



**Journal of Madenat Alelem University College**

[www.jmauc.edu.iq](http://www.jmauc.edu.iq)

E-mail: [jmauc@mauc.edu.iq](mailto:jmauc@mauc.edu.iq)

TEL: 7801099835

**Publisher:**

Madenat Alelem University College

كلية مدينة العلم الجامعة

[www.mauc.edu.iq](http://www.mauc.edu.iq)

رقم الايداع في دار الكتب والوثائق 1333 لسنة 2009

**Editor in chief**

Dr. Shaker M. Al-Jobori

**Deputy editor in Chief**

Dr. Jabbar F. Al-Maadhidi

**Editorial board**

**Lect. Isam Atta Ajaj**

**Editor**

**Dr. Saeed Selman Kamoon**

**Dr. Mousa M. Al-jobori**

**Dr. Sabah Abdul Latif Nassif**

**Dr. Usama Aladdin Ibrahim**

**Dr. Saad Abdolridha Makki**

**Dr. Abd Almonem K. Hammadi**

**Dr. Ali Mahdi**

**Dr. Hussain H. Ahmed**

**Dr. Farooq Abdul Azeez Mohammed**

**Dr. Ayaid K. Zgair**

**Advisiori Board**

**Prof. Dr. AbdolHazim Al-Rawi, Alrashed University**

**Prof. Dr. TawficNajim, Al-mammon University College**

**Prof. Dr. Ghazi Faisal, Al-Nahrin University**

**Prof. Dr. Nabil Hashim, Babel University**

**Dr. Ayad A. Al-Taweel, Ministry of Science and Technology**

**Assis. Prof. Ahmed Mossa, Technical University**

**Dr. Ammer M. Ali, MadentAlelem College**

**Dr. IbrahimKhammas, MadentAlelem College**



## INSTRUCTIONS to AUTHERS

Submitted articles to the Journal of Madenat Al-Elem University College can be published in all fields related to the Academic Departments of the College (Biology, Law, programming Engineering Sciences, Computer Techniques Engineering, Law, Medical Physics, Civil Engineering, and Accounting, and Anesthesia).

Written request for publication and signing a consent form to publish must be for articles which have not been published or submitted for publication to other journals. Three copies with CD are needed. Manuscripts should be typed on: A4 white paper, double spaced, written in Times New Roman font size 14. Margins should be 3cm from top, bottom, left and right. The main title should be in: bold Times New Roman, and font size 14. Author names should be written in the following sequence: first name, middle name, the family name, followed by the names of departments and institutions of work. A footnote accompanies the first page stating the full address of correspondence author.

Articles need to contain the following items:

- Abstract in English and Arabic not more than 300 words.
- Article includes the following items: Introduction, Materials and Methods, Results and Discussion, Conclusion and References.
- References should be numbered in the text according to the sequence appeared in the text and listed in order.
- Tables and figures should be appropriately titled with size not exceed an A4 page.

The editor reserves the right to reject or accept any article submitted.

**Publication charges:** Each accepted paper is required to pay the publication charge (100,000 Iraqi dinars). Five thousands Iraqi dinar are requested for each extra page, when the article exceed 20 printed pages.



## Distribution of ABO blood groups in beta thalassemia patients dependent on blood transfusion In Bagdad city.

*Sarah M. Mahmoud Marbut\**, *Maha A. Hamdi\*\**, *Abdulahadi M.Jumaa\*\**, *Basma Abbas Salman\*\*\** .

*Dept. of Microbiology\**, *Physiology\*\**, *College of Medicine, Tikrit University.*

*Lab. Dept. Renal dialysis, Baghdad Teaching Hospital\*\*\*.*

[Sarah1993@yahoo.com](mailto:Sarah1993@yahoo.com)

