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

IJEIE

International Journal of Electronics and Information Engineering

Vol. 6, No. 1 (Mar. 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. 1 (Mar. 1, 2017)

The Encryption Algorithms GOST28147-89-PES16-2 and GOST28147-89-RFWKPES16-2

Tuychiev Gulom
National University of Uzbekistan, Republic of Uzbekistan, Tashkent
4 Universitet St, Tashkent 100174, Uzbekistan
(Email: blasterjon@gmail.com)
(Received Jan. 12, 2017; revised and accepted Feb. 20, 2017)

Abstract

In the paper created a block encryption algorithms GOST28147-89-PES16-2 and GOST28147-89-RFWKPES16-2 based on networks PES16-2 and RFWKPES16-2, with the use the round function of the encryption algorithm GOST 28147-89. The block length of created block encryption algorithm is 128 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 PES [6] and Lai-Massey scheme developed networks PES16-2 [20] and RFWKPES16-2 [22], consisting from two round function. In the networks PES16-2 and RFWKPES16-2, similarly as in the Feistel network, in encryption and decryption process using the same algorithm. In the networks used two round function having four input and output blocks and as the round function can use any transformation.

As the round function networks IDEA4-2 [1], RFWKIDEA4-2[11], PES4-2 [10], RFWKPES4-2 [21], PES8-4 [2], RFWKPES8-4 [9] using the round function of the encryption algorithm GOST 28147-89 created the encryption algorithm GOST28147-89-IDEA4-2 [18], GOST28147-89-RFWKIDEA4-2 [27], GOST28147-89-PES4-2 [26], GOST28147-89-RFWKPES4-2 [28], GOST28147-89-PES8-4 [33] and GOST 28147-89-RFWKPES8-4 [33].

In addition, by using SubBytes(), ShiftRows(), MixColumns() and AddRoundKey() transformations of the encryption algorithm AES [5] as round functions of networks IDEA8-1 [14], RFWKIDEA8-1 [14], PES8-1 [8], RFWKPES8-1 [9], IDEA16-1 [12], RFWK IDEA16-1 [16], PES16-1 [20], RFWKPES16-1 [22], IDEA32-1 [13], RFWKIDEA32-1 [38], PES32-1 [15], RFWKPES32-1 [17], IDEA16-2 [12], RFWK

IDEA16-2 [16], PES16-2 [20], RFWKPES16-2 [22], IDEA32-4 [13], RFWKIDEA32-4 [38], PES32-4 [15], RFWKPES32-4 [17] created encryption algorithms AES-IDEA8-1 [35], AES-RFWKIDEA8-1 [37], AES-PES8-1 [36], AES-RFWKPES8-1 [19], AES-IDEA16-1 [34], AES-RFWKIDEA16-1 [30], AES-PES16-1 [32], AES-RFWKPES16-1 [32], AES-IDEA32-1 [23], AES-RFWKIDEA32-1 [31], AES-PES32-1 [24], AES-RFWKPES32-1 [24], AES-IDEA16-2 [29], AES-RFWKIDEA16-2 [29], AES-PES16-2 [29], AES-RFWKPES16-2 [29], AES-IDEA32-4 [25], AES-RFWKIDEA32-4 [25], AES-PES32-4 [3], AES-RFWKPES32-4 [39].

In this paper, applying the round function of the encryption algorithm GOST 28147-89 as round functions of the networks PES16-2 and RFWKPES16-2, developed new encryption algorithms GOST28147-89-PES16-2 and GOST28147-89-RFWKPES16-2. In encryption algorithms GOST28147-89-PES16-2 and GOST28147-89-RFWKPES16-2 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-PES16-2

2.1 The Structure of the Encryption Algorithm GOST28147-89-PES16-2

In the encryption algorithm GOST28147-89-PES16-2 length of subblocks X^0, X^1, ..., X^{15}, length of round keys $K_{24(i-1)}$, $K_{24(i-1)+1}$, ..., $K_{24(i-1)+15}$,, $i = \overline{1...n+1}$, $K_{24(i-1)+16}$, $K_{24(i-1)+17}$, ..., $K_{24(i-1)+23}$, $i = \overline{1...n}$ and K_{24n+16}, K_{24n+17}, ..., K_{24n+47} is 8-bits. The length of the input and output blocks of round functions is 32 bits. In this encryption algorithm round function of encryption algorithm GOST 28147-89 is applied twice and in each round function employed eight S-boxes, i.e. the total number of S-boxes is 16. The scheme of the encryption algorithm GOST28147-89-PES16-2 is shown in Figure 1 and the S-boxes shown in Table 1.

Table 1: The S-boxes of encryption algorithm GOST28147-89-RFWKPES4-2

	0x0	0x1	0x2	0x3	0x4	0x5	0x6	0x7	0x8	0x9	0xA	0xB	0xC	0xD	0xE	0xF
S_0	0xA	0x0	0x2	0xF	0x3	0x5	0x4	0x1	0x7	0xE	0x9	0xD	0xC	0xB	0x6	0x8
S_1	0x2	0x4	0xC	0x0	0xD	0x6	0x7	0x5	0xE	0x1	0xB	0x8	0x9	0x3	0xF	0xA
S_2	0xE	0xC	0xD	0x0	0xA	0x2	0x5	0xB	0x3	0x7	0x8	0x1	0x6	0x9	0x4	0xF
S_3	0xA	0xF	0x1	0xB	0x3	0xE	0xC	0xD	0x0	0x9	0x6	0x5	0x7	0x8	0x2	0x4
S_4	0xA	0x3	0xD	0xC	0xE	0x5	0x6	0x0	0xB	0xF	0x7	0x2	0x1	0x9	0x8	0x4
S_5	0x8	0xD	0x2	0x6	0x1	0x3	0x0	0xE	0xC	0x5	0x4	0x9	0xA	0xB	0xF	0x7
S_6	0xD	0x3	0xF	0x6	0x9	0x8	0xE	0x5	0x4	0x0	0x7	0xA	0xC	0xB	0x2	0x1
S_7	0xC	0xA	0xB	0xF	0x5	0x9	0x7	0x4	0x8	0x1	0x3	0xE	0x0	0x2	0x6	0xD
S_8	0xA	0x3	0x2	0xC	0x0	0x5	0x7	0x1	0x4	0xE	0x9	0xD	0xF	0x8	0x6	0xB
S_9	0x0	0x3	0x7	0xA	0xF	0x9	0x1	0xB	0xD	0x2	0xC	0xE	0x6	0x8	0x5	0x4
S_{10}	0xC	0xD	0xB	0x4	0x8	0x5	0x6	0xE	0x3	0x7	0x9	0x2	0x1	0xF	0x0	0xA
S_{11}	0xA	0x0	0xE	0xC	0xF	0x6	0x7	0x1	0x8	0xD	0x5	0x2	0x3	0xB	0x9	0x4
S_{12}	0xA	0xC	0x3	0x7	0x0	0x1	0x2	0xF	0xE	0x4	0x6	0x8	0xB	0x9	0xD	0x5
S_{13}	0x8	0xB	0x5	0x1	0xA	0x2	0xD	0x4	0xC	0xE	0x9	0xF	0x0	0x7	0x3	0x6
S_{14}	0xE	0x8	0x0	0xA	0xF	0xC	0x3	0x7	0x4	0x5	0x9	0x2	0xD	0x1	0xB	0x6
S_{15}	0xC	0xB	0xA	0x9	0x0	0xE	0x4	0x1	0xF	0x3	0x7	0x8	0x2	0x6	0x5	0xD

Consider the round function of encryption algorithm GOST28147-89-PES16-2. First the 8-bit sub-

blocks T^0, T^1, ..., T^7 combined from 32-bit subblocks, i.e. T_0=T^0|| T^1|| T^2|| T^3, T_1=T^4|| T^5|| T^6|| T^7. Subblocks T_0, T_1 are summed to round keys $K_{24(i-1)+16}$|| $K_{24(i-1)+17}$|| $K_{24(i-1)+18}$|| $K_{24(i-1)+19}$, $K_{24(i-1)+20}$|| $K_{24(i-1)+21}$|| $K_{24(i-1)+22}$|| $K_{24(i-1)+23}$ i.e. S^0=T_0+($K_{24(i-1)+16}$|| $K_{24(i-1)+17}$|| $K_{24(i-1)+18}$|| $K_{24(i-1)+19}$), S^1=T_1+($K_{24(i-1)+20}$|| $K_{24(i-1)+21}$|| $K_{24(i-1)+22}$|| $K_{24(i-1)+23}$). 32-bit subblocks S^0, S^1 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. Four bit subblocks s_i^0, s_i^1, $i = \overline{0...7}$ transformed into the S-boxes: R^0=$S_0(s_0^0)$|| $S_1(s_1^0)$|| ...|| $S_7(s_7^0)$, R^1=$S_8(s_0^1)$|| $S_9(s_1^1)$|| ...|| $S_{15}(s_7^1)$. The resulting 32-bit subblocks R^0, R^1 cyclically shifted left by 11 bits and we obtain subblocks Y_0, Y_1: Y_0=R^0<<11, Y_1=R^1<<11. Thereafter 32-bit subblocks Y_0, Y_1 divided into four 8-bit subblocks Y^0, Y^1, ..., Y^7 i.e., Y_0=Y^0|| Y^1|| ...|| Y^3, Y_1=Y^4|| Y^5||...|| Y^7.

Consider the encryption process of encryption algorithm GOST28147-89-PES16-2. Initially the 128-bit plaintext X partitioned into subblocks of 8-bits X_0^0, X_0^1, ..., X_0^{15}, and performs the following steps:

1) subblocks X_0^0, X_0^1, ..., X_0^{15} are summed to XOR with round key K_{24n+16}, K_{24n+17}, ..., K_{24n+31}: $X_0^j = X_0^j \oplus K_{24n+16+j}$, $j = \overline{0...15}$.

2) subblocks X_0^0, X_0^1, ..., X_0^{15} multiplied and summed with the round keys $K_{24(i-1)}$, $K_{24(i-1)+1}$, ..., $K_{24(i-1)+15}$ and calculated 8-bit subblocks T^0, T^1, ..., T^7. This step can be represented as follows: $T^j = (X_{i-1}^j \cdot K_{24(i-1)+j}) \oplus (X_{i-1}^{j+8} + K_{24(i-1)+j+8})$, $j = \overline{0...7}$, $i = 1$.

3) to 8-bit subblocks T^0, T^1, ..., T^7 applied round functions and get 8-bit subblocks Y^0, Y^1, ..., Y^7.

4) subblocks Y^0, Y^1, ..., Y^7 are summed to XOR with subblocks X_{i-1}^0, X_{i-1}^1, ..., X_{i-1}^{15}, i.. $X_{i-1}^j = X_{i-1}^j \oplus Y^{7-j}$, $X_{i-1}^{j+8} = X_{i-1}^{j+8} \oplus Y^{7-j}$, $j = \overline{0...7}$, $i = 1$.

5) at the end of the round subblocks swapped, i.., $X_i^j = X_{i-1}^{j+8}$, $X_i^{8+j} = X_{i-1}^j$, $j = \overline{1...7}$, $i = 1$.

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

7) in output transformation round keys K_{24n}, K_{24n+1}, ..., K_{24n+15} are multiplied and summed into subblocks, i.e. X_n^0, X_n^1, ..., X_n^{15}: $X_{n+1}^j = X_n^j \cdot K_{24n+j}$, $X_{n+1}^{j+8} = X_n^{j+8} + K_{24n+j+8}$, $j = \overline{0...7}$,

8) subblocks X_{n+1}^0, X_{n+1}^1, ..., X_{n+1}^{15} are summed to XOR with the round key K_{24n+32}, K_{24n+33}, ..., K_{24n+47}: $X_{n+1}^j = X_{n+1}^j \oplus K_{24n+32+j}$, $j = \overline{0...15}$.

As ciphertext plaintext X receives the combined 8-bit subblocks $X_{n+1}^0||X_{n+1}^1||...||X_{n+1}^{15}$. The lenght of ciphertext is 128-bit.

In the encryption algorithm GOST28147-89-PES16-2 with encryption and decryption is used the same algorithm, only when decryption calculated inversion round keys depending on the operations and are applied in reverse order. One important task of encryption is key generation.

2.2 Key Generation of the Encryption Algorithm GOST28147-89-PES16-2

In n-round encryption algorithm GOST28147-89-PES16-2 in each round used twenty four round keys of the 8-bit and output transformation sixteen round keys of the 8-bit. In addition, before the first round and after the output transformation we used sixteen round keys of 8-bits. Total number of 8-bit round keys is equal to $24n$+48. In Figure 1 in encryption process used encryption round keys K_i^c instead of K_i, while decryption used decryption round keys K_i^d. If n=8 then need 240 to generate round keys, if n=12, you need to generate 336 round keys and if n=16 need 432 to generate round keys.

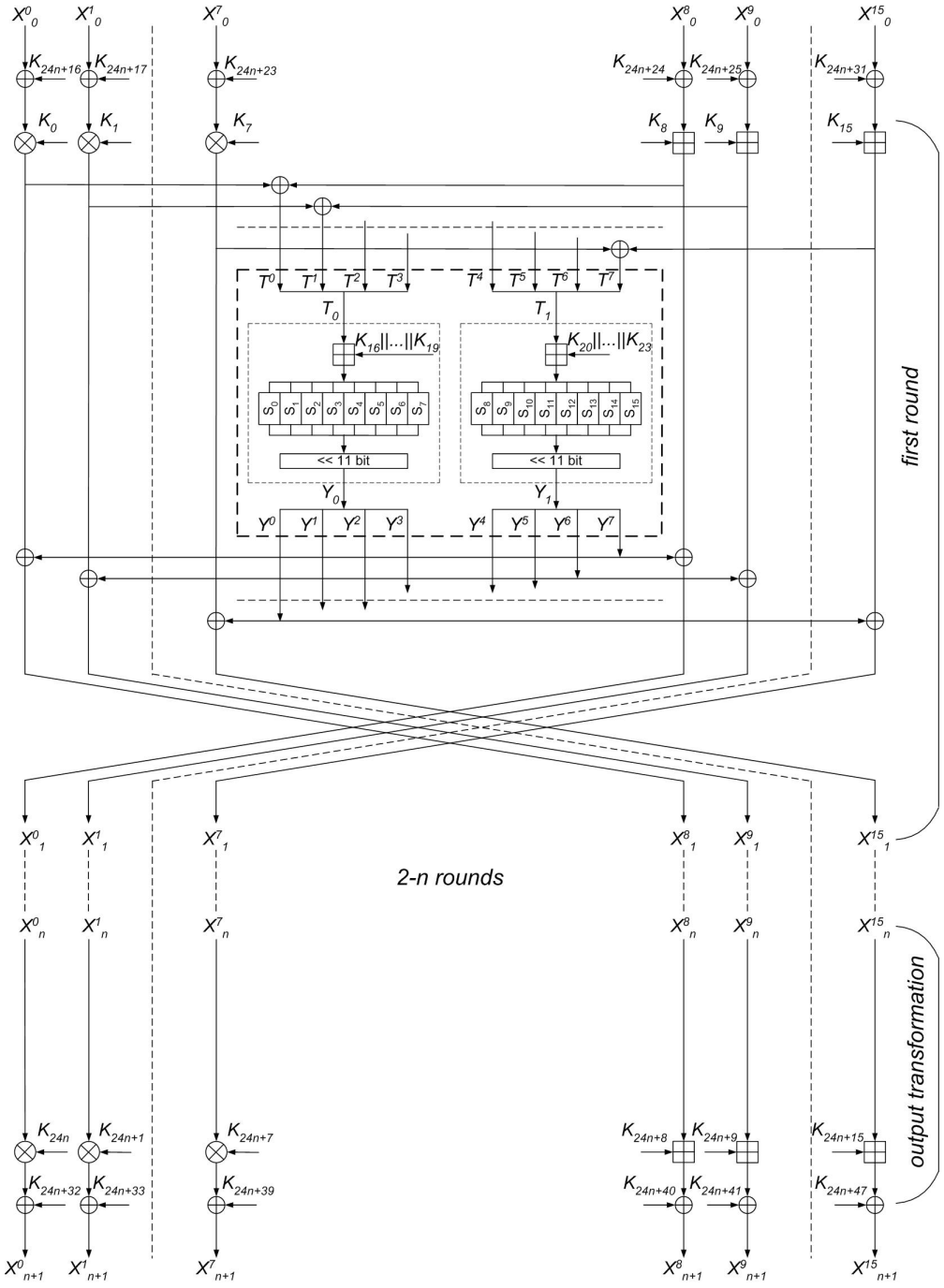

Figure 1: The scheme of encryption algorithm GOST28147-89-PES16-2

The key encryption algorithm K of length l ($256 \le l \le 1024$) bits is divided into 8-bit round keys K_0^c, K_1^c,..., $K_{Lenght-1}^c$, $Lenght = l/8$, here $K = \{k_0, k_1, ..., k_{l-1}\}$, $K_0^c = \{k_0, k_1, ..., k_7\}$, $K_1^c = \{k_8, k_9, ..., k_{15}\}$,..., $K_{Lenght-1}^c = \{k_{l-8}, k_{l-7}, ..., k_{l-1}\}$ 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 0xC5, i.e. $K_L =$ 0xC5. Round keys K_i^c, $i = \overline{Lenght...24n+47}$ are computed as follows $K_i^c = Sbox_0(K_{i-Lenght}^c) \oplus Sbox_1(RotWord8(K_{i-Lenght+1}^c)) \oplus Sbox_0(K_{i-Lenght+2}^c) \oplus Sbox_1(RotWord8(K_{i-Lenght+3}^c)) \oplus K_L$. After each round key generation the value K_L is cyclic shift to the left by 1 bit. Here, *RotWord8()*-cyclic shift to the left of 1 bit of the 11-bit subblock, *Sbox*-transformation a 8-bit subblock in the S-boxes, $Sbox_0(T) = S_0(t^0)||S_1(t^1)$, $Sbox_1(T) = S_8(t^0)||S_9(t^1)$, $T = t^0||t^1$ and t^0, t^1-four bit subblock, T-eight bit subblock, S_i-i-th S-Box.

The decryption round keys K_i^d calculated on the basis of encryption round keys K_i^c and decryption round keys of the first, second and n-round associates with the encryption round keys as follows:

$$\begin{aligned}
&(K_{24(i-1)}^d, K_{24(i-1)+1}^d, K_{24(i-1)+2}^d, K_{24(i-1)+3}^d, K_{24(i-1)+4}^d, K_{24(i-1)+5}^d, K_{24(i-1)+6}^d, K_{24(i-1)+7}^d,\\
&\quad K_{24(i-1)+8}^d, K_{24(i-1)+9}^d, K_{24(i-1)+10}^d, K_{24(i-1)+11}^d, K_{24(i-1)+12}^d, K_{24(i-1)+13}^d, K_{24(i-1)+14}^d,\\
&\quad K_{24(i-1)+15}^d, K_{24(i-1)+16}^d, K_{24(i-1)+17}^d, K_{24(i-1)+18}^d, K_{24(i-1)+19}^d, K_{24(i-1)+20}^d, K_{24(i-1)+21}^d,\\
&\quad K_{24(i-1)+22}^d, K_{24(i-1)+23}^d)\\
=\ &((K_{24(n-i+1)}^c)^{-1}, (K_{24(n-i+1)+1}^c)^{-1}, (K_{24(n-i+1)+2}^c)^{-1}, (K_{24(n-i+1)+3}^c)^{-1}, (K_{24(n-i+1)+4}^c)^{-1},\\
&\quad (K_{24(n-i+1)+5}^c)^{-1}, (K_{24(n-i+1)+6}^c)^{-1}, (K_{24(n-i+1)+7}^c)^{-1}, -K_{24(n-i+1)+8}^c, -K_{24(n-i+1)+9}^c,\\
&\quad -K_{24(n-i+1)+10}^c, -K_{24(n-i+1)+11}^c, -K_{24(n-i+1)+12}^c, -K_{24(n-i+1)+13}^c, -K_{24(n-i+1)+14}^c,\\
&\quad -K_{24(n-i+1)+15}^c, K_{24(n-i)+16}^c, K_{24(n-i)+17}^c, K_{24(n-i)+18}^c, K_{24(n-i)+19}^c, K_{24(n-i)+20}^c,\\
&\quad K_{24(n-i)+21}^c, K_{24(n-i)+22}^c, K_{24(n-i)+23}^c), i = \overline{1...n}
\end{aligned}$$

Decryption round keys of the output transformation associate with of encryption round keys as follows:

$$\begin{aligned}
&(K_{24n}^d, K_{24n+1}^d, K_{24n+2}^d, K_{24n+3}^d, K_{24n+4}^d, K_{24n+5}^d, K_{24n+6}^d, K_{24n+7}^d, K_{24n+8}^d, K_{24n+9}^d, K_{24n+10}^d,\\
&\quad K_{24n+11}^d, K_{24n+12}^d, K_{24n+13}^d, K_{24n+14}^d, K_{24n+15}^d)\\
=\ &((K_0^c)^{-1}, (K_1^c)^{-1}, (K_2^c)^{-1}, (K_3^c)^{-1}, (K_4^c)^{-1}, (K_5^c)^{-1}, (K_6^c)^{-1}, (K_7^c)^{-1}, -K_8^c, -K_9^c, -K_{10}^c, -K_{11}^c,\\
&\quad -K_{12}^c, -K_{13}^c, -K_{14}^c, -K_{15}^c).
\end{aligned}$$

Decryption round keys applied to the first round and after the output transformation associated with the encryption round keys as follows: $K_{24n+16+j}^d = K_{24n+32+j}^c$, $K_{24n+32+j}^d = K_{24n+16+j}^c$, $j = \overline{0...15}$.

