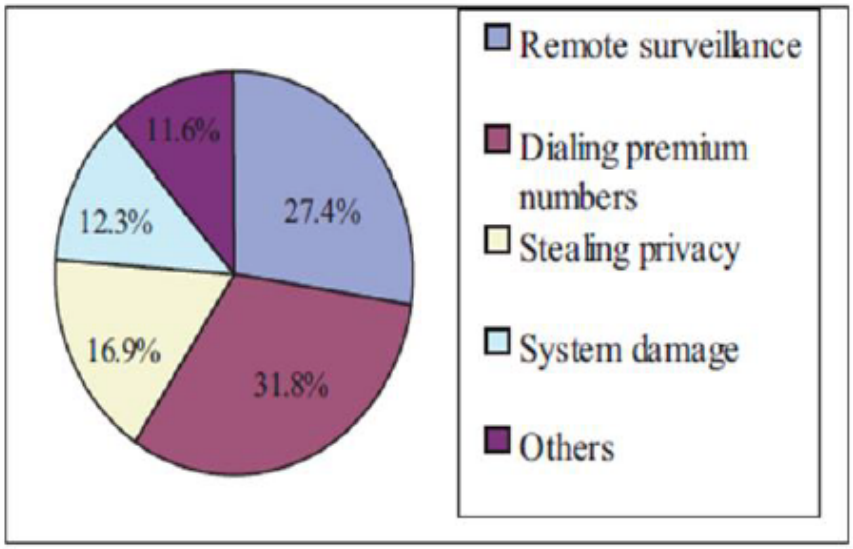


Figure 1: The mobile risks by type from January to June in 2012 [10]

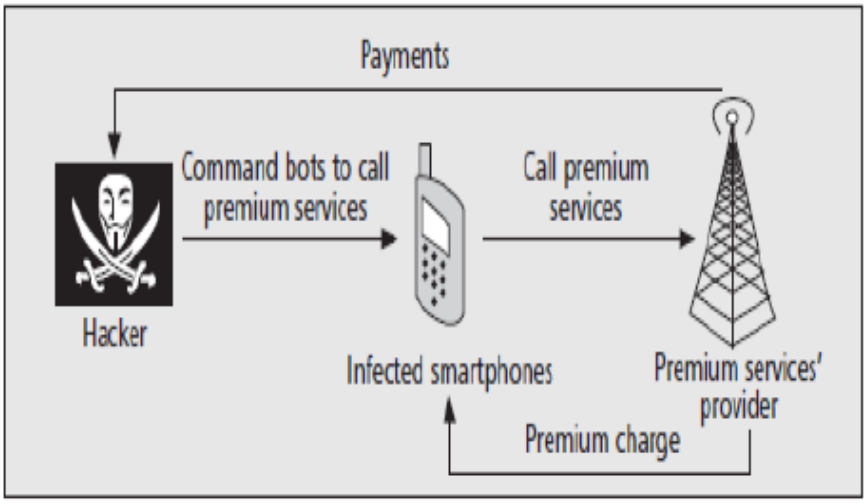


Figure 2: The dialler-ware attacks [7]

Personal data: The majority of users using smart phone to perform financial transaction and online shopping using there bank information and store these information into their phones [7]. Figure 2 shows the dialler-ware attacks [7].

Operating system: Nowadays the users attracted by android platform, the android is an open source which is make it easy to the hacker or malware writers to exploit that and create malware and use that.

User unawareness: The users are unware of the software security, the users looks into the new functions on the new app and totally ignore security measures.

There are different category for mobile malwares, the categorization are based on the propagation behavior, malicious attacks and remote control. Moreover, the propagation behavior is the way that this malware send to the victims. The malicious attacks is the way that the malware attack the user system such as attached files and the remote control behavior demonstrate the purpose of the malware to remote a server. Table 1 shows the different kinds of attacks by malware [7].