### Abstract

**Introduction:** The discovery of the ABO blood groups by Karl Land steiner was an important achievement in the history of blood transfusion followed by discovery of Rh antigen. There are differences in the distribution of ABO, and Rh (D) blood groups amongst different populations. The study of blood groups plays an important role in various genetic studies, in clinical studies for reliable geographical information and in blood transfusion practice, which will help in reducing morbidity and mortality rate. The **aim** of the study is to find out whether there is any relationship between ABO blood groups and thalassemia patients. **Methods:** The present study was done at Baghdad teaching hospital, GIT center from 1st of Jan 2017 to the end of September 2017 . A cross sectional study was done in Baghdad-Iraq, in collaboration with GIT center. A total of 200  $\beta$ - thalassemia patients without viral hepatitis. (100 males and 100 females) (the age of patients is between 7 to 24 years). Thalassemia patients ABO type beta major were tested by ABO blood group at laboratory department. Determination of ABO blood group Blood group is determined by slide haemagglutination technique. **Result:** In male and female thalassemia patients, blood group O is the highest %, followed by B blood group in in males. However, the 2nd one in female patients is (A) blood group. The lowest ABO blood group in male and female patients is AB blood group. In thalassemia patients, both gender, the highest percentage rate of Rh antigen is found in O+ , (43% in males and 38% in female patients), followed by B+, (29% in males and 21% in female patients), However, the lowest was in ABO blood group. From the present study; it could be summarized that

ABO blood groups in thalassemia patients was found mostly in O+ blood group, and almost there is no thalassemia in AB- and O- groups.

**Key words:** Thalassemia, ABO, blood group, Baghdad.

## الخلاصة

تعين مجاميع الدم في مرضى الثلاسيميا نوع بيتا و المعتمدة على نقل الدم المنتظم في مدينة بغداد

م.م ساره موسى محمود و مها ارشد حمدي و عبد الهادي محمد جمعه و باسمه عباس سلمان

توجد أنواع و فصائل دم مختلفة في البشر ، والمعروفة باسم مجاميع الدم. المستضدات في مجاميع الدم (ABO) هي مادة وراثية مصممة وتلعب دورًا حيويًا في سلامة نقل الدم و التي تم اكتشاف مجاميع الدم من قبل كارل لاند ستينر. هناك اختلافات في توزيع ABO ، ومجموعات الدم (Rh (D بين مجموعات سكانية مختلفة. وفي الدراسات السريرية للحصول على معلومات جغرافية موثوق بها وفي ممارسة نقل الدم ، والتي ستساعد في تقليل معدل المراضة والوفيات. الهدف من الدراسة هو معرفة ما إذا كانت هناك أي علاقة بين مجموعات الدم ABO و مرضى الثلاسيميا في بغداد. طرق العمل: أجريت الدراسة الحالية في مستشفى مركز الجهاز الهضمي- بغداد من 1 يناير 2017 حتى نهاية سبتمبر 2017. أجريت دراسة مقطعية مستعرضة في بغداد- العراق ، بالتعاون مع مركز GIT. ما مجموعه 200 مريض من مرضى الثلاسيميا دون الالتهاب الكبدي الفيروسي. (100 من الذكور و 100 من الإناث) (عمر المرضى بين 7 إلى 24 سنة). تم اختبار مجاميع الدم لمرضى الثلاسيميا نوع بيتا الكبرى بواسطة مجموعة ABO الدموية في قسم المختبر. تحديد مجموعة فصيلة الدم ABO يتم تحديدها عن طريق تقنية الصفيحة الدموية. النتائج: في مرضى الثلاسيميا الذكور والإناث ، تكون مجموعة الدم O هي أعلى نسبة مئوية ، تليها مجموعة الدم B في الذكور. ومع ذلك ، فإن الثانية في المرضى الإناث هي مجموعة الدم. أقل مجموعة دم ABO في المرضى الذكور والإناث هي مجموعة الدم AB. في مرضى الثلاسيميا ، كلا الجنسين ، تم العثور على أعلى نسبة مئوية من مستضد Rh في O + ، (43% في الذكور و 38% في المرضى الإناث) ، ثم تليها B + ، (29% في الذكور و 21% في المرضى الإناث)

**الكلمات المفتاحية:** مرضى الثلاسيميا، مجاميع الدم. بغداد

### Introduction:

The thalassemias (Greek: Thalassa meaning sea) are a group of single gene inherited autosomal recessive hematological disorders caused by defects in the synthesis of one or more of the hemoglobin chains that cause hemolytic anemia.<sup>1,2</sup>  $\alpha$  and  $\beta$ -thalassemias are caused due to reduced or absent synthesis of  $\alpha$  and  $\beta$  globin chains respectively, (1).