3 The Encryption Algorithm GOST28147-89-RFWKPES16-2

3.1 The Structure of the Encryption Algorithm GOST28147-89-RFWKPES16-2

In the encryption algorithm GOST28147-89-RFWKPES16-2 the length of subblocks X^0, X^1, ..., X^{15}, length of round keys $K_{16(i-1)}$, $K_{16(i-1)+1}$, ..., $K_{16(i-1)+15}$,, $i = \overline{1...n+1}$ and K_{16n+16}, K_{16n+17}, ..., K_{16n+47} are equal to 8-bits. The length of the input and output blocks of round functions is 32 bits. The structure of the encryption algorithm GOST28147-89-RFWKPES16-2 is shown in Figure 2 and the S-boxes shown in Table 1.

Consider the round function block encryption algorithm GOST28147-89-RFWKPES16-2. First the 8-bit subblocks $T^0, T^1, \ldots, T^7$ combined from 32-bit subblocks, i.e. $T_0{=}T^0||T^1||T^2||T^3$, $T_1{=}T^4||T^5||T^6||T^7$. 32-bit subblocks T^0, T^1 divided into eight four bit subblocks $T^0{=}t_0^0||t_1^0||...||t_7^0$ $T^1{=}t_0^1||t_1^1||...||t_7^1$. Four bit subblocks t_i^0, t_i^1, $i = \overline{0...7}$ transformed into the S-boxes:

$$\begin{aligned} R^0 &= S_0(t_0^0)||S_1(t_1^0)||...||S_7(t_7^0), \\ R^1 &= S_8(t_0^1)||S_9(t_1^1)||...||S_{15}(t_7^1). \end{aligned}$$

The resulting 32-bit subblocks R^0, R^1 cyclically shifted left by 11 bits and we obtain subblocks Y_0, Y_1: $Y_0{=}R^0{<<}11$, $Y_1{=}R^1{<<}11$. Thereafter 32-bit subblocks Y_0, Y_1 divided into four 8-bit subblocks Y^0, Y^1, ..., Y^7 i.e., $Y_0{=}Y^0||Y^1||...||Y^3$, $Y_1{=}Y^4||Y^5||...||Y^7$.

Consider the encryption process of encryption algorithm GOST28147-89-RFWKPES16-2. Initially the 128-bit plaintext X partitioned into subblocks of 8-bits $X_0^0, X_0^1, \ldots, X_0^{15}$, and performs the following steps:

1) subblocks $X_0^0, X_0^1, \ldots, X_0^{15}$ are summed to XOR with round key K_{16n+16}, K_{16n+17}, ..., K_{16n+31}: $X_0^j = X_0^j \oplus K_{16n+16+j}$, $j = \overline{0...15}$.

2) subblocks $X_0^0, X_0^1, \ldots, X_0^{15}$ multiplied and summed with the round keys $K_{16(i-1)}$, $K_{16(i-1)+1}$, $\ldots$, $K_{16(i-1)+15}$ and calculated 8-bit subblocks $T^0, T^1, \ldots, T^7$. This step can be represented as follows: $T^j = (X_{i-1}^j \cdot K_{16(i-1)+j}) \oplus (X_{i-1}^{j+8} + K_{16(i-1)+j+8})$, $j = \overline{0...7}$, $i = 1$.

3) to 8-bit subblocks $T^0, T^1, \ldots, T^7$ applied round functions and get 8-bit subblocks $Y^0, Y^1, \ldots, Y^7$.

4) subblocks $Y^0, Y^1, \ldots, Y^7$ are summed to XOR with subblocks $X_{i-1}^0, X_{i-1}^1, \ldots, X_{i-1}^{15}$, i.. $X_{i-1}^j = X_{i-1}^j \oplus Y^{7-j}$, $X_{i-1}^{j+8} = X_{i-1}^{j+8} \oplus Y^{7-j}$, $j = \overline{0...7}$, $i = 1$.

5) at the end of the round subblocks swapped, i.., $X_i^j = X_{i-1}^{j+8}$, $X_i^{8+j} = X_{i-1}^j$, $j = \overline{1...7}$, $i = 1$.

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

7) in output transformation round keys $K_{16n}, K_{16n+1}, \ldots, K_{16n+15}$ are multiplied and summed into subblocks, i.e. $X_n^0, X_n^1, \ldots, X_n^{15}$: $X_{n+1}^j = X_n^j \cdot K_{16n+j}$, $X_{n+1}^{j+8} = X_n^{j+8} + K_{16n+j+8}$, $j = \overline{0...7}$,

8) subblocks $X_{n+1}^0, X_{n+1}^1, \ldots, X_{n+1}^{15}$ are summed to XOR with the round key K_{16n+32}, K_{16n+33}, $\ldots$, K_{16n+47}: $X_{n+1}^j = X_{n+1}^j \oplus K_{16n+32+j}$, $j = \overline{0...15}$.

As ciphertext plaintext X receives the combined 8-bit subblocks $X_{n+1}^0||X_{n+1}^1||...||X_{n+1}^{15}$. In the encryption algorithm GOST28147-89-RFWKPES16-2 with encryption and decryption is used the same algorithm, only when decryption calculated inversion round keys depending on the operations and are applied in reverse order. One important task of encryption is key generation.

3.2 Key Generation of the Encryption Algorithm GOST28147-89-RFWKPES16-2

In n-round encryption algorithm GOST28147-89-PES16-2 in each round used sixteen round keys of the 8-bit and output transformation sixteen round keys of the 8-bit. In addition, before the first round and after the output transformation we used sixteen round keys of 8-bits. Total number of 8-bit round keys is equal to $16n{+}48$. In Figure 2 in encryption process used encryption round keys K_i^c instead of K_i, while decryption used decryption round keys K_i^d.

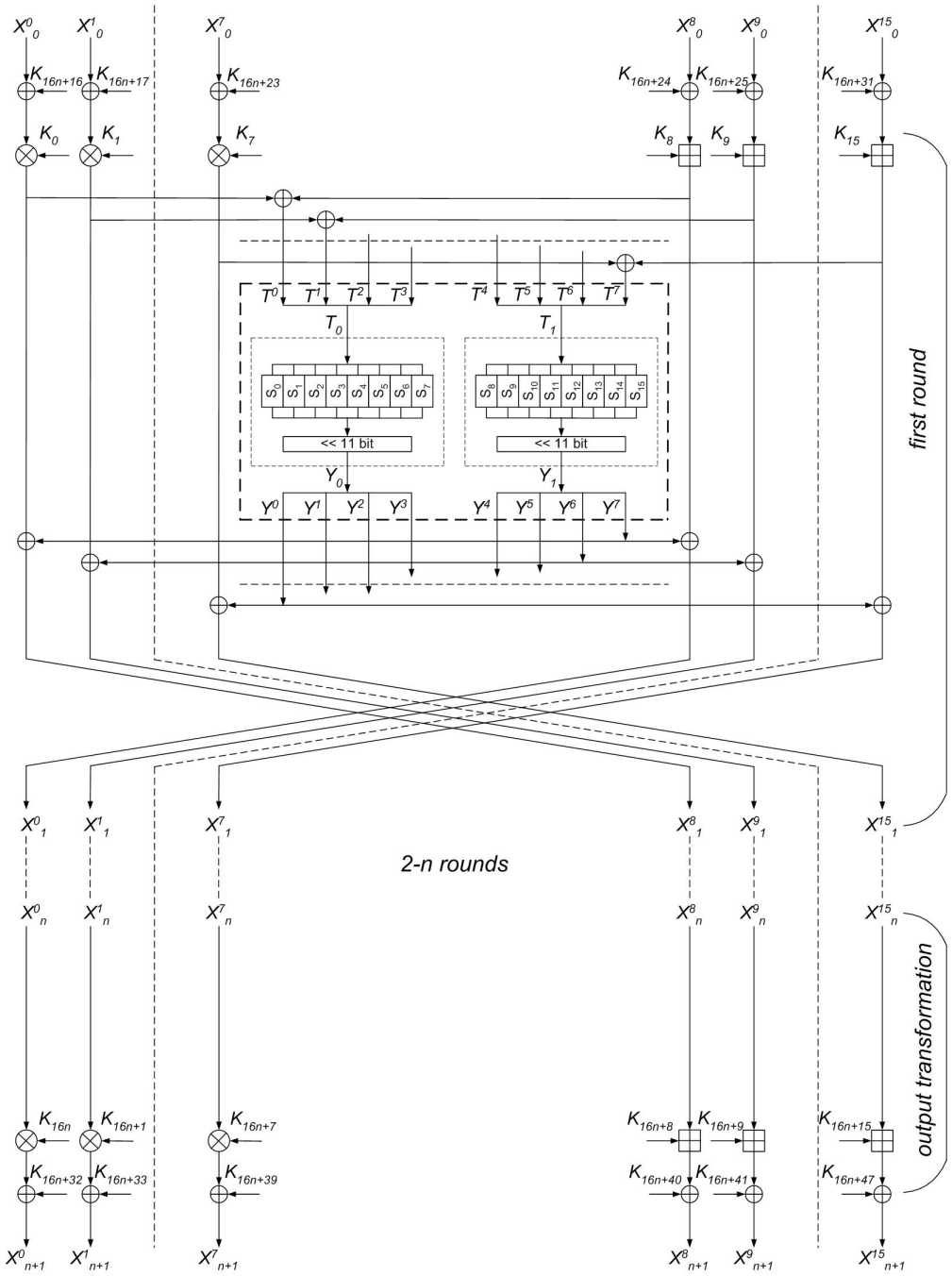

Figure 2: The scheme of encryption algorithm GOST28147-89-RFWKPES16-2

The key encryption algorithm K of length l ($256 \leq l \leq 1024$) bits is divided into 8-bit round keys K_0^c, K_1^c,..., $K_{Lenght-1}^c$, $Lenght = l/8$, here $K = \{k_0, k_1, ..., k_{l-1}\}$, $K_0^c = \{k_0, k_1, ..., k_7\}$, $K_1^c = \{k_8, k_9, ..., k_{15}\}$,..., $K_{Lenght-1}^c = \{k_{l-8}, k_{l-7}, ..., k_{l-1}\}$ 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 0xC5, i.e. $K_L =$ 0xC5. Round keys K_i^c, $i = \overline{Lenght...16n+47}$ are computed as follows $K_i^c = Sbox_0(K_{i-Lenght}^c) \oplus Sbox_1(RotWord8(K_{i-Lenght+1}^c)) \oplus Sbox_0(K_{i-Lenght+2}^c) \oplus Sbox_1(RotWord8(K_{i-Lenght+3}^c)) \oplus K_L$. After each round key generation the value K_L is cyclic shift to the left by 1 bit. Here, *RotWord8()*-cyclic shift to the left of 1 bit of the 11-bit subblock, *Sbox*-transformation a 8-bit subblock in the S-boxes, $Sbox_0(T) = S_0(t^0)||S_1(t^1)$, $Sbox_1(T) = S_8(t^0)||S_9(t^1)$, $T = t^0)||t^1$ and t^0, t^1-four bit subblock, T-eight bit subblock, S_i-i-th S-Box.

Decryption round keys K_i^d are computed on the basis of encryption round keys K_i^c and decryption round keys of the first, second and n-round associates with the encryption round keys as follows:

$$\begin{aligned}
&(K_{16(i-1)}^d, K_{16(i-1)+1}^d, K_{16(i-1)+2}^d, K_{16(i-1)+3}^d, K_{16(i-1)+4}^d, K_{16(i-1)+5}^d, K_{16(i-1)+6}^d, K_{16(i-1)+7}^d,\\
&\quad K_{16(i-1)+8}^d, K_{16(i-1)+9}^d, K_{16(i-1)+10}^d, K_{16(i-1)+11}^d, K_{16(i-1)+12}^d, K_{16(i-1)+13}^d, K_{16(i-1)+14}^d,\\
&\quad K_{16(i-1)+15}^d)\\
= \;&((K_{16(n-i+1)}^c)^{-1}, (K_{16(n-i+1)+1}^c)^{-1}, (K_{16(n-i+1)+2}^c)^{-1}, (K_{16(n-i+1)+3}^c)^{-1}, (K_{16(n-i+1)+4}^c)^{-1},\\
&\quad (K_{16(n-i+1)+5}^c)^{-1}, (K_{16(n-i+1)+6}^c)^{-1}, (K_{16(n-i+1)+7}^c)^{-1}, -K_{16(n-i+1)+8}^c, -K_{16(n-i+1)+9}^c,\\
&\quad -K_{16(n-i+1)+10}^c, -K_{16(n-i+1)+11}^c, -K_{16(n-i+1)+12}^c, -K_{16(n-i+1)+13}^c, -K_{16(n-i+1)+14}^c,\\
&\quad -K_{16(n-i+1)+15}^c), i = \overline{1...n}.
\end{aligned}$$

Decryption round keys of the output transformation associate with of encryption round keys as follows:

$$\begin{aligned}
&(K_{16n}^d, K_{16n+1}^d, K_{16n+2}^d, K_{16n+3}^d, K_{16n+4}^d, K_{16n+5}^d, K_{16n+6}^d, K_{16n+7}^d, K_{16n+8}^d, K_{16n+9}^d, K_{16n+10}^d,\\
&\quad K_{16n+11}^d, K_{16n+12}^d, K_{16n+13}^d, K_{16n+14}^d, K_{16n+15}^d)\\
= \;&((-K_0^c)^{-1}, (K_1^c)^{-1}, (K_2^c)^{-1}, (K_3^c)^{-1}, (K_4^c)^{-1}, (K_5^c)^{-1}, (K_6^c)^{-1}, (K_7^c)^{-1}, -K_8^c, -K_9^c, -K_{10}^c, -K_{11}^c,\\
&\quad -K_{12}^c, -K_{13}^c, -K_{14}^c, -K_{15}^c).
\end{aligned}$$

Decryption round keys applied to the first round and after the output transformation associated with the encryption round keys as follows: $K_{16n+16+j}^d = K_{16n+32+j}^c$, $K_{16n+32+j}^d = K_{16n+16+j}^c$, $j = \overline{0...15}$.

4 Results

As a result of this study built a new block encryption algorithms called GOST28147-89-PES16-2 and GOST28147-89-RFWKPES16-2. This algorithm is based on a networks PES16-2 and RFWKPES16-2 using the round function of GOST 28147-89. Length of block encryption algorithm is 128 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. The 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 encryption algorithms GOST28147-89-PES16-2 and GOST28147-89-RFWKPES16-2 for all S-boxes, the following equation: deg=3, NL=4, λ=0.5, δ= 3/8, SAC$\leq$2,

Table 2: Parameters of the S-boxes 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

BIC$\leq$4, i.e. resistance is not lower than the algorithm GOST 28147-89. These S-boxes are created based on Nyberg construction [4].

To the encryption algorithm applied linear cryptanalysis. Attack on 4-round GOST28147-89-PES16-2 has a data complexity of 2^{51} chosen plaintexts and on 4-round GOST28147-89-RFWKPES16-2 has a data complexity of 2^{43} chosen plaintexts.

5 Conclusions

In this way, built a new block encryption algorithms called GOST28147-89-PES16-2 and GOST28147-89-RFWKPES16-2 based on networks PES16-2 and RFWKPES16-2 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, "A proposal for a new block encryption standard," *Eurocrypt90, LNCS 473, Springer-Verlag*, pp. 389–404, 1991.

[7] GOST 2814789. National Standard of the USSR. Information Processing Systems. Cryptographic Protection. Algorithm Cryptographic Transformation.

[8] G. Tuychiev, "About networks PES8-2 and PES8-1, developed on the basis of network PES8-4," *Materials of the International Scientific Conference* .

[9] G. Tuychiev, "About networks RFWKPES8-4, RFWKPES8-2, RFWKPES8-1, developed on the basis of network PES8-4," *Materials of the International Scientific Conference.*

[10] 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.

[11] G. Tuychiev, "The networks RFWKIDEA4-2, IDEA4-1 and RFWKIDEA4-1," *Acta of Turin Polytechnic University in Tashkent*, vol. 3, pp. 71–77, 2013.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.
[27] 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.
[28] 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.
[29] 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.
[30] G. Tuychiev, "The encryption algorithm AES-RFWKIDEA16-1," *Infocommunications: Networks-Technologies-Solutions*, vol. 34, no. 2, pp. 48–54, 2015.
[31] 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.
[32] 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.
[33] 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.
[34] 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.
[35] 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.
[36] 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.
[37] 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.
[38] 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.
[39] 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.

Tuychiev Gulom candidate technical Sciences (Ph.D.), National University of Uzbekistan.

A New Muzzle Classification Model Using Decision Tree Classifier

Ibrahim El-Henawy[1], Hazem. M. El Bakry[2], Hagar M. El Hadad[3]
(Corresponding author: Hagar M. El Hadad)

Faculty of Computer and Information Sciences, Zagazig University Zagazig, Egypt[1]
Department of Information Systems, Faculty of Computer and Information Sciences[2]
Mansoura University, Mansoura, Egypt
Department of Information Systems, Faculty of Computer Science and Information Sciences[3]
Beni-Suef University, Beni-Suef, Egypt
(Email: elhadad.hager@yahoo.com)
(Received Feb. 8, 2016; revised and accepted Apr. 9 & Dec. 10, 2016)

Abstract

Animal agricultures pays a great attention for saving the cattle because of the rapid growth of the livestock. The critical point in this research is to classify large different groups of cattle's with high degree of accuracy. This paper presents cattle classification model depending on decision tree classifier. Such model consists of three parts; pre-processing part, texture feature extraction part and classification part. Pre-processing part consists of histogram equalization used to increase image contrast and mathematical morphology filtering to remove noise. The second part uses two different algorithms in order to extract image features; box-counting and segmentation based fractal texture analysis (SFTA). Then the decision tree is applied for the classification process. The used data base consists of fifty-two different cattle with sixty different images for each cattle. Experimental results prove the advancement of decision tree classifier than other classifiers techniques. The result of decision tree is compared with artificial neural networks (ANNs). For fifty-two different groups the accuracy rate in case of decision tree is 96.39% compared to 14% in case of using ANN classification system.

