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

IJEIE

International Journal of Electronics
and Information Engineering

Vol. 4, No. 2 (June 2016)

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. 4, No. 2 (June 1, 2016)

A Construction of Some Group Codes

Nacer Ghadbane, Douadi Mihoubi
(Corresponding author: Nacer Ghadbane)

Laboratory of Pure and Applied Mathematics, Department of Mathematics, University of M'sila, Algeria
(Email: Nacer.ghadbane@yahoo.com)

(Received Nov. 29, 2015; revised and accepted Apr. 27, 2016)

Abstract

Let Σ^* be the free monoid over a finite alphabet Σ and H a subgroup of a given group G. A group code X is the minimal generator of X^* with $X^* = \Psi^{-1}(H)$, where Ψ is a morphism from the free monoid Σ^* to the group G. In general, it is not obvious to detect if a subset X of Σ^* is a code or not. In this paper, we use the fact that the syntactic monoid $M(X^*)$ of X^* is isomorphic to the transition monoid of the minimal automaton recognizing X^*, to giving some examples of groups codes based on the following two results from [1]: 1) The subset X of Σ^* is a group code if and only if the monoid $M(X^*)$ is a group. 2) Let $X \subseteq \Sigma^*$ be a finite code, the syntactic monoid $M(X^*)$ is a group if and only if $X = \Sigma^n$ for some a positif integer n. And in this case, the group $M(X^*)$ is a cyclic group of order n.

Keywords: The free monoid and relatives, finite automaton and syntactic monoid

1 Introduction

Let Σ be an alphabet. A subset X of the free monoid Σ^* is a code over Σ if for all $m, n \geq 1$ and x_1, x_2, $\cdots$, x_n, y_1, y_2, $\cdots$, $y_m \in X$, the condition:

$$x_1 x_2 \cdots x_n = y_1 y_2 \cdots y_m$$

implies $n = m$ and $x_i = y_i$ for $i = 1, 2, \cdots, n$.

In other words, a set X is a code if any word in X^+ can be written uniquely as a product of words in X [1]. It is not always easy to verify a given set of words is a code. For this purpose, it is shown [1, 6], Let G be a group and H a subgroup of G. Let $\psi : \Sigma^* \longrightarrow G$ be a surjective morphism, $\psi^{-1}(H)$, it is generated by a biprefix code called a group code.

The remainder of this paper is organized as follows. In Section 2, some mathematical preliminaries. In Section 3, we use the fact that the syntactic monoid $M(X^*)$ of X^* is isomorphic to the transition monoid of the minimal automaton recognizing X^*, to giving some examples of groups codes and we show that the submonoid X^* generated by the group code X is a complete submonoid of the free monoid Σ^*. Finally, we draw our conclusions in Section 4.

2 Preliminaries

A monoid is a set M equipped with an associative binary operation and has a neutral element. The operation is usually written multiplicatively. The neutral element is unique and is denoted by 1_M, or simply by 1.

An alphabet Σ is any finite set, the elements of an alphabet are called letters or symbols. A finite word over Σ is a finite sequence of symbols $w = (\sigma_1, \sigma_2, \cdots, \sigma_n)$ of elements of Σ denoted by the concatenation $w = \sigma_1 \sigma_2 \cdots \sigma_n$. The integer $n = |w|$ is the length of the word w. For example, the finite sequences 00110 and 110 are two words over the binary alphabet $\{0, 1\}$ with $|00110| = 5$ and $|110| = 3$. The empty sequence () of length 0 is called the empty word and is denoted by ϵ. The set Σ^* of all words over Σ equipped with the operation of concatenation has a structure of a monoid with the empty word ϵ as a neutral element, called the free monoid on Σ. We denote by $\Sigma^+ = \Sigma^* - \{\epsilon\}$ the free semigroup over Σ [7].

For example, $\{0, 1, 2\}^* = \{\epsilon, 0, 1, 2, 00, 01, 02, 11, 12, 20, 21, \cdots\}$. If σ is a letter of the alphabet Σ, for any word $w = \sigma_1 \sigma_2 \cdots \sigma_k$ of Σ^*, we denote by $|w|_\sigma = \text{Card } \{i = 1, 2, \cdots, k : \sigma_i = \sigma\}$, the number of the occurrences of σ in the word w. For example $|00110|_0 = 3$ and $|00110|_1 = 2$. A language L over Σ^* is any subset of Σ^* [2].

A submonoid of Σ^* is a subset M which is stable under the operation and which contains the neutral element of Σ^*, i.e., $MM \subset M$ and $\epsilon \in M$.

A congruence on a monoid M is an equivalence relation $\equiv$ on M compatible with the operation of M i.e., for all $m, m' \in M, u, v \in M$

$$m \equiv m' \Longrightarrow umv \equiv um'v.0.$$

Let L be a language over Σ, the syntactic congruence of L denoted by $\equiv_L$ is defined by:

$$u \equiv_L v \Longleftrightarrow (\forall x, y \in \Sigma^* : xuy \in L \Longleftrightarrow xvy \in L)\,.$$

The quotient of Σ^* by $\equiv_L$ is, by definition, the syntactic monoid of L denoted $M(L)$, i.e., $M(L) = \Sigma^* / \equiv_L$ [4].

The right congruence of $L \subseteq \Sigma^*$, denoted by $\sim_L$, is defined by:

$$\forall w_1 \in \Sigma^*, \forall w_2 \in \Sigma^*, (w_1 \sim_L w_2) \Longleftrightarrow (\forall x \in \Sigma^*, w_1 x \in L \Longleftrightarrow w_2 x \in L)\,.$$

For $L \subset \Sigma^*$, the minimal automaton for L, in terms of the congruence $\sim_L$, is $A = (Q, q_0, F, \Sigma, \delta)$ with:

1) The set state of A are $Q = \{\overline{w}, w \in \Sigma^*\}$;

2) The initial state of A is $q_0 = \overline{\epsilon}$;

3) The set of final state of A is $F = \{\overline{w,}w \in L\}$, and finally

4) The transition function of A is defined by: $\delta\left(\overline{w}, \sigma\right) = \overline{w\sigma,}$ with $\overline{w} \in \Sigma^* / \sim_L$ and $\sigma \in \Sigma$ [4].

A morphism from a monoid Σ^* into a monoid Γ^* is a function $h : \Sigma^* \longrightarrow \Gamma^*$ which satisfies, for all, $u, v \in \Sigma^*$, $h(uv) = h(u)h(v)$ and furthermore $h(\epsilon) = \epsilon$.

Note that, the homomorphism h is completely determined by the images of letters of Σ in Γ^*, i.e., $h(\sigma)$ for any σ belong to Σ.

A bijective monoid homomorphism is called a monoid isomorphism. Two monoids are said to be isomorphic if there is a monoid isomorphism between them.

Recall this important result which states that the syntactic monoid $M\left(L\right)$ of a language L is isomorphic to the transition monoid of the minimal automaton recognizing L.

A set $X \subset \Sigma^*$ is a code if any word in X^+ can be written uniquely as a product of words in X, that is, has a unique factorization in words in X, i.e., if for all $m, n \geq 1$ and $x_1, \cdots, x_n, y_1, \cdots, y_m \in X$, the condition:

$$x_1 x_2 \cdots x_n = y_1 y_2 \cdots y_m$$

implies $n = m$ and $x_i = y_i$ for $i = 1, 2, \cdots, n$ [8, 9].

The code X is a maximal if, for any code X' such that $X \subseteq X'$ then we have $X = X'$.

A subset X of Σ^* is called prefix (suffix) if $X \cap X\Sigma^+ = \emptyset$ (resp. $X \cap X\Sigma^+ = \emptyset$). A subset X of Σ^* is biprefix if it is both suffix and prefix [3].

3 Group Codes

Let us first recall the definition of a group code. Let G be a group and H a subgroup of G. Let $\psi : \Sigma^* \longrightarrow G$ be a surjective morphism, $\psi^{-1}(H)$ it is generated by a biprefix code called a group code.

The following propositions from [1] gives a methods to construct group codes.

Proposition 1. *[1] Let G be a group and H a subgroup of G. Let $\psi : \Sigma^* \longrightarrow G$ be a morphism.*

Let $X^* = \psi^{-1}(H)$ with X is the minimal generator of the set X^*. Then X is biprefix code. Furthermore, if ψ is surjective, then X is a maximal biprefix code. In the last case the set X is called a group code denoted by $X\left(G, H\right)_{\psi}$.

Example 1.

1) *Consider the morphism of monoids $\psi : \{a, b\}^* \longrightarrow (\mathbb{Z}, +)$ defined by:*
$\psi(a) = 1, \psi(b) = -1, \psi(\epsilon) = 0.$

And then, $\forall w \in \{a, b\}^*$ we have $\psi(w) = |w|_a - |w|_b$.

The mapping ψ is surjective because $\forall m \in \mathbb{Z}, \exists w \in \{a, b\}^*$ such that $\psi(w) = m$.

In fact:

a. *If $m = 0$ then $\psi(\epsilon) = 0$.*

b. *If* $m > 0$ *then* $\psi(a^m) = m \cdot \psi(a) = m \cdot 1 = m$.

c. *If* $m < 0$ *then* $\psi(b^{-m}) = -m \cdot \psi(b) = -m \cdot (-1) = m$.

Let $H = \{0\}$ *be the trivial subgroup of* $(\mathbb{Z}, +)$. *Then* $X^* = \psi^{-1}(\{0\}) = \{w \in \{a,b\}^* : \psi(w) = |w|_a - |w|_b = 0\} = \{w \in \{a,b\}^* : |w|_a = |w|_b\}$.

The set X *is infinite because* X *contains at least the set of all the words of the form* $a^n b^n, n > 0$, *which is evidently infinite. Finally, according to the proposition* 1 *the set* X *is a maximal biprefix code.*

2) *Let* $\psi : \Sigma^* \longrightarrow (\mathbb{Z}/n\mathbb{Z}, \oplus)$ *be the morphism of monoids defined by:* $\psi(\sigma) = \overline{1}$ *for all* $\sigma \in \Sigma$, *and* $\psi(\epsilon) = \overline{0}$.

Then, we have for all $w \in \Sigma^* : \psi(w) = |w| \bmod (n)$. *We have the mapping* ψ *is surjective because for all* $\overline{m} \in \mathbb{Z}/n\mathbb{Z}$, *the word* $w = \sigma^m \in \Sigma^*$, *for all* $\sigma \in \Sigma$, *satisfies the condition* $\psi(\sigma^m) = \overline{m}$. *And if* $X^* = \psi^{-1}(\{\overline{0}\}) = \{w \in \Sigma^* : |w| \equiv 0 \, mod \ (n)\}$ *then,* $X = \Sigma^n$. *According to Proposition* 1. *This shows that the set* X *is a maximal biprefix code.*

Proposition 2. *[1] Let* $X \subseteq \Sigma^*$ *be a subset of* Σ^*. *Then* X *is a group code if and only if the syntactic monoid* $M(X^*)$ *is a group.*

Example 2. *Consider the morphism* $\psi : \{a,b\}^* \longrightarrow (\mathbb{Z}/2\mathbb{Z}, \oplus)$ *defined by:* $\psi(a) = \overline{0}, \psi(b) = \overline{1}, \psi(\epsilon) = 0$.

We take $X^* = \psi^{-1}(\{\overline{0}\}) = \{w \in \{a,b\}^* : |w|_b \equiv 0\}$ *[5]. The minimal generator of* X^* *is the set* $X = ba^*b \cup \{a\}$.

To show that the monoid syntactic $M(X^*)$ *of* X^* *is group, it suffices to show that the transition monoid of the minimal automaton of* X^* *is group.*

The minimal automaton of $X^* A = (Q, q_0, F, \Sigma, \delta)$ *is constructed using the congruence* $\sim_{X^*}$ *with,*

1) $Q = \{\overline{w}, w \in \{a,b\}^*\}$;

2) $q_0 = \overline{\epsilon}$;

3) $F = \{\overline{w}, w \in X^*\}$;

4) $\delta(\overline{w}, \sigma) = \overline{w\sigma}, \overline{w} \in \{a,b\}^* / \sim_{X^*}, \sigma \in \{a,b\}^*$.

The right congruence relation $\sim_{X^*}$ *defined by:*

$$\begin{aligned}
\forall w_1 \in \{a,b\}^*, \forall w_2 \in \{a,b\}^*, (w_1 \sim_{X^*} w_2) &\iff (\forall x \in \{a,b\}^*, w_1 x \in X^* \iff w_2 x \in X^*) \\
\forall w_1 \in \{a,b\}^*, \forall w_2 \in \{a,b\}^*, (w_1 \sim_{X^*} w_2) &\iff (\forall x \in \{a,b\}^*, |w_1 x|_b \equiv 0 \iff |w_2 x|_b \equiv 0) \\
\forall w_1, w_2 \in \{a,b\}^*, (w_1 \sim_{X^*} w_2) &\iff (\forall x \in \{a,b\}^*, |w_1|_b + |x|_b \equiv 0 \iff |w_2|_b + |x|_b \equiv 0).
\end{aligned}$$

Then the monoid quotient contains two classes $\{a,b\}^* / \sim_{X^*} = \{X^*, \overline{X^*}\}$, *where* $\overline{X^*}$ *is the complement of* X^* *with respect to* $\{a,b\}^*$. *The transition function of the automaton* $A = (Q, q_0, F, \Sigma, \delta)$ *is defined by:*

$$\begin{aligned}
\delta(X^*, a) &= X^*, \delta(X^*, b) = \overline{X^*} \\
\delta(\overline{X^*}, a) &= \overline{X^*}, \delta(\overline{X^*}, b) = X^*.
\end{aligned}$$

The transition monoid of the automaton $A = (Q, q_0, F, \Sigma, \delta)$ *is* $\varphi(\{a,b\}^*)$ *with,* $\varphi : \{a,b\}^* \longrightarrow Q^Q$. *And* Q^Q *is the monoid of all mappings from* Q *to* Q *equipped with the operation of composition.*

Then,

$$\begin{aligned}
\varphi(a) &= \begin{pmatrix} X^* & \overline{X^*} \\ X^* & \overline{X^*} \end{pmatrix} = id_Q, \\
\varphi(b) &= \begin{pmatrix} X^* & \overline{X^*} \\ \overline{X^*} & X^* \end{pmatrix} = (X^* \overline{X^*}).
\end{aligned}$$

Then $\varphi(\{a,b\}^*) = \langle \varphi(b) \rangle$, *since* $\varphi(b)$ *is a permutation of order* 2 *then* $\varphi(\{a,b\}^*)$ *is a group of order* 2. *Finally, we obtain* $M(X^*)$ *is a group of order* 2. *(see Figure 1)*

Proposition 3. *[1] Let* $X \subseteq \Sigma^*$ *a finite code. Then* $M(X^*)$ *is a group if and only if* $X = \Sigma^n$. *And in this case, the syntactic monoid* $M(X^*)$ *is cyclic group of order* n.

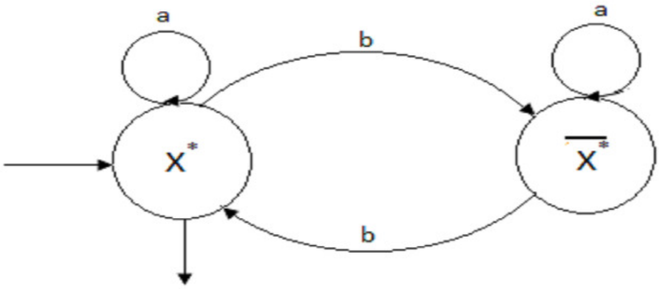


Figure 1: The minimal automaton of X*

Example 3. *Consider the morphism of monoids* $\psi : \{a,b\}^* \longrightarrow \mathbb{Z}/3\mathbb{Z}$ *defined by:* $\psi(a) = \psi(b) = \overline{1}, \psi(\epsilon) = \overline{0}$.

We take $X^* = \psi^{-1}(\{\overline{0}\}) = \{w \in \{a,b\}^* : |w| \equiv 0\}$ *[6].*

Then $X = \{a,b\}^3 = \{aaa, abb, aab, aba, baa, bbb, bab, bba\}$.

To show that the monoid syntactic $M(X^*)$ *of* X^* *is cyclic group of order* 3*, it suffices to show that the transition monoid of the minimal automaton of* X^* *is cyclic group of order* 3.

The minimal automaton of X^* $A = (Q, q_0, F, \Sigma, \delta)$ *is constructed using the congruence* $\sim_{X^*}$ *with,*

1) $Q = \{\overline{w}, w \in \{a,b\}^*\}$;

2) $q_0 = \overline{\epsilon}$;

3) $F = \{\overline{w}, w \in X^*\}$;

4) $\delta(\overline{w}, \sigma) = \overline{w\sigma}, \overline{w} \in \{a,b\}^* / \sim_{X^*}, \sigma \in \{a,b\}^*$.

The right congruence relation on $\sim_{X^*}$ *is defined by:*

$$\begin{aligned}
\forall w_1 \in \{a,b\}^*, \forall w_2 \in \{a,b\}^*, (w_1 \sim_{X^*} w_2) &\iff \left(\forall x \in \{a,b\}^*, w_1 x \in X^* \iff w_2 x \in X^*\right). \\
\forall w_1 \in \{a,b\}^*, \forall w_2 \in \{a,b\}^*, (w_1 \sim_{X^*} w_2) &\iff \left(\forall x \in \{a,b\}^*, |w_1 x| \equiv 0 \iff |w_2 x| \equiv 0\right). \\
\forall w_1, w_2 \in \{a,b\}^*, (w_1 \sim_{X^*} w_2) &\iff \left(\forall x \in \{a,b\}^*, |w_1| + |x| \equiv 0 \iff |w_2| + |x| \equiv 0\right).
\end{aligned}$$

Then we have: $\{a,b\}^* / \sim_{X^*} = \{X^*, L_1, L_2\}$ *with* $L_1 = \{w \in \{a,b\}^* : |w| \equiv 1\}$ *and* $L_2 = \{w \in \{a,b\}^* : |w| \equiv 2\}$.

The transition function of the automaton $A = (Q, q_0, F, \Sigma, \delta)$ *is defined by:* $\delta(X^*, a) = L_1, \delta(X^*, b) = L_1, \delta(L_1, a) = L_2, \delta(L_1, b) = L_2, \delta(L_2, a) = X^*, \delta(L_2, b) = X^*$.

The transitions monoid of automaton $A = (Q, q_0, F, \Sigma, \delta)$ *is* $\theta(\{a,b\}^*)$ *with* $\theta : \{a,b\}^* \longrightarrow Q^Q$.

Then

$$\begin{aligned}
\theta(a) &= \begin{pmatrix} X^* & L_1 & L_2 \\ L_1 & L_2 & X^* \end{pmatrix} = (X^* L_1 L_2) \\
\theta(b) &= \begin{pmatrix} X^* & L_1 & L_2 \\ L_1 & L_2 & X^* \end{pmatrix} = (X^* L_1 L_2).
\end{aligned}$$

Then $\theta(\{a,b\}^*) = \langle \theta(a) \rangle = \langle \theta(b) \rangle$. *Since* $\theta(a)$ *is cycle of length* 3 *then* $\theta(\{a,b\}^*)$ *is a cyclic group of order* 3.

Finally, $M(X^*)$ *is a cyclic group of order* 3 *(see Figure 2).*

In following proposition, we show that the submonoid X^* generated by the group code X is a complete submonoid of the free monoid Σ^*.

Proposition 4. *Let* G *be a group and* H *a subgroup of* G. *Let* $\psi : \Sigma^* \longrightarrow G$ *be a surjective morphism.*

The submonoid $X^* = \psi^{-1}(H)$ *with* X *is the minimal generator of the set* X^**is a complete submonoid in the free monoid* Σ^*.

Proof. To obtain the desired result, we show that for all $w \in \Sigma^*$, there exists $u, v \in \Sigma^*$ such that $uwv \in X^*$.

Since $X^* = \psi^{-1}(H)$, then we have

$$uwv \in X^* \text{ if and only if } \psi(uwv) \in H \text{ if and only if } \psi(u)\psi(w)\psi(v) \in H.$$

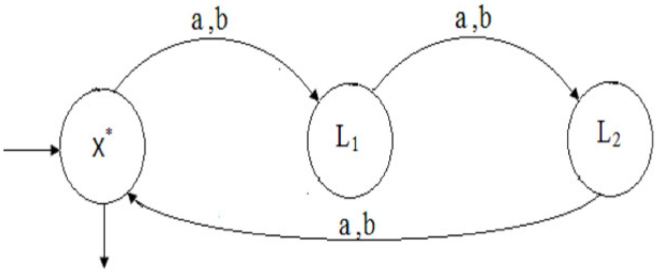


Figure 2: The minimal automaton of X*

Then it suffices to take $\psi(u) = (\psi(w))^{-1}$ ($(\psi(w))^{-1}$ denote the inverse of $\psi(w)$ in the group G), i.e., $u \in \psi^{-1}\left(\left\{(\psi(w))^{-1}\right\}\right)$ with $\psi^{-1}\left(\left\{(\psi(w))^{-1}\right\}\right)$ is the inverse image of $\left\{(\psi(w))^{-1}\right\}$ and $v = \epsilon$.