Table 1: Different types of attacks lunched by malware [7]

Attacks	Description
Phishing	Users' credentials such as account details and credit card numbers are collected by means of apps, emails, or SMS, which seem to be genuine.
Spyware	Users' activities on the smart phones are being monitored, which means personal information is extracted or inferred. Compared to surveillance attacks, spyware does not have specific targeted victims.
Surveillance attacks	A specific user is under surveillance by means of his/her infected smart phone, making use of the built-in sensors.
Diallerware attacks	Users' money is stolen using the malware that makes hidden calls to premium numbers or SMS services.
Financial malware attacks	Such attacks aim to steal users' credentials from the smart phones or perform man-in-the-middle attacks on financial applications.
Worm-based attacks	A worm is a malware that duplicates itself, typically propagating from one device to another, using different means through an existing network without users' intervention.
Botnets	A botnet is a set of zombie devices that are infected by malware so that a hacker can remotely control them.

2.3 Defence Methodology

In order to defend against malware, there are two level strategy followed to achieve that. The first level prevent the malware to reach into your phone. The second level is enable a detection to any malware and delete it [15]. Figure 3 shows the classification techniques for mobile malware detection techniques [5].

According to [5] this study is to investigate ability to design defence strategy from two perspectives the warfare kill zone and the airport check-in system named as a third line defence strategy. The purpose of that is to solve the issue of SMS-based malware in the Android-based Smartphones. The designed framework has the ability to use the security features of the android operating system in order to block and detect the malicious SMS [5].

2.4 The Third Line of Defence Strategy and Based on Three Main Aspects

Kill zone warfare analogy: this technique is called the ambush tactic which is force the enemy to enter into an area called "kill zone" then handle that enemy and restrict him from escape [2].

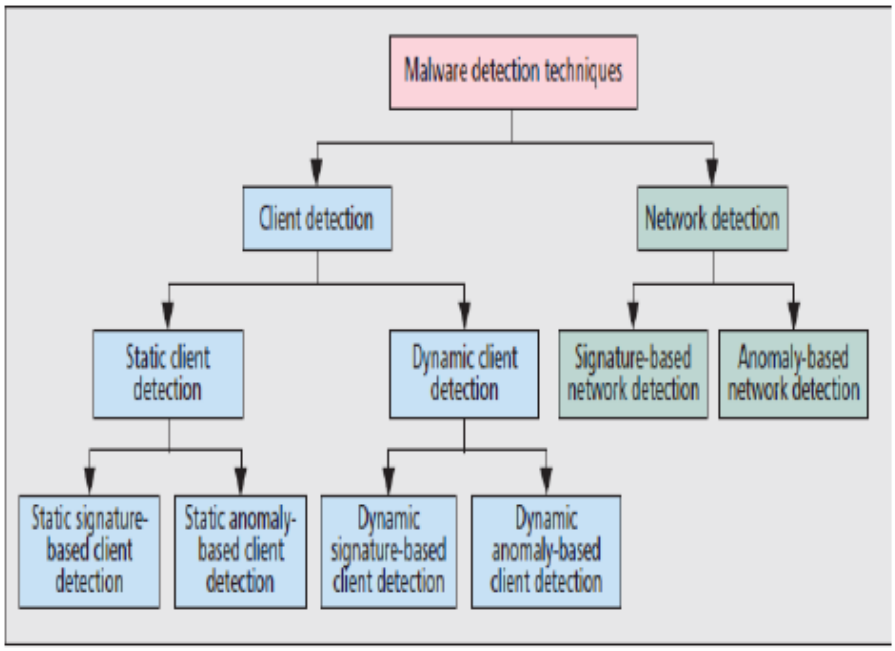


Figure 3: The classification techniques for mobile malware detection techniques [5]

2.5 From the Kill Zone to The Third Line of Defence

The design of the previous architecture is secured to restrict the sms-based malware. The defence strategy, the intrusion detection systems (IDSs) and firewalls are the first and the second line defence. While the third line is the implementation of SMS-bot concept, the main purpose is to intercepts the SMS before the SMS reaches any application on the system to prevent any damages [3]. Figure 4 shows the third line of defence mechanism of mobile phones [5].

According to [11] Since the mobile malware witness a dramatically increase to harm the smart phones and android devices, There are a highly demands to increase the protection of user privacy and prevent all kind of mobile malware malicious that infect the systems. This study aims to propose a cloud based malware detection approach in order to increase the mobile security [11].

2.6 Cloud-based Mobile Malware Detection Framework

The task of increase the mobile security to detect new thread consider of the most challenging tasks, due to lack of security solutions. Therefore, in this study could- based framework is proposed to detect and prevent mobile malware [6].