Keywords: Artificial Neual Networks (ANNs); Box-counting Algorithm; Cattle Classification; Decition Tree; Segmentation Based Fractal Texture Analysis (SFTA)

1 Introduction

Today, Veterinarians pay great effort to help the animal agriculture to save the cattle products. The rapid growth of the livestock products is the critical point. So veterinarians and animal agriculture do their best to trace all animal specially cattle which has diseases infections. A lot of traditional techniques used to solve this critical problem but all cannot solve the deceitful farmers' problems. This paper aims to build a model that helps the veterinarians and animal agriculture to know the deceitful farmer. A lot of farmers able to remove the cattle ear tag easily after these cattle die and use this ear tag to another animal. The proposed intelligent model can help is achieve justice. This research help the end used to save the animal record and connect it with the cattle muzzle print. This model helps

veterinaries in tracing the non-healthy animals. The important point in this paper is to identify each cattle in this crowd. Cattle muzzle print play a great role in cattle observation and monitoring especially in cattle diseases beginning, cattle ownership assignment and traceability, vaccination management and production management [31].

The traditional classification system such as muzzle ink printing, ear notching, ear tags, Electronic Identification and Radio Frequency Identification (RFID) [26], Freeze branding and hot iron branding, tattooing, Neck Chains and Barcode, blood test or hair sample (DNA) and Nose printing. These traditional techniques are not satisfied in case of classification and identification especially in cases such as cattle repetition and farmer fraudulent. So, animal agriculture forces no using more accurate and reliability systems to get rid of the disadvantages and defects of all the traditional tracing techniques. Fingerprint is the human identifier and in cattle muzzle print is its biological identifier. In human, hair cover skins except some parts of the body like fingerprint. In cattle muzzle consists of distribution of valleys and ridges over it. Researchers such as Baranov and his team discovered the muzzle print for the cattle is discovered the asymmetry between the two halves and hereditable [18]. Cattle muzzle print is considered as a biometric identifier because the uniqueness [6, 10, 11]. The essential key to identify each animal or individual is the biometric depending on the behavioral features [12, 14].

Cattle classification models must have the following characteristics; reliability, acceptability, accuracy, uniquely identifies each cattle and solves the fraudulent problems [15]. Since 1921, the cattle muzzle is considered as a unique biological identifier like fingerprint in human case [13]. The traditional techniques such as ink print that is paper based technique and it was the earlier technique for animal identifying by animal agriculture. The disadvantages of using ink print technique are holding the animal still, build up wetness on the cattle noses, and use a lot of ink which case in wasting time. From this point of view the animal agriculture starts to search for new intelligent techniques that solve the traditional techniques disadvantages.

The new intelligent techniques based on the using digital image processing in cattle classification and identification systems. The new intelligent techniques use digital cameras instead of the ink print techniques. The advantages of these techniques depend on using different factors such as the growth of the availability of using workstations and microcomputers with large capability in saving livestock and working with large data base. These factors help in reducing the cost of computation and image acquisition and the rapid increasing in the image processing applications because they improve and increase the capabilities of image equipment and display devices [5]. So the first critical part in this paper is to collect a live cattle database based on different captured image for each cattle. The difference between what really automatically extracted from feature extracted technique and human observation is known scientifically by a Semantic Gap Problem [8].

The second critical part is the number of features in each feature vector that visually represents each captured cattle muzzle image contains. The feature vectors were used to solve the semantic gap problem [7]. The new researches in texture feature extraction field is used to increase the ability of differentiate between each cattle muzzle images [16]. The first technique that used for image texture feature extraction is box-counting, which is used in the second part in the proposed model in this paper [9]. This paper used two different techniques in the texture feature extraction part; box-counting and SFTA techniques then it used to compare between the accuracy rate after the classification part which is depend on the number of features in the feature vector. Many studies saw that box-counting technique is the common used technique for fractal calculations, so box-counting selected to use for fractal extraction among different techniques [3]. Chaudhuri and Sarker improved box-counting algorithm to differential Box-Counting [2, 17, 23, 30]. Box-counting algorithm feature vector consists of eight different features for each muzzle image and the accuracy for the classifier depends on number of features for each image. So a SFTA technique is used in order to increase number of feature for each cattle muzzle image. SFTA technique depends on decompose the gray scale muzzle image in to a set of

binary images from the fractal dimension of the regions to describe the segmented cattle muzzle texture patterns. The last part in this paper is the classification using decision tree technique.

Decision tree technique is a robust statistical technique for interpretation, prediction, data manipulation and classification which is used in many research fields. Decision tree technique commonly used in data mining in case of classification systems that depends on different attributes and to advanced prediction method for the target variable. Decision tree depends on classifies the problem in to branch such as segments that form the inverted tree with a root node, internal nodes and leaves nodes [25]. The proposed model in this research used box-counting and SFTA algorithms in the feature extraction part and the compare between the accuracy rates of the classification process that depends on these two different techniques. This is the fourth contribution for the authors and this is the best accurate technique if the authors compare between this model and the previous models accuracy rates.

The rest of the paper is organized as follows. Backgrounds are discussed in Section 2. Section 3 presents the proposed the cattle classification model in detail. Experimental results are discussed in Section 4. Conclusions and future work are discussed in Section 5.

2 Background

2.1 Histogram Equalization Algorithm

Histogram equalization algorithm (HEQ) is responsible for redistributing the gray level values in order to get on regular image histogram. HEQ algorithm replaces each pixel in the gray image with the integral histogram of the image in that pixel [27]. Histogram equalization used to adjust the image contrast. The adjustment of the intensities after applying histogram equalization on image makes the cattle muzzle image intensity better than the original one. This allow the low contrast area of image to become high contrast by spreading the most frequently intensity values [21, 22]. HEQ algorithm is discussed in details in the following lines.

2.2 Mathematical Morphology Filtering Algorithm (MMF)

Mathematical morphology filter based on the geometric shapes texture features. MMF used to remove noise from images. MMF consists of four operations opening, closing, dilation and erosion. This research depends on opening operation followed by closing operation to remove image noise. The two main elementary operations on which the mathematical morphology filtering operations depend are opening and closing. The Dilation Operation depends on replacing the gray values by the maximum weight of its neighborhood gray value. The erosion operation replaces the gray values by the minimum weight of its neighborhood gray value [24].

2.3 Box-Counting Algorithm

The first texture feature extraction algorithm which use in this paper is the Box-Counting Algorithm. Also, there is more than one algorithm to calculate the texture fractal dimension for each cattle muzzle print images. More than one study showed that box-counting is the common algorithm for calculating fractal dimensions [4] where this algorithm depends on counting the number of boxes that cover area of interest.

2.4 Segmentation-Based Fractal Texture Analysis or SFTA Algorithm

SFTA Algorithm applies multi-thresholding level Otsu on the gray scale cattle image to decompose the segmented cattle image to several parts. The pairs of upper threshold (tu) and lower threshold (t1) are

Algorithm 1 Box-counting algorithm D_b to any subset A in R^n (Euclidean space)

1: To calculate $D_b(A)$ set the value of $N_r(A)$ to the smallest number of r set that cover cattle muzzle area as

$$D_b(A) = \lim_{r \to 0} \frac{\log(N_r(A))}{\log(1/r)}$$

2: Dividing R^n in lattice Sub of grid size $r \times r$ where r is continually reduced
3: Set grid number of elements that divide Db(A) and $N_r(A)$ to $N'_r(A)$ as

$$D_b(A) = \lim_{r \to 0} \frac{\log(N'_r(A))}{\log(1/r)}$$

4: box counting $N_r(A)$ and $D_b(A)$ are related by relation power law shown in the following equation:

$$D_r(A) = \frac{1}{r^{D_b(A)}}$$

5: Place the bounded set A to the grid that created from boxes size $r \times r$
6: Continue this algorithm by alter r to gradually small size and each time calculate $N_r(A)$.

selected by using the Two Threshold Binary Decomposition (TTBD) technique. The resulted feature vector elements are: means gray level, fractal dimension, size of cattle area image etc [1].

Algorithm 2 Segmentation Based Fractal Texture Analysis algorithm (SFTA)

1: Covert cattle muzzle image from RGB to Gray scale I, where I is cattle Grayscale image
2: Set number of threshold n_t
3: assign Multi Level Otsu (I, n_t) function to variable T
4: Set $T_A == \{\{t_i, t_i + 1\} : t_i, t_i + 1 \in T,\ i \in [1 \cdots |T| - 1]\}$
5: Set $T_B == \{\{t_i, n_l\} : t_i \in T,\ i \in [1 \cdots ||]\}$, where n_l denote gray level range and T is the set of threshold values
6: Set $i == 0$
7: **for** $\{\{t_l, t_u\} : \{t_l, t_u\} \in T_A \cup T_B\}$ **do**
8: where t_l, t_u denote lower threshold, upper threshold respectively
9: VSFTA[i] == BoxCounting ($\triangle$), where ($\triangle$) border of cattle muzzle image
10: VSFTA[i+1] == MeanGrayLevel(I, I_b)
11: VSFTA[i+2] == PixelCount(I_b) where I_b is cattle binary image
12: VSFTA[i+2] == PixelCount(I_b) where I_b is cattle binary image
13: **end for**
14: return VSFTA, where VSFTA denote the extracted SFTA feature vectors

2.5 Decision Tree Algorithm

Decision tree is widely used in expert systems to represent knowledge. Decision tree classifiers formed to classify the feature vector for each muzzle with Boolean or categorical class labels [19]. Breiman et al propose the classification and regression tree (CART) [20] structure which called as Hierarchical Optimal Discriminate Analysis (HODA). CART is not a parametric decision tree that produces either regression or classification trees based on the reliant variable is numeric or categorical. The term binary

means that node in a decision tree can split into two groups only. CART depends on gini index which used as cheating measure for selecting cattle muzzle image patters attribute. The process of splitting the nodes depends on using the attributes with the large reduction in the population. CART uses categorical and numerical values and also solves the problem of missing values. It is useful to use cost complexity refinement and generate regression tree.

Algorithm 3 Decision tree induction algorithm

```
Tree(E, F)
if stop-condition(E,F)=true, then then
    Set leaf ← create-node ().
    leaf-label ← classify (E).
    return leaf.
else
    Root ← create-node ().
    Root.test.condition= find.best-split(E,F).
    Let V ← {U—U is a possible outcome of Root.test.condition}.
    for each U ϵ V do
        E_U ← {e—Root.test.condition(e)=U and e ϵ E}.
        child ← Tree(E,F).
        add child as descendent of root and label the edge (root → child) as U.
    end for
end if
return root
```

3 Proposed Cattle Muzzle Identification Model

In this research, the proposed model contains three parts: pre-processing part that is the first and critical initial part. The pre-processing part contains both histogram equalization and mathematical morphology filtering to increase image contrast and remove noise form image respectively. The texture feature extraction is the second part of the proposed model in which we use box-counting algorithm and SFTA algorithm to extract the feature vector of each cattle muzzle image that reflects each image contents. The decision tree is the third and the last part in the proposed model to classify cattle muzzle pattern image. These three parts are discussed in this section. The feature characteristics for each part are described in Figure 1.

3.1 Pre-processing Part

Pre-processing part is the first and critical part. The proposed model in this paper contains histogram equalization and mathematical morphology filter. The histogram equalization is used to increase image contrast because it depends on the distributing the intensity of pixels. Histogram equalization increases the contrast of cattle muzzle image because it depends on representing the used data by closed value of contrast which means that the image area of the lower contrast becomes a higher contrasted area. Mathematical Morphology Filtering is used to remove noise from cattle muzzle image. The main four operations that mathematical morphology depends on are: closing, erosion, opening and dilation where the main part is opening and closing. Open and close operations contain erosion and dilation. Dilation used in case of maximizing the object values. After cattle muzzle image, the dilation operation increases its intensity and becomes brighter than the original gray scale one. The erosion process is opposite to

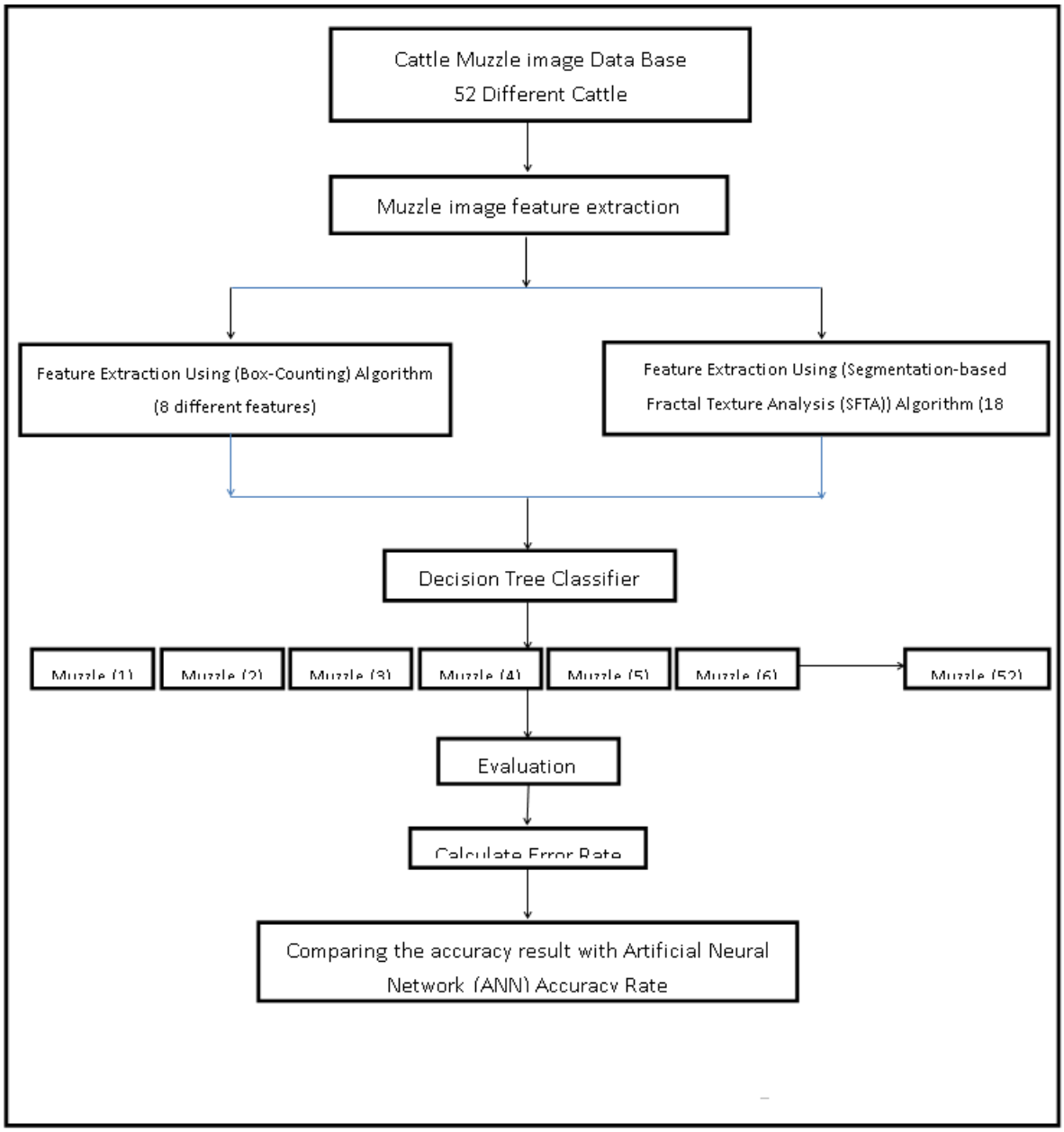

Figure 1: The proposed cattle classification model based of decision tree classifier

dilation because it minimizes the values of the muzzle image. The proposed model first implemented histogram equalization nut then implemented the mathematical morphology filtering operations to remove noise from the cattle muzzle image. In the mathematical morphology operation, we first open the image the resultant from this step is closed.

3.2 Texture Feature Extraction Part

The Texture Feature Extraction is the second and critical part of the proposed model. This part is still the challenging point in cattle muzzle identification because it depends on a number of features which are contained in the feature vector. More elements in the feature vector, lead to more accurate identification results. In this part, we implement box-counting algorithm and Segmentation-based Fractal Analysis Algorithm (SFTA).

3.3 Box-Counting Algorithm

The Texture feature vector, after implementing the box-counting algorithm, contains only eight features for each muzzle image. Depending on Box Count Function (C), where (D) is the dimensional array represented by C (where D = 1, 2, 3, 4, $\cdots$). Count (N) number of boxes of size (R) and of dimension (D). Box size is calculated by Power two, $R = 1, 2, 3, 4, \cdots, 2^P$, where P is the small integer as $MAX(SIZE(C)) <= 2^P$. If size (C) over each dimension is smaller than 2^P, then (C) is stuffed with zeros to size 2^P over each dimension.

3.4 Segmentation-based Fractal Texture Analysis Algorithm (SFTA)

Texture Feature Vector, after implement SFTA, contains only eighteen features for each muzzle image. The SFTA Algorithm is divided into two main parts; the first part is the decomposed input muzzle image to set of binary images. The Two Thresholds Binary Decomposition (TTBD) is the first technique that we use in case of decomposing the input cattle image. The resulted binary muzzle image is used to compute its feature fractal dimension for its regions' boundary.

3.5 Decision Tree Classifier

Decision tree classification part depends on the following steps:

Step 1: select training cattle muzzle image dataset for learning.

Step 2: find mapping between every individual feature vector attribute to cattle classes.

Step 3: find all possible values for every features and that equivalent possible cattle classes.

Step 4: then count values of each feature which belongs to unique cattle class.

Step 5: Make root node to that feature which have minimum number of values having unique cattle class.

Step 6: Likewise select other feature for next level in decision tree from residual feature the basis of minimum number of values having unique cattle class.

4 Experimental Results

A. Cattle Muzzle Print Database

The first challenge in this research was the lack of the real live printed cattle muzzle database. Therefore, the critical point in this research was to collect a muzzle image database which consists of fifty-two cattle each with sixty different muzzles image for each cattle. A sample printed muzzles for two different individual cattle are shown in Figure 2 where during the capturing part, a special care was made for the quality of collected cattle muzzles.

Figure 2: A sample of different cattle printed images. This figure represents print images for cattle muzzle that have taken from two different cattle

The identification scenarios: 3, 5, 10, 14, 20, 25 and 30 groups of cattle muzzle each group with 20 and 60 different muzzle images used in the training phase to calculate the accuracy of implementing the decision tree classification model. The use of decision tree comes after extracting the feature vector of each cattle image by using Box-counting algorithm and SFTA Algorithm. The cattle muzzle in the testing phase is correctly classified if it is found that the similarity between input images feature vector equals the tested image feature vector.

B. Evaluated Results

First: the accuracy rate after using Box-counting algorithm in the second part for feature extraction and decision tree in the classification part.

1) Twenty cases for each cattle:
As Table 1 show that the accuracy rate increases especially in cases that use large number of different cattle groups. By comparing this accuracy rate with the accuracy rate of the authors' previous work in which they use the artificial neural network (ANN) instead of

Table 1: Accuracy rate in case of using 3, 5, 10, 15, 25 and 52 different groups of muzzle, each group has 20 Cases. (Box-Counting algorithm and decision tree classifier)

	3 groups	5 groups	10 groups	15 groups	25 groups	52 groups
20 cases	96.29%	95.55%	91.2%	86.87%	82.47%	85.49%

Table 2: Accuracy rate in case of using 3, 5, 10, 15, 25 and 52 different groups of muzzle, each group has 20 Cases. (Box-Counting algorithm and Artificial neural network classifier)

	3 groups	5 groups	10 groups	15 groups	25 groups	52 groups
20 cases	100%	80%	48%	40%	36%	14%

decision tree classifiers the accuracy rate was very bad in the large number of the cattle groups. Table 2 shows the accuracy rate in the same cases that studied in this paper. The following statistical representation shows the big difference of using the decision tree in the classification part instead of artificial neural network. The accuracy rate in case of using decision tree and ANN to classify and differentiate between 52 different cattle groups are 85.49% and 14% respectively. But the accuracy rate in case of using 3 different group decreases to 96.29% after using decision tree classifier.