We have then $\psi(u)\psi(w)\psi(v) = (\psi(w))^{-1}.\psi(w).\psi(\epsilon) = 1_G.1_G = 1_G \in H$. □

Example 4. *We show that the submonoid $X^* = \left\{w \in \{a,b\}^* : |w|_a = |w|_b\right\}$ cited in Example 2, is a complete submonoid in the free monoid $\{a,b\}^*$. Let $w \in \{a,b\}^*$, there is only three cases to be considered.*

Case 1. *If $w \in X^*$ then $|w|_a = |w|_b$ in this case it suffices to take $u = v = \epsilon$ and we have $\epsilon w \epsilon = w \in X^*$.*

Case 2. *If $|w|_a < |w|_b$ then it suffices to take $u = a^{|w|_b - |w|_a}$ and $v = \epsilon$. We have then $|uwv|_a = |u|_a + |w|_a + |v|_a = |w|_b - |w|_a + |w|_a + 0 = |w|_b$. In the same case $|uwv|_b = |u|_b + |w|_b + |v|_b = 0 + |w|_b + 0 = |w|_b$. Then, we have $|uwv|_a = |uwv|_b$ and consequently $uwv \in X^*$.*

Case 3. *If $|w|_a > |w|_b$ then it suffices to take $u = b^{|w|_a - |w|_b}$ and $v = \epsilon$. We have then $|uwv|_a = |u|_a + |w|_a + |v|_a = 0 + |w|_a + 0 = |w|_a$. In the same case $|uwv|_b = |u|_b + |w|_b + |v|_b = |w|_a - |w|_b + |w|_b + 0 = |w|_a$. Then, we have $|uwv|_a = |uwv|_b$ and consequently $uwv \in X^*$.*

Finally, the three cases shows that the submonoid X^* is a complete submonoid of $\{a,b\}^*$.

4 Conclusion

In this paper, we give some examples of group code and we show that the submonoid X^* generated by the group code X is a complete submonoid of the free monoid Σ^*.

References

[1] J. Berstel, D. Perrin, *Theory of Codes*, Academic Press, 1984.

[2] S. Marcel,” Langage formels et monoïdes finis”,*Séminaire Dubreil, Algèbre et théorie des nombres*, vol. 23, no. 2, pp. 1–3, 1970.

[3] N. Maurice, ”Eléments de la théorie général des codes”, Université de Paris, 1966.

[4] R. Michel, ”Théorie des automates et langages formels”, *Université Liège*, 2010.

[5] D. Perrin, ”Le degrè minimal du groupe d’un code biprèfixe fini”, *Journal of Combinatorial Theory, Series A*, vol. 25, no. 2, pp. 163–173, Sept. 1978.

[6] D. Perrin, G. Rindone, ”On syntactic groups”, *Bulletin of the Belgian Mathematical Society-Simon Stevin*, vol. 10, no. 5, pp. 749–759, 2003.

[7] A. Salomma, ”Jewels of Formal Language Theory”, Computer Science Press, 1981.

[8] H. J. Shyr, ”Free monoids and languages”, Department of Mathematicsn, Soochow University, Taipei, Taiwan, 1979.

[9] G. Viennot, ”Factorisations des monoides libres, bascules, et algebres de Lie libres”, *Seminaire Dubreil. Algebre*, vol. 25, no. 2, pp. 1–8, 1973.

Nacer Ghadbane is a research scholar in the Laboratory of Pure and Applied Mathematics, Department of Mathematics, University of M’sila, Algeria.

Douadi Mihoubi is a professor in the Laboratory of Pure and Applied Mathematics, Department of Mathematics, University of M’sila, Algeria.

Hybrid Rough-Fuzzy Classifier for Liver Disease Diagnosis

Shimaa Abd Allah Ibraheem, Hatem Mohamed Abd Elkader, Ibraheen Selim, and Reda Hussein

(Corresponding author: Shimaa Abd Allah Ibraheem)

Computer Science Department, Higher Technological Institute
Next to Small Industries Complex, Industrial Area2, Ash Sharqia Governorate, Egypt
(Email: earth_sky_2014@yahoo.com)

(Received Mar. 25, 2016; revised and accepted May 5, 2016)

Abstract

The intelligent classification techniques have been widely used in the medical field for accurate classification. The liver disease diagnoses is one of the most commonly tasks in medical diagnosis area. However liver disease in most cases does not cause any symptoms at earlier stage. Symptoms partly rely on the type and the degree of liver disease. In this paper an intelligent technique called a hybrid rough- fuzzy classifier (HRFC) is proposed for liver disease diagnosis. Rough sets are used to generate and reduce classification rules which are used in fuzzy set to enhance the classification accuracy of liver diseases diagnoses. The classification is implemented by two phases. In the first phase, rough set rules generated using LEM2 algorithm is applied to generate minimal classification rules. The rule induction is used to improve the accuracy. In the second phase fuzzy inference system is applied to identify the types or risks of the liver disease. The results shows that the proposed model gives 99.1 percent classification accuracy in rule generation.

Keywords: Classification techniques, fuzzy inference system, LEM2 algorithm, liver diseases, rough set

1 Introduction

The liver disease is considered one of the most dangerous diseases in the world. Because this disease may not cause any symptoms at earlier stage. It also may cause vague symptoms, as loss of energy and weakness. Liver diseases are diagnosed based on the liver functional test [10]. So the intelligent systems especially classification techniques are playing an important role to early detect and diagnose this disease. In general, the classification aims to classify, objects and mini real world state into classes. Each of the represented objects is original and this mean that classification is a real degree of generalization. Computational intelligence (CI) techniques are considered efficient tools for implementation of a classification model. CI models composed of several models like, neural networks, fuzzy sets, evolutionary computation algorithms, and rough sets. CI methods have two commonalities which are the non-symbolic representation of pieces of knowledge [2] and "bottom-up" architecture where the structures and paradigms appear from an unordered beginning [23]. Hybrid models which combine several CI models in one framework are very effective and can easily enhance the classification and prediction accuracy in many areas.

The rough sets theory (RST) [11] is based on the research of information system logical properties [14], and uncertainty [15] in it is expressed by a boundary region. RST were used for a definition of IF-THEN rules and fuzzy system FSs were applied in RFC as a fuzzy inference system (FIS). FIS have been successfully applied in fields such as modelling of municipal creditworthiness, automatic control, decision analysis, data analysis, decision systems or expert system [4, 7]. Classification techniques are considered the cornerstone of all intelligent medical diagnoses tools. In case of liver disease, patients are not easily discovered in early stages as it will be functioning normally even when it is partially damaged. An early diagnosis of liver disease will save the patient's life [8].

There are many technologies interested in establishing a liver diseases diagnosis. Gulia et al. "liver patient classification using intelligent techniques" [8] and Karthik, et al. "Classification and Rule Extraction using Rough Set for Diagnosis of Liver Disease and its Types" [10] liver disease". And Satarkar1, et al. "fuzzy expert system for the diagnosis" [20]. In this work we built a hybrid CI approach called hybrid rough-fuzzy classifier (HRFC), which combines rough sets (RSs) and fuzzy systems (FSs) in one framework. The aim of this study is to build a robust expert system for liver disease diagnosis based on of proposed HRFC hybrid model. RSs are used to generate and reduce classification rules which are employed in fuzzy set to provide accurate classification of liver diseases diagnose.

The rest of this paper is represented as follows: Section 2 presents rough set concepts. Section 3 introduces Fuzzy inference system concepts; Section 4 presents the architecture of proposed hybrid model. Section 5 presents the experimental results of the proposed model. Section 6 outlines the conclusion of this paper.

2 Rough Set Concept

RST is relying on IS "information system" concept. IS can be considered as an information/decision table [5, 11] which data set is reviewed as set of rows and columns, each row refer to object case and every column refer to an attribute which are measures for each object. IS is defined by the following functions:

$$IS = (U, A, V_a, f_a)\forall a_i \in A, i = 1, 2, \ldots, n. \tag{1}$$

Where, the finite set of objectives (Universe) is defined as $U = x_1, x_2, \cdots, x_m$ An attributes finite set is defined as $A = a_1, a_2, \cdots, a_n$, The attributes domain of is V_a, where $V_a = v_1 1, x_1 2, \cdots, v_m 1, \cdots, v_m n$, $f_a : U \rightarrow V_a$ is an information function like $f(x, a) \in V_a$ for each $a \in A$, $x \in U$.

Data set is appeared as an information table, where attributes are represented by column and each row represents an object. Descriptor is needed for object and attribute. Descriptor is exact and precise value of attribute. When the attribute values prevent generally their precise classification this means limited objects discernibility [21]. RST can deal and support uncertainty and vagueness data. RST approach deal with data/field uncertainty or vagueness by a lower and the upper approximation. The objects as supposition can be only seen by the available information about them, so that knowledge has idol structure. So some objects seem as similar and undiscerned. All objects which certainly belong to the vague concept is composed of the lower Approximation, upper approximation of all objects which may be belong to the concept The boundary region is difference between the upper and lower approximation [9].

Assume we have two non-empty and finite sets U and A, U is called the universe and A is a set of attributes. For each attributes $a \in A$ is associate a set of V_a (value set) called the domain of a. Any subset B of A determines a binary relation $IND(B) on U$ which will be called an indiscernibility relation is an equivalence relation and is called B-indiscernibility relation [11].

$$IND(B) = (x, y) \in U \mid \forall a \in Ba(x) = a(y), \tag{2}$$

$If (x, y) \in IND(B)$, then x and y are B-indiscernible (indiscernible from each other by attributes from B).this refer to $IND(B)$ is an equivalence relation and is called B-indiscernibility relation. The equivalence classes of the Indiscernibility relation will be denoted $B(x)$. Basic concept of RST is defined by The Indiscernibility relation. Let IS be define Equation (1) and let and $B \subseteq A$ and $X \subseteq U$. We can approximate X using only the information contained in B by constructing lower approximation and upper approximation of X on the following way:

$$\underline{B}(X) = (x \in U : B(x) \subseteq X) \tag{3}$$

$$\overline{B}(X) = (x \in U : B(x) \cap X \neq \phi) \tag{4}$$

The upper approximation objects are classified as possible members of X on the basis of knowledge in B. when the lower approximation objects can be surly classified as members of X on the basis of knowledge in B:

$$BNB(X) = (\overline{B}(X) - \underline{B}(X)), \tag{5}$$

Equation (5) refers to the boundary region of X which composed of these objects that never classify into X on the basis of knowledge B. Based on rough sets [11, 14, 16]. If the boundary region is empty, this mean the set X is crisp with respect to B. If the boundary region is not empty, the set X is rough with respect to B.

3 Fuzzy System Concept

The theory of fuzzy has advanced in a variety of ways and in many disciplines [17] fuzzy theory is used in many application and support many concept. Fuzzy set theory offers an important contribution to data mining leading to fuzzy data mining [13]. Fuzzy sets are sets whose elements have degrees of membership. It introduced by Zadehand Dieter Klaua. It is one on most important approach that deals with uncertainty, vagueness. Fuzzy set theory support membership degree (membership function) concept for each element in set; this membership function valued in the real unit interval [0, 1]. Fs process sets theory as each element whether is or is not a set member. Molding and describing reality problem lead to a certain conflict. It enhance two issues,first issue which there is mathematical methods precision by which a certain problem is described and, the second issue, there is consequent inaccuracy of model because a more complicated reality blackmail a range of simplifications.

We find Disproportionate increase of the number of terms and constraints as result of effort to maximize accuracy. The principal of incompatibility in [24] is formulated: If the complexity of a system increased, our ability to formulate accurate and significant judgments about its behavior decreases, and the boundary is reached behind which accuracy and relevance are practically mutually exclusive characteristics.

Let X is a variable which have values from universe set U and real number N be assigned to each element $u \in U$ where $N_(u) \in [0,1]$. While $N_(u)$ show that the possibility degree of variable X take just value u. In FSs theory, function set is defined by membership function (x) on universe U by. X does not belong to function N when $\mu_N(x) = 0$, otherwise $\mu_N(x) = 1$ then x belongs to function set N, but x partially belongs to function set N when $\mu_N(x) \in (0,1)$, in other term it is not possible to be identify certainly if X belongs to function set [24, 25]. Fuzzy inference system is one of most approaches deal with j uncertainty and vagueness which uses concept of FSs for diagrammatic representation and mapping from a given input to a driven output. FIS have two types of FIS Mamdani and Sugeno. Architecture of FIS consist of inputs, fuzzification process, and input/output membership functions, base rules, fuzzy logic/FSs operators, implication and aggregation, defuzzification and output.

4 Ootline Architecture for HRFC

Combin Rough and fuzzy theory is a suitable for dealing with noisy, vague, uncertain, or inexact information [3, 1, 22, 12, 19]. The rough set and fuzzy set theory have generated a great deal of interest among more and more researchers [6]. While Rough Set theory has been conceived as a tool to conceptualize, organize and analyze various types of data [18], but fuzzy for mathematics operation. The problem of classification in our model is composed of two phases: first is rules generation using Rough Sets approaches and second phase is using FIS based on FSs to classify or diagnose liver diseases. The Goal is to generating conditional rules to create and analyze a hybrid HRFC data classifier model. LEM2 algorithm is used for rules generation which exploits RST. LEM2 (Learning from Examples Module version-2) is used as modules in the algorithm LERS for learning from examples based on rough set These rules were used in Mamdani type of fuzzy inference system which represents importance of HRFC. Figure 1 show how these two phases are combined or hybrid in RFC model.

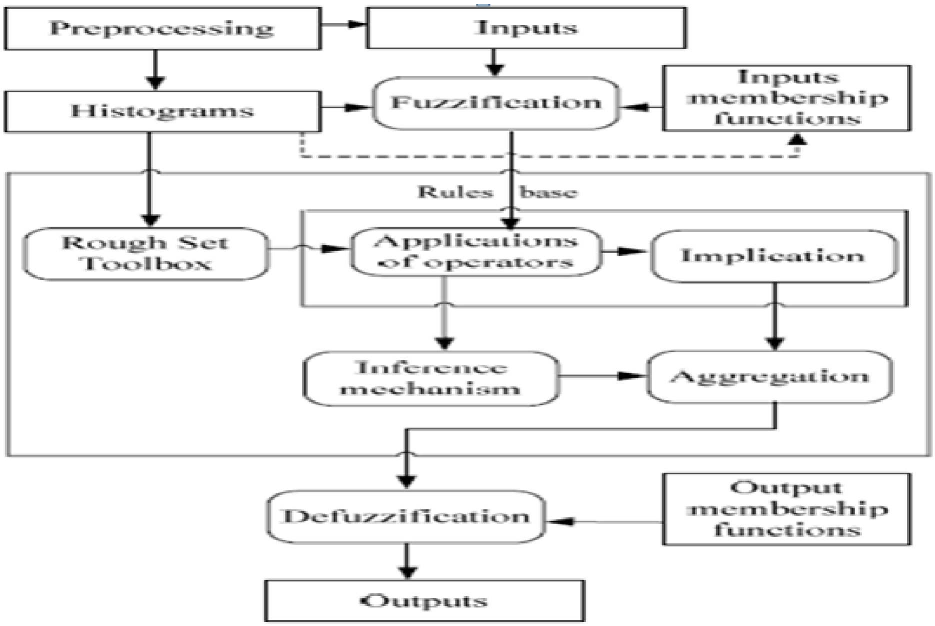


Figure 1: Hybrid Rough - fuzzy classifier model

5 The Experimental Results

The purpose of this study is build robust expert system for diagnoses diseases of liver based on hybrid model rough-fuzzy set (HRFC) through two phases.

5.1 First Phase

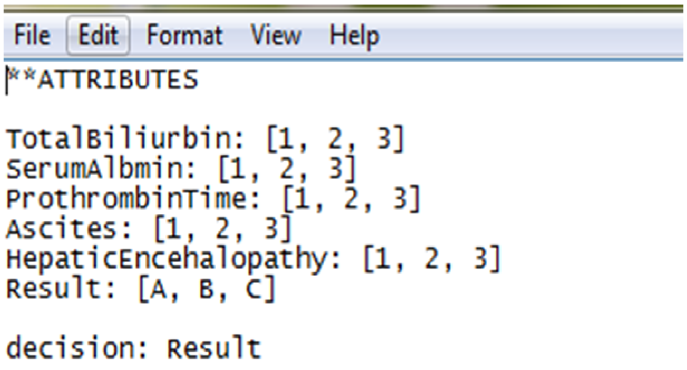
In this phase we apply rough set technique (RST) for extract and generate minimal cover rules using LEM2 algorithm.

5.1.1 Data Set

The dataset were collected from medical data warehouse of liver disease patients.our dataset consist of 466 cases, it consists of five measured variables as follow:

- Total Bilirubin: in the total bilirubin there are three values (1, 2, and 3) referring to low, medium and high.
- Serum Albumin: the Serum albumin is one of most important factor that effect on liver which has three values (1, 2, and 3) that refer to high, medium and low.
- Prothrombin Time (PT): is a blood test that measures the time it takes for the blood plasma to coagulate. It has three values which refer to high, medium and low.
- Ascites: It usually occurs when the liver stops working correctly; this lead to pain and swelling in the abdomen, and Queasiness.it is done when fluid fills the space between the abdominal lining and the organs. It has three values that refer to high, medium and low.
- Hepatic Encephalopathy: it decrease function of brain that come as a result of highly liver disease, it has three values refer to high, medium and low.

Figure 2 shows attributes and its values, (Result) is the decision attribute, it is mean liver risk levels which its value (A, B, C) refer to low risk, moderate risk, high risk.