Phenotypically  $\beta$ -thalassemia is of three types.  $\beta$ -thalassemia minor is a heterozygous state in which there is around 50% decrease in synthesis of  $\beta$ -globin protein, causing mild to moderate microcytic anemia. Affected individuals are usually asymptomatic.  $\beta$ -thalassemia intermedia shows mild to moderate anemia, (2).

The ABO system is one of the most important blood group systems in transfusion medicine.

The ABO system consists of A antigens, B antigens, and antibodies against these antigens. Landsteiner discovered the ABO system in 1900. As opposed to many other blood group systems such as the Rh system, in this system the presence of “naturally occurring” antibodies against A and B antigens in individuals who do not express those antigens (Landsteiner’s Law) causes an adverse and occasionally fatal outcome at the first mismatched transfusion, (3,4).

The existence of ABO and Rhesus (Rh) antigens is clinically very important as it plays a major role in blood transfusion and organ transplantation. Though all the population of the world have same blood group system, but the frequency of ABO and Rh antigens is found to vary amongst all populations, (5).

A relationship between blood groups and certain diseases in human was well established, it may not is of great genetic importance. Such a study found that the blood group B was more susceptible to hypertension, (6).

Previous Studies showed an association between ABO blood group and severity of chronic periodontitis. Investigators concluded that the patients with group B were found to be at greater risk of developing more severe form of periodontitis, (7).

Also, a previous study found that blood type A was more directly related with insulin resistance, while those with type O are less directly related with insulin resistance in Turkish population, (8). A large number of studies have examined the association between ABO blood groups and variety of diseases or conditions, (9, 10).

The **aim** of the study is to find out whether there is any relationship between ABO blood groups and thalassemia patients and the study also attempt to explore any relationship between blood group antigens and  $\beta$ -thalassemia so that it will become very easy to predict the type of population which is more prone or resistant to  $\beta$ -thalassemia.

### **Patients and Methods**

The present study was done at GIT center in Baghdad, from 1<sup>st</sup> of Jan 2017 to the end of September 2017. A cross sectional study was done in Baghdad-Iraq, in collaboration with GIT center. A total of 200  $\beta$ -thalassemia patients without viral hepatitis. (100 males and 100 females) (the age of patients is between 7 to 24 years)

Thalassemia patients ABO type beta major were tested by ABO blood group at laboratory department.

## **Determination of ABO blood group**

Blood group is determined by slide haemagglutination technique. 2.5% suspension of red blood cells was prepared in normal saline (0.85g/dl sodium chloride in distilled water) preparation method given below. Mix one drop of blood with 1 ml of normal saline. This provided the red suspension. On one half of glass slide, one drop of Anti A human poly clonal or murine monoclonal blood grouping serum was placed. On the other half a glass slide one drop of Anti B (yellow color) human polycolonal or murine monoclonal blood grouping serum was placed. Using a Pasteur pipette one drop of red blood cell suspension was added to each half of the slide. With separate applicator; the serums

was well mixed back and forth and observe for agglutination, (1).

## **Statistical analysis**

All data were presented as a mean & standard deviation (S.D). Un paired student T test was used to compare between mean of variables. P value less than 0.05 was accepted as a significant value.

## **Results**

The distribution of ABO blood group in male and female thalassemia patients is presented in table 1. In male and female thalassemia patients, blood group O is the highest %, followed by B blood group in in males. However, the 2<sup>nd</sup> one in female patients is A blood group.

The lowest ABO blood group in male and female patients is AB blood group.

**Table 1** Distribution of ABO blood groups in thalassemia patients according to gender.

<b>Gender</b>	<b>A</b>	<b>B</b>	<b>AB</b>	<b>O</b>	<b>Total</b>
<b>Males</b>	19	34	3	44	100
<b>Females</b>	31	21	7	41	100
<b>Total</b>	50	55	10	85	200

While table 2 show the result of Rh antigen in thalassemia patients.