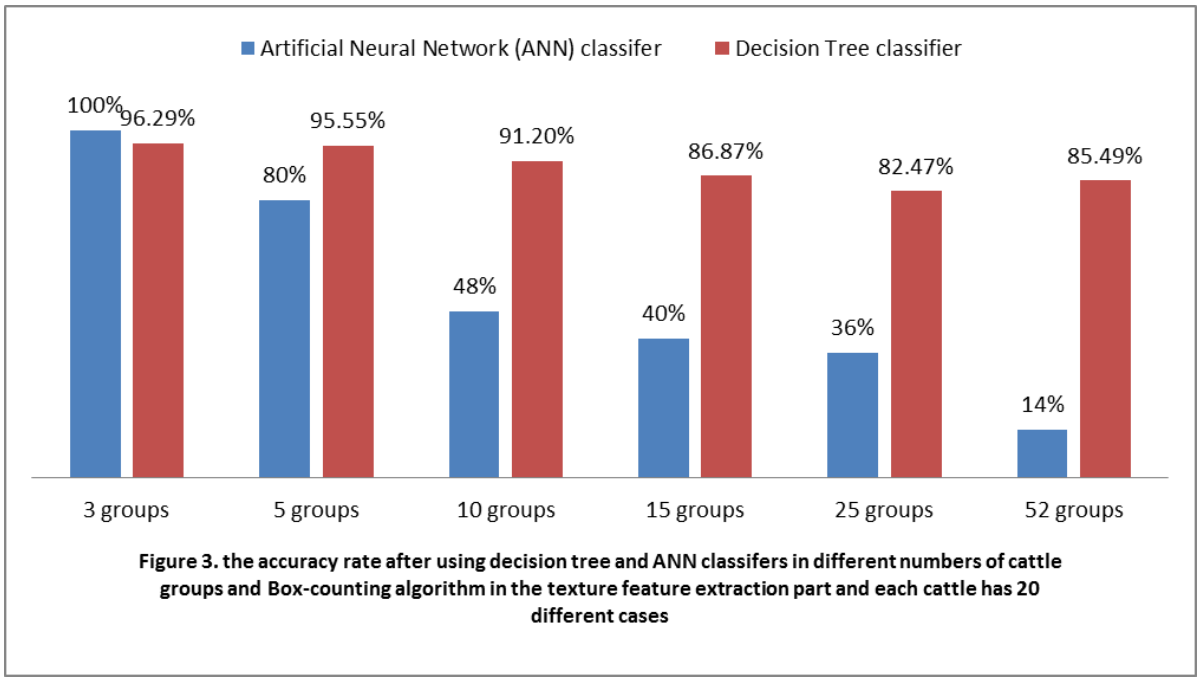

Figure 3. the accuracy rate after using decision tree and ANN classifers in different numbers of cattle groups and Box-counting algorithm in the texture feature extraction part and each cattle has 20 different cases

As shown in Figure 3 the decision tree classifier made a huge difference in the accuracy rate also, the number of the features in the texture feature vector extracted after using box-counting algorithm is eight but the accuracy rate is very good than ANN accuracy

rate for the same cases.

2) Sixty cases for each cattle:

Table 3: Accuracy rate in case of using 3, 5, 10, 15, 25 and 52 different groups of muzzle, each group has 60 Cases. (Box-counting algorithm and decision tree classifier)

	3 groups	5 groups	10 groups	15 groups	25 groups	52 groups
60 cases	100%	100%	100%	97.87%	95.51%	96.39%

Table 3 shows that the accuracy rate after using 60 different cases in the training phase, help the decision tree classifier to increase the accuracy rate to 100% in cases of using 3, 5 and 10 different groups and also increases in 15, 25 and 52 cases to 97.87%, 95.51% and 96.39% respectively. Comparing this accuracy rate with the same cases but replace decision tree with ANN classifier the accuracy rate is illustrated in the following statistical representation Figure 4.

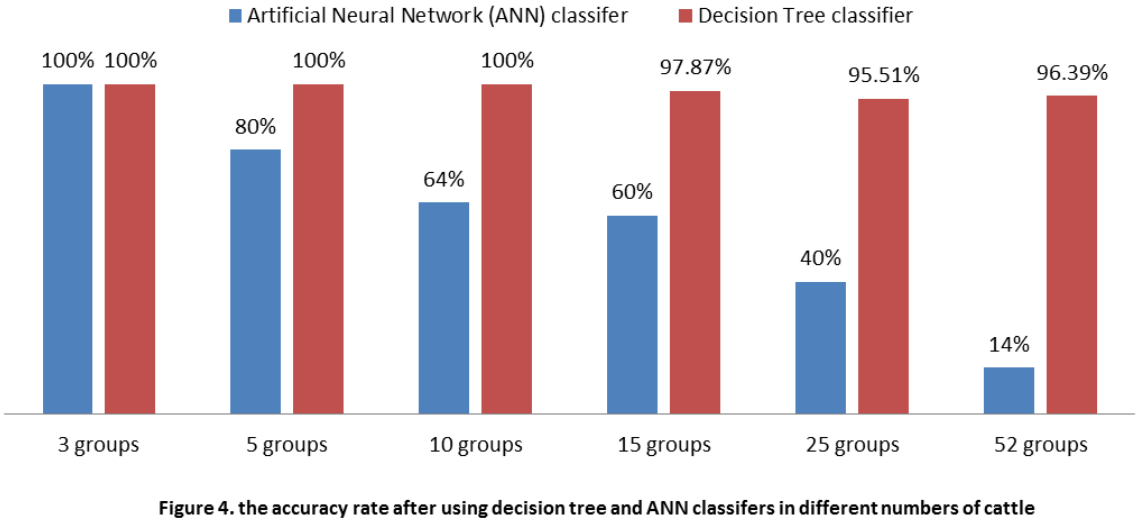

Figure 4. the accuracy rate after using decision tree and ANN classifers in different numbers of cattle groups and Box-counting algorithm in the texture feature extraction part and each cattle has 60 different cases

Second: the accuracy rate after using SFTA algorithm in the second part for feature extraction and decision tree in the classification part.

Table 4 shows that by increasing the number of feature in the feature vector extracted from SFTA algorithm to eighteen and the accuracy rate increases. This means that by increasing the number of feature in the feature vector for each muzzle image the accuracy rate increases. A comparison between the accuracy rates after using decision tree classifier with the SFTA algorithm and using ANN with SFTA algorithm is shown in Table 5.

Table 4: Accuracy rate in case of using 3, 5, 10, 15, 25 and 52 different groups of muzzle, each group has 20 and 60 cases. (SFTA algorithm and decision tree classifier)

	3 groups	5 groups	10 groups	15 groups	25 groups	52 groups
20 cases	96.29%	97.77%	97.22%	91.30%	91.612%	93.93%
60 cases	100%	100%	100%	93.47%	96.12%	93.93%

Table 5: A comparison between using decision tree and ANN in the classification part based on the features extracted from SFTA algorithm

	3 groups	5 groups	10 groups	15 groups	25 groups	52 groups
ANN 60 cases	100%	100%	80%	60%	56%	18%
Decision tree 60 cases	100%	100%	100%	93.47%	96.12%	93.93%

5 Conclusions and Future Work

A new cattle classification model that depends on the image of cattle muzzle has been presented. Such model consists of three different parts. The pre-processing part has used histogram equalization and mathematical morphology filter in order to enhance cattle muzzle image contrast and remove noise respectively. The second important part is the texture feature extraction which has employed two different algorithms (box-counting/SFTA) to extract different vectors. The feature vector of the box-counting algorithm consists of eight elements while SFTA feature vector contains eighteen features. Then, the decision tree classifier has been used to get the final result. The comparisons of accuracy have proven the advancement of SFTA algorithm over the box-counting one. Furthermore, it has been shown that accuracy of our proposed model out performs all previous models presented in [18, 28, 29]. It's recommended to increase number of features in feature vector to increase the accuracy rate.

References

[1] P. Anand, T. Ajitha, M. Priyadharshini, and M. G. Vaishali, "Content based image retrieval CBIR using multiple features for texture images by using SVM classifier," *International Journal of Computer Science & Communication Networks*, vol. 2, no. 2, pp. 33–42, May 2014.

[2] A. R. Backes, C. Casanova and O. M. Bruno, "Color texture analysis based on fractal descriptors", *Pattern Recognition*, vol. 45, pp. 1984–1992, Nov. 2011.

[3] A. G. R. Balan, A. G. M. Traina, C. Traina Jr., P. M. Azevedo-Marques, "Fractal analysis of image textures for indexing and retrival by content", in *Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems*, 2005.

[4] A. G. R. Balan, A. J. M. Traina, A. J. M. Traina, and P. M. Azevedo-Marques, "Fractal analysis of image textures for indexing and retrival by content," *18th IEEE Symposium on Computer-Based Medical Systems*, pp. 581–586, 2005.

[5] A. S. Baranov, R. Graml, F. Pirchner, and D. O. Schmid, "Breed differences and intra-breed genetic variability of dermatoglyphic pattern of cattle," *Journal Of Animal Breeding & Genetics*, vol. 110, no. 5, pp. 385–392, 1993.

[6] U. G. Barron, *Muzzle Pattern as a Biometric Identifier for Cattle*, The BioTrack Project. December 6th, 2005.

[7] A. F. Costa, G. Humpire-Mamani, A. M. Traina, "An efficient algorithm for fractal analysis of texture", in *25th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI'12)*, pp. 39–46, 2012.
[8] T. M. Deserno, S. Antani, and R. Long, "Ontology of gaps in content-based image retrieval," *Journal of digital imaging*, vol. 22, no. 2, pp. 202–15, 2009.
[9] I. El-Henawy, H. M. El Bakry and H. M. El Hadad, "Bovines muzzle identification using box-counting", *International Journal of Computer Science and Information Security*, vol. 12, no. 5, pp. 29–34, May 2014.
[10] I. El-Henawy, H. M. El Bakry, H. M. El Hadad, "Cattle identification using segmentation-based fractal texture analysis and artificial neural networks," *International Journal of Electronics and Information Engineering*, vol. 4, No. 2, pp. 82–93, 2016.
[11] I. El-Henawy, H. M. El Bakry, H. M. El Hadad, "Muzzle classification using neural networks," Accepted and Under Publication in the *International Arab Journal of Information Technology*.
[12] R. Giot, M. El-Abed, and C. Rosenberger, "Fast computation of the performance evaluation of biometric systems: Application to multibiometrics," *Future Generation Computer Systems*, Special Section: Recent Developments in High Performance Computing and Security, vol. 29, no. 3, pp. 788–799, 2013.
[13] A. Ismail, A. E. Hassanien, and H. M. Zawbaa, "A cattle identification approach using live captured muzzle print images," in *Advances in Security of Information and Communication Networks Communications in Computer and Information Science*, vol. 381, pp. 143–152, 2013.
[14] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction to Biometrics*, Springer, 2011.
[15] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for VideoTechnology*, vol. 14, no. 1, pp. 4–20, 2004.
[16] U. Kandaswamy, D. Adjeroh, and M. Lee, "Efficient texture analysis of SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 9, pp. 2075–2083, 2005.
[17] H. Nagahashi, M. Yamaguchi, M. Sakamoto, and A. Hashiguchi, "Multifractal feature based cancer detection for pathological images", in *IEEE 5th International Conference on Bioinformatics and Biomedical Engineering (iCBBE'11)*, pp. 1–4, May 2011.
[18] A. Noviyanto, and A. M. Arymurthy, "Beef cattle identification based on muzzle pattern using a matching refinement technique in the SIFT method", *J. Anim. Breed Genet*, vol. 99, no. C, pp. 77–84, 2013.
[19] N. Patel, D. Singh, "An algorithm to construct decision tree for machine learning based on similarity factor", *International Journal of Computer Applications*, vol. 111, no. 10, Feb. 2015.
[20] N. Patil and R. Lathi, "Comparison of C5.0 & CART classification algorithms are using pruning technique", *International Journal of Engineering Research & Technology*, vol. 1, no. 4, 2012.
[21] D. N. Ponraj, M. E. Jenifer, P. Poongodi, J. S. Manoharan, "A Survey on the Preprocessing Techniques of Mammogram for the Detection of Breast Cancer," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, no. 12, pp. 656–664, Dec. 2011.
[22] Z. S. Rahman, K. Ahmed, "Flag identification using support vector machine", *Journal of Information Technology*, vol. 2, pp. 11–16, June 2013.
[23] N. Sarker, B. B. Chauduri, "An efficient box-counting approach to compute fractal dimension of image", *IEEE Transactions on Syst. Man Cybernet*, vol. 24, pp. 115–120, 1994.
[24] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, London, 1982.
[25] Y. Y. Song, L. U. Ying, "Decision tree methods: applications for classification and prediction", *Shanghai Archives of Psychiatry*, vol. 27, no. 2, pp. 130–135, 2015.
[26] C. Sun, F. Jiang, and S. H. Jiang, "Research on RFID applications in construction industry," *Journal of Networks*, vol. 8, no. 5, pp. 1221–1228, May 2013.
[27] K. Thangavel, R. Roselin, "Mammogram mining with genetic optimization of ant-miner parameters", *International Journal of Recent Trends in Engineering*, vol. 2, no. 3, pp. 67–69, Nov. 2009.

[28] A. Tharwat, G. Tarek, A. E. Hassanien, H. A. Hassanien, F. M. Tolba, "Cattle identication using muzzle print images based on texture features approach." in *Advances in Intelligent Systems and Computing*, pp. 217–227, 2014.
[29] A. Tharwat, T. Gaber, and A. E. Hassanien, "Cattle identification based on muzzle images using gabor features and SVM classifier", in *Communications in Computer and Information Science*, pp. 236–247, 2014.
[30] N. Theeta-Umpon, "Fractal dimension estimation using modified differential box-counting and its application to MSTAR target classification", in *Proceedings of the IEEE International Conference on System, Man and Cybernetics*, vol. 2, pp. 537–541, 2002.
[31] M. Vlad, R. A. Parvulet, and M. S. Vlad, "A survey of livestock identification systems," in *Proceedings of the 13th WSEAS International Conference on Automation and Information, (ICAI'12)*, Iasi, Romania: WSEAS Press, pp. 165–170, 2012.

Ibrahim El-Henawy received the M.S. and Ph.D. degrees in computer science from State University of New York, USA in 1980 and 1983, respectively. Currently, he is a professor in computer science and mathematics department, Zagazig University. His current research interests are mathematics, operations research, statistics, networks, optimization, Intelligent Computing, Computer Theory, digital image processing, and pattern recognition.

Hazem M. El-Bakry (Mansoura, EGYPT 20-9-1970) received B.Sc. degree in Electronics Engineering, and M.Sc. in Electrical Communication Engineering from the Faculty of Engineering, Mansoura University - Egypt, in 1992 and 1995 respectively. Dr. El-Bakry received Ph. D degree from University of Aizu - Japan in 2007. Currently, he is associate professor at the Faculty of Computer Science and Information Systems - Mansoura University - Egypt. His research interests include neural networks, pattern recognition, image processing, biometrics, cooperative intelligent systems and electronic circuits. In these areas, he has published many papers in major international journals and refereed international conferences. According to academic measurements, now the total number of citations for his publications is 2817. The H-index of his publications is 28. Dr. El-Bakry has the United States Patent No. 20060098887, 2006. Furthermore, he is associate editor for journal of computer science and network security (IJCSNS) and journal of convergence in information technology (JCIT). In addition, is a referee for IEEE Transactions on Signal Processing, Journal of Applied Soft Computing, the International Journal of Machine Graphics & Vision, the International Journal of Computer Science and Network Security, Enformatika Journals, WSEAS Journals and many different international conferences organized by IEEE. Moreover, he has been awarded the Japanese Computer & Communication prize in April 2006 and the best paper prize in two conferences cited by ACM. He has also been awarded Mansoura university prize for scientific publication in 2010 and 2011. Dr. El-Bakry has been selected is who Asia 2006 and BIC 100 educators in Africa 2008.

Hagar Mohamed Reda El Hadad graduated from Faculty of Computers and Information, Minia University, Minia, Egypt in 2008. Hagar received her master degree in 2011 in Information Systems from the Faculty of Computers and Information, Mansoura University, Mansoura, Egypt. Hagar is teaching assistant in faculty of computer and information systems, Beni-Suef University Beni-Suef, Egypt. Hagar main research interests are in the areas of data mining such as (text - numbers - Images).

An Empirical Evaluation of Security tips in Phishing Prevention: A Case Study of Nigerian Banks

Abdul Abiodun Orunsolu[1], A. S. Sodiya[2], A. T. Akinwale[2], B. I. Olajuwon[3], M. A. Alaran[1], O. O. Bamgboye[1], and O. A. Afolabi[4]

(Corresponding author: Abdul Abiodun 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]
Department of General Studies, Moshood Abiola Polytechnic, Abeokuta, Nigeria[4]

(Received Dec. 21, 2016; revised and accepted Feb. 20, 2017)

Abstract

To shield users from phishing scams, various online brands send security tips as email, SMS and online posts to their customers. This paper presents the first empirical evidence about the effectiveness of the security tips in phishing prevention from customers' perspective in Nigerian financial sector. We developed anti-phishing questionnaire which captured the basic essence of most security tips messages and formulate two hypotheses. We then test our hypotheses using an experimental method with 247 participants. The experimental method was divided into a Pretest which evaluates our first hypothesis and a Posttest which evaluates our second hypothesis. The results illustrate that most customers do not understand the security tips at statistical confidence interval of 95% using the Mann Whitney Test.

Keywords: Anti-phishing; Electronic Commerce; Phishing Cues; Security Tips; User Awareness

1 Introduction

During the last two decades, the Internet and the computers have dramatically transformed the global society. Most people are attracted to computers and online services due to the advantages of speed, automatic availability of resources, zero-delay in service delivery, etc. In recent times, however, the threat of cybercrime has become an important factor for the global economy [23, 30]. The attractiveness of cybercrimes may not be unconnected with the fact that most online users have poor understanding of internet-based services [4, 9, 29]. In addition, most users see security as a secondary issue when accessing Internet resources. In the same vein, inadequate cross-border legislation and anonymity on the part of con-artists contributed to the unpopular incidences of cybercrimes. These cybercrimes come in various forms and shades. Whereas some of these cybercrimes merely constitute nuisance to users, other cybercrime methods result in huge financial and brand damages to both service providers and users. One example of the latter is phishing attacks which has attracted extensive press coverage [22].

Phishing has become a crippling problem for many of today's Internet users as the attack continuously escalate in number and sophistications [11]. Phishing is a social engineering attack that describes an attempt to deceptively acquire personal and financial information via electronic communication with malicious intent [20]. A typical phishing attack begins with unauthenticated message crafted by con-artists. This crafted message is then sent to a large number of internet users in form of email, SMS, e-chat and web post. One of the core features of this message is its deceptive view. The user falls for a phish by actively following the instruction in the message through performing a click or download action. In the end, the user's action results in the execution of the phisher's payload. The payload usually involves financial damages on the part on unsuspecting users as well as service providers [23, 33].

To respond to the phishing threats, various anti-phishing solutions have been proposed. These countermeasures are either software enhancement methods or anti-phishing education methods. The software enhancement methods involve the use of heuristics-based techniques, blacklist approaches, whitelist approaches, multi-channel authentication mechanisms, etc. In general, the software enhancement methods are categorized as classification techniques and non-classification techniques [16]. On the other hand, anti-phishing education method refers to user-awareness/education and government legislation designed to mitigate phishing attacks [2, 21]. However, in spite of the existence of these countermeasures, the incidences of phishing attacks continue to defeat the current anti-phishing techniques [4, 11]. For instance, COREN online security report indicated an increase of 51% in phishing sites in the first quarter of 2015. In a similar vein, a white paper from the Central Bank of Nigeria estimated that about $250 million was lost to cybercrime in 2013 [26]. In addition, RSA's online fraud report showed estimated losses of over $5.9 billion by global organizations in nearly 450,000 attacks in 2013.