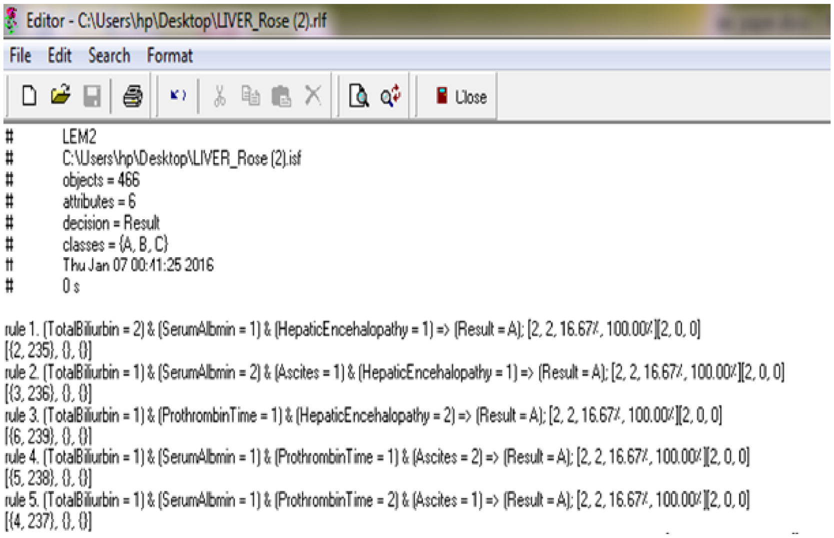


```
**ATTRIBUTES

TotalBiliurbin: [1, 2, 3]
SerumAlbmin: [1, 2, 3]
ProthrombinTime: [1, 2, 3]
Ascites: [1, 2, 3]
HepaticEncehalopathy: [1, 2, 3]
Result: [A, B, C]

decision: Result
```

Figure 2: Attributes with its values

5.1.2 Rule Generation

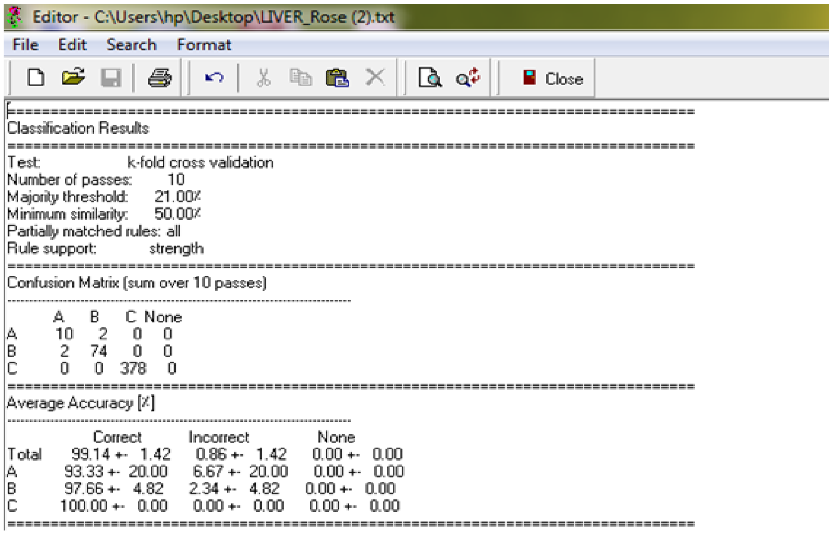
We establish a set of rules, the "minimum cover rules" this mean that the set of rules does not contain any redundant rules, as well as it is a certain rules, such that there are a total of 31 rules generated from the data. The following Figure 3 shows sample of the minimum cover rules obtained.



```
#       LEM2
#       C:\Users\hp\Desktop\LIVER_Rose (2).isf
#       objects = 466
#       attributes = 6
#       decision = Result
#       classes = {A, B, C}
#       Thu Jan 07 00:41:25 2016
#       0 s

rule 1. (TotalBiliurbin = 2) & (SerumAlbmin = 1) & (HepaticEncehalopathy = 1) => (Result = A); [2, 2, 16.67%, 100.00%][2, 0, 0]
[{2, 235}, {}, {}]
rule 2. (TotalBiliurbin = 1) & (SerumAlbmin = 2) & (Ascites = 1) & (HepaticEncehalopathy = 1) => (Result = A); [2, 2, 16.67%, 100.00%][2, 0, 0]
[{3, 236}, {}, {}]
rule 3. (TotalBiliurbin = 1) & (ProthrombinTime = 1) & (HepaticEncehalopathy = 2) => (Result = A); [2, 2, 16.67%, 100.00%][2, 0, 0]
[{6, 239}, {}, {}]
rule 4. (TotalBiliurbin = 1) & (SerumAlbmin = 1) & (ProthrombinTime = 1) & (Ascites = 2) => (Result = A); [2, 2, 16.67%, 100.00%][2, 0, 0]
[{5, 238}, {}, {}]
rule 5. (TotalBiliurbin = 1) & (SerumAlbmin = 1) & (ProthrombinTime = 2) & (Ascites = 1) => (Result = A); [2, 2, 16.67%, 100.00%][2, 0, 0]
[{4, 237}, {}, {}]
```

Figure 3: Sample of minimal covered rules

This rules generated using LEM2 algorithm with classification accuracy 99.1. Figure 4 show the classification accuracy of rules generated using k-fold cross validation technique.



```
==========================================================================
Classification Results
==========================================================================
Test:                k-fold cross validation
Number of passes:        10
Majority threshold:     21.00%
Minimum similarity:     50.00%
Partially matched rules: all
Rule support:           strength
==========================================================================
Confusion Matrix (sum over 10 passes)
--------------------------------------------------------------------
        A     B     C  None
A      10     2     0     0
B       2    74     0     0
C       0     0   378     0
==========================================================================
Average Accuracy [%]
--------------------------------------------------------------------
            Correct          Incorrect          None
Total     99.14 +-  1.42    0.86 +-  1.42    0.00 +-  0.00
A         93.33 +- 20.00    6.67 +- 20.00    0.00 +-  0.00
B         97.66 +-  4.82    2.34 +-  4.82    0.00 +-  0.00
C        100.00 +-  0.00    0.00 +-  0.00    0.00 +-  0.00
==========================================================================
```

Figure 4: RST result of cross-fold validation technique

5.2 Second Phase

In this phase we construct FIS for liver diseases through using rules generate by RST for diagnose or classify Pathological stages (risks) of the patient in liver diseases diagnose system. We represent the fuzzy expert system by designing Membership functions, fuzzy rule base, fuzzification and defuzzifaction.

There are five input variables and one output variable. So we construct the membership functions of all the variables are designed. At first the input variables with their membership functions are described. In second step output variable with its membership functions are described. In fuzzy rule base step Rules are employed from the first phase. Then result is generated in last step of fuzzification and defuzzification.

5.2.1 Input Variables

Fuzzy set range of Total Bilirubin is shown in Table 1. Membership functions for fuzzy sets of Total Bilirubin are trapezoidal and triangular and are shown in Figure 5.

Table 1: Fuzzy sets of total bilirubin

Input field	Range	Fuzzy set
Total Bilirubin	<34(<2)	1
	34-50(2-3)	2
	>50 (>3)	3

Table 2: Fuzzy sets of serum albumin

Input field	Range	Fuzzy set
Serum Albumin	>3.5	1
	2.8-3.5	2
	<2.8	3

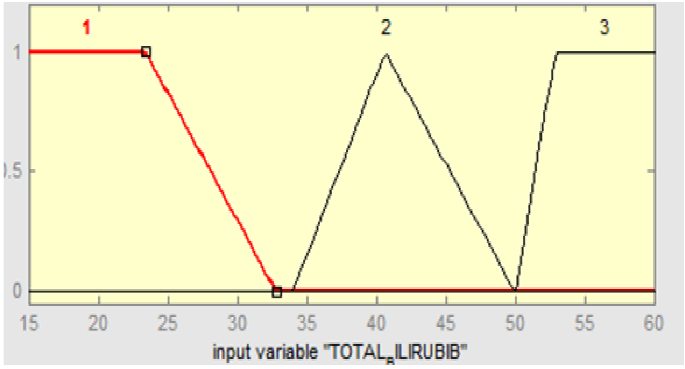


Figure 5: Membership functions for total bilirubin

Fuzzy set range of Serum Albumin is shown in Table 2. The membership functions for fuzzy sets are trapezoidal and triangular and are shown in Figure 6.

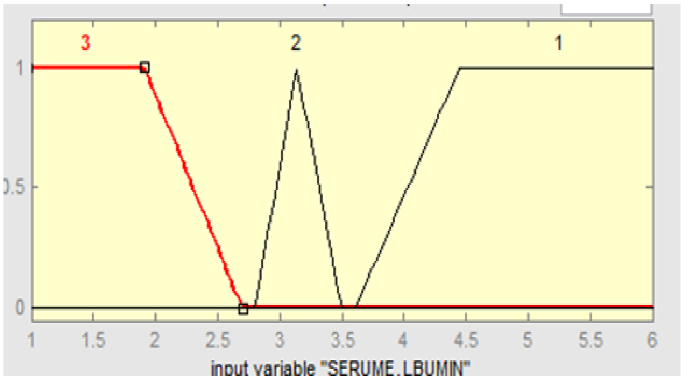


Figure 6: Membership functions for serum albumin

Fuzzy set range of Prothrombin Time is shown in Table 3. Membership functions for fuzzy sets are trapezoidal and triangular and are shown in Figure 7.

Table 3: Fuzzy sets of prothrombin time

Input field	Range	Fuzzy set
Prothrombin Time	<4.0	1
	4.0-6.0	2
	>6	3

Table 4: Fuzzy sets of ascites

Input field	Range	Fuzzy set
Ascites	none (<2)	1
	mild (2.0-3.0)	2
	Moderate to severe (>3)	3

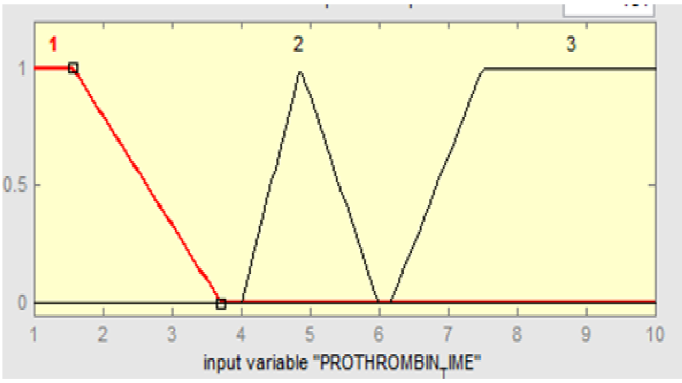


Figure 7: Membership functions for prothrombin time

Fuzzy set range of Ascites is shown in Table 4. Membership functions for fuzzy sets are trapezoidal and triangular and are shown in Figure 8.

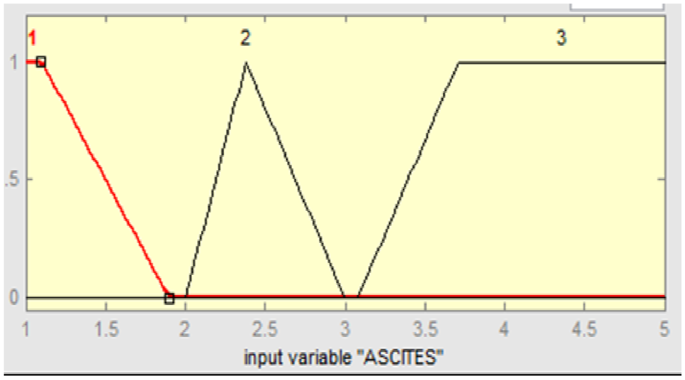


Figure 8: Membership functions for ascites

Fuzzy set range of Hepatic Encephalopathy is shown in Table 5. Membership functions for fuzzy sets are trapezoidal and triangular and are shown in Figure 9.

Table 5: Fuzzy sets of hepatic encephalopathy

Input field	Range	Fuzzy set
Hepatic Encephalopathy	None	1
	Grade I-II (2-3)	2
	Grade III IV(>3)	3

Table 6: Fuzzy sets of output results

Input field	Range	Fuzzy set
Result	5-6	A
	(7-9)	B
	(10-15)	C

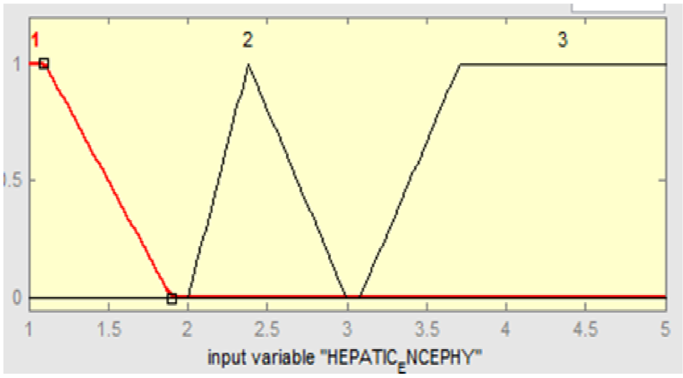


Figure 9: Membership functions for hepatic encephalopathy

5.2.2 Output Variables

The aim of the system is to identify risk status of liver. The output is a value from 0 to 15 Representing fuzzy set A, B, C which refer to Low risk, moderate risk and High risk. These fuzzy sets and its ranges are shown in Table 6. The Membership functions of these fuzzy sets are triangular and are shown in Figure 10.

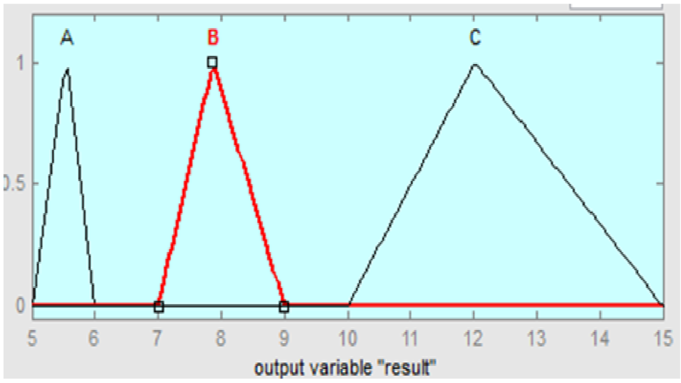
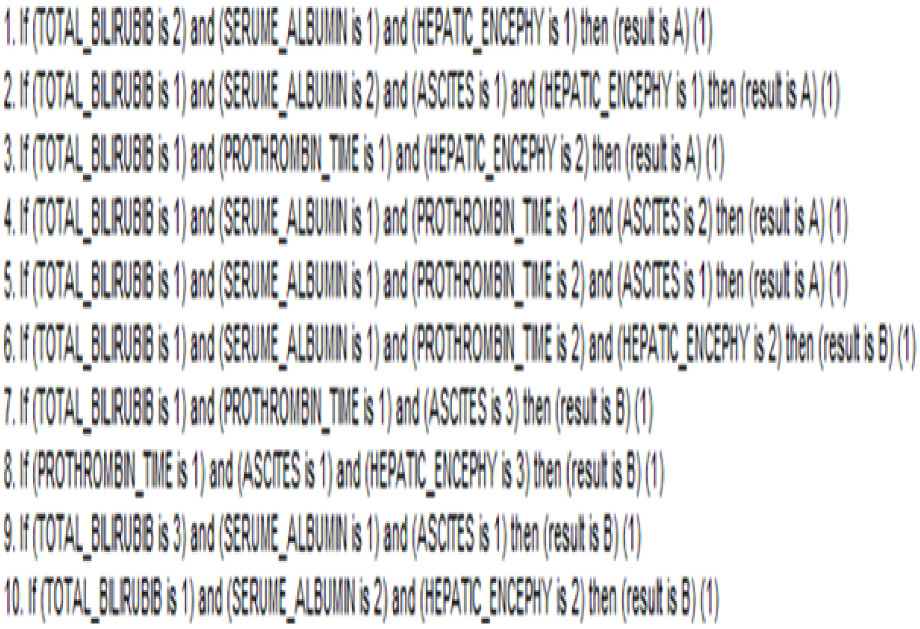


Figure 10: Membership functions for output result

5.2.3 Fuzzy Rule Base

The rule base which determined before with RST using LEM2 algorithm are entered to FIS system as fuzzy rule base. The rule base consists of 31 well defined rules that determine the risk status. Sample of rule base in FIS is shown in Figure 11.



1. If (TOTAL_BILIRUBB is 2) and (SERUME_ALBUMIN is 1) and (HEPATIC_ENCEPHY is 1) then (result is A) (1)
2. If (TOTAL_BILIRUBB is 1) and (SERUME_ALBUMIN is 2) and (ASCITES is 1) and (HEPATIC_ENCEPHY is 1) then (result is A) (1)
3. If (TOTAL_BILIRUBB is 1) and (PROTHROMBIN_TIME is 1) and (HEPATIC_ENCEPHY is 2) then (result is A) (1)
4. If (TOTAL_BILIRUBB is 1) and (SERUME_ALBUMIN is 1) and (PROTHROMBIN_TIME is 1) and (ASCITES is 2) then (result is A) (1)
5. If (TOTAL_BILIRUBB is 1) and (SERUME_ALBUMIN is 1) and (PROTHROMBIN_TIME is 2) and (ASCITES is 1) then (result is A) (1)
6. If (TOTAL_BILIRUBB is 1) and (SERUME_ALBUMIN is 1) and (PROTHROMBIN_TIME is 2) and (HEPATIC_ENCEPHY is 2) then (result is B) (1)
7. If (TOTAL_BILIRUBB is 1) and (PROTHROMBIN_TIME is 1) and (ASCITES is 3) then (result is B) (1)
8. If (PROTHROMBIN_TIME is 1) and (ASCITES is 1) and (HEPATIC_ENCEPHY is 3) then (result is B) (1)
9. If (TOTAL_BILIRUBB is 3) and (SERUME_ALBUMIN is 1) and (ASCITES is 1) then (result is B) (1)
10. If (TOTAL_BILIRUBB is 1) and (SERUME_ALBUMIN is 2) and (HEPATIC_ENCEPHY is 2) then (result is B) (1)

Figure 11: fuzzy rules that generated by RST

5.2.4 Fuzzification and Defuzzifacation

This system depends on MAMDANI model for inference mechanism Figures 12 and 13 shows the result values of FIS. Figure 12 show that the result (liver risk) has value 12.3 when total bilirubin is 43.3, Serum Albumin is 3.2, Prothrombin Time is 7.82, ascites is 3.76 and Hepatic Encephalopathy is 2.48 as tested data.

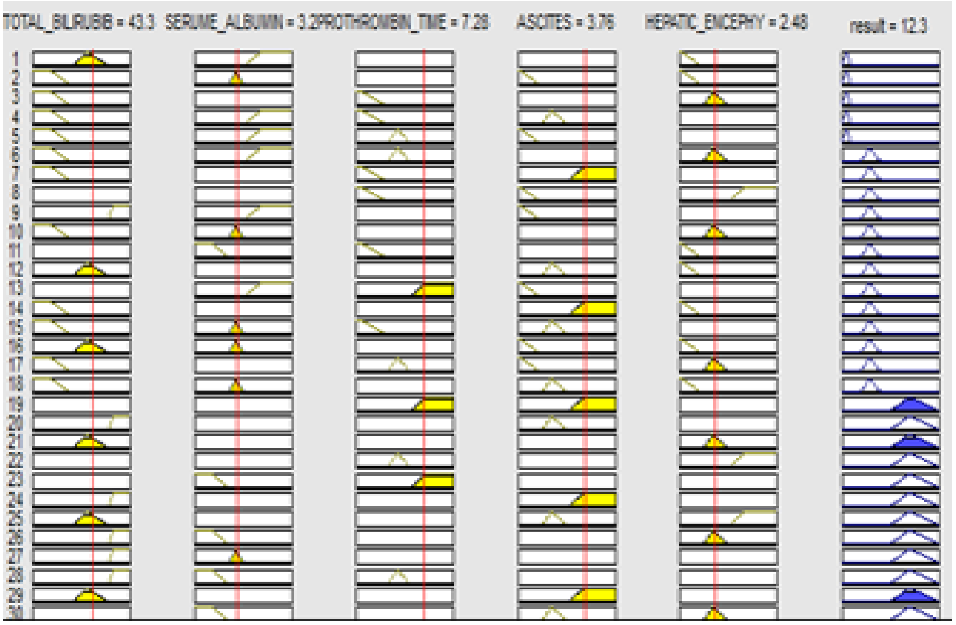


Figure 12: Result value of rules viewer

Figure 13 shows the result value of surface viewer that represents the relation between ascites and Hepatic Encephalopathy as tested value.

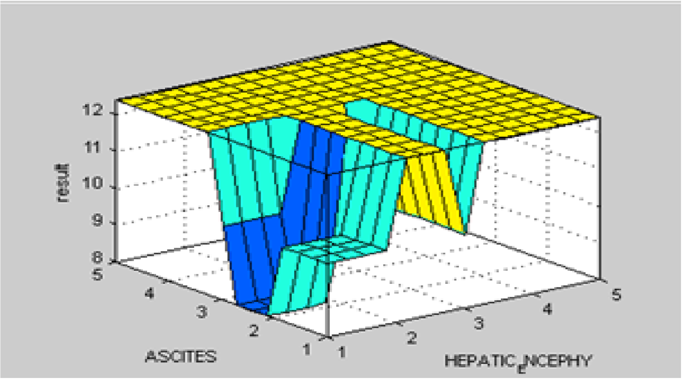


Figure 13: result value of surface viewer

6 Conclusion

In this paper a hybrid rough fuzzy classifier model (HRFC) is proposed for liver disease diagnoses. Since using RSs or FSs alone cannot provide accurate result and require more knowledge and experience. The proposed model has two stages, in first stage the RSs are used mainly to automatic generation of minimal cover rules. In the second stage, construct fuzzy inference system FIS for liver diseases through using rules generate by RST in first stage for diagnose or classify Pathological stages (risks) of the patient in liver diseases diagnose system.The experimental results presented in this paper showed that the proposed HRFC model is better accuracy than single rough sets or fuzzy classification systems.

References

[1] J. Anuradha and B. K. Tripathy, "An optimal rough fuzzy clustering algorithm using particle swarm optimization", *International Journal of Data Mining, Modeling and Management*, vol. 7, no. 4, 2015.

[2] J. C. Bezdek, "What is computational intelligence?" in *Computational Intelligence: Imitating Live*, pp. 1–12, 1994.

[3] E. Al Daoud, "An efficient algorithm for finding a fuzzy rough set reduct using an improved harmony search", *International Journal of Modern Education and Computer Science*, vol. 2, no. 1, pp. 16–23, 2015.

[4] D. Dubois and H. Prade, *Fuzzy Information Engineering and Soft Computing: A Guided Tour of Applications*, New York: John Wiley & Sons, 1997.

[5] G. Düntsch and I. Gediga, *Rough Set Data Analysis - A Road to Non-invasive Knowledge Discovery*, Angor: Methodos, 2000.

[6] M. Gamal, A. El-Fetouh, and S. Barakat, "A fuzzy rough rule based system enhanced by fuzzy cellular automata", *Annals of Fuzzy Mathematics and Informatics*, vol. 4, no. 5, pp. 1–11, 2015.

[7] S. Greco, B. Matarazzo, and R. Slowinski, "The use of rough sets and fuzzy sets in MCDM", in *Multicriteria Decision Making: Advances in MCDM Models, Algorithms, Theory, and Applications*, vol. 22, no. 5-6, pp. 14–59, 1999.