**Table 2** Distribution of Rh antigen in Thalassemia patients

<b>ABO types</b>	<b>Males</b>		<b>Females</b>		<b>Total</b>	<b>%</b>
	Number	%	Number	%		
<b>A+</b>	19	19%	30	30	49	24.5
<b>A-</b>	0	0	1	1	1	0.5
<b>B+</b>	29	29	21	21	50	25
<b>B-</b>	5	5	0	0	5	2.5
<b>AB+</b>	3	3	7	7	10	5
<b>AB-</b>	0	0	0	0	0	0
<b>O+</b>	43	43	38	38	81	40.5
<b>O-</b>	1	1	3	3	4	2
<b>Total</b>	100	100%	100	100%	200	100%

In thalassemia patients, both gender, the highest percentage rate of Rh antigen is found in O+ , (43% in males and 38% in

female patients), followed by B+, (29% in males and 21% in female patients), However, and the lowest was in ABO blood group.

The distribution of Rh antigen in both male and female patients was found as follows;

The highest % was in O+, (40.5%), followed by B+, (25%) and A+, (24.5%). However, the lowest one in AB-, (0%), and A-, (0.5%).

From table 1 and table 2, the study summarized that thalassemia found mostly in O+ blood group, and almost there is no thalassemia in AB- and O- groups.

### **Discussion**

In the present study, in both gender of thalassemia patients, blood group O is the highest percentage in thalassemia patients, followed by B blood group in in males. However, the 2nd one in female patients is A blood group. The present study also found that the largest proportion of Beta thalassemia major patients was of group O+,

followed B+, then A+. While, the lowest percentage of ABO blood group in thalassemic patients is B blood group.

A previous study was done by Abid Al-Kader Abbas (2013), during 2013 in Kirkuk found nearly similar results except that , his study revealed blood group "O+" was the most common group (48.4%) in his patients' sample, (11).

In previous study done in thalassemia unit in Mumbai, India; it was found that the most common blood group getting affected by the disease  $\beta$ -thalassemia is O +ve with the same people having higher chances of family history of the same disease. Within the family members who have the positive history of the disease, most common was O +ve blood group again, (12).

Also, in study done during (2014) in Kirkuk showed that

blood groups "O+ " is the most common among blood group types equal to 48.4% followed by B+ (24.2%) and "A" equal to 18.1% among thalassaemic patients, (13).

As compare with normal population, a previous study, done in Tikrit city by Marbut *et al*, (2008). A total of 4309 subjects were investigated in this study. 2411 blood donors (2357 men & 54 women) & 1898 blood recipient patients (759 men & 1139 women). The highest frequency of blood group in blood donors subjects was blood group O+ (41.5%) & the lowest frequency of blood group was blood group AB- (0.29%). (35.8%) and the lowest frequency of blood group was blood group AB- (1%). (14).

ABO blood group system is one of the most commonly used factor in different investigation especially in human population

genetics for its important role and easy availability as compared with other tissues of the human body, (1).

From the Table 1 and 2, it can be concluded that the most common blood group observed in patients of  $\beta$ -thalassemia is O +ve in 40.5% of the patients. This finding agree with previous result in normal population in Iraq, (14).

With 36% B+ve becomes the common blood group in the normal population. AB-ve and O –ve people are less likely to get affected by the disease.

Comparing it with common population, O –ve becomes a blood group presenting in lesser number of people.

Mohssin MY *et al*, studied of frequency distribution of hemoglobin variant and ABO blood groups among thalassaemia patients from Ibn-Al-Baladi hospital in Baghdad/Iraq stated

the same fact of O blood group being common incidence (59.1%) and AB with the least common occurrence of the disease  $\beta$ -thalassemia, (2).

Previous study done in Iran about the association of ABO blood group and complication of multiple blood transfusion in beta thalassemia, it was found that the prevalence of hepatitis-C and related factors among  $\beta$ -thalassemia major patients in Southern Iran which stated that HCV rate in Overall distribution of ABO frequency in India shows the group B to be the commonest blood group in northern and western part of India whereas in eastern, southern and central part O is the most prevalent blood group. Cumulatively, O is the dominant blood group among the Indian population (15). While in previous study,  $\beta$ -thalassemia patients was seen more in patients of blood group O, (16,17,18).

According to this result, we **recommended** that ABO blood group system should be included in future investigations related to blood disorder diseases.