The reality of this arms race between the phishers and the security community demands a new countermeasure to address the existing research gaps. New anti-phishing countermeasures are proposed every day to match the sophistication of newer phishing attacks [23, 29]. Various global organizations are coming up with strategies to reduce the incidences of phishing attacks. Today, most financial institutions provide security tips to their customers to enable them identify online scams. For instance, Nigerian banks send security tips via SMS, email and online security centers to provide anti-phishing education to their customers. Do bank customers understand these security tip messages? Are the security tips messages effective for users to identify a typical online scam? Motivated by these questions, we conducted and reported a two-staged simulated study to evaluate the effectiveness of the security tips. The simulated study consists experiments in which users are asked to judge the benignity of interfaces bearing phishing cues discussed in security tips message. In the first stage, a hypothesis was proposed to evaluate the effectiveness of the security tips messages. In the second stage, a hypothesis was proposed to evaluate the effectiveness of an enlightenment phase introduced into the study. The two hypotheses were validated using two-tailed statistics at α=0.05.

The rest of the paper is organized as follows: Section 2 presents related work. The data collection and experimentation of our proposed study are discussed in Section 3. In Section 4, the results of our findings are presented. Conclusions and future works are presented in Section 5.

2 Related Works

In this section, we begin by examining related works on why people fall for phishing. Later, we provided related works from software enhancement methods and anti-phishing education methods. These methods are commonly referred to as technical and non-technical approaches respectively.

2.1 Why Phishing Works?

Researchers have identified a number of reasons why people become victims of phishing attacks. One of the earliest works that investigated this problem was Dhamija et al. [9] The authors identified lack of computer system knowledge, lack of knowledge of security and security indicators, visual deception and bounded attention as reasons why people fell for phishing. 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 another development, Jagatic et al. [17] researched into the concept of spear-phishing attacks. In this form of phishing scams, phishers used specific knowledge of individuals and their organizations to launch attack. For instance, whereas a typical phishing message comes with "Hello dear Customer", a spear-phishing scams uses "Hello dear Tom" which is the actual name of the target. The authors found out people were 4.5 times more likely to fall for phish sent from an existing contact over standard phishing attacks. This is why social networking sites like Facebook are now more patronized by phishers.

Jakobsson et al. [18] provided useful insights on why phishing works using demographic data. The authors revealed users' sensitivity to variety of common trust indicators such as logos, padlock icons, etc. when navigating web pages.

On gender vulnerability to phishing attacks, Sheng et al. [34] revealed that women are more vulnerable than men due to their less exposure to technical knowledge.

Appealing to people's sense of greed is an ancient technique now adapted to the digital world especially in phishing scams. This kind of phishing scam may look like online survey in which unsuspicious users are promised some financial returns for participating in the survey exercise. In a similar vein, phishers might pose as reputable relief agency asking for help for victims of recent natural disasters to appeal to people's sense of emotion. Most unsuspecting users may not suspect anything negative even when asked to provide their financial details because of some gory pictures that usually accompanied such campaigns [30].

2.2 Technical Approaches to Defeat Phishing

Technical approaches are anti-phishing defense mechanisms in which software are enhanced through design upgrades to mitigate phishing attacks. They are often deployed as client or server-side defenses. The main motivation for this method is to bridge the gap that is left due to the human error or ignorance. For example, PhishNet [31] proposed an active blacklist approach in which new malicious URLs can be effectively predicted from the existing blacklist entries. This is achieved by processing blacklisted URLs and producing multiple variations of the same URL using IP address equivalence, query string substitution, brand name equivalence, directory structure similarity and top level domain replacement. In this way, multiple variations of the same URL called children are obtained. In order to filter non-existent children URLs, the system performed DNS query, TCP connect, HTTP header response and content similarity. The approach achieved remarkable results during real-time blacklist feeds against new malicious URLs. However, the problem of false positives still exists.

PhishZoo [3] built profiles of trusted websites based on fuzzy hashing techniques in a whitelisted-based approach. The approach also used blacklisting and heuristic approaches to warn users about malicious sites. This approach compared the stored profile of authentic sites with the content of sites under investigation. The approach achieved significant accuracy rate of about 96% with the possibility of defeating zero-day attack. However, there is lack of generalization to new phishing due to human interventions.

Han et al. [13] developed an Automated Individual White-List (AIWL) in which the record of well-known benign sites visited by users is kept. In this way, AIWL maintains a record of every URL along with its Login User Interface information where the user input his or her details to prevent unhealthy

disclosure of confidential information to malicious sites. The LUI information maintains by AIWL for any suspicious website include the URL, the Input Area and the IPs. The URL refers to the Unified Resource Locator of the website. The input area includes the form username path and password path. The IPs is a list of legitimate IP addresses mapping to a URL. This method is very effective against pharming and dynamic phishing attacks.

Gowtham et al. [12] present a dynamic defense approach in which direct and indirect links associated with a malicious page is generated. In this way, the target domain set is constructed as input into Target Identification algorithm to recognize a phishing page. Using DNS lookup and IP address resolution, the suspicious page can be predicted without the use of machine learning algorithms or existing restriction lists. However, the prediction of this approach is largely dependent on the TF-IDF algorithms, search engine speed and DNS lookup.

Huang et al. [15] presented a practical authentication service in which the need for preset user password is eliminated during information flow between the client and the server. The proposed approach involves two processes namely a registration process and a login process with four participating entities: websites, instant messaging service provider, users and phishers. In the registration process, a user chooses a unique account name, selects a login password, fills in all the required information fields, completes an additional IM account registration and provides at least one type of personal contact information. The approach cannot detect XSS attacks and phishing sites hosted on compromised domains.

2.3 Non-Technical Approaches to Defeat Phishing

Non-Technical approaches are methods or training or government policies that are directed to users to help them understand the dangers of phishing attacks. This approach is substantiated by the submission of Hong [14] that "it doesn't matter how many firewalls, encryption software, certificates, or two-factor authentication mechanisms an organization has if the person behind the keyboard falls for a phish". This statement confirms the age-long belief that the computer users are the weakest link in the field of information security.

Phishing education is meant to protect individual users against phishing threats and are focused on online training materials, testing, and situated learning. Online training materials have been published by government organizations, non-profits security institutions and businesses. These materials explain what phishing is and provide tips to prevent users from falling for phishing attacks. The security tip message is an example of this method and thus, the investigation of its adoption by the Nigerian banks is the main purpose of this work.

Kumaraguru et al. [24] conducted research works which focuses on educating users about phishing and helping them make better trust decisions. They identified a number of challenges for end-user security education in general and anti-phishing education in particular. They developed an e-mail-based anti-phishing education system called "PhishGuru" and an online game called "Anti-Phishing Phil" that teach users how to use cues in URLs to avoid falling for phishing attacks. Their test result suggests that while automated detection systems should be used as the first line of defence against phishing attacks, user education offers a complementary approach to help people better recognize fraudulent e-mails and web sites.

Similarly, Sheng et al. [34] conducted a role-play survey among 1,001 online respondents to study both the relationship between demographics and phishing susceptibility and the effectiveness of several anti-phishing educational materials. Their work shows that educational materials reduced users' tendency to enter information into phishing web pages by 40 percent. However, some of the educational materials they tested also slightly decreased participants' tendency to click on legitimate links.

Arachchilage and Love [6] proposed and developed a game framework which enhanced user avoidance

behavior through motivation by protecting users from phishing attacks. A theoretical model derived from Technology Threat Avoidance Theory (TTAT) was used in the game design framework. The TTAT identified the issues that the game design framework needed to address by developing threat perceptions that motivated individuals to avoid phishing attacks and use safeguarding measures. The study emphasized that avoidance motivation for phishing attack was significantly determined by the interaction of perceived threat and safeguard effectiveness. In addition, the study bridged the gap in the software-based anti-phishing approaches through user awareness model using a game design approach. However, the study did not provide evidence to address the interaction of perceived threat and safeguard effectiveness in the game design framework.

In recent studies, Mohammed et al. [29] 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. Similarly, Kathryn et al. [20] conducted a role play experiment of people's ability to differentiate between phishing and genuine emails. In the study, the authors explicitly informed half of the participants about the goal of the study. Their results indicated that informed users were significantly better at discriminating between phishing and genuine emails than the uninformed participants. The main goal of user awareness approach is to reduce the vulnerabilities of human factor because it does not matter how many defense mechanisms is available if the user behind the keyboard falls for a phish.

On government legislations to defeating attacks, Larson [25] recommended that courts should consider either large scale damages against individual phishers or secondary liability against Internet Service Providers under the areas of either intellectual property or unfair competition law. The addition of secondary liability to anti-phishing efforts might motivate ISPs to become actively involved in anti-phishing efforts since ISPs are best positioned to prevent phishing scams.

Lovet [27] highlighted the judicial challenges and recommended for expedient international cooperation and harmonization of cyber-criminal offences amongst legal systems beyond borders.

3 Method

3.1 Participants

A total of 427 participants were involved in our study (245 male and 182 female). These participants were recruited from three different academic institutions within Ogun State, South West Nigeria. The three institutions namely Tai Solarin College of Education Omu Ijebu, Gateway ICT Itori and Moshood Abiola Polytechnic Abeokuta are located in the three different Senatorial Districts of the State. The actual institutions used for the research were selected based on the ease of accessibility of the researcher and participants were selected using judgment sampling technique - a non-probability sampling method where the researcher selects units to be sampled based on their knowledge and professional judgment. Informed consent of selected participants was obtained prior to their participation and participants were assured that participation is voluntary and they can withdraw from the study at will without any negative consequence. In addition, young and middle-aged adults can be easily found in tertiary institution environment and this demographic strata are most suitable for technological driven-research like security of e-commerce transaction. The participants' demographics is provided in Table 1. From Table 1, majority of the respondents are in the 18 - 25 years age bracket. About 58% of the respondents are male and about 63% rated themselves as having average computer literacy. Over 90% of the respondents have tertiary education qualifications.

Table 1: Summary of respondents demographics

Age (in years)	Frequency	%
< 18	11	2.58
18 - 25	331	77.52
26 - 35	58	13.58
36 - 50	26	6.09
50+	1	0.23
	427	100
Gender		
Male	245	57.51
Female	182	42.49
	427	100
Computer Literacy		
None	3	0.47
Low	31	7.28
Average	269	63.15
High	124	29.11
	427	100
Highest Academic Achievement		
SSCE/WASC	38	8.94
NCE/OND	235	55.29
HND/BSC	129	30.35
MSC/PH.D	25	5.41
	427	100

3.2 Materials

3.2.1 The Anti-phishing Questionnaire

In this study, we designed an Anti-Phishing Questionnaire (APQ) to test user's ability to detect fake SMS, email and webpages which security tips messages have provided some information about. The objective of the APQ is to investigate our two-staged experimental procedure in the next session. The APQ consists of Data Collection phase, Site Selection phase, Pretest phase, Enlightenment phase and Posttest phase. Figure 1 shows some participants during the experimental session. In the Data Collection phase, personal details of respondents such as gender, educational qualification, computer literacy, types of online services used, receipt of bank alerts etc. are captured. Based on the user's responses, the APQ categorizes users as informed or non-informed. To determine this, the application assigned a value 1 or 0 to each data supplied by the user. The value 1 indicates that the data has significant influence on the user's knowledge of computer/Internet and its services. On the other hand, the value 0 indicates that the data does not significantly influence the users' knowledge. For example, age, gender and Local Government of Residence do not indicate ones' knowledge of computer and its associated services. For a user to be classified as informed, the user must score 5 points on all the significant data values. Thus, an informed user is a person that has appreciable knowledge of how computers/Internet works and a non-informed user is one without adequate knowledge of the computer/Internet.

After the status of respondent has been determined, the user is allowed to select a number of online

services or webpages they are accustomed to. In this way, the application is able to customize the user's test messages in the Pretest phase to the webpages that are known to them. For example, if a user selects First Bank Nigeria PLC and OLX as brands that are known to him, then the application will ask such user to identify the Home page of such online brands. Table 2 presents the statistics of participants on the type of bank services used.

Table 2: Types of bank services used by the participants

Type of Service used n = 427	Frequency	%
ATM	390	91.33
MOBILE BANKING	218	51.05
INTERNET BANKING	155	36.30
ONLINE BANKING	174	40.75
ONLINE SHOPPING	170	39.81
QUICKTELLER	97	22.72

The Third phase of the application is the Pretest phase where the users are asked to judge the status of randomized 5 messages consisting of email, SMS and Webpages with varying phishing cues. In each of the five test messages, certain phishing cues are embedded in the messages to test the understanding of the security tips from the users' perspective. Each message provides a Yes or No answer displayed at the upper part of the page for users to judge. The following parameters are tested in each of the messages:

1) In Link;

2) URL with more than three dots;

3) URL using IP Numbers;

4) Fake Bank Verification Number (BVN) message;

5) Fake Customer Care email message.

The first three parameters are common phishing cues which are mostly used by phishers to deceive online users. For instance, In Link is used when a URL within a webpage is directed to another domain. Cyber criminals use this method to divert user's attention to their fake page without the user being aware. In addition, fake websites with IP based address and URL with more than three dots were created by using the logo and layout of the corresponding real bank sites or online shopping stores. The remaining parameters are mostly used as SMS or email ploy to deceive bank users. This SMS or email message contains private number and uses sense of urgency to deceive users. In addition, bad grammars are common in this message. In order to protect the respondents from real-world phishing messages, we downloaded these sites and messages for offline use and hosted them on our local machine.

Our experiment is designed to ask users identify fake messages and then determine users' performance based on their understanding of the security tips. This leads to the formulation of our first hypothesis as:

H0: Security tips are well understood by bank customers;

H1: Security tips are not well understood by bank customers.

In the fourth stage, the APQ application trains the users by connecting them to a middleware layer application. The middleware layer application contains information on the implications and meanings of the phishing cues encountered by the users in the Pretest session. This is to provide enlightenment to users on why their judgments is correct or not. The middleware layer application was implemented using the HTML Agility Pack (HAP). It is a .NET code library that allows parsing of HTML files.

In the final stage of the study, users are presented with a new set of messages to evaluate their understanding of the enlightenment stage and this leads us to formulate the second hypothesis as:

H0: Enlightenment session does not assist bank customers to defeat phishing attacks;

H1: Enlightenment session assists bank customers to defeat phishing attacks.

The purpose of the second hypothesis is to evaluate if the embedded enlightenment phase can assist users to defeat phishing attacks. This hypothesis determines the efficacy of the enlightenment session. This final stage is called Post-test. The Post-test evaluates the same five parameters used in the pretest but with a new set data.

3.2.2 The Experimental Method

The experimental method is a two-staged process consisting of Pretest and Posttest. The Pretest process is used to evaluate our first hypothesis on the security tip message. Table 3 indicates that the problem is worthy of investigation as about 91.80% of the participants claimed to have received such message at one time or another. On the other hand, the Posttest is used to evaluate our second hypothesis on the embedded enlightenment phase in the study. Each participant commenced a session by going through a welcome page. The welcome page contains the instruction on the experiment. Then, the participant is asked to fill in personal data and the number of online services/webpages they patronized. Most participants took an average of 3.4 minutes to complete this preliminary stage. In the Pretest stage, participants are asked to judge 5 randomized messages with varying phishing cues.

Figure 1: Some participants during the experimental session

These messages consisting of one (1) fake BVN message, one (1) fake email messages and three (3) webpages of the banks/online shopping sites used by the participants. The fake BVN message contains bad grammars and private number call center as portrayed in most of such unsolicited fake messages. The fake email message asked the user to update their details. On the other hand, the three fake

webpages used the logo and visual similarity of known financial brand and online shopping sites. The first fake webpage used In-Link attributes as a phishing cue. The second webpage used an IP-based URL with look and feel of known financial institution and the third fake webpage used the three dots attributes. Each trial consisted of a message (e.g. webpage or BVN or email) shown for as long as the participant responded. The response consists of a panel with a question," Is this a legitimate website" or "Is this a legitimate BVN message" or "Is this a legitimate email" and a "Yes" or "No". This panel is displayed at the header part of each investigative page. Each participant took about 5.6 minutes to complete the Pretest process. The participants' responses are recorded into our database and exported in a spreadsheet format for easy statistical analysis. In the next phase, an enlightenment user interface is displayed prompting users to check the correctness of their answers.

Table 3: Types of messages received by the respondents

Type of message received	Frequency	%
Security Message	392	91.80
Transaction Alert	386	90.40
Unrequested Messages	258	60.42
BVN Update Message	215	50.35

This is to provide instant education to users on why their judgment is correct or not. Figure 2 shows the enlightenment user interface. Each participant took an average of 2 minutes to check the enlightenment user interface. In the final stage, a new set of 5 randomized messages with varying phishing cues are displayed for users to identify. These 5 messages are equally based on the parameters tested in the Pretest. It is interesting to note that while the messages in the Pretest are customized based on the user's response in Data collection and Site selection phase, the Posttest interfaces are shuffled without regards to the conditions of the Pretest. Thus, the Posttest is application-oriented. For instance, if a participant selects GT Bank as his financial institution in data collection phase, the application will ask the user to identify a GT Bank home page. However, in Posttest the same participant may be asked to identify the Home page of First Bank Nigeria PLC or any online brands within database of the local machine. This is because we assume that the enlightenment session should be enough for user to navigate and identify any messages or website. Most participants completed the Posttest process faster than the Pretest using an average of 1.99 minutes.

4 Analysis and Results

The statistical results in this paper are reported at the significance level (α) of 0.05 using two-tailed Test. Figure 3 shows the summary of the respondents, performance to each phishing cue tested in the Pre-test. About 58.91% failed on the tree dots attributes, 58.59% failed on the use of IP attribute (represented as Ipadd on the graph) while 58.73% failed on the In link. In addition, 76.71% could not correctly identify a phished BVN Update message while, 74.24% could not correctly identify a fake e-mail message. The Mann-Whitney U test (the nonparametric equivalent of the independent t-test) was used to ascertain if there is a significant difference between participants who passed and those participants who failed on each of the test attributes. Furthermore, the Wilcoxon W Test was used to examine the difference in each of the attribute. This analysis show that in spite of the security messages the participants claimed to have received from their banks, they were still unable to correctly identify malicious messages or online scams.

Figure 2: Enlightenment User Interface (EUI)

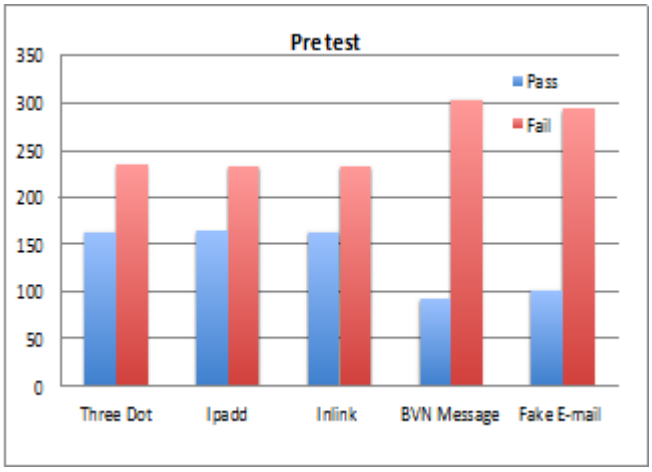

Figure 3: Evaluation analysis of the pretest session

The result in Table 4 shows that those who failed the Pretest are significantly more than those who passed. This suggests that the participants do no really understand the security messages they receive from their banks. Hence, we reject the null hypothesis that security tips are well-understood by bank customers. Table 4 presents the Test statistics for the Pretest process.

Table 4: Test Statistics for the Pretest on the five phishing cues

	Three Dots	IP-based URL	In Link	BVN Msg	Fake E-mail
Mann-Whitney U	12539.500	11384.000	11233.000	14920.500	15073.000
Wilcoxon W	40034.500	38879.000	38494.000	42415.500	42568.000
Z	-6.735	-7.919	-7.619	-4.618	-4.479
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.000

Furthermore, the effectiveness of the short enlightenment session carried out after the Pretest was subjected to statistical evaluation. Figure 4 presents the comparative analysis of the Pretest and Posttest results. In the results, there is about 13.51% increase in the number of respondents who passed the Posttest on three dots attribute compared to those who passed the Pretest. For the IP attribute (called Ipadd), there is a 13.42% in the pass rate, and a 15.87% increase in the pass rate for In link attribute. 45.49% more respondents were able to correctly identify a phished BVN update message after the phishing enlightenment exercise, while 43.67% more respondents have been able to identify a fake e-mail after the enlightenment exercise.