[8] A. Gulia, R. Vohra, and P. Rani, "Liver patient classification using intelligenttechniques", *International Journal of Computer Science and Information Technologies*, vol. 5, no. 4, pp. 5110–5115, 2014.

[9] K. Jir, J. Pavel, "Classification model based on rough and fuzzy sets theory", in *6th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics*, pp. 198–202, Tenerife, Spain, Dec. 2007.

[10] S. Karthik, A. Priyadarishini, J. Anuradha, and B. K. Tripathy, "Classification and rule extraction using rough set for diagnosis of liver disease and its types", *Advances in Applied Science Research*, vol. 2, no. 3, pp. 334–345, 2011.

[11] J. Komorowski, L. Polkowski, A. Skowron, "Rough sets: A tutorial", in *Rough Fuzzy Hybridization: A New Trend in Decision-making*, pp. 3–98, 1999.

[12] J. Krupka and P. Jirava, "Rough-fuzzy classifier modeling using data repository sets", in *18th Annual Conference on KES-2014*, Gdynia, Poland, Sept. 2014.

[13] C. Marsala, B. Bouchon-Meunier, "Fuzzy data mining and management of interpretable and subjective information", *Fuzzy Sets and Systems*, vol. 281, pp. 252–259, 2015.

[14] Z .Pawlak, “A. rough sets”, *International Journal of Informationand Computer Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[15] Z. Pawlak, “A primer on rough sets: A new approach to drawing conclusions from data”, *Cardozo Law Review*, vol. 22, no. 5-6, pp. 1407–1415, 2001.

[16] L. Polkowski, “Rough sets, mathematical foundations”, *Advances in Soft Computing Physica*, Springer-Verlag, vol. 1, no. 1, 2001.

[17] L. V. Popli and B. P. Singh, “Fuzzy sets and artificial intelligence: A survey”, *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 10, pp. 308–313, 2014.

[18] M. Pushpalatha and V. Anuratha, “A survey: Rough set theory in incomplete information systems”, *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 3, no. 8, pp. 7183–7186, 2015.

[19] A. Saha and A. Mukherjee, “Soft interval-valued intuitionistic fuzzy rough sets”, *Annals of Fuzzy Mathematics and Informatics*, vol. 9, no. 2, pp. 279–292, 2015.

[20] S. L. Satarkar and Dr. M. S. Ali, “Fuzzy expert system for the diagnosis of common liver disease”, *International Engineering Journal for Research & Development*, vol. 1, no. 1, pp. 1–7, 2014.

[21] J. Stefanowski and R. Sowinski, “Rough set reasoning about uncertain data”, *Fundamenta Informaticae*, vol. 22, no. 5-6, pp. 229–243, 1996.

[22] X. Yang, W. Xu, and Y. She, “Theory and application on rough set,fuzzy logic ,and granular computing”, *The Scientific World Journal*, vol. 2015, no. 1, 2015.

[23] L. A. Zadeh, “The roles of fuzzy logic and soft computing in the conception, design and deployment of intelligent systems”, *BT Technology Journal*, vol. 14, no. 4, pp. 32–36, 1996.

[24] L. A. Zadeh, “Outline of a new approach to the analysis of complex systems and decission processes”, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 1, pp. 28–44, 1973.

[25] L. A. Zadeh, “Fuzzy sets”, *Information and Control*, vol. 8, no. 1, pp. 338–353, 1965.

Shimaa Abd Allah Ibraheem was born on November 12, 1981 in benha, kaluobya, Egypt. She received the B.S from Faculty of Computers and Informatics, Zagazig University, Egypt in 2003 with grade very good, and submitted for master degree from October 2015. she is working in higher technology institute Egypt as teaching assistance at computer science department.

Hatem Mohamed Abd Elkader is vice Dean of Faculty of Computers and Information, Menoufia university, Shebin Elkom, Egypt. Prof Hatem obtained his BSC. And M.SC. (by research) both in Electrical Engineering from the Alexandria University, Faculty of Engineering, Egypt in 1990 and 1995 respectively. He obtained his Ph.D. degree in Electrical Engineering also from Alexandria University, Faculty of Engineering, in 2001 specializing in neural networks and applications. Since 2009 he is the Head of the department of Information Systems (IS). Prof. Hatem has published more than 100 papers in international journals, international conferences, local journals and local conferences.

Ibraheen Selim is an assistant in computer science of higher technology institute , tenth of Ramdan city,Egypt.

Reda Hussein works in information system department at faculty of computers and information, menoufya University ,Egypt.

Forensic SIM Card Cloning Using Authentication Algorithm

Nuril Anwar[1], Imam Riadi[2], Ahmad Luthfi[1]
(Corresponding author: Nuril Anwar)

Islamic University of Indonesia[1]
Jl. Kaliurang KM 14,5 Yogyakarta 55584
Ahmad Dahlan University[2]
Jl. Prof. Dr. Soepomo, S.H. Janturan Yogyakarta 55164, Indonesia
Email: anwar_nuril@yahoo.co.id

(Received Sept. 20, 2015; revised and accepted Jan. 11, 2016)

Abstract

Crime in the telecommunications sector increasingly, especially in the mobile security system found several security flaws of data outside of the network. Clone SIM card is a major problem in the SIM card device. Research cloning SIM card can be presented in the form of analysis algorithms A3 SRES, and A8 RAND to get Ki AUC for the investigation process digital forensic cloning SIM card, testing scheme SIM card cloning used parameter "Due Under Test" (DUT) and "Trial and Error" with the following phases ; identification, preservation, collection, examination, anally and presentation. Conclusion SIM card cloning and analysis in the form of percentage of success then conducted a forensic investigation to cloning SIM card with the matching algorithm A8 (RAND) contained in each SIM card which produces authentication Ki as contained in the investigation file structure SIM card. Memory capacity has advantages and disadvantages, which is 32kb SIM card Ki produced a success rate of 100% success, 64kb SIM card cloning success rate of 25% to 50%. Research cloning SIM card with forensic investigations have been successfully cloned.

Keywords: Authentication, Cloning, Forensic, RAND, SIM card

1 Introduction

This SIM card storing information relating to the network that is used for authentication and user identification. The most important data is the number of identity card (ICCID Integrated Circuit Card ID), the number of international users (IMSI, International Mobile Subscriber Identity), a key authentication (Ki, Authentication Key), area code (LAI, Local Area Identity), and number emergency call operator. SIM card also store numbers for the SMS service center (SMSC, Short Message Service Center), service provider name (SPN Service Provider Name). When the SIM card is oriented as a smart card, it opens the possibility of security that resonate far beyond the world that is mobile [7]. SIM card containing electronic components as well as consist of various sizes as shown in Figure 1.

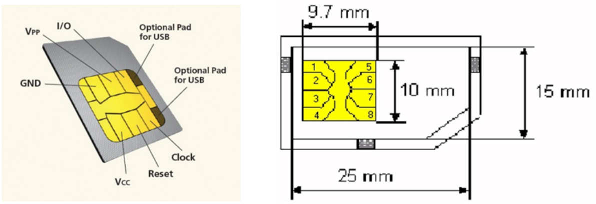


Fig. 1 Model SIM card [12]

(Ki, Authentication Key), a GSM SIM card using cryptography to reduce fraud against the confidentiality of the user. Before it is released to customers SIM card programmed in advance for the purposes to authentication.

Being required to read a special computer algorithm that runs internally on SIM card. KI copy is also stored by the network operator in the Authentication Center (AuC).

SIM card produced on the basis of algorithms COMP128v1, SIM card used now mostly still use in the development of algorithms COMP128v1. COMP128v1 algorithm contained in the coding system which consists of a SIM card GSM algorithms A3 and A8. A3 authentication algorithm is an algorithm in the GSM security model. A3 function is to generate response better known as the SRES in response to the random challenge known as the Random Number Generator (RAND) in other words SRES and RAND are algorithms that are on the network or provide on the network. while the A8 algorithm is an algorithm that serves to generate a session key, Kc or KI in SIM card, by looking at the random challenge, RAND received from the MSC and the secret key Ki contained in the SIM card [13].

COM128v1 has the major advantage that there are two systems of encoding or algorithms A3 and A8, A3 refers to the security of the network is being A8 refers to the security SIM card but on the other hand there is a shortage caused by the algorithm A8 which includes encryption security in the form of Authentication Key (KI) contained in SIM card.

SIM cards are available in various data capacities, from 8 kb to 128 kb however, from various SIM card memory capacity existing generation SIM card 32kb, 64kb and 128kb more in the market. Related SIM card memory usage associated with it play a role in determining the success of cloning SIM card it self more memory contained on the original SIM card then the longer the process of crack Ki A8 algorithm on SIM card.

Problems that arise from the above background related to the presence A8 algorithm embedded in every SIM card used by telecommunication users allowing copying, or cloning SIM card harm either side of the privacy and security of mobile telephone users. The goal of research cloning SIM card this is to give a warning to the security of users and provide dedicated SIM card to handling criminal investigations cloning SIM card along with its misuse of data.

SIM card include Subscriber Authentication Based on IMSI (Stored on SIM) and Random Number Generator/RAND (Provided by Network), it will be investigated further about SIM card cloning authentication by matching the customer's network login response to the mobile service network. Random Number Generator (RAND) contains an algorithm A3 (Provide by Network) so that in the process of cloning SIM card RAND participate in the process of matching algorithms contained on SIM card A8 to A3 algorithms contained on linked network authentication user data.

2 Literature Review

Related research studying the possibility of using the SIM card cloning subscriber identity module (SIM), Universal Mobile Telecommunications System. It also explores how the mobile system can find the SIM card cloning as soon as possible and how to reduce the possibility of using cloning a SIM card in the mobile network. Illegal mobile station is attached to a mobile network can be detected by the location in Area Update, update the location of the area periodically, and by calling out is removed from the original phone. Analytic model was developed to investigate the effects of location area updates and outgoing calls issued by the original phone on illegal cell phone use. Mobility management, such as registration, cancellation, and entry and exit procedures for legal and illegal users will be investigated and analyzed. Analytical models to determine the effect of the arrival of outgoing calls, and regional location of the residence time on detected illegal users have been presented. This study sought to improve the security of communication by avoiding deception of phone cloning by proposing solutions to accelerate the detection of SIM card cloning [8].

Mobile computing and mobile commerce are the most popular now days because of the services offered for mobility. Mobile computing has become a reality today than the wireless market. Mobile is rapidly increasing. Quality and speed available in the mobile environment must be in accordance with the fixed network if network convergence of fixed and mobile wireless communication occurs in the rea sense. The challenge for mobile network located within a very large footprint providing mobile service with high speed and security. Online transactions using mobile devices must ensure high security for user credentials and possibly to abuse. M-Commerce is the electronic trading is done by using a mobile device. Since a user's credentials to be kept secret, a high level of security must be ensured [9].

Exploration of digital evidence on SIM card scheme case of SIM card cloning in this case to find out more about the characteristics of the data and the digital evidence to the SIM card, imaging techniques, collecting and analyzing data, as well as exploration SIM card and investigative efforts SIM card in general [10].

Studies and Comparative Security GSM and CDMA GSM security system based on the exchange of data between the HLR (Home Location Register) with the SIM card in the MS (Mobile Station) RAND, MSC to the BTS and then MS. Ki and Kc is used to encrypt messages between base stations with MS. RAND, SRES Authentication in GSM is using A3 algorithm with a key Ki with the method Challenge and Response. Authentication using unique challenge procedure [4].

Analysis clone SIM card on IM3 smart and use Elliptic Curve Cryptosystem, this research cloning SIM card and cryptographic methods to analyze the combination of ECC (Elliptic Curve Cryptography) algorithms A3, A5 and A8 to get quality better security. It was found that the method can only be combined with the ECC algorithms A3 and A8 as well as the ECC method was not effective when combined with the algorithm A5 is due to differences between the two systems and procedures [5].

Forensic Software Tools for Cell Phone Subscriber Identity Module forensic specialists in making appropriate and inspection data. For the Global System for Mobile Communications (GSM), This paper gives an overview of the state of forensic software for SIM card. Forensic examination tool translating the data into a format and structure that can be understood by the examiner in identifying and recovering digital evidence with advantages and disadvantages [13].

Forensics and the GSM mobile telephone system Senior Investigator, this paper briefly describes the basics of the GSM system. The items of evidence that can be obtained from the Mobile Equipment, SIM and explored the core network to develop better forensic procedures. GSM SIM card conclusion that imitation is indeed possible for anyone who could. Forensic analysis methods are still physical contact with a mobile phone to access the stored information [13].

Validating Tools for Cell Phone Forensics This paper presents preliminary research in creating a basis for testing forensic tools and observed that some phones with information stored in the subscriber identity in the SIM card exactly in store logs on a T-Mobile SIM card standard SIM standard T-Mobile locked, it is still possible changes and modifications related to data protection [2].

Analysis SIM card Cloning With Algorithm Random Number Generator. This study discusses cloning SIM card along with stages ranging from crack SIM card cloning testing stage along with further analyze the effects of cloning media are bought and sold freely. Conclusions of research in the form of analysis SIM card after cloned and a series of early stage research and will be further examined for forensic investigation SIM card cloning research [1].

In this study focused on cloning SIM card forensic authentication algorithm using a random number generator (RAND). Forensic investigations against cloning and SIM card original with matching algorithms A3 and A8 to get Authentication Key (KI) in SIM card. Results are expected to be a description of the process of forensic investigation related to cloning SIM card authentication and authentication algorithm analysis Random Number Generator (RAND).

3 Research Methods

The subjects of this research is focused on authentication SIM card to stage cloning for further development of forensic investigation with SIM card cloning as evidence. The process of investigation will be conducted at the Laboratory of IT Centrum Islamic University of Indonesia, which is part of the Center for the Study of Digital Forensics. Related research methods to analyze forensic SIM card based on the theory that the focus of research in accordance with the facts related field evidence handling cloning SIM card for further analysis to prove that the hypotheses raised in accordance with the criteria. The final stage of the analysis of the study will be presented influence cloning SIM card, SIM card log cloning based clones with original SIM card log. Analysis of the research include:

- Attack and scenario testing;
- Design and cloning SIM card forensic proposal;
- Testing cloning SIM card;
- RAND Authentication;
- Forensic Analysis SIM card cloning.

3.1 Attack and Scenario Testing

Scenario testing is emphasized in the process of cloning of SIM card cloning it self include the success of the process generate authentication key (Ki) RAND A8 towards SIM card original and then tested the response A3 SRES to the network when SIM card cloning direct contact with SIM card original while attack scenarios attack inflicted post SIM card cloned in the form of attacks or duplication of such communication access short messages (SMS), call and access the data from SIM card cloning as if there is the same number as the original SIM card further here in after known effects [11]. Flowchart SIM card cloning attack shown in Figure 2.

3.2 Design and Cloning SIM Card Forensic Proposal

Based on forensic proposal SIM card cloning in Figure 3 above it can be concluded that the design refers to the RAND topic as the subject of research with the authentication key Ki as a sub topic of research.

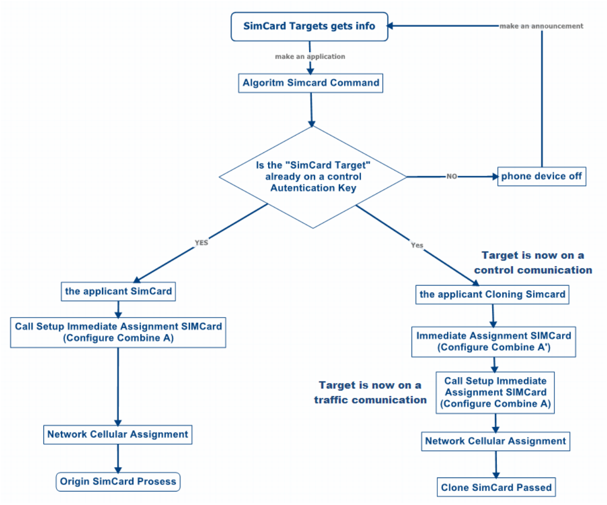

SimCard Targets gets info
make an announcement
make an application
Algoritm Simcard Command
Is the "SimCard Target" already on a control Autentication Key
NO
phone device off
YES
Yes
Target is now on a control comunication
the applicant SimCard
the applicant Cloning Simcard
Call Setup Immediate Assignment SIMCard (Configure Combine A)
Immediate Assignment SIMCard (Configure Combine A')
Target is now on a traffic comunication
Call Setup Immediate Assignment SIMCard (Configure Combine A)
Network Cellular Assignment
Network Cellular Assignment
Origin SimCard Prosess
Clone SimCard Passed

Fig. 2 Flowchart Attack SIM card Cloning

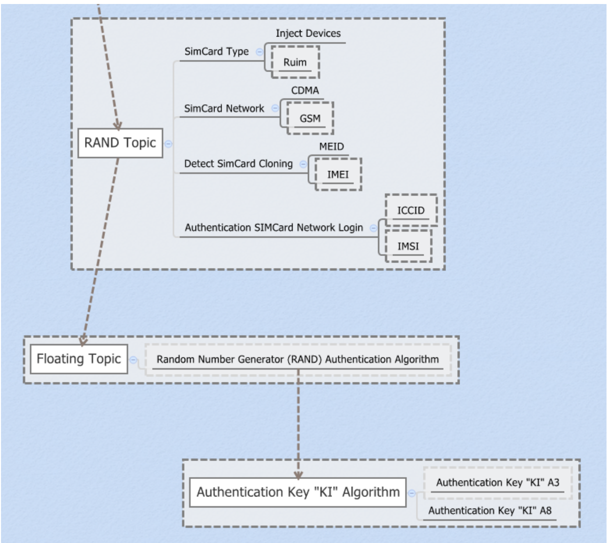

Inject Devices
SimCard Type
Ruim
CDMA
SimCard Network
GSM
RAND Topic
MEID
Detect SimCard Cloning
IMEI
ICCID
Authentication SIMCard Network Login
IMSI
Floating Topic
Random Number Generator (RAND) Authentication Algorithm
Authentication Key "KI" Algorithm
Authentication Key "KI" A3
Authentication Key "KI" A8

Fig. 3 Forensic Proposal SIM card Cloning

3.3 Testing Cloning SIM card

Authentication Cloning SIM card in scenario testing refers to authentication key (Referred to as Ki SIM card GSM), which consists of IMSI and ESN number, the test trial-and-error, giving input different SIM card and observe responses of both SIM card both include:

- Response to post providers network SIM card cloned;
- Accumulate between RAND and the authentication response SRES;
- The possibility of modifying the RAND algorithm or simply distribute authentication Ki to the media cloning;
- The effect of traffic from mobile users to the base station SIM card has been cloned to the original SIM card.

3.4 RAND Authentication

Another important aspect is the strength of the GSM authentication algorithm SIM card or often called A3. In principle, A3 owned by certain mobile operators, but the use of inter-operator algorithms tend to be the same. The statement further from the two algorithms in comparison RAND and SRES Authentication based on GSM security. If Ki can be extracted from the SIM, the user will be able to make a duplicate driver's license. Algorithms A3 and A8 determine the input (RAND and Ki) and output (SRES and Kc) of each algorithm [6].

3.5 Forensic Analysis SIM card Cloning

SIM card is a smart card, which contains a processor and non-volatile memory. In GSM, a SIM card which is used as a data storage device customers. The sole purpose of this procedure is to apply the mechanism of access and security features. The SIM card can be accessed by mounting the card the reader with the smart card reader whereas the standard required to access the software as a card reader or access SIM card. SIM card consisting of structures containing binary data file. Best forensic procedures will overview the entire contents is to download the entire SIM memory and compute the hash value memory is often called the acquisition of evidence, were to do this it takes forensic tools to access the file [3].

4 Results and Discussion

4.1 Analysis SIM Card Cloning

Ability SIM card cloner counter market is extremely diverse and has the advantage of each other, but from a variety of power applications has generated Ki cover or scan generated different. The percentage success rate of cloning SIM card consists of variables including mobile operators, generation, application cloner and SIM card memory capacity can be presented in Table. 1

Table 1: The success percentage cloning SIM card

No	Provider Name	Gen.1^{st}	Gen.2^{nd}	Gen.3^{th}
		Before 2011	2011 2014	2015 - Now
1.	Telkomsel	100%	100%	100%
2.	Indosat	100%	50%	0%
3.	XL	100%	25%	0%
4.	3	-	25%	-
Memory		32kb	64kb	128kb
Signal Coverage		1G	2G/3G	3G/4G

Inter-generation SIM card obtained from the sample between the users SIM card to register early for the network is divided into classifications include:

- 1st Generation = Before the year 2011;
- 2nd Generation = Between the years 2011-2014;
- 3rd Generation = After 2015 Now.