## References

- 1-Rai, J. and Singh, B. Distribution of ABO Blood Groups and Rhesus Factor Percentage Frequencies Amongst the Populations of Sikkim, India.
- 2- Mohssin MY, Mahmood AE, Kamal SB, Batah EH. Frequency distribution of hemoglobin variant and ABO blood groups among thalassemia patients from Ibn-Al-Baladi pediatric hospital in Baghdad/Iraq. W J Pharma Pharmaceut Sci. 2015;4(11):31-9.
- 3- Landsteiner K, Weiner AS. An agglutinable factor on human blood recognized by immune sera for rhesus blood. Proceedings of the society for experimental Biology 1940; 43:223.

- 4- Koeppen, BM and Stanton, BA. Berne & Levy Physiology. 6<sup>th</sup> edition. Mosby. Ny. 2010: 287-290.
- 5- Neil D. Avent and Marion E. Reid. The Rh blood group system: a review. 2000; 95: 375-387 .
- 6- Sadiq, HN, Anjum, R., Shaikh, SM *et al.* A study on the correlation of ABO blood group system and hypertension. International J. of Applied Dental Sciences 2017; 3(4): 38-41.
- 7- Ghamdi AA. Association between ABO Blood groups and severity of Chronic Periodontitis. JKAU Med. Sci. 2009; 16:31-41.
- 6- Hasna Amer Mouhaus , Saleemh Hameed Abbas , Azhar Salih Musa, and Haider Kassim Mahawi. A study of ABO blood group and Rhesus factor distribution among sample of Missan province population. Journal of Basrah Researches ((Sciences)). 2010; 36(5): 48-53.
- 7- Watkins WM. The ABO blood group system: historical background. Transfus Med 2001;11:243- 65.
- 8- Aykas, F., Denizavci, L., Ferhatarik, F, Çetinkaya. AL. *et al.* There is a relation between blood subgroups and insulin resistance. Acta Medica Mediterranea, 2017, 33: 987.
- 9-O'Donnell, J & M.A. Laffan. (2001)."The relation ship between ABO blood group, factor VIII and von will brand factor". Transfuse Med. ; 11(4) : 343-351.
- 10- Reddy, VM; M. Daniel ; E. Bright ; S.R Broad and A.A. Moir. (2008) "Is there an association between blood group O and epistaxis". J. laryngology & Otology , 122 : 366-368.
- 11- Abbas AM. Iron overload in thalassemia and its effect on

gonads. M. Sc Thesis submitted to College of Medicine, Tikrit university 2013.

12- Pranoti A. Sinha<sup>1</sup>, Sachin H. Mulkutkar<sup>1</sup>, J. B. Bhavani. Study of distribution of ABO blood groups in  $\beta$ -thalassemia patients. *Int J Res Med Sci.* 2017 Aug;5(8):3479-3483.

13- Azeez, FS. Iron Overload in Beta- thalassaemia Major patients and its effect on Pituitary Gland. M.Sc Thesis submitted to the college of Medicine, Tikrit university. 2014.

14- Marbut, MM., Aseel H. Ali, Tamara A. Mujeed, Zahraa K. Subhy. Physiological and hematological characteristics of blood recipients and donors subjects attending Tikrit teaching hospital from 2006 to end of 2007. *TMJ.* 2008; 14(2): 222-234.

15-Shekhar H, Kaur A, Jadeja P, Parihar P M and Mangukiya K K. Frequency and distribution of ABO blood group and Rh (D) factor in southern Rajasthan *International Journal of Science & Nature.* 2014; 5 494-497.

16- Mohammadali F, Pourfathollah A. Association of ABO and Rh blood groups to blood borne infections among blood donors in Tehran-Iran. *Iranian J Public Health.* 2014;43(7):981-9.

17- Awad MH. Homozygous beta thalassaemia in Mosul. M. Sc Thesis submitted to University of Mosul. Iraq. 1999.

18- 1-Ghori MR, Tayyab M, Raziq F. Frequency of ABO and Rh D blood groups in transfusion dependent patients. *J Postgraduate Med Institute.* 2003;17(2);177-83.