To test the second hypothesis, we used the Wilcoxon's Signed ranked test (the non-parametric equivalent of paired t-test). The result of the analysis is presented in Table 5. The result in Table 5 shows that for all the attributes, the posttest result is significantly higher than the pretest. This suggests that the enlightenment is effective and hence we rejected the H0. On the basis of this rejection, we can conclude that security enlightenment can assist bank customers to defeat phishing attacks.

Table 5: Test Statistics[a]

	ThreeDots(Post) - Three Dots (Pre)	Ipadd (Post) - Ipadd (Pre)	In link (Post) - In link (Pre)
Z	-4.630[b]	-4.287[b]	-5.345[b]
Asymp. Sig. (2-tailed)	.000	.000	.000
	BVN MSG (Post) - BVN MSG (Pre)	FAKE E-Mail (Post) - FAKE E-Mail (Pre)	
Z	-3.109[b]	-1.659[b]	
Asymp. Sig. (2-tailed)	.002	.027	

a. Wilcoxon Signed Ranks Test;
b. Based on positive ranks.

It is interesting to note some basic differences between our enlightenment process used in the study and the security tips messages sent by Nigerian banks. Whereas most participants do not understand the import of the security tips as evident in their performance in the Pretest, the Posttest results showed the enlightenment session primed some of the participants to achieve better performance. Similarly, the enlightenment session was able to prime the participants that the look and feel of a message or website is not enough to determine the benignity of such message or website. There is no other way that this can

be achieved than through a knowledge process used in our study. A security tip may not be sufficiently enough to demonstrate this process through a textual presentation. Thus, the enlightenment session further lends credence to earlier research (e.g. [6, 20]), that users' performance increases when primed to identify fake messages through active process.

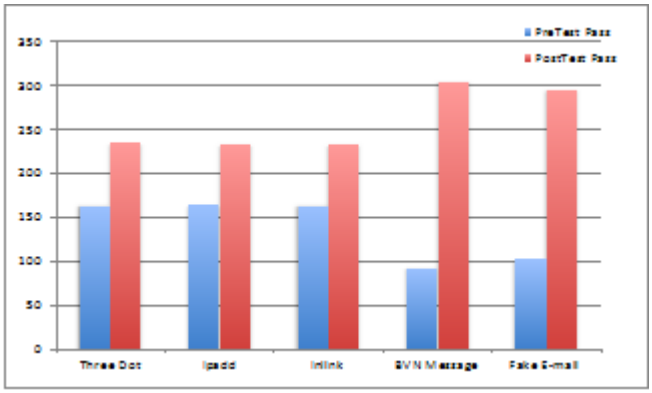

Figure 4: Comparative analysis of Pretest and Posttest

5 Discussions

This study re-affirms the evidence of earlier researches that people are poor at identifying phishing messages [7, 9, 29] This is because most people lack the basic understanding of the Internet technology. Phishing and online scams exploit this lack of knowledge by using forged email header, fraudulent URLs, absence of https or closed padlock, etc. This is evident as most respondents incorrectly labeled phishing pages as genuine during the Pretest session despite the presence of conspicuous incorrect or unrelated URL. The visual deception cue also contributed to the participants' poor judgments as about 59% failed to recognize fake webpages. From our observation in the Pretest session, most participants focused on the logo and images of familiar brands as evidence of benignity of a webpage. The implication of this is that most banking customers in Nigeria may still fall for deceptive online scams where logos and images of known brands are employed unless an enlightenment process through multimedia is adopted. Although, Dhamija et al. [9] demonstrated that users were easily fooled through visually similar images, logos, texts, homographic attacks, etc. about a decade ago, our findings show that the reality is still the same in Nigeria. Whereas the success rate of Dhamija et al. [9] was 58%, our study recoded an abysmally 41.26% success rate on visually deceptive cue. However, Mohammed et al. [29] shows that participants that are primed to understand phishing attacks scored 64% success rate in their own study. This result is consistent with our Posttest process as participants performed better (average success rate of 55.56%) as a result of the priming process.

However, participants in our study are still more likely to be deceived by phishing scams as their performance p in Posttest on In Link, IP address and three dot attribute were not very impressive i.e. $13\% < p < 16\%$. From our observation, phishers are likely to employ these attributes than SMS or e-mail with bad grammars or private call number. In the light of this, regular enlightenment session on these phishing tricks need to be available to users. In fact, most respondents in our study agreed to this as they acknowledged that the study has really impacted positively on their judgments of online products, SMS and email.

Our findings with regard to email and BVN showed that participants were adequately primed through the enlightenment process as their performance were better on these two metrics. This will assist the participants to defeat such malicious messages that now flood mobile-networked-enable devices [7]. This is impressive as most users will now respond to these messages on sound of notification peeps with adequate knowledge. This is supported by Kathryn et al. [20]. In their work, the authors classified participants as informed and non-informed based on the priming the participants with the purpose of the study. Interestingly, most primed participants significantly identified phishing emails than the unprimed participants. This is consistent with our study as the Posttest results provide a better performance than the Pretest.

5.1 Contribution to Knowledge

This study, arguably represents the first phishing study from the non-technical perspective from West Africa. The study quantified and empirically demonstrated the effectiveness of the security tips message. It is our belief that our findings have important implications for the future design of security tips messages. This study shows that the current security tips have not assisted the participants to identify malicious message. This study is timely now that most online users respond to online message instantly due to the proliferation of network-enabled mobile platforms.

This study is unique as the data collected for the research analysis is not restricted to a single academic environment as common with most anti-phishing studies. The data collection was done in three different academic environments in three different major towns within Ogun State. This means that our result is significant and our generalizations may reflect the performance of an average bank customer.

Our findings also showed that in spite of large numbers of OND/NCE and HND/BSC in the population sample, the failure rate in the Pre-test study was very high. The implication of this is that most of these individuals have poor knowledge of internet in spite of their patronage of online services and products.

6 Conclusions and Future Work

The primary goal of this research work was to examine the effectiveness of the security tips used by the Nigerian banks as a form of education for their customers. In these messages, information are provided on how users can identify online scams and which actions the users should avoid. The work provided an empirical analysis of the effectiveness of these messages through a stimulated phishing study using malicious cues embedded in SMS, email and webpages. A total number of 427 respondents participated in the study and our findings showed that most respondents still failed to recognize a phishing scams in spite of having received a security tip. However, there was an appreciable increase in the success rate in the Posttest which was preceded by an enlightenment session.

In the future, we hope to investigate how gender affects the users' evaluation and the role of their discipline. In addition, we hope to provide a useful framework on how enlightenment message should be constructed as against the security tip messages.

References

[1] N. Abdelhamid, A. Ayesh, F. Thabtah., "Phishing detection based associative classification data mining", *Journal of Expert Systems with Applications*, Elsevier Press, 2014.

[2] N. Ajaya, R. Md Lutfor, S. Nitesh, H. Leane, M. Clerc, "A multi-modal neuro-physiological study of phishing detection and malware warning", in *Proceedings of CCS*, Denver, Colorado, USA, 2015.

[3] A. Afroz, R. Greenstadt, "PhishZoo detecting phishing websites by looking at them", in *Proceedings of IEEE Fifth International Conference on Semantic Computing*, pp. 368–375, 2011.
[4] S. Antonio and P. Xavier, "Phishing secrets: History, effects and countermeasures", *International Journal of Network Security*, vol. 11, 2010.
[5] S. Aparna, K. Muniasamy, "Phish indicator: An indication for phishing sites", *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, pp. 481–487, 2015.
[6] N. Arachchilage, S. Love, "A game design framework for avoiding phishing attacks", *Journal of Computers in Human Behavior*, vol. 29, 2016.
[7] V. Arun, "Mobile device affordance:Explicating how smartphones influence the outcome of phishing attacks", *Computer in Human Behaviour*, vol 63, pp 198–207, 2016.
[8] K. Chen, J. Chen, C. Huang, C. Chen, "Fighting phishing with discriminative keypoint features", *IEEE Internet Computing*, vol. 13, no. 3, pp. 56–63, 2009.
[9] R. Dhamija, J. D. Tygar, M. Hearst, " Why phishing works", in *Proceedings of the IGCHI Conference on Human Factors in Computing Systems*, ACM Press, pp. 581–90, 2006.
[10] J. S. Downs, M. B. Holbrook, L. F. Cranor, "Decision strategies and susceptibility to phishing", in *Proceedings of the Second Symposium on Usable Privacy and Security*, 2006.
[11] R. Gowtham, I. Krishnamurthi, "A comprehensive and efficacious architecture for detecting phishing webpages", *Journal of Computers and Security Journal*, Elsevier, 2014.
[12] R. Gowtham, I. Krishnamurthi, S. Kumar, "An efficacious method for detecting phishing webpages through target domain identification", *Journal of Decision Support Systems*, Elsevier Press, 2014.
[13] 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.
[14] J. Hong, "The state of phishing attacks", Contributed Articles in the Communication of the ACM. Vol 55 No 1, 2012
[15] C. Huang, S. Ma, K. Chen, "Using one-time passwords to prevent password phishing attacks", *Journal of Network and Computer Applications*, Elsevier Press, 2011.
[16] R. Islam, J. Abawajy, "A multi-tier phishing detection and filtering approach", *Journal of Network and Computer Applications*, 2013.
[17] T. Jagatic, N. Johnson, M. Jakobson, F. Menczer, "Social phishing", *Communications of the ACM*, vol. 50, no. 10, 2007.
[18] M. Jakobsson, S. Myers, "Phishing and countermeasures: Understanding the increasing problem of identity theft", *Introduction to Phishing*, pp. 1–2, New York: John Wiley & Sons, Inc., 2007.
[19] Kapersky Lab, *Annual Security Report*, 2007.
[20] P. Kathryn, M. Agata, P. Malcolm, B. Marcus, J. G. Jakobsson, S. Myers, "The design of phishing studies: Challenges for researchers", *Journal of Computers and Security*, 2015.
[21] M. Khonji, Y. Iraqi, A. Jones, "Phishing Detection", *A Literature Survey*, IEEE, 2013.
[22] E. Kirda, C. Kruegel, "Protecting users against phishing attacks", *The Computer Journal*, vol. 49, no. 5, pp. 554–561, 2006.
[23] C. Konradt, A. Schilling, B. Werners, "Phishing: An economic analysis of cybercrime perpetrators", *Journal of Computers and Security*, vol. 58, pp 39–46, 2016.
[24] P. Kumaraguru, Y. W. Rhee, A. Acquisti, L. Cranor, J. Hong, "Protecting people from phishing: The design and evaluation of an embedded training email system", in *Proceedings of CHI*, 2010.
[25] J. S. Larson, "Enforcing intellectual property rights to deter phishing", *Intellectual Property and Technology Law Journal*, 2010.
[26] T. Longe, "Ensuring Information Security Assurance through Policy Framework", in *Proceedings of First National Cyber Security Forum*, Lagos, Nigeria, 2014.
[27] G. Lovet, "Fighting cybercrime: technical, juridical and ethical challenge", Proc. of the Virus Bulletin Conference, 2009
[28] M. Maurer, L. Hofer, *Sophisticated Phishers Make More Spelling Mistakes: Using URL Similarity Against Phishing*, Springer, 2012.

[29] A. Mohammed, A. Furkan, C. Sonia, "Why phishing still works: User strategies for combating phishing attacks", *International Journal of Human-Computer Studies*, vol. 82. pp. 70–82, 2015.
[30] Panda Labs, *2012 Annual Report Pandalabs*, Retrieved, Sept. 23, 2013.
[31] P. Prakash, M. Kumar, R. R. Kompella, M. Gupta, "Phishnet: Predictive blacklisting to detect phishing attacks", in *Proceedings of the 29th Conference on Information Communications*, San Diego, CA, USA, pp. 346–50, 2010.
[32] V. Ramanathan, H. Wechsler , "Phishing detection and impersonated entity discovery using conditional random field and latent dirichlet allocation", *Journal of Computers and Security*, 2013.
[33] RSA Anti-Fraud Command Center, em RSA Monthly Online Fraud Report, 2014.
[34] S. Sheng, M. Holbrook, P. Kumaraguru, L. F., J. Downs, "Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions", in *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, USA, 2010.

Orunsolu A. Abiodun is a Lecturer in the Department of Computer Science, Moshood Abiola Polytechnic, Abeokuta, Nigeria. He currently undergoes his Ph.D degree at the Federal University of Agriculture, Abeokuta. His research interests include 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.

Alaran M. Adunni is a Lecturer in the Department of Computer Science, Moshood Abiola Polytechnic, Abeokuta, Nigeria. She is presently on her Ph.D degree at the Federal University of Agriculture, Abeokuta where her research interest is on Similarity measurement. She is a member of Nigeria Computer Society and Computer Professional Registration Council of Nigeria.

O. O. Bamgboye is a Lecturer in the Department of Computer Science, Moshood Abiola Polytechnic, Nigeria. He holds a Master degree in Computer Science from the Federal University of Agriculture Abeokuta. He is currently on his Ph.d programme at the Napier University, Scotland.

O. A. Afolabi is an Associate Lecturer at the General Studies Department of Moshood Abiola Polytechnic, Abeokuta, Nigeria.

A Process of Security Assurance Properties Unification for Application Logic

Faisal Nabi[1], Muhammad Mustafa Nabi[2]
(Corresponding author: Muhammad Mustafa Nabi)
Al-Maqsood Institute of Islamic and Modern Sciences, Karachi, Pakistan[1]
(Email: nabifaisal11@gmail.com)
Department of Computer Science, Preston University, Karachi, Pakistan[2]
(Email: mustafa.nabi2003@gmail.com)
(Received Dec. 22, 2016; revised and accepted Feb. 20, 2017)

Abstract

It is clear with evidences that for last one decade security experts only giving importance to the traditionally available information security practices, such as secure protocols (SSL/TSL) and Intrusion Detection Systems like Honey Pot. These can only secure network level problems in the security issues, but all these are security functional requirement based techniques, not security assurance requirements. It is also observed that current vulnerability analysis techniques are focused on traditional old methods, those were in practice for last one decade for traditional software engineering, such as penetration of white Box and Black Box. These practices can only tailor threats based on a check list of security policies to analyze vulnerability by auditors that can short fall identifying design flaws through traditionally available intrusion detection tools or vulnerability analyzing tools. This research paper focuses on representing "A Process of Security Assurance Properties Unification"for Component-ware Risks related to application logic to deal logical vulnerability in system.

Keywords: Application Logic; Business Logic; CBS; Design Flaws; E-commerce System Assurance; Logical Attacks; Software Architectural Flaws

1 Introduction

Present-days system security is more dependent on security mechanisms for example, secure protocols, cryptographic schemes/techniques, parameter security, Intrusion detection systems [10, 11], etc, but ground reality is that, these all security features of technology, currently in practical e-commerce systems [3] are higher rated target for hackers.

Although so much research has been carried out in this domain (on web application security) that use a large share of the present software, but on the other side (CBS) "Component Based Software" in the middle tier of e-commerce application server, which rapidly develop business application logic, also open security breaching opportunities [1]. Existing methods and approaches only target technical vulnerabilities of web applications that either on detecting, blocking, sanitizing, static and dynamic analysis approaches for web based attacks and using vulnerability analysis tools to identify security problems [9]. The vulnerability analysis tools cannot detect design flaw and logical flaw problems, because these approaches can only deal with the technical vulnerabilities that came into existence at

Implementation level, based on software specification Faults,Errors, Program failure and Bugs, such as web server IIS software Bug, that caused the security breach and later it was patched to eliminate the vulnerability, same as that on the application level Coding bugs or fault such as SQL Injection [12] or Cross-Site Scripting are technical vulnerability.

There is a common "Signature" associated with technical vulnerabilities that can be detected by analysis tools and fixed, but on the other hand there are some business logic vulnerabilities, those never been discovered or identified, that at later stage covert these threats into logical attacks in the application layer and these are "Business Component based-Rapidly Developed" web software application design flaw, the point of business logic attacks.

According to the Purdue University researchers, it is observed that experience reveals that attack method and profile both are strongly connected with several boundary conditions, these conditions could be based on three well defined areas:

1) Environmental (faults discovered by Krsul [6]);
2) Coding (fault);
3) Configuration errors (Krsul [6] & Aslam [2]). whereas in 2004 University of Luton researcher, Nabi first time identified [7], that design flaw could also be a cause of attack profile boundary condition of design specification in e-commerce systems [8].

Examples designed by contract component-ware flaw.

Case 1: The Failure of Ariane 5-ECA Rocket Space Mission - European Space Agency.

The failure of Ariane 5-ECA was not caused by defective management, Processes, Implementation or Testing but Rather a Reuse Specification mismatch that was violation of "design by contract" caused flaw.

Reasons: Reused component difference between "Component-Realization Specification vs Design by Contract Specification caused Flaw in design. because deployment was static state of Component not dynamically.

Because of Component Usage Specification based process integration, depends on logical component's interface specification information model of component, it refers to a particular Service or Task function/processing logic, specification of a logical component.

European Space Research Center omitted to follow this principle in CBSRD Integration based on Design by contract is approach to providing components that can be used & reused in dynamically varying contexts.

In design by contract components have executable pre-conditions, post-condition & invariants. Exceptions are raised when any of these predicates are violated in an execution.

Case 2: Mars Polar Landing Mission Feb 7,1994.

Reason of Project Failure:

Touchdown Monitor (TDM) (Touchdown Monitor) component failed to meet the requirement Specification as compare to its functional specification based on design by contract interface driven specification integration, which gave birth to design flaw in the MPL system & mission Failed.

Requirement Modeled of TDM:

TDM is a software component of MPL system that monitor the state of three landing Legs during Two stages of descent.

Logical Component Information Processing:

Multi-Tasking executive Calls TDM module at a rate of 100 times per second, receives information on the leg sensors from a second module.

These two module establish interface to TDM.During First stage, starting 5 KM above Mars Surface, TDM software Monitors the three touchdown Legs.

Application Logic of Component:

At First Stage start reading at approximately 5 Km above Mars surface, TDM monitors touchdown legs, one sensor each leg to determine touchdown.

Processing Logic design:

Developer assumed when Legs lock into deployed position, it was a known possibility sensor might indicate an erroneous touchdown signals.TDM software was to handle this potential event by Marking Leg that generates a spurious signal on two consecutive sensor-reads as having a bad sensor.

Second Stage:

Starts about 40 meters above surface, TDM was to monitor the remaining Good sensor. When a sensor had 2 consecutive reads indicating Touchdown, TDM software was to command the descent Engine to Shutdown.

What happened:

One or more sensor did have 2 Consecutive reads before40 meter Point, Leg-sensor information was stored in TDM Component Memory. When MPL crossed 40 meter point,TDM changed states & read the memory associate with the leg-sensor during first stage of descent. Result shutdown Engine.

Conclusion:

Developer could have designed & implemented the requirement in many ways, but the Essence of design flaw, is components predicates (pre-condition, Post-condition & Invariants)violated an execution of state of bad sensor information retained by the programme variables.

Hence, its proved that problem was not in the implementation logic but rather design by contract application logic of logical component, Requirement Specification as compare to its Functional, Specification based on design by contract interface driven specification integration, which gave birth to design flaw in the MPL system & mission Failed.

2 Component-based Application's Vulnerability

2.1 Logical Vulnerability in Application Layer

This research study is to focus on web software application logic problems & identify vulnerabilities that are because of mismatch between design and Architecture, while developing web software application. Our approach is more specific to what components can pinpoint vulnerability in a system design. We will only target application Logic vulnerabilities.