Based on the list of tables SIM card along with the providers and the application of cloning that accompanies it can be concluded while that object and focus SIM card cloning taken SIM card with a provider "Telkomsel" on the grounds of the generation SIM card most still use generation 1st and all 2nd which allows cloning SIM card with generated relatively short time compared to other providers, assuming the less memory is embedded in the SIM card allows the crack / authentication generated key that is relatively short. Powered by multiple applications SIM card cloner after test-generated Ki that SIM card with mobile operators "Telkomsel" assessed tend to be quicker to obtain the crack of authentication key Ki.

Here is a comparison between algorithms which emphasized the role of RAND and SRES algorithms, the algorithms are interconnections where A8 RAND in contact with A3 AUC SIM card Network so as to obtain the process flow as shown in Figure 4.

Formation of cloning SIM card to put the results of the algorithm generated Ki A8 of the original to be copied to the device subsequent cloning SIM card write with certain specifications that can serve the same cloning SIM card when making contact with the communications network. From the figure above it can be stated that Ki1, A31 and A81 are similar to variable SIM card original, then the cloning process is forwarded to the mobile station network or SRES Authentication act as authentication of customer data provider, if the data contained on the device SIM card (Ki and Random Number Generator) is considered matched with the central database will be given access to communications. Formation of cloning as shown in Figure. 5

Based on the formation and flow of the process of RAND and SRES can be concluded that the role of algorithms. RAND and SRES in this case is that the RAND as an algorithm of random formation of Ki authentication key and play a role in the authentication process of post SIM card performed the cloning was the role of algorithms SRES as challenge response to a

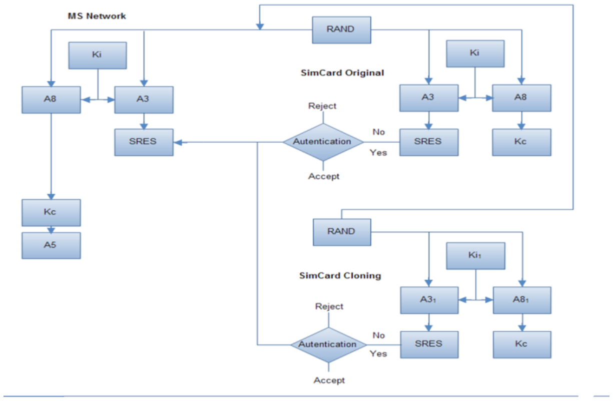


Fig. 4 RAND and SRES Cloning

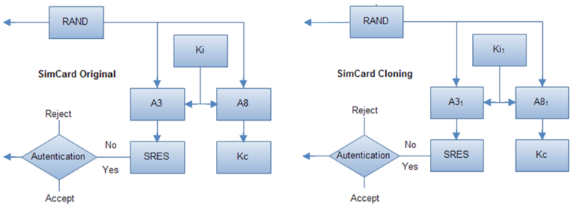


Fig. 5 Formation of Cloning

4.2 Forensic SIM Card Cloning

The next stage after the SIM card cloning is analyzed by algorithms auentikasi random number generator will be studied further forensic investigations related to cloning SIM card that will be on the exploration of the data contained in the findings of SIM card. Stages forensic SIM card cloning refers to the table network of contacts and response when there is more than one SIM card have a common authentication key Ki. SRES algorithm in this case in particular for the benefit of forensic investigations can not be explored by the analysis device SIM card because the algorithms contained in the communication service provider (provide by the network).

Investigative Process for Digital Forensic Science (DFRWS) technical report further adapted to the handling of evidence obtained by cloning SIM card table refer investigations to stage SIM card handling device as shown in Table. 2.

Based on the stage table above is obtained stages more emphasis on research related to the investigation of evidence SIM card that stage of examination will be conducted data acquisition, investigations, from the exploration of evidence with stages examination described above, can then be obtained data findings can combined until eventually the data findings goods evidence can be presented in table form table classification as a test case as shown in Table. 3:

Results of comparative investigation of data can be argued that the acquisition of evidence in the form of analysis and its SIM card can be obtained using forensic software and the conclusion is valid in accordance with the original SIM card which is owned by the victim.

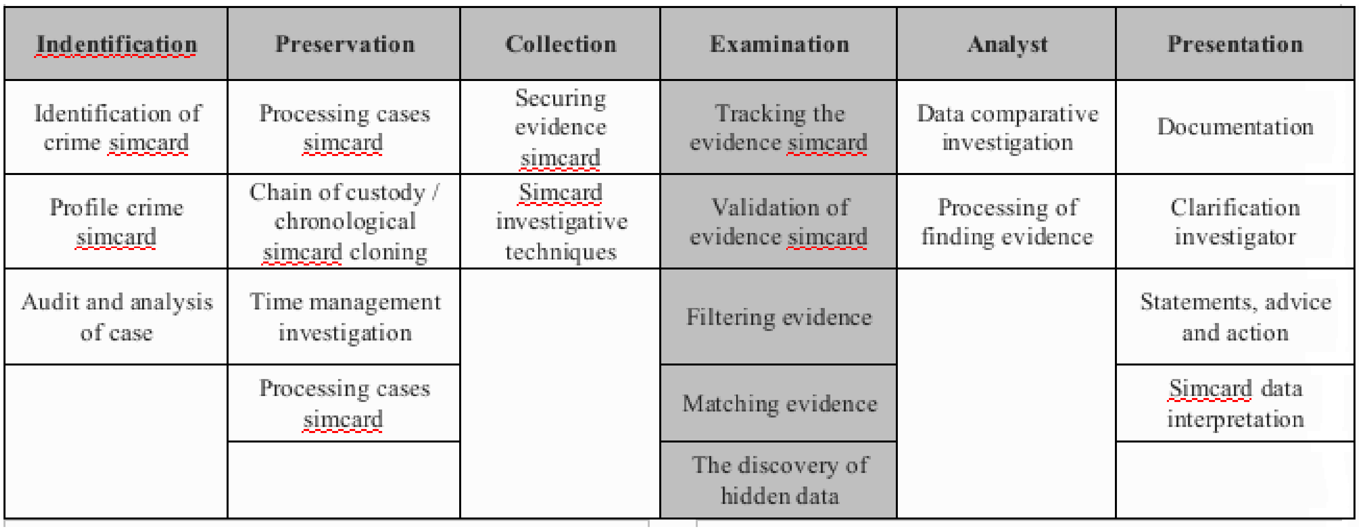
4.3 Discussion

The discussion on the SIM card cloning and analysis of variables that can affect the success rate of cloning and consists of

- Generation SIM card
  SIM card generational differences can affect the success rate in cloning SIM card actors, that the SIM card with 1st generation and 2nd generation was the one that lets done authentication crack or generated key.

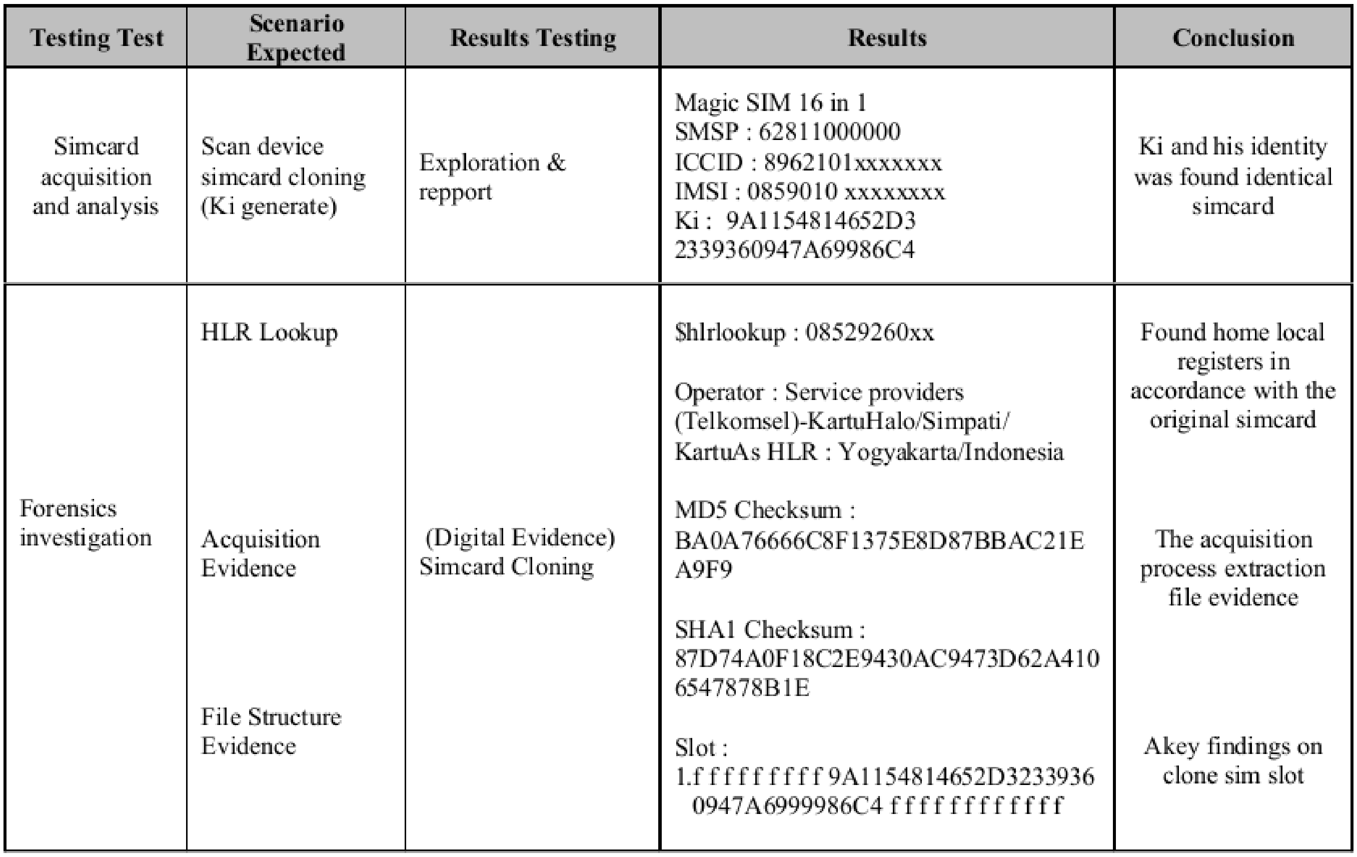
- Memory SIM card
  The memory is pinned by the provider of the service provider to each provider based generation also affects

Table 2: SIM card Investigation Process Cloning



Indentification	Preservation	Collection	Examination	Analyst	Presentation
Identification of crime simcard	Processing cases simcard	Securing evidence simcard	Tracking the evidence simcard	Data comparative investigation	Documentation
Profile crime simcard	Chain of custody / chronological simcard cloning	Simcard investigative techniques	Validation of evidence simcard	Processing of finding evidence	Clarification investigator
Audit and analysis of case	Time management investigation		Filtering evidence		Statements, advice and action
	Processing cases simcard		Matching evidence		Simcard data interpretation
			The discovery of hidden data		

Table. 3 Test Case Result (Forensics Tools)



Testing Test	Scenario Expected	Results Testing	Results	Conclusion
Simcard acquisition and analysis	Scan device simcard cloning (Ki generate)	Exploration & repport	Magic SIM 16 in 1 SMSP : 62811000000 ICCID : 8962101xxxxxxx IMSI : 0859010 xxxxxxxx Ki : 9A1154814652D3 2339360947A69986C4	Ki and his identity was found identical simcard
Forensics investigation	HLR Lookup	(Digital Evidence) Simcard Cloning	$hlrlookup : 08529260xx Operator : Service providers (Telkomsel)-KartuHalo/Simpati/ KartuAs HLR : Yogyakarta/Indonesia	Found home local registers in accordance with the original simcard
	Acquisition Evidence		MD5 Checksum : BA0A76666C8F1375E8D87BBAC21E A9F9 SHA1 Checksum : 87D74A0F18C2E9430AC9473D62A410 6547878B1E	The acquisition process extraction file evidence
	File Structure Evidence		Slot : 1.f f f f f f f f 9A1154814652D3233936 0947A6999986C4 f f f f f f f f f f f	Akey findings on clone sim slot

related literacy cloning application, the greater the SIM card memory ranging from 32kb, 64kb to 128kb latter will affect the reading process and Ki crack at the target SIM card cloning.

The next stage after that SIM card analysis of forensic investigation phase to the final destination in the form of a series of stages evidence handling cloning SIM card along with the findings contained or hidden in order to represent data. Based on the research include analysis of forensic SIM card cloning and cloning it can get the gist of related research SIM card. That the motives that made the perpetrator in committing of crimes targeting SIM card cloning is copying the data on devices SIM card in the form of results generated authentication key (Ki) henceforth be copied to media SIM card cloner that can be acquired or traded on the market freely and where actors can clone SIM card then can certainly add to the long list of motives of crime, especially mobile phone, while the results of forensic investigation SIM card cloning to explore the evidence can be found the file structure of a SIM card cloning containing partially identical data such as authentication key (Ki) obtained during generates a random number on the device SIM card along with data cloning victims. From the discussion above may be obtained several sub discussion include:

- Research Focus
  Forensics SIM card along with analysis is the focus of this study refers to the SIM card cloning scheme further testing scenario SIM card combine against cloning according to research methods. Based on the scenario and the process can be obtained SIM card cloning research focus as in Figure. 6

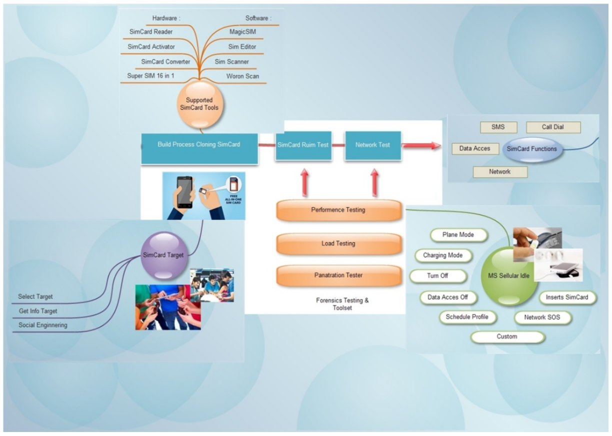


Fig 6: Focus SIM card Research Cloning

- Resume Research
  Based on the results and discussion of research related to the presence of SIM card and has been analyzed along with the working principles of cloning on SIM card subsequently conducted the investigation based on the stage of the investigation SIM card cloning which is emphasized in the process of examination of evidence SIM card to obtain detailed findings of digital evidence from the SIM card, then the results can be obtained and verification as in Table 4:

The next stage after that SIM card analysis of forensic investigation phase to the final destination in the form of a series of stages evidence handling cloning SIM card along with the findings contained or hidden in order to represent data.

The results of forensic investigation SIM card cloning to explore the evidence can be found of the file structure that contains a SIM card cloning partially identical data such as authentication key (Ki) obtained during a random number generated on the device along with the SIM card data cloning victims. The file structure attached to both SIM card original or cloning has similar characteristics but in the interests of the investigation needed further examination to be able to distinguish the characteristics of the file system of each SIM card. Classification of the file structure along with the sub file system can be seen in Figure. 7

The file structure that consists of a:

- Master File (MF),
- Directory File (DF) and
- Elementary File (EF),

Each component sub-system has file different capacity. In the interest of forensic investigations related evidence cloning SIM card will be differentiated based hierarchical file system for combined and match against the original SIM card. SIM card with each provider both 1st generation, 2nd and 3rd with different memory capacity when done scanning the file system will obtain the same file hierarchy. Comparison between the original SIM card and SIM card cloning discovered that differentiate file system that is contained in the file system:

- Elementary file (EF)

Table 4: Resume SIM card Research Cloning

Testing Test	Scenario Expected	Results Testing	Results	Conclusion
	SimCard Origin Telkomsel/AS	Synonymous with genuine SIM card (RAND, Ki)	Sim Number 085292608008 ICCID 89621019924260800080F IMSI 085901012924060880 Ki(A) Origin 9A1154814652D32339360947A69986C4 Ki(A') MagisSim 16 in 1 9A1154814652D32339360947A69986C4	Valid
SIM card Acquisition and Analysis	Scan device SIM card cloning (Ki generate)	Exploration & Repport (Digital Evidenc) SIM card Cloning	Magic SIM 16 in 1 SMSP: 62811000000 ICCID: 8962101xxxxxxxxx IMSI: 0859010 xxxxxxxxxx Ki: 9A1154814652D32339360947A69986C4	Valid
	HLR Lookup		$hlrlookup: 08529260xxxx Operator: Service providers (Telkomsel) - KartuHalo/Simpati/ KartuAs, HLR: Yogyakarta/Indonesia	Valid
	Acquisition Evidence			Valid
			MD5 Checksum: BA0A76666C8F1375-E8D87BBAC21EA9F9	Valid
	File Structure Evidence		SHA1 Checksum: 87D74A0F18C2E943 0AC9473D62A41065 47878B1E Slot: 1.ffffffff9A1154814 652D32339360947 A69986C4ffffffff	Valid
			2.Last Dial Number +6221500046	Valid
			3.Phone Book Copy 0163827648	Valid
			4.Pesan Text: Status: Read From: +6285105870607	Valid

– Elementary File (EF_Ki)

Elementary file (EF) with sub-system Elementary File (EF_Ki), with a storage slot containing EF_Ki Ki according to media cloning SIM card capability.

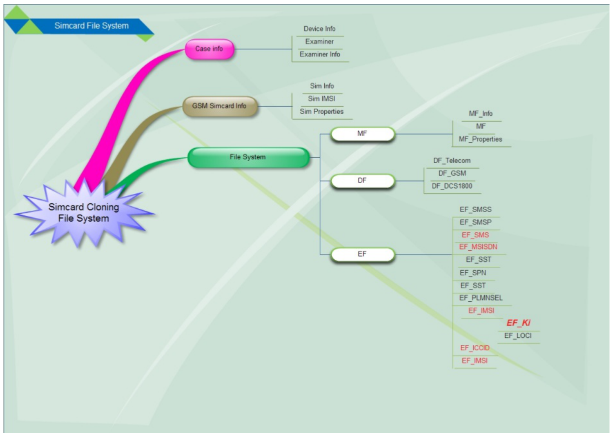


Fig. 7 File System SIM card

5 Conclusion

The process of matching algorithm or Random Generator / RAND against SIM card cloning in this research is to produce a RAND algorithm called A8 or SIM card cloning Ki to the media. Analysis SIM card cloning able to provide early warning for communications service, and can provide knowledge to the user (end user) in maintaining the security of mobile devices. SIM card generation in this case consists of generation 1st , 2nd , and 3rd respectively with memory capacity between 32kb, 64kb and 128kb. After Ki generated between generations conclude that the only SIM card with generation 1st, 2nd generation 32kb and 64kb able to perform the cloning and cloner applications that are sold freely in the latest. Newes generation SIM card with Ki can not be generated due to the limitation to read and scan Ki with larger memory , Random Number Generator (RAND) A8 and Sign Response (SRES) A3 participate in the process of forensic SIM card cloning. RAND to authenticate against Ki contained in SIM card while SRES act on network authentication network (mobile station). Both algorithms intertwined when it comes in contact SIM card to mobile networks. Forensic investigation SIM card cloning refers to the process table investigation published by the Digital Forensics Research Conference which emphasizes on the stage of the examination process. SIM card evidence of the exploration process of cloning can be concluded that matching or authentication algorithm A8-based Random Number Generator (RAND) is very helpful in the investigation related to the presence of cloning SIM card. The process of matching algorithm Random Number Generator (RAND) on cloning SIM card can be used as a step acquisition of evidence related to the SIM card cloning.

Contribution related forensic investigators SIM card cloning using a random number generator algorithm is to contribute to the analysis of SIM card cloning and its impact, further investigation and exploration forensic evidence SIM card cloning to be matched between a data SIM card cloning of the original SIM card.

6 Acknowledgments

This study was supported by the Laboratory of IT Centrum Indonesian Islamic University under Centre for the Study of Digital Forensics. The authors gratefully acknowledge the Reviewers for reviews their valuable comments.

References

[1] N. Anwar, I. Riadi, A. Luthfi, “Analysis SIM card Cloning With Algorithm Random Number Generator”, *Jurnal Buana Informatika*, vol. 7, no. 2, pp. 143–150, April 2016.

[2] N. Bhadsavle, J. A. Wang, “Validating Tools for Cell Phone Forensics”, in *Southeast Section Conference*, South Marietta Parkway: ASEE Southern Polytechnic State University, 2009.

[3] G. Palmer, “A Road Map for Digital Forensic Research”, Technical Report DTR-T001-01, Digital Forensic Research Workshop (DFRWS 2001), Nov. 2001.

[4] M. F. Fauzan, “Studies and Comparative Security GSM and CDMA”, Informatics Engineering Program, Bandung Institute of Technology, 2013.