2.7 Futuristic Mobile Malware Security Strategies

Anomaly/Heuristics Based Detection: The anomaly mechanism is designed to detect and monitor the apps from any malicious activity that can harm the system, then it can alert the user that the phone is sending and receive data [1].

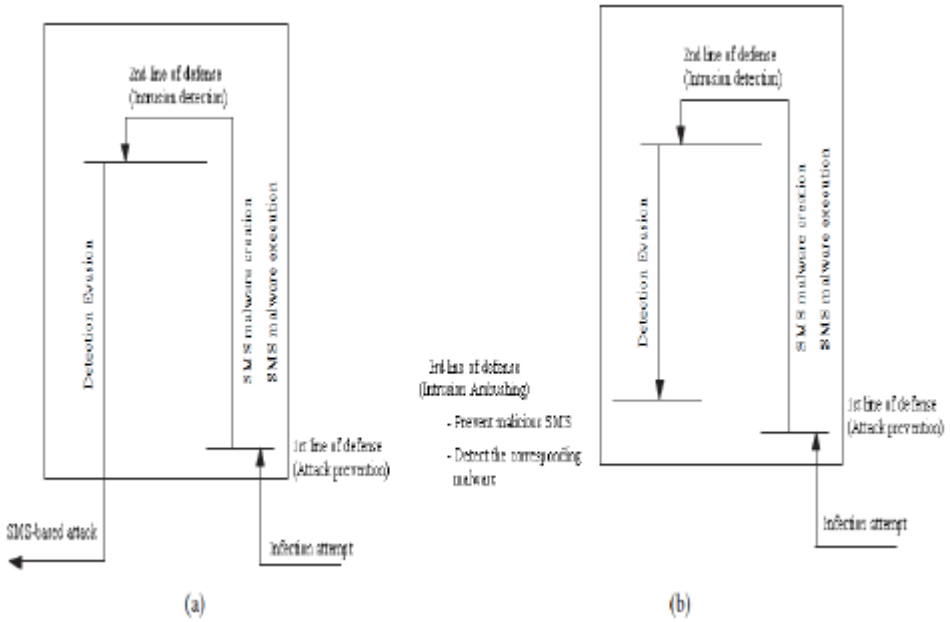


Figure 4: The third line of defence mechanism of mobile phones [5]

App Ranking System: Another way to enhance the detection methods is to use the ranking system, which is consider the users reviews, researchers review and analysed reviews the same way that google play follow to rank the application functionality [13].

Cloud-based Detection: The benefits of the cloud-based detection are a quick and effective for mobile security.

Cloud-based Detection: is an approach to enhance could service to detect the malware apps attempts to attack the systems. The details of the framework shown in Figure 5.

The most advantages of using a cloud based detection approach is that all the malicious activity performs outside the mobile device, after analyzing the app the approach will allow the mobile device to use the application. Therefore, this framework is a significant approach toward mobile security.

3 Conclusion

The rapid increase of smart phone users and the existing of huge amount of information makes them a wanted target for the hackers. The mobile malware obviously increases which is indicates that these is a lack of the defending systems on the smart phones.

The reason that encourages the malware writers to attack user can refer to that there is not a perfect methods to defend the smart phone system from malware attacks. In this study, we highlighted the reason that make the users and system vulnerable, and how the hacker can exploits these points to attack the users. In addition, a recent proposed framework is summarized to defend the user system form different kind of malware apps.