2.2 Application Logic Attacks Operation

Unlike, Common application technical attacks, such as SQL injection or Buffer Overflow, Each application logic attack is usually unique, since its not been mentioned or part of any Taxonomy of web application attacks, and since it has to exploit a function or feature that is specific to the application.

Application logic attacks are not based on characteristics like Buffer Overflow which can be characterize them as other technical vulnerabilities in the web application (SQL, SSI or buffer overflow). This makes it more difficult for automated vulnerabilities testing TOOLS to identify or detect such vulnerability class of attacks because they are caused by the flaws in Logic & not necessarily flaws in the actual Code [8].

The logical attacks focus on the exploitation of a web application's logic flow. Application logic is the expected procedural flow used in order to perform a certain action. A web site may require a user to correctly perform a specific multi- step process to complete a particular action. An attacker may be able to circumvent or misuse these features to harm a web site and its users.

2.3 Case Study Gala Biscuit Manufacturing Company

Gala Biscuit Company is one of the leading companies in the Asian market (French Licensed in French biscuit classic trademark) .This case study focus on the Business Logic Flaw in a web-based enterprise resource planning application used within a manufacturing company for B2B [8]. In other words, business logic flaw while defining business policy/rule in application's function, which cause design flaw logic defining into application function.

The Application Functionality:

Finance personnel had the facility to perform funds transfers between various bank accounts owned by the company and their key customers and suppliers. As a precaution against fraud, the application prevented most users from processing transfers with a value greater than ϵ20,000. Any transfer larger than this required a senior manager's approval.

The Design Logic of Application:

The wrapping component code responsible for implementing this check within the application was extremely simple:

```
boolCAuthCheck::RequiresApproval(int amount)
{
    if (amount <= m_apprThreshold)
        return false;
    else return true;
}
```

The developer assumed that this transparent check was bulletproof. No transaction for greater than the configured threshold could ever escape the requirement for secondary approval.

Attack possibilities:

The developer's assumption was flawed because he had completely overlooked the possibility that a user would attempt to process a transfer for a negative amount. Any negative number will clear the approval test, because it is less than the threshold. However, the banking module of the application accepted negative transfers and simply processed them as positive transfers in the opposite direction. Hence, any user wishing to transfer ϵ40,000 from account A to account B could simply initiate a transfer of - ϵ40,000 from account B to account A, which had the same effect and required no approval. The antifraud defences built into the application could be trivially by passed.

Cause of such mistakes:

Many kinds of web applications employ numeric limits within their business logic, For example:

- A retailing application may prevent a user from ordering more than the number of units available in stock.
- A banking application may prevent a user from making bill payments that exceed her current account balance.
- An insurance application may adjust its quotations based on age thresholds.

Result:

Finding a means of violate such limits will represent a security compromise of the application business policy rules itself. However it may have serious business consequences and represent a breach of the controls that the owner is relying on the application to enforce.

2.4 Scientific Justification in the Light "State of Art CBSD"

Component-based-software development uses (RDA) rapid development approach by using components; each component has functional rules, which develops application logic into the component. Since business application logic is developed with business components, it uses business component-ware; each component's logic makes a process, by integrating these components to make "logical component-ware" that translates business application logic in the e-commerce system. Reuse of component means in e-commerce application, a business component that has a complete function, business logic rules. A designer starts design a business component that can be reused, but it is also very important that designer must have complete knowledge about previous designed business logic in the system, so that no logical flaw could provide opportunity to intruder to violate the business application logic in e-commerce system.

This case study describes, its complete design based on component - oriented process and how they made wrong integration, while ignoring business logic of application's current functionality, which led this towards wrong analysis of previous component-based business application logic.

3 Security assurance process properties unification Artifacts

Components are specifically designed to be combined into systems, and, in fact, it is these systems that ultimately need security assurance [4].

Composing security assurance specifications into specifications for larger systems is not only a non trivial task, but considered one of the hard unsolved problems in computer security [5].

To deal this problem in component-based (logical component-ware interface-focused design) business logic in e-commerce application, and to deal the issue of business logic separately from implementation logic. We need of security assurance process properties unification, for logical component-ware based rapid development, and gradually changing business processing logic in e-commerce systems.

We have also experienced that security architecture shows highest level picture of the system.

Where, in the beginning step of a trustworthy system design, is defined by its components in the system, based on their trusted and un-trusted suitability that isolate boundaries between them. This makes us able to perform security analyses of system designs, before under proceeding to fully implementing and verifying the system. Moreover, the architectural analysis drives subsequent steps to the overall process. Where assurance activity of a whole system is specified through the trusted and un-trusted components in the system, and interconnections among them that reflects the security-related properties (see Figure 1).

We define that architecture of an information system is specified in two levels of refinement (see Figure 2):

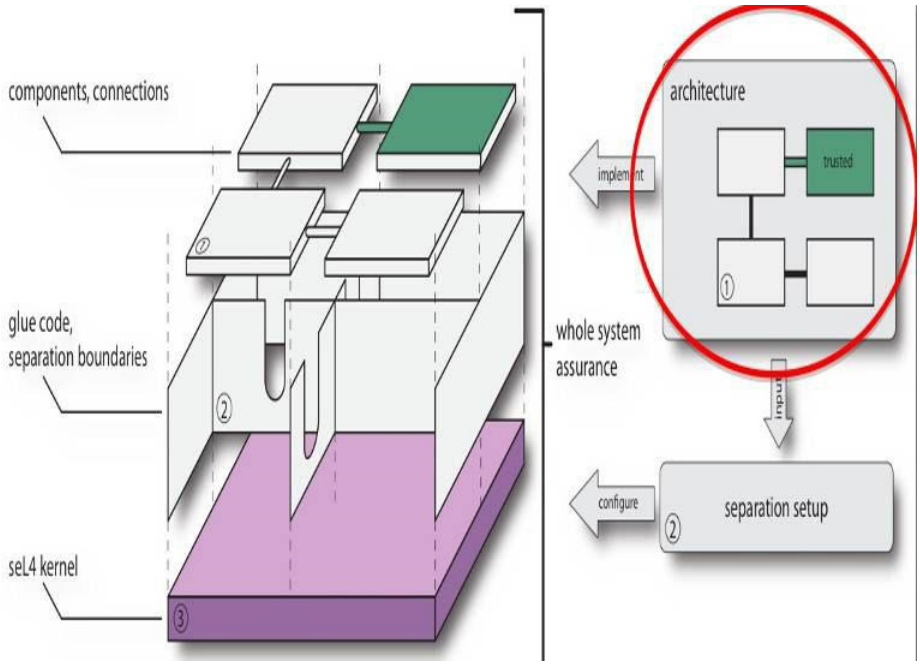

Figure 1: The core element of security assurance process properties unification

1) The System architectural level describes style to be applied at highest level abstraction the layers architectural style.
2) A logical component Architecture specification.

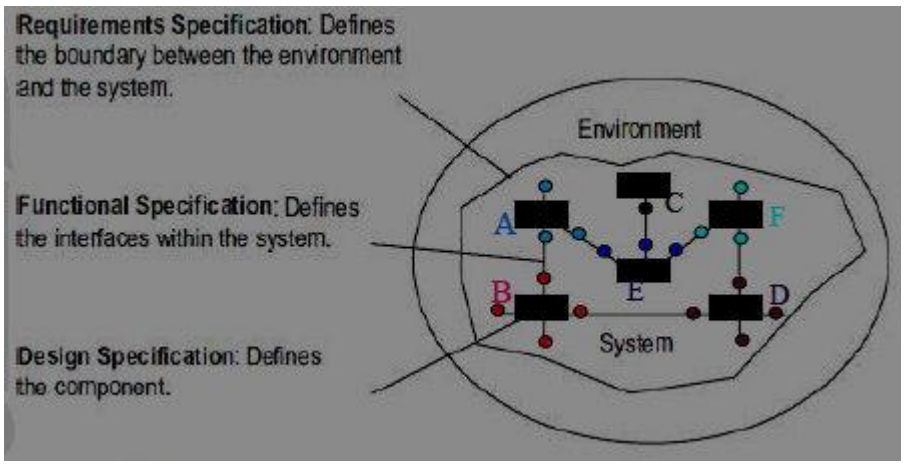

Figure 2: Refinement process layer wise

Therefore, Component-based development can be addressed more precisely in two distinctive different

levels.

1) One is regarding the technology used for system implementation.

2) Second is a more abstract level, (completely focus on interface as the main design abstraction that encourages designers to take into account system behavior more abstractly), where the focus is to set the logical structure of the solution and where technological issues are not considered (separation of business logic from implementation logic).

This is the stage where logical problem occur due to short fall assurance process related to logical structure of the solution, which then translates at further stage design flaw of product that causes business logic vulnerability.

In order to assure secure application, need to require applicability of security by design approach, by in co-operating above mentioned core elements of security assurance process, rather than by adding an other layer into system after being built. Therefore, design by security can be derived by following course of assurance method.

We are considering multi-tier of specification, these will be system tier and the component tier.

1) The system level tier represents a top view of the product.

2) While the component level tier will account for the design, test, and diagnostics specification of the individual component that make up the system.

This idea can be extended that a system can be nested as a component in the next higher order system.

The purpose of defining a design for test strategy, each tier will require a model based method for capturing design, test requirements, and diagnostic information.

Artifacts help to understand design and apply Model-based-Testing Approach for component-based e-commerce applications assurance (see Figure 3).

Application of extracted test design for an e-commerce system should follow a complete plan.

1) Scenario-based-approach for modelling business scenario to generate test scenarios from extracted Test design.

2) Selecting integration strategies of components matching with requirement specifications & their offered and used interface, design drivers and stubs.

3) Derive test scenarios from business scenarios and business flows of components, and derive test cases by analyzing business data.

Model-based approach in component-based-software development process is considered as a cost-effective and high quality of assurance for platform - independent design. The Model-based approach is a means of realizing the "correct by construction philosophy", whereby flaws in a product design are discovered early at the design stage, such as components integration flaws. By using the Model-based approach, we can extract the integration flaw/fault existing between the components interfaces interacting with each other in the system, in order to deliver a service, which is composed with the components integration on the bases of their "business process functionality" within the application.

4 Conclusion

In this research, we investigated the problem of business logic vulnerability in the component-based rapidly developed e-commerce distributed system's application. We proposed security assurance process

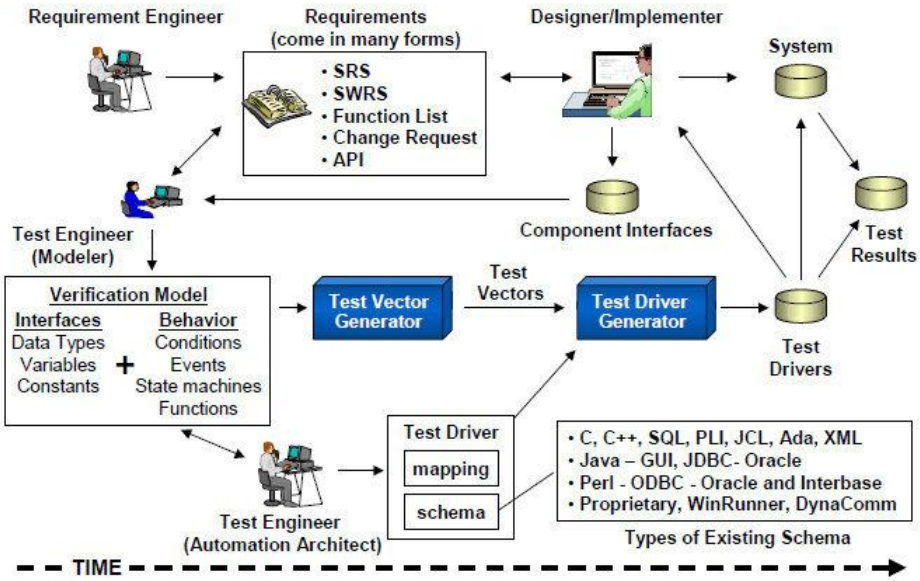

Figure 3: Model based approach application process for applicability assurance properties unification

properties unification, on Component-based and compositional interpretations of hazards, verification, and assurance arguments targeting security assurance logical component behavior specification approach to formalize and design a solution for business logic vulnerability phenomena. Therefore, common sense is an appropriate tool while designing your web application software and deploying component based business logic into that system , must focus on security beside the functionality because this functionality can be productive only when it works as per and within its functional control defined business policy into the e-commerce systems.

References

[1] A. Anurag, "Network neutrality: Developing business model and evidence based net neutrality regulation," *International Journal of Electronics and Information Engineering*, vol. 3, no. 1, pp. 1–9, 2015.

[2] T. Aslam, *A Taxonomy of Security Faults in the Unix Operating System*, Purdue University, 1995.

[3] T. M. El-Sakka and M. Zaki, "Using predicate-based model checker for verifying e-commerce protocols", *International Journal of Network Security*, vol. 16, no. 2, pp. 90–101, 2014.

[4] J. Gregoire, K. Buyens, B. De Win, R. Scandariato, W. Joosen, "On the secure software development process: CLASP and SDL compared", in *Third IEEE International Workshop on Software Engineering for Secure Systems (SESS'07)*, Minneapolis, MN, USA, May 2007.

[5] G. B. Jeong, G. B. Kim, "A framework for security assurance in component based development", in *International Conference on Computational Science and Its Applications (ICCSA'04)*, Lecture Notes in Computer Science, vol. 3043, pp. 587–596, 2005.

[6] I. Krsul, *Software Vulnerability Analysis*, Ph.D. Purdue University, West Lafayette, 1998.

[7] F. Nabi, "Secure business application logic e-commerce", *Computer & Security*, vol. 24, no. 3, pp. 208–217, May 2005..

[8] F. Nabi, "Designing a secure framework method for secure business application logic integrity in e-Commerce systems", *International Journal of Network Security*, vol. 12, no. 1, pp. 29–41, 2011.
[9] E. U. Opara, O. A. Soluade, "Straddling the next cyber frontier: The empirical analysis on network security, exploits, and vulnerabilities," *International Journal of Electronics and Information Engineering*, vol. 3, no. 1, pp. 10–18, 2015.
[10] Q. S. Qassim, A. M. Zin, M. J. Ab Aziz, "Anomalies classification approach for network-based intrusion detection system", *International Journal of Network Security*, vol. 18, no. 6, pp. 1159–1172, 2016.
[11] V. M. Shah, A. K. Agarwal, "Reliable alert fusion of multiple intrusion detection systems", *International Journal of Network Security*, vol. 19, no. 2, pp. 182–192, 2017.
[12] M. Stampar, "Inferential SQL injection attacks", *International Journal of Network Security*, vol. 18, no. 2, pp. 316–325, 2016.

Faisal Nabi is a Junior scientist in the field of information and computer security was born in Karachi city. He is specialized in e commerce and information security his MSc by research from university of Bedfordshire and honorary PhD from Brock university Canada.

Muhammad Mustafa Nabi is a researcher in the field computer security and model based application, has undertaken his Mcs from school of computer science Preston University Karachi and BA in economics from Karachi university .He is also member of Distributed system Lab at Preston university.

Active Monitoring & Postmortem Forensic Analysis of Network Threats: A Survey

Anshul Tayal, Nishchol Mishra and Sanjeev Sharma
(Corresponding author: Anshul Tayal)

School of Information Technology, UTD-RGPV Bhopal
DLake Perl Garden, Airport Road Bhopal
(Email: tayal.anshul14@gmail.com)
(Received Nov. 23, 2016; revised and accepted Feb. 21, 2017)

Abstract

In this era of Science and Technology, need of internet connectivity in the business organizations, banking institutions, universities etc. is growing exponentially. Establishing the computer network at the workplace has now become common and sometimes compulsory for making the business process easy and effective. As a result of this connectivity explosion, risk of potential vulnerability exploitation by attackers, malicious codes and network threats are becoming the headache of the network administrator. This situation brought us a brand new field of research called Network Security. Network security is an area of concern for most of the organizations nowadays. A lot of the network environments keep on facing ever increasing security threats in the form of Trojan, worm attacks and viruses that can damage the computer system and communication channels. This study is precisely highlighting the previous methods, techniques and mechanisms of active monitoring (real time analysis) and postmortem forensic analysis. As well as examining the downsides of the previously used methods. And dedicated in the search of the solution of those downsides.

Keywords: Active Monitoring; Log File Analysis, Networking Threats; Postmortem Forensic Analysis; Real Time Analysis

1 Introduction

Computers and networks are involved in virtually all activities today. People use them to communicate, to create intellectual property, to shop, to perform business transactions, to plan trips, and much more. Networks afford users the opportunity to continuously use computers through cell phones, personal digital assistants (PDAs), wireless connectivity, and the ubiquitous Internet. Any computer can be used for many purposes, just because a computer is located in the workplace does not mean that the computer is used only for work. The pervasive nature of computers and networks means that they are increasingly connected to incidents and crimes. And wherever crime took place these mysterious questions comes under the mind of investigator or network administrator, who is involved in this? Who is responsible for obtaining this information? How can relevant information be obtained from computers to support criminal, civil, or disciplinary action? And The Network Forensics is the answer of all these questions. You might hear the term cyber forensics or digital forensics; they usually refer to network forensics, not computer forensics. And "Network forensics is a process of detecting and analyzing the attacks on the network that jeopardize the Confidentiality, Integrity and Availability of an IT system".

The Confidentiality, Integrity and availability can be compromised by any of networking threats like Viruses, Worms, and Trojans or by any networking attacks like Intrusion, Denial of Service (DoS) etc. There are number of antivirus tools and Intrusion Detection System (IDS) are available, for handling the viruses, worms & Intrusions. Mostly these type of tools are based on the concept of signature based detection, either it is behavioral or anomaly based. And signature based methods have proven to be effective when a known pattern can be tested for. "Signature offers unparalleled precision in detection & forensics" but there is a downside also that traditional signature based tools are fail to detect zero day attacks or targeted attacks. That's why active monitoring or real time analysis is necessary for keeping eyes on today's highly dynamic malwares.

The process of network forensic is all about, examining network traffic in order to collecting the digital evidences. This examination of network traffic can be done by analyzing router logs, firewall logs or eavesdropped data from the network.

The most important source of evidence in the network forensics is log files, because any event occurring in an organization information technology system or network is recorded with various entries in a log file. The process of recording log files is known as logging. We can conduct efficient investigation and gather useful information by log file analysis.

The rest of this paper is organized as follows. Section 2 provides an overview of Log Files and what exactly the log file contains. Section 3 is all about the Log File Analysis, where we look at the categories of log file analysis as well as some previous work in the field of log file analysis. Section 4 outlines the related research work done in the field, and finally Section 5 concludes the paper.

2 Overview of Log Files

The process of recording events in a file during the execution of the operating system, process, network, virtual machine, or application is called "logging" and the file is called a "log file" [5]. A Log File is composed of log entries and each log entry contains useful information associated with events that occur in the system, network, virtual machine, or application. However Log file entries differ with respect to their types and requirements. Log Messages are a nice way for any application to convey messages about its current actions to human users, consequently logging is considered an essential means of security control which helps the investigators in identifying, answering, and precluding operational issues, incidents, violations, and fraudulent activities [5]. Fig. 1 shows the overview of log files. Ordinarily multiple software and hardware maintains log files, but here we are getting an overview of log files of our concern only.

2.1 Router Logs

Routers work on the network layer (OSI layer 3), so they only capture the traffic as a man in the middle. But the routers can store information from the higher layers to the log, such as the TCP header from the transport layer (OSI layer 4). Due to the large amount of data that is going through a high-end router, it is not that realistic to store all the data, not even the data in the TCP/IP headers. Similar packets are therefore aggregated into records, so that they take up less space (in most cases) [11].

Typical information found in the router logs:

- Export timestamp.
- Start and end timestamp.
- Source and destination IP address.
- Type of Services (ToS).