[5] C. Hayat, “Analysis of the Clone In IM3 SIM card Smart And Use Ellptic Curve Cryptosystem To Improve Network Security GSM”, Depok, Indonesia: Department of Information Systems, University Gunadarma, 2004.

[6] M. Isomaki, "The relationship between GSM security paramenters and functions", Security in the Traditional Telecommunications Networks and in the Internet, Nov. 1999.

[7] W. Jansen, R. Ayers, "Forensic Software Tools For Cell Phone Subcriber Identity Modules", in *Procedings of the Conference on Digital Forensics, Security and Law*, pp. 93-106, 2006.

[8] M. A. Al-Fayoumi, F. Nidal, "Cloning SIM cards usability reduction in mobile networks", *Journal of Network and Systems Management*, vol. 22, no. 2, pp. 259–279, Apr. 2014.

[9] K. Prakash and Balachandra, "Security issues and challenges in mobile computing and M-commerce", *International Journal of Computer Science & Engineering Survey*, vol. 6, no. 2, pp. 29–45, Apr. 2015.

[10] Y. Prayudi, F. Rifandi, *Digital Evidence on SIMCard Exploration*, Digital Forensika Study Center, SESINDO FTI Islamic University of Indonesia, Dec. 2013.

[11] D. P. Tomcsanyi, "The big GSM write-up, how to capture, analyze and crack GSM", Oct. 13, 2013. (https://domonkos.tomcsanyi.net/?p=418)

[12] C. Velazco, *SIM Card Maker Gemalto Investigates Spy Agencies' Hack Attack*, May 2016. (http://www.engadget.com/2015/02/20/gemalto-investigates-spy-hacks/)

[13] S. M. Willassen, "Forensics and the GSM mobile telephone system", *International Journal of Digital Evidence Spring*, vol. 2, no. 1, pp. 1–7, 2003.

Nuril Anwar earned a BA from the Department of Informatics, University of Ahmad Dahlan (UAD) in 2012 and he is currently studying a Master of Computer Science with Digital Forensic interest of Megister Department of Information Engineering, Islamic University of Indonesia (UII). Email: anwar_nuril@yahoo.co.id.

Imam Riadi earned his Doctoral Program from the Department of Computer Science, University of Gadjah Mada (UGM) in 2014. Currently he is a lecturer at Ahmad Dahlan University (UAD) with interest in Network Engineering with concentration and interest in Internet Forensics. Email: imam.riadi@is.uad.ac.id.

Ahmad Luthfi He has got a Master of Computer Science from the Department of Computer Science, University of Gadjah Mada (UGM) in 2005. Today, He is staff and lecturer at Islamic University of Indonesia (UII) with concentration and interest in Mobile Forensics. Email: ahmad.luthfi@uii.ac.id.

Cattle Identification Using Segmentation-based Fractal Texture Analysis and Artificial Neural Networks

Ibrahim El-Henawy[1], Hazem. M. El Bakry[2], Hagar M. El Hadad[3]

(Corresponding author: Hazem. M. El Bakry)

Faculty of Computer and Information Sciences, Zagazig University Zagazig, Egypt[1]

Faculty of Computer and Information Science Department of Information System, Mansoura University[2]
Mansoura, Egypt

Faculty of Computer Science and Information Science, Department of Information System, Beni-Suef University[3]
Beni-Suef, Egypt

Email: helbakry5@yahoo.com

(Received Dec. 26, 2015; revised and accepted Feb. 5, 2016)

Abstract

Cattle Feature Extraction is the critical point in this paper which is considered as a continual research for authors. The biometric identifier of cattle is its muzzle. Today, Veterinarians search for new technologies to save cattle's livestock. This paper presents Artificial Neural Networks (ANNs) as the identification model. The proposed model contains the following three parts: pre-processing, Feature Extraction and Cattle Identifications. Pre-processing techniques are histogram equalization and mathematical morphology filtering. The proposed model compares between the following two feature extraction algorithms: Box-Counting Algorithm and Segmentation-based Fractal Texture Analysis (SFTA). Box-Counting Algorithm gives a feature vector of eight features and SFTA gives eighteen features for each cattle image. For achieving more accurate results in the identification part, ANNs have been used. This paper also uses the supervised learning technique in which the main factor is external teacher. The experimental results showed that SFTA Algorithm has achieved the best accuracy among all other identification techniques and our approach is superior than the existed work as our work achieves 99.97% identification accuracy.

Keywords: Artificial Neural Networks (ANNs), Box-counting, Histogram equalization, Image processing, Muzzle Identification, Segmentation-based Fractal Texture Analysis (SFTA)

1 Introduction

Nowadays, Animal Agriculture pays a great attention for saving animal products. The critical point which they faced is the rapid growth of livestock products. Veterinarians do their efforts in tracing each cattle in case of diseases infections. In this research we aim to build a robust model that helps the animal agriculture and veterinarians to discover the deceitful farmer because some farmer can easily remove the cattle ear tag after these cattle die and use it for other cattle so, our intelligent model can help in achieving Justice. So by this artificial model the end user can save each cattle record and connect it with its muzzle print to trace the non-healthy cattle. Therefore, the first important step is to identify each cattle in the flock. Cattle muzzle identification takes the great attention to monitor or observe the cattle disease outbreak, production management, vaccination management, cattle ownership assignment and traceability [30]. The traditional identification systems such as ear notching, muzzle ink printing, tattooing, Freeze branding and hot iron branding, ear tags, Neck Chains and Barcode, Electronic Identification and Radio Frequency Identification (RFID) [24], Nose printing, and blood test or hair sample (DNA). These traditional techniques are not enough to identify cattle in case of identification and identification due to fraudulent and repetition. So, veterinarians and animal agriculture tend to use more reliability and accurate systems to solve the defects of the traditional tracing systems.

In the case of humans, we use fingerprint as identifier because in mammals hair covers skins except some parts of the body. Cattle muzzle is formed by the distribution of ridges and valleys over it. Baranov et al. [18] discovered that the cattle muzzle pattern is hereditable and discovered the asymmetry between the two halves. In cattle, due to the uniqueness, the cattle muzzle print can consider as biometric identifier [6]. In general biometrics is the essential key to identify each individual or animal depending on their behavioral features [11, 14].

Muzzle Cattle Identification Model must have the following characteristics: Accuracy, acceptability, reliability, solve the fraudulent problem and uniquely identify each cattle [15]. Since Year 1921 till now, muzzle of cattle is considered as a unique identifier such as human fingerprint [13]. One of the traditional identification techniques is ink print based on paper model. It was the earliest technique used by animal agriculture. In order to use ink print model veterinarians first need to hold the cattle still then build up wetness on the animal noses, use too much ink, and waste time. From this point onwards the veterinarians start to search for more interactive identification and identification techniques that depend on some intelligent techniques. These techniques focus on using digital image processing in cattle muzzle identification and identification field depending on some factors such as working with large data base and the growth of the availability of using workstations and microcomputers with large capability in saving livestock. These factors led to a rapid increase in the image processing applications because they increase and improve capabilities of image equipment, display devices and reduce the cost of computation and image acquisition [5]. So, the critical step in this research is to collect a live cattle database to evaluate the cattle identification and identification model.

The difference between human user image observation and what really automatically extracted from any feature extracted algorithm is called a Semantic Gap Problem [9]. The important point in the feature extraction task is the set of features (feature vector) that represent visual content of cattle image. In many applications, feature vectors were used to solve the semantic gap problem [8]. The new research in texture feature extraction field is confirmed to increase the distinguishing ability of the feature extracted from the cattle muzzle images [16].

Box-Counting Algorithm is the first algorithm which the researchers use for the feature extraction part which is the second important part in the identification model for each muzzle image [10]. Box-counting is the selected algorithms that are used for fractal dimension extraction between several algorithms because many studies saw that box-counting algorithm is the common algorithm for fractal calculations [3]. Scientists such as Sarker and Chaudhuri improved this algorithm to Differential Box-Counting [2, 17, 22, 29]. In the first paper, the authors [10] after implementing box-counting algorithm to classify more than ten groups, the box-counting result was very weak. Therefore, the authors search for more accurate algorithm to use in part of the image feature extraction. New successful algorithm used for solving the weakness of box-counting algorithm that the researchers face is SFTA where this extraction algorithm feature depends on decomposing input cattle muzzle into a set of binary muzzle images from the fractal dimension of the regions to describe the segmented muzzle texture patterns [22].

ANNs have been used in many real word identification problems. The main applications on which ANNs were implemented are: Pattern recognition, face detections, bioinformatics, supervised and unsupervised learning, hand written recognition etc. In this paper we use ANNs in identification after using box-counting algorithm and SFTA to compare between these two techniques for feature extraction. The mechanism that the neural networks uses resemble that used in the human brain were it takes the structure of the biological neural system. After the network train on using supervised learning technique, it can be used to classify the test cattle images [25]. The proposed model in this paper compares between box-counting algorithm and SFTA to calculate the similarity between the input cattle training image and the tested image. The superiority of the proposed model is the confirmed muzzle cattle identification system validity provided by combining ANNs with box-counting features extraction algorithm and ANNs and SFTA for robust cattle muzzle matching. After implementing these techniques, we make a comparative study between these feature extraction algorithms.

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

2 Preliminaries

2.1 Histogram Equalization Algorithm

Histogram Equalization Algorithm (HEQ) is used to re-distribute the gray levels over muzzle image to obtain a Regular Histogram. After applying the histogram equalization algorithm, each pixel in the original cattle muzzle image is replaced by the integral of the histogram of image in that pixel [26]. To adjust the image contrast HEQ algorithm each cattle uses an image's histogram. This adjustment makes the best distribution of the intensity on the image histogram. It allows that the less contrast area to be more contrasted because it spreads out the most intensity values that were frequently distributed [20]. HEQ is explained in details in Algorithym 1.

Algorithm 1 Histogram Equalization algorithm (HEQ)

```
1: Consider the cattle muzzle image gray levels full in the range $[0, M-1]$;
2: Calculate the Probability Distribution Function (PDF) of each muzzle using the following Equation (1):
       $$PDF(r_h) = \frac{n_h}{N}, \quad H = 0, 1, \cdots, M-1 \tag{1}$$
   Where $n_h$ is the number of pixels in the image having gray level $r_h$ and $r_h$ the Hth is gray level.
3: Calculate the Cumulative Distribution Function (CDF) for each cattle image using Equation (2):
       $$CDF(n_h) = \sum_{i=0}^{h} P(r_i), \quad h = 0, 1, \cdots, M-1, 0 \leq C(n_h) \leq 1. \tag{2}$$
4: Calculate Histogram Equalization (HE) gray level to gray level for each input cattle muzzle image according to Equation (3):
       $$S_h = (M-1)CDF(r_h). \tag{3}$$
5: The changes in gray level $S_h$ can be calculated using histogram equalization algorithm according to Equation (3).
6: return Gray Level Value for each cattle muzzle.
```

2.2 Mathematical Morphology Filtering Algorithm

Mathematical Morphology Filtering (MM) [7] depends on the geometric shape feature. The main operations that MM depend on are erosion, dilation, opening and closing. In this research the (MM) Algorithm is very suitable to help in removing cattle image noise where we first implement an open operation followed by a close operation. 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 [23].

2.3 Box-Counting Algorithm

The first texture feature extraction algorithm which the authors use is the Box-Counting Algorithm. Also, there is more than one algorithm to calculate the texture fractal dimension for each cattle muzzle image. 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 (See Algorithm 2).

Algorithm 2 Box-counting algorithm Db to any subset A in (Euclidean space)

```
1: To calculate Db(A) set the value of $N, (A)$ to the smallest number of $r$ set that cover cattle muzzle area as in Equation (4):
       $$Db(A) = \lim_{r \to 0} \frac{\log(N_r(A))}{\log(1/r)} \tag{4}$$
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 in Equation (5):
       $$Db(A) = \lim_{r \to 0} \frac{\log(N'_r(A))}{\log(1/r)} \tag{5}$$
4: Box counting $N_r(A)$ and $Db(A)$ are related by relation power law shown in Equation (6):
       $$N_r(A) = \frac{1}{r^{D_b}(A)} \tag{6}$$
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 $N_r(A)$.
```

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 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] (See Algorithm 3).

Algorithm 3 Segmentation Based Fractal Texture Analysis algorithm (SFTA)

```
Covert cattle muzzle image from RGB to Gray scale $I$, where $I$ is cattle Grayscale image;
Set number of threshold $n_t$;
Assign Multi Level $Otsu(I, n_t)$ function to variable $T$;
Set $T_A == \{\{t_i, t_i + 1\} : t_i, t_i + 1 \in T, i \in [1 \cdots |T| - 1]\}$;
Set $T_B == \{\{t_i, nl\} : t_i \in T, i \in [1 \cdots |||]]\}$, where $nl$ denote gray level range and $T$ is the set of threshold values;
Set $i == 0$;
For $\{\{t_l, t_u\} : \{t_l, t_u\} \in TA \cup T_B\}$ do, where $t_l$, $t_u$ denote lower threshold, upper threshold respectively;
$VSFTA[i] == BoxCounting(\triangle)$, where $\triangle$ border of cattle muzzle image;
$VSFTA[i + 1] == MeanGrayLevel(I, I_b)$;
$VSFTA[i + 2] == PixelCount(I_b)$ where $I_b$ is cattle binary image;
$VSFTA[i + 2] == PixelCount(I_b)$ where $I_b$ is cattle binary image;
End For
Return $VSFTA$, where $VSFTA$ denotes the extracted SFTA feature vectors.
```

2.5 ANNs Algorithm

The third challenge that the veterinarians and animal agricultures face is the identification part. In order to classify each muzzle image, use the texture feature vector of image that results from box-counting or SFTA. The Neural Network depends on comparing the tested muzzle vector with all the training vectors. Depending on the rule of ANN, weights of the network are adapted during training phase according to the relation between error function and weights $(\vartheta_E/\vartheta_w)$ must be smaller than the given threshold.

3 Proposed Cattle Muzzle Identification Model

In this paper, 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 to increase image contrast and mathematical morphology filtering to remove noise form image. 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 ANN 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 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.

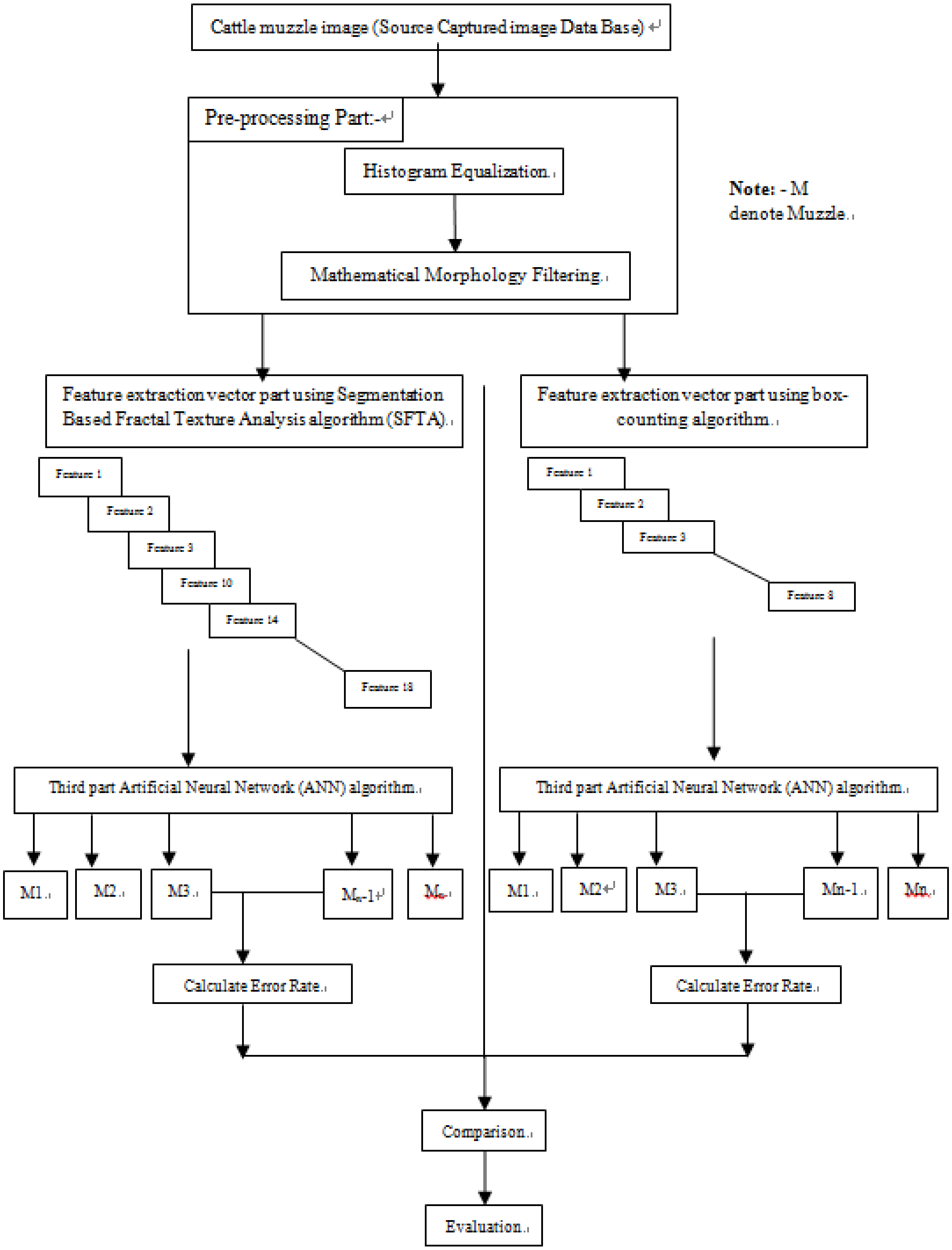

Cattle muzzle image (Source Captured image Data Base)
Pre-processing Part:-
Histogram Equalization
Mathematical Morphology Filtering
Note: - M denote Muzzle
Feature extraction vector part using Segmentation Based Fractal Texture Analysis algorithm (SFTA)
Feature extraction vector part using box-counting algorithm
Feature 1
Feature 2
Feature 3
Feature 10
Feature 14
Feature 18
Feature 1
Feature 2
Feature 3
Feature 8
Third part Artificial Neural Network (ANN) algorithm
Third part Artificial Neural Network (ANN) algorithm
M1
M2
M3
Mn-1
Mn
M1
M2
M3
Mn-1
Mn
Calculate Error Rate
Calculate Error Rate
Comparison
Evaluation

Figure 1: The proposed cattle muzzle image processing part

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 $SIZE(C) \leq 2^P$. 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 Artificial Neural Network (ANN)

We use the ANN technique in the identification step because it solves the complexity problems. The ANN adapts itself by sequential training algorithm and by its architecture and connected weights. This paper uses feed forward and gives more accurate results. This paper also uses the supervised learning technique in which the main factor in supervised learning technique is external teacher i.e. (the ability of using target vector). In this paper we start with using three classes that mean's three different cattle and the maximum number of classes is thirty classes. The neural network uses its weight to adjust error signal and train vectors. The procedure that we follow when implementing neural network is:

1) Use the input feature vector data set as the neural network input layer. In this research we use SFTA. The accuracy result for neural network depends on the number of training cases in the learning step. The target in our research is to use the ANN to identify and differentiate between cattle if we have 3, 5, 10, 14, 20, 25 and 30 different cattle.

2) Create a network. We use a pattern recognition network, which is a feed-forward network with tan-sigmoid transfer functions in both the hidden layer and the output layer. As in the function-fitting example, use 20 neurons in one hidden layer.

The network has three, five, and ten, etc. output neurons in case of using three, five, ten, etc. different cattle, because there is the muzzle categories associated with each input vector. Each output neuron represents one cattle category. When an input vector of the appropriate cattle muzzle category is applied to the network, the corresponding neuron should produce a 1, and the other neurons should output a 0.

1) Train the network: in this research the pattern recognition network uses the Scaled Conjugate Gradient algorithm for training the cattle images network. The neural network application divides the input vectors and target vectors (in the training case) into three sets; 60% are used for training, 20% which are used in the validation step that the network is generalizing and it is also used to stop the neural network training before over fitting and the last 20% are used as a completely independent test of the bovine's images neural network generalization.

2) Test the network: this is the step in which we used input vector without target and test the neural network.

4 Experimental Results

The experimental results in this research were conducted using Intel Core^{TM} Duo CPU laptop with 3 GB of RAM and running at 1.86 GHz. Matlab R2009b is the Authorized Version which the experiments implement.

4.1 Cattle Muzzle Print Database

The first challenge in preparing this research was the lack of the 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 twenty muzzles. 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.