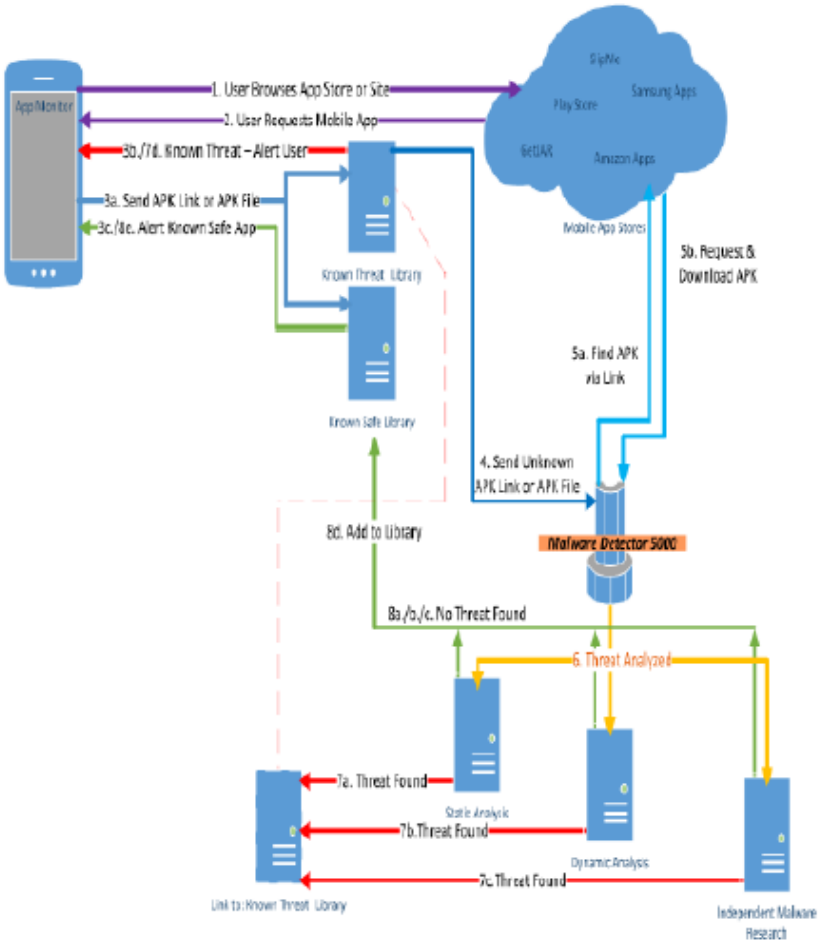


Figure 5: The framework for the cloud based mobile malware detection [13]

Table 2: Summary of latest related research

Authors	Method	Strength
Peng et al. 2014	Investigate the recent studies	Explore some weakness in recent studies
He et al. 2015	Mobile malware detection techniques	Prevent the malware to reach into your phone. The second level is enable a detection to any malware and delete it.
Derhab et al. 2014	The third line of defence strategy	The kill zone tactic used to kill the malware
Penning et al. 2014	Cloud based mobile malware detection	Detect all malicious activity outside the mobile device

References

[1] A. Al-Khatib, R. Hassan, "Performance evaluation of AODV, DSDV, and DSR routing protocols in MANET using NS-2 simulator," in *The 2nd International Conference of Reliable Information and Communication Technology (IRICT'17)*, 2017.

[2] M. Christodorescu and S. Jha, "Testing malware detectors," in *Proceedings of the ACM SIGSOFT International Symposium on Software Testing and Analysis*, 2004.

[3] M. Christodorescu, S. Jha, S. Seshia, D. Song, and R. Bryant, "Semantics-aware malware detection," in *IEEE Symposium on Security and Privacy*, pp. 32–46, 2005.

[4] G. Delac, M. Silic, J. Krolo, "Emerging security threats for mobile platforms," in *Proceedings of the 34th International Convention*, pp. 1468–1473, 2011.

[5] A. Derhab, K. Saleem, A. Youssef, "Third line of defense strategy to fight against SMS-based malware in android smart phones," in *IEEE International Conference on Wireless Communications and Mobile Computing (IWCMC'14)*, pp. 542–547, 2014.

[6] R. Hassan, A. Al-Khatib, W. Hussain, "A framework of universiti Kebangsaan malaysia patent: UKM patent," in *19th International Conference on Advanced Communication Technology (ICACT'17)*, pp. 232–236, 2017.

[7] D. He, S. Chan, M. Guizani, "Mobile application security: malware threats and defenses," *Wireless Communications*, vol. 22, no. 1, pp. 138–144, 2015.

[8] D. Maslennikov, *Mobile Malware Evolution: Part 6*, 2013. (`http://www.securelist.com/en/analysis/204792283/Mobile_Malware_Evolution_Part_6`)

[9] S. Peng, S. Yu, A. Yang, "Smartphone malware and its propagation modeling: A survey," *Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 925–941, 2014.

[10] S. C. Peng, "A survey on malware containment models in smart phones," *Applied Mechanics and Materials*, vol. 263, pp. 3005–3011, 2012.

[11] N. Penning, M. Hoffman, J. Nikolai, Y. Wang, "Mobile malware security challenges and cloud-based detection," in *International Conference on Collaboration Technologies and Systems (CTS'14)*, pp. 181–188, 2014.