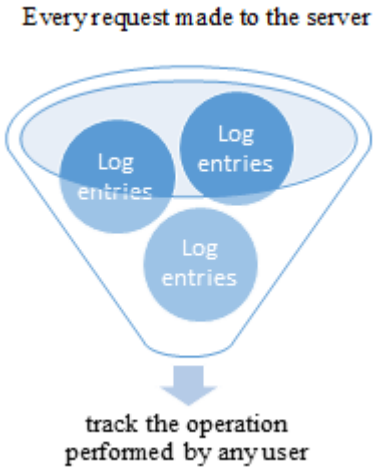

Figure 1: Overview of Log File

- Packet and Byte count.
- Input and output interface number.
- TCP flags.
- Routing information etc.

A flow is uniquely identified on the router by the following tuple:

$$
\begin{aligned}
Src_{addr} &= Source\ Address;\\
Src_{port} &= Source\ Port;\\
dst_{addr} &= DestinationAddress;\\
dst_{port} &= DestinationPort;\\
src_{if} &= SourceInputInterface;\\
t_{os} &= TypesofServices;\\
\\
flow &= (Src_{addr}, Src_{port}, dst_{addr}, dst_{port}, ip_{port}, src_{if}, t_{os}).
\end{aligned}
$$

2.2 Windows Registry

Windows Registry contains much information about user activities. Some, such as Most Recently Used (MRU) lists, contain information that can be directly extracted, while other information, such as the meaning of the order of MRU items through time, must be inferred. The Investigators observe these traces of evidence on a system and naturally make inferences as to their meanings based on their knowledge of the system and past experiences [4].

For example, consider the core traces from Internet explorer (generalized):

%SystemRoot%\Prefetch\IEXPLORE.EXE-%s.pf

%HomeDrive%%HomePath%\Local Setting \
Application Data\Microsoft\Feeds Cache\index.dat

HKEY_USERS\%SID\Software\Microsoft\CTF\TIP

HKEY_USERS\%SID\Software\Microsoft\Internet
Explorer\Security\AntiPhising\%s

These traces are updated every time, and only, when the user action is executed.

2.3 Proxy Server Logs

Proxy server maintains number of log files by that the investigator has lots of information regarding the scene and has more chances to gather lots of evidence. Generally the proxy server maintains a cache log, access log, referrer log, user agent log, HTTP server log, etc. and following are the information that can be found:

- Timestamp.
- Origin of request.
- Size of packet.
- Method of request.
- Requested URL.
- Username.
- Content type of reply.
- Browser related information. Etc.

Access.log file contains all the crucial information, that's why it is a highly important log file. The log messages in the access.log file are not as readable as messages in the cache.log file, but once we understand what the different fields mean, it's easy to interpret the log messages.

Some tuples from the access.log file are shown below:

1284565351.509 |114 |127.0.0.1 |TCP_MISS/302 |781 |GET |http://www.google.com/ | | -
FIRST_UP _ PAR ENT/proxy.example.com |text/html

1284565351.633 |108 |127.0.0.1 |TCP_MISS/200 |6526 |GET |http://www.google.co.in/ | | -
FIRST_UP _PARENT/proxy.example.com |text/|html

1284565352.610 |517 |127.0.0.1 |TCP_MISS/200 |29963 |GET | http://www.google.co.in/
images/srpr/ nav_logo14.png | | - FIRST_UP_PARENT /proxy.exa mple.com |Image/|png

In the previous example of a log message, the first column (here columns are separated by vertical bar '—') represents the seconds elapsed since a UNIX epoch, which can't be interpreted by human users. The second column represents the response time in milliseconds. The third column represents the client's IP address. The fourth column is a combination of Squid's request status and the HTTP status code. The fifth column represents the size of the reply including HTTP headers. The sixth column in the log message represents the HTTP request method which will be GET most of the time, but may also have values such as POST, PUT, DELETE, and so on. The seventh column represents the request URL. The eighth column is the username, which is blank in this case because the request was not authenticated. The ninth column is a combination of the Squid hierarchy status and IP address or peer name of the cache peer. The last column represents the content type of the replies.

3 Log File Analysis

Log file analysis is basically the process of network forensic investigation to collect digital evidence. Bill Nilsson et al. stated in their book "Guide to Computer Forensics & Investigations", any method or technique can be used for investigation as long as it comes under the law and legislation. So for better understanding forensic analysis of logs can be split up into two categories:

- Postmortem forensic analysis, and
- Real- time analysis or active monitoring.

Postmortem analysis of logs is the investigation of something that already has happened, and which one cannot do anything about now. The purpose of this analysis is therefore to find out what has happened. Real-time analysis is an ongoing process, which returns results with a low latency, so that the system or operators can respond to the attacks. Postmortem analysis can therefore be more exhaustive than real-time analysis. Real-time analysis needs to find the attacks quickly to be effective. Postmortem analysis can be used to examine the attack in more detail and give a more thorough result/report. Another aspect that differs between both the analysis methods, is that real-time analysis can only go through the log data once, whereas a postmortem analysis could go through the file many times, and examine interesting flows it had found in previous runs [11].

Here are some research work done in the field of Postmortem Forensic Analysis.

3.1 Analyzing Log Files for Postmortem Intrusion Detection

In [3] Karen A. Garcia et al. approached postmortem intrusion detection using an anomaly, learning-based, host-based intrusion detection model, where the working hypothesis is that an attack takes the form of an unusual sequence of system calls. In our setting, an attacker has already bypassed the online intrusion detection system, if any. They assumed that there is a set of logs, where evidence of the intrusion has (hopefully) been recorded, in the form of a trace of system calls. They emphasize that our method forms a profile of postmortem intrusion detection out of a log file comprising multiple sessions, each of which may contain dozens, of processes.

The mechanism they used for the postmortem intrusion detection:

- Test Log Session.
- Reduction Model.
- Sliding Window.
- Detection Model.

- Classification Output.

3.2 Signature Based Detection of User Events for Postmortem Forensic Analysis

Joshua Issac James et al. [4] used the novel approach to user event reconstruction by showing the practicality of generating and implementing the signature based analysis methods to reconstruct high-level user actions from a collection of low level traces found during the postmortem forensic analysis of the system.

For deriving user action signature they followed these steps:

- Determining traces for the signature.
- Analysis of timestamp updates.
- Signature generalization.
- Creation of the signature for user action.

3.3 Post-Attack Intrusion Detection using Log File Analysis

The system follows the learning based approach. Assume that all the activities related to system are monitored and logged in several log files, and behavior of an intruder is significantly different from that of legitimate users. The aim of the system to find out a particular activity performed by attacker. This System has to deal with various key issues like huge size of log file and creating a profile of normal behavior only with some part of log file instead of analyzing complete log file [10].

3.4 Forensic Examination of Log Files

In the research work, Joan Petur Petersen [11] developed the system that can identify malicious traffic in the router logs on a log entry level. The system is implemented using feature extractors and a classifier based on a neural network. Their system could easily be extended to detect other kinds of malicious traffic, such as Denial of Service (DoS) attack.

Table 1 shows the critical analysis of the above discussed research work.

4 Related Work

Research on Log File Analysis has been very prolific, consequently providing a reasonable overview of existing log file analysis techniques actually is a research issue. We shall, therefore confine ourselves to overview research work done on the network threats and its solutions, log file analysis, and some user profiling related research works. Actually it is a hierarchy, means suppose, in any network some threats are floating, that jeopardizes integrity, availability or confidentiality of an IT system. That indicates the crime happened and needs to be investigated and for investigation log file analysis comes under the consideration. Now if evidences are collected, and investigator found any new traces at the digital crime scene the user profiling took place, for future use.

Table 1: Overview of log file analysis

S. No.	Reference	Technique	Log Source	Advantage	Disadvantage
1.	Karen A. Garcia et al. [3]	Postmortem forensic analysis	Router	High Performance, Good accuracy	Time Consuming, Needs high technical proficiency
2.	Joshua Issac James et al. [4]	Signature based	Windows Registry	Easy to conduct	Platform Dependent, Data can be modified by user.
3.	Apurva S. Patil et al. [10]	Postmortem forensic analysis	Router	Needed less effort	Less accuracy
4.	Joan Petur Petersen [11]	Feature extractor and classifier based	Router	Easy to extend to detect other kinds of attacks also, such as DoS.	Can be used for Cisco routers only.
5.	Mike Thelwall [17]	Search engine query extraction	Web logs	Provides useful information	Dependent on user actions

4.1 Network Threats

Sachin Taluja et al. [16] provides the detailed information about various types of network threats in network security, and their solutions by the use of different firewalls phenomena. This paper helped us to understand the concept of real-time analysis, through firewalls. But still a huge effort is required to construct new security strategies or security mechanisms.

Similarly, Craig Smith et al. [14], provides a complete picture of the anatomy of computer worms, of how worm's behavior can be stealthy and how to detect that behavior. They discussed multiple issues that may be encountered at different stages ranges from evasions to detection of the worm attack.

On the other hand Usman Asghar et al. [13], mainly focuses on the techniques of intrusion detection & prevention and contains lots of information of different intrusion detection systems, intrusion prevention systems, their pros, their cons. But the downfall of this research is that none of the IDS or IPS can work properly if they don't have the definition of new signatures of attacks.

Besides of these, when low level services suffer attacks, the high level services that depends on them will suffer from indirect threats. Most evaluation methods do not consider the dependency relationship between services, upon these problems an evaluation method that based on dependency analysis is presented by Xiangdong Cai et al. [1].

Table 2 shows the critical analysis of the above discussed research work.

4.2 Log File Analysis

A thorough examination of log files is needed to reveal the hidden actions of criminals in computer network. The model by Himal Lalla et al. [7], specifies the steps that forensic investigator can follow with regard to the extraction and examination of digital evidence from log files.

Table 2: Critical Analysis - Classification: Network Threats

Objective	Method/Approach	Tool/Technology /Platform	Reference
To provide detailed information about various types of network threats in network security, and their solutions by the use of different firewalls phenomena.	N/A	Firewall	Sachin Taluja et al. [16]
To provide a complete picture of the anatomy of computer worms, of how worm's behavior can be stealthy and how to detect that behavior.	Target discovery and detection, Distribution mechanism and detection, Activation mechanism and detection, worm payload detection.	LaBrea, Honeycomb, Bro/Snort	Craig Smith et al. [14]
Mainly focused on the techniques of intrusion detection & prevention and contains lots of information of different intrusion detection systems, intrusion prevention systems, their pros, their cons.	N/A	IDS, IDPS, HIDPS	Usman Asghar Sandhu et al. [13]

Joan Petur Petersen [11], provides the system that can identify malicious traffic in the router logs on a log entry level. They uses the concept of neural network, feature extractors and a classifier. The study describes about the log file analysis in detail, but still their system is mainly focused on postmortem forensic analysis, nothing is done in the field of real-time analysis.

In the field of postmortem forensic analysis, Karen A. Garcia et al. [3], uses the approach which is the combination of Hidden Markov Model (HMM) with k-means, to factor out repetitive behavior, thus, speeding up the process of locating the execution of an exploit. They also used the entropy based approach in addition, that speeds up the construction of a profile for system behavior. The disadvantage of this approach is that, it is too complicated as well as time consuming and needs lots of expertise knowledge in the field.

A different approach based on the signature based analysis was proposed in [4], in which Joshua Isaac James et al. introduces a novel approach to user event reconstruction by showing the practicality of generating and implementing the signature based analysis method, to construct high level user action from a collection of low level traces found during postmortem forensic analysis of a system. They used windows registry as a source of evidence, but this is the major disadvantage of this approach, because registry files can be modified and timestamps are totally in the control of user.

Similarly, Mike Thelwall in [17] practically described that web log files are a useful source of information about visitor, site use, navigation behavior, etc. log files can also reveal the existence of both web pages and search engine queries that are sources of new victors. But the study covers only one website and, therefore, its finding does not generalize to the whole web.

Table 3 shows the critical analysis of the above discussed research work.

Table 3: Critical Analysis - Classification: Log File Analysis

Objective	Method/Approach	Tool/Technology /Platform	Reference
To specify the steps that forensic investigator can follow with regard to the extraction and examination of digital evidence from log files.	Extraction, Analysis, and Correlation	N/A	Himal Lalla et al. [7]
To provide the system that can identify malicious traffic in router logs on a log entry level. They uses the concept of neural network, feature extractors and a classifier.	Feature extractor and classifier based on neural network	DK-CERT	Jon Petur Petersen [11]
Used the approach which is the combination of Hidden Markov Model (HMM) with k-means, to factor out repetitive behavior, thus, speeding up the process of locating the execution of an exploit.	Factoring out repetitive behavior, Good enough is enough, Abstracting out system behavior, Anomaly detection model- HMM and k-means	Fedora 8.0, Red Hat 9.0, Ubuntu 5.0	Karen A. Garcia et al. [3]
To introduce a novel approach to user event reconstruction by showing the practicality of generating and implementing signature based analysis method, to construct high level user action from a collection of low level traces found during post-mortem forensic analysis of a system.	Signature Based	Windows Registry	Joshua Isaac James et al [4]
To practically describe that web log files are a useful source of information about visitor, site use, navigation behavior, etc.	Identifying Link Source Page, Extracting Search Engine Queries	NetTracker	Mike Thelwall [17]

Table 4: Critical Analysis - Classification: User Profiling

Objective	Method/Approach	Tool/Technology /Platform	Reference
To adopt an anomaly detection approach by detecting possible intrusions based on user profile, built from normal usage data.	HMM, Minimum Likelihood, Minimum Cross Entropy, Novelty Detection	N/A	Dit-Yan Yeung et al. [18]
To derive user profile from observing user behavior.	Implicit acquisition and Interpretation of user data	Formula Based	Gerald Stermsek et al. [15]
To detect frequent patterns from log files to build the normal user profile.	Line pattern detection, Frequent pattern mining, Anomaly detection	N/A	Xin Cheng et al. [2]

4.3 User Profiling

Dit-Yan Yeung et al. [18] adopted an anomaly detection approach by detecting possible intrusions based on user profile, built from normal usage data. The user profiling is basically based on two behavioral models. The dynamic modelling approach is based on Hidden Markov Model (HMM), while static modelling approach in based on event occurrence frequency distribution. The major disadvantage of this study is, they are using the number of approaches, but still the research is limited to, host based intrusion. The network based intrusion is un-touched.

In other research work, an approach to derive user profile from observing user behavior is used. In particular Gerald Stermsek et al. [15] uses web server log files and metadata describing the page contents to extract user interests. This study helps us in understanding the concept and technique of user profiling.

Similarly, Xin Cheng et al. [2], used the novel approach, which detects frequent patterns from log files to build the normal profile, and then to identify the anomalous behavior in log files. Because detecting failure and diagnosing their root cause in a timely manner is essential. Fast and accurate detection of these failures can accelerate problem determination.

Table 4 shows the critical analysis of the above discussed research work.

5 Conclusion

In this paper, we discussed about different threats and attacks on network security. Moreover, we have looked at different techniques, methods and approach of log file analysis, postmortem forensic analysis and real-time analysis to overcome the same. On the other hand, attackers are discovering new techniques and ways to break the security policies. The only way to beat them is to know about their methods and techniques of attack. Knowing about the attacker's technique is called the signature based investigation. The paper discussed number of techniques that are based on the same, but have some downfalls. In desire to overcome those downfalls, In future we are going to investigate log file of the proxy server and try to collect evidence and signature more appropriately.

References

[1] X. Cai, J. Yang, H. Zhang, "Network security threats situation assessment and analysis technology study," in *IEEE International Conference on Measurement, Information and Control (ICMIC'13)*, pp. 643–646, Aug. 2013.

[2] X. Cheng and R. Wang, "Communication network anomaly detection based on log file analysis," in *International Conference on Rough Sets & Knowledge Technology*, Springer International Publishing, pp. 240–248, Oct. 2014.

[3] K. A. Garcia, R. Monroy, L. A. Trejo, C. Mex-Perera, and E. Aguirre, "Analyzing log files for postmortem intrusion detection," *IEEE Transactions on Systems, Man, and Cybernetics - PART C Applications and Review*, vol. 42, no. 6, pp. 1690–1704, Nov. 2012.

[4] J. I. James, P. Gladyshev, and Y. Zhu, *Signature Based Detection of User Events for Postmortem Forensic Analysis*, Center for Cybercrime Investigation, University College of Dublin, 2007.

[5] K. Kent, M. Souppaya, *Guide to Computer Security and Log Management: Recommendations of National Institute of Standards and Technology*, US Department of Commerce, Technology Administration, NIST, 2006.

[6] S. Khan, A. Gani, A. Wahid, A. Wahab, M. A. Bagiwa, M. Shiraz, S. U. Khan, R. Buyya, and A. Y. Zomaya, "Cloud log forensics: Foundations, state of the arts, and future directions," *ACM Computing Surveys*, vol. 49, no. 1, May 2016.

[7] H. Lalla, S. Flowerday, T. Sanyamahwe, and P. Tarwiregi, "A log file digital forensic model," in *IFIP International Conference on Digital Forensics*, Springer Berlin Heidelberg, pp. 247–259, Jan. 2012.
[8] B. Nelson, A. Phillips, and C. Steuart, "Guide to computer forensics and investigation," *Cengage Learning*, 2014.
[9] M. Noor, S. Annapurna, and H. S. Bhadauria, "Taxonomy on security attacks on self-configurable networks," *International Journal of Electronics and Information Engineering*, vol. 3, no. 1 pp. 44–52, Sept. 2015.
[10] A. S. Patil, and D. R. Patil, "Post attack intrusion detection using log file analysis," *International Journal of Computer Application*, vol. 127, no. 18, pp. 19–21, Oct. 2015.
[11] J. P. Petersen, *Forensic Examination of Log Files*, Informatics and Mathematical Modeling Technical University of Denmark, Jan. 2005.
[12] R. U. Pratap, and N. K. Singh, "Weighted role based data dependency approach for intrusion detection in database," *International Journal of Network Security*, vol. 19, no. 3, PP. 358–370, May 2017
[13] U. A. Sandhu, S. Haider, S. Naseer, O. U. Ateeb, "A survey of intrusion detection & prevention techniques," in *International Conference on Information Communication and Management (IPCSIT'11)*, vol. 16, pp. 66–70, 2011.
[14] C. Smith, A. Matrawy, S. Chow, and B. Abdelaziz, "Computer worms: Architecture, evasion, strategies, and detection mechanisms," *Journal of Information Assurance and Security*, vol. 4, pp. 69–83, 2009.
[15] G. Stermsek, M. Strembeck, G. Neumann, "A user profile derivation approach based on log-file analysis," *IKE*, vol. 2007, pp. 258–264, June 2007.
[16] S. Taluja, and R. L. Dua, "Survey on network security, threats and firewalls," *International Journal of Advanced Research in Computer Engineering & Technology*, vol. 1, no. 7, pp. 53–58, Sept. 2012.
[17] M. Thelwall, "Web log file analysis: backlinks and queries," *ASILB Proceedings*, vol. 53, no. 6, pp. 217–223, Mar. 2013.
[18] D. Y. Yeung and Y. Ding, *User Profiling for Intrusion Detection Using Dynamic and Static Behavioral Model*, Hong Kong Innovation and Technology Commission under Project AF/223/98 and the Hong Kong University Grant Committee under Area of Excellence Research Grant AoE98/99.EG01.

Anshul Tayal obtained his master's degree in Cyber Forensics from School of Information Technology - UTD RGPV Bhopal, Madhya Pradesh, India. He has completed his Bachelor of Engineering in Computer Science and Engineering from Malwa Institute of Science & Technology Indore, Madhya Pradesh, India. He has research interests in computer forensics, data security, network security, information security & privacy.

Nishchol Mishra is currently an Assistant Professor at the School of Information Technology - UTD RGPV Bhopal, Madhya Pradesh, India. He obtained his Ph.D. degree in Computer Science and has 15 years of experience in the field of research and teaching. His research interests include data mining, database security, information security & privacy, and big data analytics.

Sanjeev Sharma is currently an Associate Professor and Head of the department at the School of Information Technology - UTD RGPV Bhopal, Madhya Pradesh, India. He obtained his Ph.D. degree in Information Technology and has 20 years of experience in the field of research and teaching. His research interests include cloud computing, information security & privacy, and big data analytics.

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.