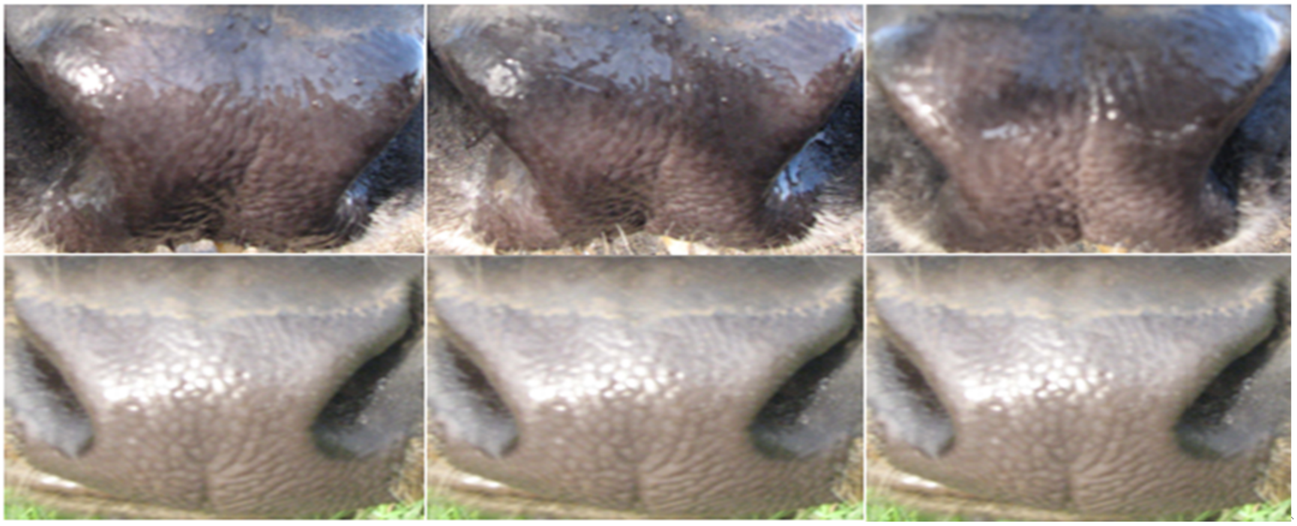


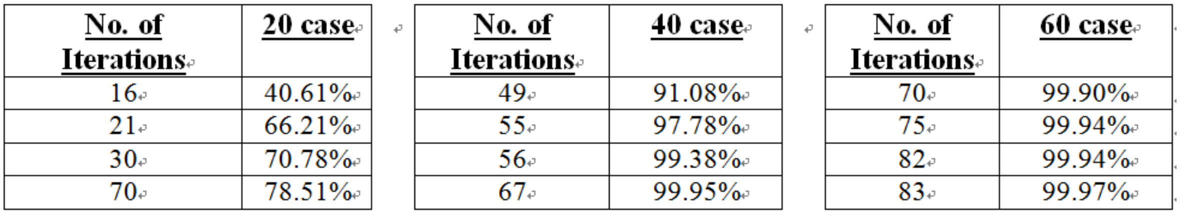
Figure 2: A sample of different cattle printed images. This figure represents print images for cattle muzzle that have taken from two different cattle.

The identification scenarios: 3, 5, 10, 14, 20, 25 and 30 groups of cattle muzzle each group with 20, 40, and 60 different muzzle images used in the training phase to calculate the accuracy of implementing the ANN identification mode. The use of ANN 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.

4.2 Evaluated Results

First: Classify cattle according to the following three groups using the feature vector extracted from: 1. Box-Counting Algorithm; 2.Segmentation based Fractal Texture Analysis algorithm (SFTA).

Table 1: Accuracy rate in case of three groups of muzzle, each group has 20, 40 and 60 cases.

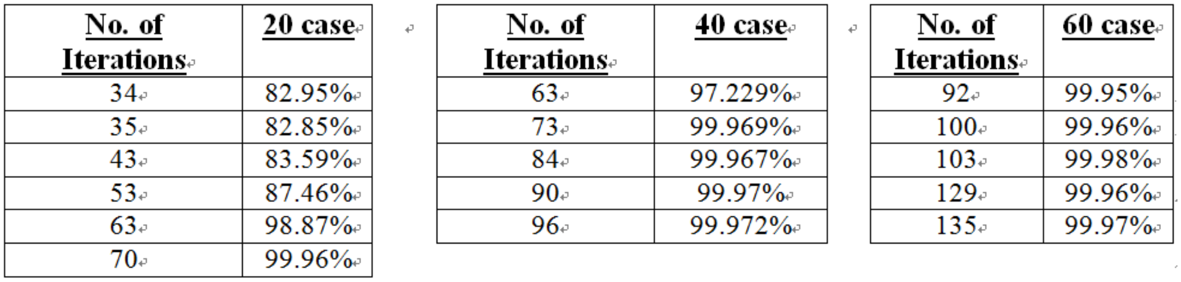


No. of Iterations	20 case	No. of Iterations	40 case	No. of Iterations	60 case
16	40.61%	49	91.08%	70	99.90%
21	66.21%	55	97.78%	75	99.94%
30	70.78%	56	99.38%	82	99.94%
70	78.51%	67	99.95%	83	99.97%

In Table 1 we made a comparison between three cases to identify only three different cattle but for each cattle in the training step we use 20, 40 or 60 different images for each cattle muzzle. As noted when we learn the ANN in the training step with only 20 different muzzles image for each cattle the accuracy rate increase to 78.51%, when we increase number of cases for each cattle to 40 and 60 the accuracy rate increased to 99.95%. In Table 2 we use the SFTA feature vector as you note the accuracy rate is 87.46%, so this percentage we can got it in the case when number of muzzles image is 20. And the accuracy rate increases to 99.96% in case of using 40 and 60 different cases for each cattle. This concept is also for all the data in the followed tables.

Second: Classify cattle in five groups using the feature vector extracted from: 1. Box-Counting Algorithm (See Table 3); 2. Segmentation based Fractal Texture Analysis Algorithm (SFTA) (See Table 4).

Table 2: Accuracy Rate in case of three groups of muzzle, each group has 20, 40 and 60 cases.

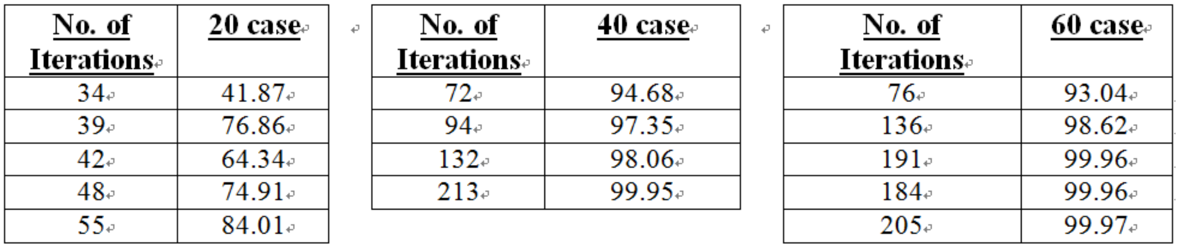


No. of Iterations	20 case
34	82.95%
35	82.85%
43	83.59%
53	87.46%
63	98.87%
70	99.96%

No. of Iterations	40 case
63	97.229%
73	99.969%
84	99.967%
90	99.97%
96	99.972%

No. of Iterations	60 case
92	99.95%
100	99.96%
103	99.98%
129	99.96%
135	99.97%

Table 3: Accuracy Rate in case of five groups of Muzzle, each group has 20, 40 and 60 cases.

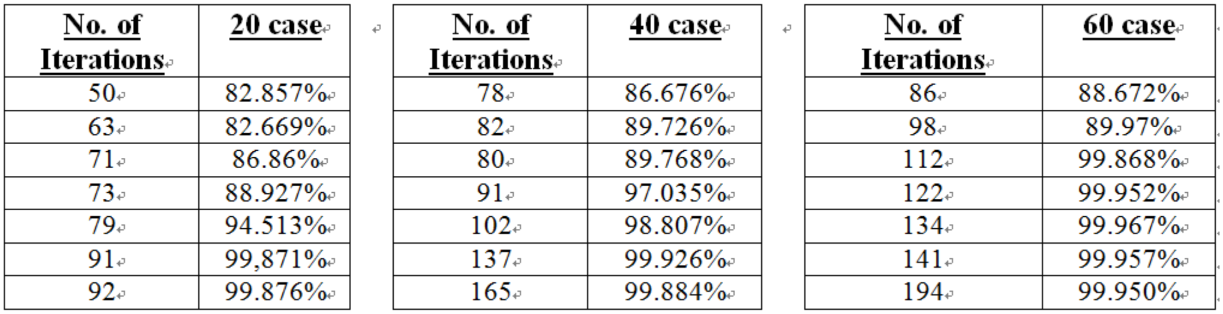


No. of Iterations	20 case
34	41.87
39	76.86
42	64.34
48	74.91
55	84.01

No. of Iterations	40 case
72	94.68
94	97.35
132	98.06
213	99.95

No. of Iterations	60 case
76	93.04
136	98.62
191	99.96
184	99.96
205	99.97

Table 4: Accuracy Rate in case of five groups of Muzzle, each group has 20, 40 and 60 cases.



No. of Iterations	20 case
50	82.857%
63	82.669%
71	86.86%
73	88.927%
79	94.513%
91	99,871%
92	99.876%

No. of Iterations	40 case
78	86.676%
82	89.726%
80	89.768%
91	97.035%
102	98.807%
137	99.926%
165	99.884%

No. of Iterations	60 case
86	88.672%
98	89.97%
112	99.868%
122	99.952%
134	99.967%
141	99.957%
194	99.950%

Third: Classify cattle in ten groups using the feature vector extracted from: 1. Box-Counting Algorithm (See Table 5); 2. Segmentation based Fractal Texture Analysis Algorithm (SFTA) (See Table 6).

Table 5: Accuracy Rate in case of ten groups of Muzzle, each group has 60 cases.

No. of Iterations	60 Cases
95	54.53
121	74.12
123	77.86
146	78.86

Table 6: Accuracy Rate in case of ten groups of Muzzle, each group has 60 cases.

No. of Iterations	60 Cases
125	95.398
129	97.152
156	98.32
161	98.7
252	98.83

Box-counting algorithms accuracy decrease in case of increase number of classified group. In case of using box-counting algorithm for classify more than ten groups the accuracy rate became 0%. But Segmentation based Fractal Texture Analysis algorithm (SFTA) still work till twenty groups and twenty-five. As shown in the following.

Fourth: Classify cattle in five groups using the feature vector extracted from: 1. Segmentation based Fractal Texture Analysis Algorithm (SFTA) (See Figure 3).

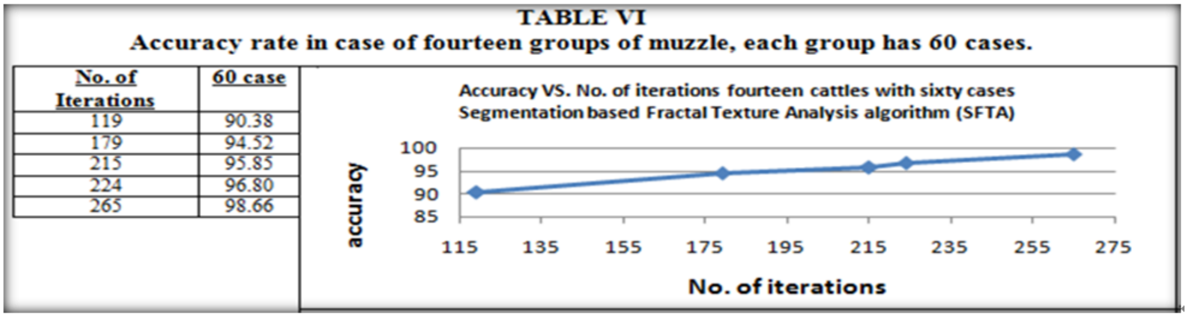


TABLE VI

Accuracy rate in case of fourteen groups of muzzle, each group has 60 cases.

No. of Iterations	60 case
119	90.38
179	94.52
215	95.85
224	96.80
265	98.66

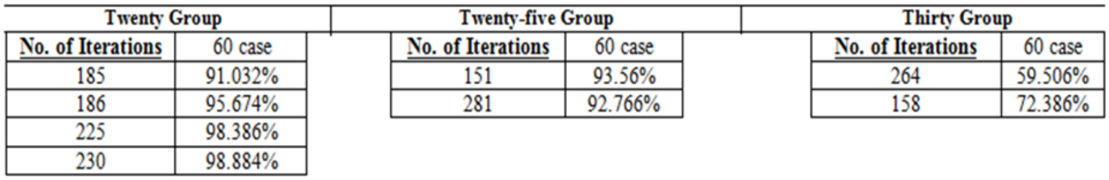
Figure 3: Classify cattle in five groups using the feature vector extracted from Segmentation based Fractal Texture Analysis Algorithm (SFTA)

Fifth: Classify cattle in twenty, twenty-five and thirty groups using the feature vector extracted from: 2. Segmentation based Fractal Texture Analysis Algorithm (SFTA) (See Table 7.

5 Conclusions and Future Work

This paper has presented two models for Muzzle identification using printed cattle muzzle images as input. The First Model has used box-counting algorithm for texture feature extraction for each cattle images; then used ANNs for the identification and matching feature for the tested pattern with the training muzzles. The Second Model has employed SFTA for cattle muzzle feature extraction for each cattle; then used these feature vectors in the training phase for ANNs. The Box-Counting algorithm has given eight features for each muzzle image which resulted in

Table 7: Classify cattle in twenty, twenty-five and thirty groups using the feature vector extracted from Segmentation based Fractal Texture Analysis Algorithm (SFTA).



Twenty Group		Twenty-five Group		Thirty Group	
No. of Iterations	60 case	No. of Iterations	60 case	No. of Iterations	60 case
185	91.032%	151	93.56%	264	59.506%
186	95.674%	281	92.766%	158	72.386%
225	98.386%				
230	98.884%				

the weakness of neural network when the number of the classified group increased to ten groups. SFTA algorithm has given feature vector with length eighteen features for each cattle muzzle image. In case of using the numbers of identification groups of three, five, ten, fourteen, twenty, twenty-five and thirty; then the box-counting accuracy results were: 99.97%, 99.97%, 78.86%, 0%, 0%, 0% and 0% respectively. In case of using SFTA algorithm, the accuracy results were: 99.97%, 99.950%, 98.83%, 98.66%, 98.88%, 92% and 59.506% respectively. Firstly, the accuracy of our proposed model to identify cattle animals using muzzle print images has achieved excellent results comparing to all previous models in [19, 27, 28]. Secondly, the experimental results have shown that the SFTA algorithm is a more accurate algorithm used for classifying such cattle muzzle image database. Therefore, it's recommended to increase the number of features in feature vector to increase the accuracy rate. In the future work, authors tend to use algorithm that give more feature in feature vector to increase identification accuracy in case of using large number of group. We shall search for minimizing the running time because it grows when the number of groups used in the identification became large.

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, no. 2012, pp. 1984–1992, Nov. 2011.

[3] A. G. R. Balan, A. G. M. Traina, C. T. 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*, pp. 581–586, 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," in *Proceedings of 18th IEEE Symposium on Computer-Based Medical Systems*, pp. 581–586, June 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, Dec. 6th, 2005.

[7] S. Beucher, "Segmentation d'images et morphologie mathematique," Ph.D. Thesis, Ecole des Mines de Paris, Juin, 1990.

[8] A. 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 2012)*, pp. 39–46, 2012.

[9] 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.

[10] 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, USA, May 2014.

[11] 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.

[12] F. Gnther, S. Fritsch, "Neuralnet: Training of neural networks," *The R Journal*, vol. 2/1, June 2010.

[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, A. and Hashiguchi, “Multifractal feature based cancer detection for pathological images,” in *IEEE 5th International Conference on Bioinformatics and Biomedical Engineering, (iCBBE 2011)*, pp. 1–4, May 2011.

[18] A. Noviyanto, A. M. Arymurthy, “Beef cattle identification based on muzzle pattern using a matching refinement technique in the SIFT method,” *Computers and Electronics in Agriculture*, vol. 99, no. C, pp. 77–84, 2013.

[19] A. Noviyanto and A. M. Arymurthy, “Beef cattle identification based on muzzle pattern using a matching refinement technique in the SIFT method,” *Computers and Electronics in Agriculture*, vol. 99, pp. 77–84, Nov. 2013.

[20] N. D. Ponraj, E. M. Jenifer, P. Poongodi, and ManoharanS, “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, 2011.

[21] Z. Rahman, M. Shameemmhossain, and K. Ahmed, “Flag identification using support vector machine”, *Journal of Information Technology*, vol. 2, pp. 11–16, 2013.

[22] N. Sarker, B. B. Chauduri, “An efficient box-counting approach to compute fractal dimension of image,” *IEEE Transactions on System Man Cybernet*, vol. 24, pp. 115–120, 1994.

[23] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, London, 1982.

[24] C. Sun, F. Jiang, and SH. Jiang, “Research on RFID applications in construction industry,” *Journal of Networks*, vol. 8, no. 5, pp. 1221–1228, May 2013.

[25] L. H. Thai, T. S. Hai, N. T. Thuy, “Image classification using support vector machine and artificial neural network,” *Information Technology and Computer Science*, vol. 4, no. 5, pp. 32–38, 2012.

[26] K. Thangavel and 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, 2009.

[27] 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

[28] 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.

[29] 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.

[30] 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 2012)*, 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 2757. 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 in 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).

Window Method Based Cubic Spline Curve Public Key Cryptography

Addepalli V. N. Krishna[1], Addepalli Hari Narayana[2], K. Madhura Vani[3]
(Corresponding author: Addepalli V. N. Krishna)

The Department of Computer Science & Engineering, Faculty of Engineering, Christ University[1]
Bangalore, Karnataka - 560 029, India
The Department of Electrical Engineering, Indian Institute of Technology[2]
Indore, Madhya Pradesh- 453 331, India
The Department of Computer Science & Engineering, Shreyas Institute of Engineering & Technology[3]
Hyderabad, Telangana - 500 068, India
Email: hari_avn@rediffmail.com

(Received Jan. 16, 2015; revised and accepted Aug. 28, 2015)

Abstract

In this work, a cubic spline curve is considered for Asymmetric mode encrypting data. A steady state, one dimensional equation is integrated over a control volume. The derivatives of above equation form piece wise linear profile, which leads to discretization equation with corresponding weights. These weights are initially solved to generate global variables. Those global variables are used to calculate public key which in turn use the ElGamal mode of encryption. The proposed algorithm supports the features like Authenticity of users, Security & Confidentiality of data transmitted. Going by the construction of the algorithm, Encryption is being done on blocks of data for which it consumes less computing resources. Going by complexity of the algorithm, the key length needed is about 120 bit length to provide sufficient strengths against cryptanalysis.

Keywords: Cubic Spline Curve, Public Key Cryptography, Window Method

1 Introduction

Any symmetric encryption scheme uses a private key for secure data transfer [12]. In their work on a new Mathematical model on encryption scheme for secure data transfer [6], the authors considered not only key but also time stamp and nonce values to increase the strength of sub key generated. In addition the nonce value can also be used for acknowledgement support between participating parties. The model can be further improved by considering a non linear model where the key values vary with the data generated [7].

Most of the products and standards that use public-key cryptography for encryption and digital signatures use RSA today [8]. Recently, Elliptic Curve Cryptography has begun to challenge RSA. The principal attraction of ECC, compared to RSA, is that it appears to offer better security for a smaller key size, thereby reducing processing overhead.

Some recent works on application of ECC are cited here. [4] explains the engineering of ECC as a complex interdisciplinary research field encompassing such fields as mathematics, computer science and electrical engineering. [1] presents a high performance EC cryptographic process for general curves over GF(p). [5] specifies the standard specifications for public key cryptography. Encryption to data supports the very important features like security, Confidentiality to data & Authentication of users. [10, 13] discussed the features of Numerical data analysis which helps in building a mathematical model. [14] discusses computing complexity of ECC and [9, 11] identifies additional features with ECC which makes it more secure.

The work is divided into following modules: 1) Converting Mathematical models to Cryptosystems; 2) Generating Public and Private Keys; 3) Encryption and Decryption Process; 4) Cryptanalysis and Complexity of the Model; 5) Conclusion.