[12] M. Polla, F. Martinelli, D. Sgandurra, "A survey on security for mobile devices," *Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 446–471, 2013.

[13] V. Rastogi, Y. Chen, X. Jiang, "Catch me if you can: Evaluating android anti-malware against transformation attacks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 1, pp. 99–108, 2014.

[14] S. H. Seo, A. Gupta, A. M. Sallam, E. Bertino, K. Yim, "Detecting mobile malware threats to homeland security through static analysis," *Journal of Network and Computer Applications*, vol. 38, pp. 43–53, 2014.

[15] Y. Wang, K. Streff, S. Raman, "Smartphone security challenges," *Computers*, vol. 45, no. 12, pp. 52–58, 2012.

Biography

Abdullah Abdulrahman graduated from Al-Ahgaff College Hathramout-Yemen in 2009. He works in Communtiy College from 2009. He enrolled to study Master in University Kebangsaan Malaysia (UKM) in Computer Sceince (Network Technology) in 2015. His research interests are in Software Defined Network (SDN) and IP Security (IPSec).

Waleed A. Hammood was born in 1992. He got bachelor from Baghdad University, he got his master from University Kebangsaan Malaysia (UKM) in Computer Sceince (Network Technology) in 2017, He enrolled to study PhD in University Malaysia Pahang (UMP).

Guide for Authors
International Journal of Electronics and Information Engineering

International Journal of Electronics and Information Engineering (IJEIE) will be committed to the timely publication of very high-quality, peer-reviewed, original papers that advance the state-of-the art and applications of Electronics and Information Engineering. Topics will include, but not be limited to, the following: Algorithms, Bioinformatics, Computation Theory, AI, Fuzzy Systems, Embedded Systems, VLSI/EDA, Computer Networks, Information Security, Database, Data Mining, and Information Retrieval, Digital Content, Digital Life, and Human Computer Interaction, Image Processing, Computer Graphics, and Multimedia Technologies, Information Literacy, e-Learning, and Social Media, Mobile Computing, Wireless Communications, and Vehicular Technologies, Parallel, Peer-to-peer, Distributed, and Cloud Computing, Semiconductor, Software Engineering and Programming Languages, Telecommunication, etc.

1. Submission Procedure

Authors are strongly encouraged to submit their papers electronically by using online manuscript submission at http://ijeie.jalaxy.com.tw/.

2. General

Articles must be written in good English. Submission of an article implies that the work described has not been published previously, that it is not under consideration for publication elsewhere. It will not be published elsewhere in the same form, in English or in any other language, without the written consent of the Publisher.

2.1 Length Limitation:

All papers should be concisely written and be no longer than 30 double-spaced pages (12-point font, approximately 26 lines/page) including figures.

2.2 Title page

Title page should contain the article title, author(s) names and affiliations, address, an abstract not exceeding 100 words, and a list of three to five keywords.

2.3 Corresponding author

Clearly indicate who is willing to handle correspondence at all stages of refereeing and publication. Ensure that telephone and fax numbers (with country and area code) are provided in addition to the e-mail address and the complete postal address.

2.4 References

References should be listed alphabetically, in the same way as follows:
For a paper in a journal: M. S. Hwang, C. C. Chang, and K. F. Hwang, ``An ElGamal-like cryptosystem for enciphering large messages," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 2, pp. 445--446, 2002.

For a book: Dorothy E. R. Denning, *Cryptography and Data Security*. Massachusetts: Addison-Wesley, 1982.

For a paper in a proceeding: M. S. Hwang, C. C. Lee, and Y. L. Tang, ``Two simple batch verifying multiple digital signatures," in *The Third International Conference on Information and Communication Security (ICICS2001)*, pp. 13--16, Xian, China, 2001.

In text, references should be indicated by [number].

2.5 Author benefits

No page charge is made.

Subscription Information

Individual subscriptions to IJEIE are available at the annual rate of US$ 200.00 or NT 6,000 (Taiwan). The rate is US$1000.00 or NT 30,000 (Taiwan) for institutional subscriptions. Price includes surface postage, packing and handling charges worldwide. Please make your payment payable to "Jalaxy Technique Co., LTD." For detailed information, please refer to http://ijeie.jalaxy.com.tw or Email to ijeieoffice@gmail.com.