2 Converting Mathematical Model to Cryptosystems

2.1 Introduction To Cubic Spline Curve

The basic idea behind cubic spline is to draw smooth curves through number of points [3]. The spline consists of weights attached to a flat surface at the points to be connected. A flexible strip is then bent across each of these weights resulting in a pleasant smooth curve. These weights are the coefficients on cubic polynomials used to interpolate the data. The essential idea is to fit a piece wise function of the form

$$S(x) \quad = \quad \begin{cases} S_1(x) & x_1 \le x \le x_2 \\ S_2(x) & x_2 \le x \le x_3 \\ \vdots & \\ S_{n-1}(x) & x_{n-1} \le x \le x_n \end{cases}$$

where S_i is a third degree polynomial defined by

$$S_i(x) = a_i(x-x_i)^3 + b_i(x-x_i)^2 + c_i(x-x_i) + d_i \;\; \text{for} \;\; i = 1, 2, 3, \cdots, n-1.$$

The first & second derivations of degree $n-1$ equations are fundamental to this process. They are

$$\begin{aligned} S_i^{'}(x) &= 3a_i(x-x_i)^2 + 2b_i(x-x_i) + c_i \\ S''{}_i(x) &= 6a_i(x-x_i)^2 + 2b_i, \;\; \text{for} \;\; i = 1, 2, 3, \cdots, n-1. \end{aligned}$$

On summing the equation based on the fact that $S'(x)$ & $S''(x)$ are continuous across interval, the coefficients for $(n-1)$ equations can be calculated from $a_i\, b_i\, c_i\, d_i$ and $S_i''(x_1) = M_i$ which can be represented in matrix form as follows. We can consider three cases,

1) Natural splines, $M_1 = M_n = 0$;

2) Parabolic runout spline:

$$\begin{aligned} M_1 &= M_2; \\ M_{n-1} &= M_n. \end{aligned}$$

3) Cubic runout spline:

$$\begin{aligned} M_1 &= 2M_2 - M_3; \\ M_n &= 2M_{n-1} - M_{n-2}. \end{aligned}$$

2.2 Numerical Models of Steady State Equations

Steady state one-dimensional equation is given by $\frac{\partial}{\partial x}\left(K.\frac{\partial T}{\partial x}\right) + s = 0.0$, where $K\&s$ are constants. To derive the discretization equation we shall employ the grid point cluster. We focus attention on grid point P which has grid points E, W as neighbors. For one dimensional problem under consideration we shall assume a unit thickness in y and z directions. Thus the volume of control volume is delx*1*1. Thus if we integrate the above equation over the control volume, we get $\left(K.\frac{\partial T}{\partial X}\right)_e \left(K\frac{\partial T}{\partial X}\right)_w + \int S\,\partial X = 0.0$. If we evaluate the derivatives $\partial T/\partial X$ in the above equation from piece wise line ar profile , the resulting equation will be

$$K\,e\,\frac{(T_e - T_p)}{(\partial X)_e} - K\,W\frac{(T_p - T_W)}{(\partial X)_e} + S^{*}\mathrm{del}\,x = 0.0$$

where S is average value of s over control volume.

2.2.1 Linear Data Flow Problem

Dividing the problem area into M number of points, and for simplicity by assuming data of the 1^{st} and M^{th} grid points are considered to be known and constant. For the grid points $2, M-1$, the coefficients can be represented by considering the conservation equation,

$$\frac{\alpha}{\partial x}(T_{I+1}^{n+1} - T_I^{n+1}) + \frac{\alpha}{\partial x}(T_I^{n+1} - T_{I-1}^{n+1}) = \frac{\partial x}{\partial t}(T_I^{n+1} - T_I^{n}) \tag{1}$$

where T_i represents data value for the considered grid point for the preceding delt, T_{I+1}^{n+1} & T_I^{n+1} represents data values for the preceding and succeeding grid points for the current delt.

Considering α for the given model, the coefficients are obtained for each state (grid point) in terms of $A(I)$ refers to data value of the corresponding grid point, $C(I)$ and $B(I)$ refers to data values of preceding and succeeding grid points for the current delt, $D(I)$ refers to data value of the considered grid point in the preceding delt.

$$\begin{aligned} A(I) &= 1+2\alpha\frac{\mathrm{del\,t}}{(\mathrm{del\,x})^2} \\ B(I) &= -\alpha\frac{\mathrm{del\,t}}{(\mathrm{del\,x})^2} \\ C(I) &= -\alpha\frac{\mathrm{del\,t}}{(\mathrm{del\,x})^2} \\ D(I) &= T_I^n, \end{aligned}$$

where α is a constant value, the model generated is a linear model. For $I = 1, 2, 3, \cdots, n_i$. Thus the data value T is related to neighboring data values T_{i+1} and T_{i-1}. For the given problem $C_1 = 0$ and $B_n = 0$. The coefficients are arranged in matrix form. As both sides of the curve are maintained at same data points, the coefficients are represented as

$$\begin{bmatrix} a_2 & b_2 & & & & \\ & a_3 & b_3 & c_3 & & \\ & & a_4 & b_4 & c_4 & \\ & & & \vdots & & \\ & & & & c_{n-1} & a_{n-1} \end{bmatrix} = \begin{bmatrix} d_2 \\ d_3 \\ d_4 \\ \vdots \\ d_{n-1} \end{bmatrix}$$

These conditions imply that T_1 is known in terms of T_2. The equation for $I = 2$, is a relation between T_1, T_2, & T_3. But since T_1 can be expressed in terms of T_2, this relation reduces to a relation between T_2 and T_3. This process of substitution can be continued until T_{n-1} can be formally expressed as T_n. But since T_n is known we can obtain T_{n-1}. This enables us to begin back substitution process in which $T_{n-2}, T_{n-3}, \cdots, T_3, T_2$ can be obtained. For this tri-diagonal system, it is easy to modify the Gaussian elimination procedures to take advantage of zeros in the matrix of coefficients [2].

2.2.2 Modelling of Cubic Spline Curve Problem

Global Parameters:

$$\begin{aligned} T_1 &= T_N = \text{First \& last data points} \\ w &= \text{wit=dht of the curve considered} \\ G &= \text{base Sequence considered} \\ t &= \text{private key considered} \\ r &= \text{random number generated} \\ p &= \text{field considered.} \end{aligned}$$

For the 2 points on the curve,

$$\begin{aligned} B_2 &= -\alpha\frac{\Delta t}{\Delta x^2} \bmod P \\ A_2 &= 1+2\alpha\frac{\Delta t}{\Delta x^2} \bmod P \\ D_2 &= Y(2)+\alpha\frac{\Delta t}{\Delta x^2} * D(1) \bmod P. \end{aligned}$$

For the 3 to $n-2$ points on the curve,

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

For the $n-1$ point on the curve,

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

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

2.2.3 Scalar Multiplication

Left to Right Multiplication:

The iteration value of the process is represented by its equivalent binary value. The reading starts from left to right positions. It starts with G as its initial input stream. If the process encounters a 0 in the bit pattern, the output of earlier iteration forms the input for the current iteration else the output of earlier iteration is summed up with Value of G which in turn forms the input of current iteration (See Algorithm 1).

1) Data Doubling: The output of earlier iteration forms the input of current iteration.
2) Data Addition: The output of earlier iteration is summed up with Value of G which in turn forms the input of current iteration.

Algorithm 1 Scalar Multiplication

```
Q := 0
for i from m to 0  do
   Q := 2Q ()(Using Data Doubling)
   if di = 1 then
      Q := Q + G (using Data addition)
   end if
end for
Return Q
```

I.E. If t is 13, its equivalent binary bit pattern is 1101. Then the iteration process continues with G as input in the first iteration, the output of earlier iteration is summed up with Value of G which in turn forms the input of current iteration, then followed with the output of earlier iteration forms the input for the current iteration and finally the output of earlier iteration is summed up with Value of G which in turn forms the input of final iteration.

Window Method (See Algorithm 2):

Algorithm 2 Window Method

```
Q := 0
for i from m to 0 do
   Q := $2^w$Q (using repeated data doubling)
   if $d_i > 0$ then
      $Q := Q + d_1P$ (using a single data addition with the pre-computed value of $d_1P$)
   end if
end for
Return Q
```

Sliding windowed method (See Algorithm 3):

Algorithm 3 Sliding Windowed Method

```
Q := 0
for i from 0 to m do
  if $d_i = 0$ then
    Q := 2Q (point double)
  else
    Grab up to $w - 1$ additional bits from $d$ to store into
    $Q := Q + d_1 P$ (using a single data addition with the pre-computed value of $d_1 P$ (including $d_i$) $t$ and decrement $i$ suitably)
    if (If fewer than w bits were grabbed) then
      Perform double-and-add using $t$
      Return Q
    else
      $Q := 2^w Q$ (repeated point double)
      $Q := Q + tP$ (point addition)
    end if
  end if
end for
Return Q
```

3 The Proposed Scheme

There are two phases of the proposed scheme: Generating Public and Private Keys and Encryption & Decryption Processes.

3.1 Generating Public and Private Keys

The sender executes the following steps (See Figure 1):

1) Sender chooses the Cubic Spline curve with both sides of the curve maintained at same data values i.e $T_S = T_N$.

2) Sender chooses the following parameters $G, T_S, T_N, \alpha, \Delta x, \Delta t$ as Global.

3) Sender chooses an integer t as private key.

4) Sender chooses a large Prime Number P to calculate the Field.

5) Sender calculates (G^t) as Public Key.

6) Private Key considered is t.

7) Public key $= (G)^t = P_E = G_2$.

3.2 Encryption and Decryption Processes

Encryption: Sender selects P_m as her Plain text. He/She then calculates a pair of texts as Cipher texts for chosen random number r.

$(P_m + P_B^r), G_r =$	$(P_m + G_3), G_1 =$	C_1, C_2

Decryption: Receiver after receiving C_1, C_2, Calculates P_m, the Plain Text using the following Formula,

$C_1 - C_2^t =$	$P_m + G_3 - G_3 =$	P_m

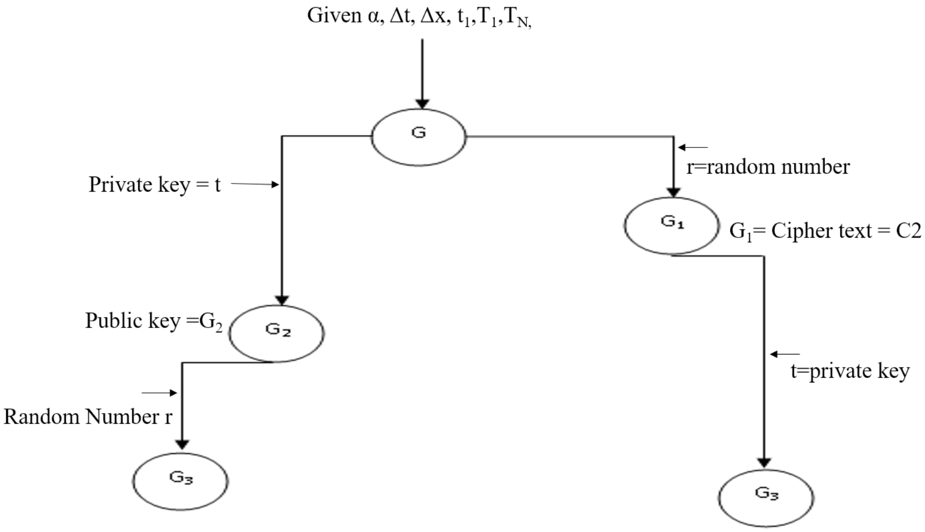


Figure 1: The generating public and private keys phase

Table 1: The parameters of an example

$\alpha = 2$	$\Delta t = 2$	$\Delta x = 2$	$P = 17$

t=2	G	3	12	16	12	6	11	1	16	3
t=4	$C_2 = G_1 = G^r$	3	9	1	15	12	6	8	16	3
t=8	$G_2 = P_B$	3	3	4	16	1	9	5	11	3
t=12	$G_3 = P_B^r = C_2^t$	3	11	14	9	16	10	1	13	3

Example 1. *The following parameters $P, \alpha, \Delta x, \Delta t$ and t are gave as Table 1.*

Mapping the alphabets of English with numerical values, like $a = 0$, $b = 1$, $x = 23$, $y = 24$, $z = 25$.

Encryption: *The plaintext of the example is ASYMMETRY. The ciphertext is in Table 2.*

Table 2: The example of encryption phase

Plain text	A	S	Y	M	M	E	T	R	Y
Alpha numeric	0	19	24	12	12	4	20	18	24
P_B^r	3	11	14	9	16	10	1	13	3
$C_1 = P_m + P_B^r$	3	30	38	21	28	14	21	31	27
Mod 26	3	5	13	21	3	14	21	6	2
C_1 In Alpha numeric	3	5	13	21	3	14	21	6	2
C_1	d	F	n	V	d	o	v	g	C
$G^r = C_2$ In Alpha numeric	3	9	1	15	12	6	8	16	3
C_2	d	J	b	P	m	g	i	q	D

Decryption: *The ciphertext C_1 of the example is 3 5 13 21 3 14 21 6 2. The plaintext is in Table 3.*

Table 3: The example of decryption phase

C_1	3	5	13	21	3	14	21	6	2
C_2^t	3	11	14	9	16	10	1	13	3
$C_1 - C_2^t$	0	-6	-1	12	-13	4	20	-7	-1
Add 26 if Negative	0	19	24	12	12	4	20	18	24
Plain text	A	S	Y	M	M	E	T	R	Y

The simulation: *The simulation of the example is shown in Figure 2.*

4 Cryptanalysis and Complexity of the Model

In this section, we show that cryptanalysis and complexity of the model.

4.1 Cryptanalysis

1) Depending on the width of the curve considered say n a block of plain text will be converted to block of cipher text. So the computing resources needed will be mapped per block rather than per character. So the amount of computing resources needed will be less when compared with algorithms like RSA & ECC.

2) Going by the construction of the algorithm; Known the first and last data points; For the data points 3 to $n-1$ (TDMA Algorithm),

$$\begin{aligned} R &= C(i)/A(i-1) \bmod P; \\ A(i) &= A(i)R(B(i-1))| \bmod P; \\ D(i) &= D(i)R(D(i-1))| \bmod P. \end{aligned}$$

Since $D(n-1)$ is known in terms of $D(n)$, $D(n-1)$ is calculated as

$$D(j) = \frac{D(j) - B(j) * D(j+1)}{A(j)} \bmod \quad \text{where} \ \ j = n-1. \tag{2}$$

By the back substitution process, the data values like $D(n-2)$, $D(n-3)$, $\cdots$, $D(2)$ are calculated.

Thus going by the construction of the algorithm, 4 modulus operations are calculated and 2 inverse operations are calculated per bit value in each step of encryption, where as in ECC it is 3 modulus operations and 2

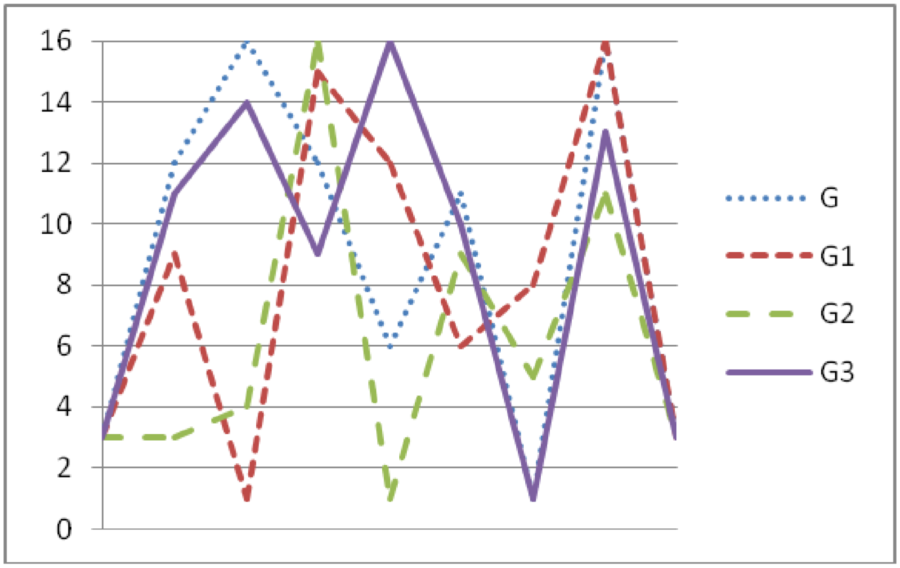


Figure 2: The simulation of the example for $t = 0$, $t = 2$, $\cdots$, $t = 16$

inverse operations per bit value in each step of encryption process. Thus the complexity of Cubic Spline is 4/3 times of ECC. Thus the sufficient key length in Cubic Spline Curve for providing necessary security to the data transmitted is roughly 3/4 of key length of ECC. Thus the key length for cubic spline curve is roughly 120 bit which is necessary to provide sufficient strength against crypto analysis.

3) The windowed method provides the benefit of using fewer point additions.

4) This algorithm has the benefit that the pre-computation stage is roughly half as complex as the normal windowed method while also trading slower point additions for point doubles. In effect, there is little reason to use the windowed method over this approach.

5) As a random number is used for every block of encryption of data, it is free from differential side channel attacks.

4.2 Complexity

Consider the equation,

$$P_B \equiv g^x (\mathrm{mod} P) \tag{3}$$

where P_B is the public key generator, g is the generator, P be the field, x is the private key.

Given g, x, and P, it is easy to calculate P_B. But given y, g, P it is very difficult to calculate x (given the differential logarithm problem). Thus the asymptotically fastest known algorithm for taking discrete logarithm modulo prime number of the order of

$$e^{\left((ln\,P)^{1/2} ln(lnP)\right)^{2/3}} \tag{4}$$

Which is not feasible for large primes.

5 Conclusion

In this work a Numerical model based cubic spline curve public key cryptography algorithm is developed. The algorithm is based on ElGamal mode of Encryption & Decryption. This works on asymmetric block cipher mode. Going by the construction of the algorithm, the key length needed for sufficient security to date is less than key length needed for ECC algorithm. Since a random number is used for Encryption process, it is free from differential side channel attacks.

The work is carried out for the boundary condition $T_1 = T_n$, i.e., both the boundaries of curve are maintained at same data values. The work can also be carried out for other boundary conditions of spline curves. The present work handles data encryption at block level of plain text. The work can also carried out for encryption of data at character level of plain text.

Acknowledgment

The authors of this paper would like to acknowledge Dr. Somanchi VSSNVG Krishna Murthy, Head, Department of Applied Mathematics, Defence Institute of Advanced Technology, Pune for supporting this work.

References

[1] R. C. C. C. Cheng, N. J. Baptiste, W. Luk, and P. Y. K. Cheung, "Customizable elliptic curve cryptosystems," *IEEE Transactions on VLSI Systems*, vol. 13, no. 9, pp. 1048–1059, 2005.

[2] S. D. Conte and C. deBoor, *Elementary Numerical Analysis*. New York: Mc Graw Hill, 1972.

[3] CSC-93-09, *The interpolating random spline cryptosystem.*

[4] W. Diffie, "The first ten years of public key cryptography," *Proceedings of IEEE*, vol. 76, no. 5, pp. 560–577, 1988.

[5] IEEE, "Standard specifications for public key cryptography," *IEEE Standard*, pp. 1363, 2000.

[6] A. V. N. Krishna, "A new non linear model based encryption scheme with time stamp & acknowledgement support," *International Journal of Network Security*, vol. 13, no. 3, pp. 202–207, 2007.

[7] A. V. N. Krishna and A. V. Babu, "A new model based encryption scheme with time stamp & acknowledgement support," *International Journal of Network Security*, vol. 11, no. 3, pp. 172–176, 2010.

[8] A. V. N. Krishna and A. V. Babu, "A new non linear, time stamped & feed back model based encryption mechanism with acknowledgement support," *IJANA*, vol. 2, no. 5, pp. 191–198, 2010.

[9] S. Moon, "A binary redundant scalar point multiplication in secure elliptic curve cryptosystems," *International Journal of Network Security*, vol. 3, no. 2, pp. 132–137, 2006.

[10] R. Ramanna, "Numerical methods," pp. 78–85, 1990.

[11] R. R. Ramasamy, M. A. Prabakar, M. I. Devi, and M. Suguna, "Knapsack based 11 ECC encryption and decryption," *International Journal of Network Security*, vol. 9, no. 3, pp. 218–226, 2009.

[12] W. Stallings, "Cryptography and network security," *Prentice Hall, 4th Edition.*

[13] V. P. Suhas, "Numerical heat transfer and fluid flow," pp. 11–75, 1991.

[14] N. Sun, T. Ayabe, and K. Okumura, "An animation engine with cubic spline interpolation iih msp-08," *Proceedings of 2008 International Conference on Intelligent Information Hiding & Multimedia Signal Processing*, pp. 109–112, 2008.

A.V.N. Krishna has a total of 22 years of teaching and research experience. The author has Published his research work in the National and International Journals f repute. Presently the author is working as Professor in the Department of Computer Science & Engineering, Faculty of Engineering, Christ University, Bangalore, Karnataka, India. The author is presently guiding 4 doctoral students in the field of Cryptography.

Addepalli Hari Narayana is in his 3^{nd} year B.Tech in Electrical Engineering from IIT-Indore, Indore.

K. Madhura Vani is presently working as Assistant Professor, CSE in Shreyas Institute of Engineering & Technology and working in Network Security area.

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.