## English Character Recognition from Video Stream based on Bag of Visual Words (BOVW)

*Ekhlas Falih Nasser*

*Dr. Abdulameer A. Karim*

[ekhlas\\_uot1975@yahoo.com](mailto:ekhlas_uot1975@yahoo.com)

[ameer\\_aldelphi@yahoo.com](mailto:ameer_aldelphi@yahoo.com)

### Abstract

Numerous digital images are available for printing the documents. The discussions are continuing for arriving to best algorithm for identifying the English letters. The suggested method has four steps. Firstly, apply wavelet transform on images of letters using Haar filter. Secondly interest points were detected using features from accelerated segment test (FAST) corner detection. Thirdly those points were described using Speeded up Robust Features (SURF). Fourthly, the clustering algorithm of moving k-means is employed to obtain bag of visual words (BOVW) and then build vocabulary and a histogram from visual words. The features of each visual word for video images and test image are matched using Manhattan distance measure. The suggested system was tested on three types of English letters font's databases (Time New Roman, Arial Black and Calibri). Experimental outcomes show that the suggested method is more efficient and fast for matching and recognizing a letter than seven moment's method. The recognition time for BOVW is less than the seven moment's time and the BOVW accuracy depends on number of correct character recognition. BOVW have optimal accuracy in the process of recognition of letters.

**Keywords:** Haar filter, FAST, SURF, moving k-means clustering, Manhattan distance

تمييز الحرف الأنكليزي من سلسلة الفيديو بالاعتماد على حقيبة من الكلمات المرئية

م.أخلاق فالح ناصر

أ.م. د. عبدالأمير عبدالله كريم

الجامعة التكنولوجية / قسم الحاسوب

### الخلاصة

هنالك الكثير من الصور الرقمية المتوفرة للوثائق المطبوعة والبحث مازال مستمر للوصول الى افضل خوارزميه للتعرف على الحروف الأنكليزية. الطريقة المقترحة تتكون من أربع خطوات. اولاً يتم تطبيق تحويل المويجه على الصور الحرفيه باستخدام Haar فلتر. ثانياً يتم اكتشاف النقاط المهمه باستخدام كاشف زاوية الصفات لأختبار المقطع السريع (FAST). ثالثاً يتم وصف تلك النقاط المهمه باستخدام الواصف تسريع الصفات القوي (SURF). رابعاً خوارزميه k-means moving تستخدم لتجميع الصفات بشكل عناقيد (clusters) للحصول على حقيبة من الكلمات المرئيه وبعدها يتم بناء المعجم و مخطط للكلمات المرئيه. صفات كل كلمه مرئيه في صور الفيديو والصوره المختبره يتم مطابقتها باستخدام مقياس المسافه منهاتن. تم اختبار النظام المقترح على ثلاثة أنواع من قواعد بيانات الحروف الإنكليزية (تايم نيو رومان، أريال بلاك و كالبري). تبين النتائج التجريبية أن الطريقة المقترحة هي أكثر كفاءة وسرعة لمطابقة وتمييز الحرف مقارنة مع استخدام طريقة العزوم السبعه. الوقت الذي تأخذه حقيبة الكلمات المرئيه لتمييز الحرف يكون اقل من الوقت الذي تاخذه العزوم السبعه، دقة حقيبة الكلمات المرئيه يعتمد على عدد تمييز الحروف بصوره صحيحه. لذا تعتبر حقيبة الكلمات المرئيه ذات الدقه الأمثل الدقه المثلى في عملية تمييز الحروف.

**الكلمات المفتاحية:** تسريع الصفات القوي و تمييز الحروف و فلتر هار

## 1. Introduction

Employing digital documents images were increased in the recent years and they can be considered as resources for management and education in many applications. They are not dealing directly with digital images contents; because of they are complete digital images. For document's text accessing, it must be recognized text contents and divided it into symbols and letters, which the document wrote. The task which can be performed daily was called recognition of an object. It can be performed in the more varied circumstances, navigation to the sources of food, migration, predators' identification, mates' identification, etc. with efficiency remarkable [1]. Methods can be developed, which are capable for emulating different recognition of object forms that are evolving along with the need for constructing systems of an intelligent automated. Today's technology's main direction in industry and other activity domains also in these objects systems are represented in a convenient path for processing type, and these representations can be called *patterns*. The scientific discipline that deals with object classification and description methods is called *Recognition of Pattern (RoP)*. Recognition of pattern is

therefore a fruitful region of research, with doubled links to numerous other disciplines; including professionals from many regions [2].

## 2. Moment Invariants

Moment invariants are employed in numerous applications for recognition of pattern. Invariants of image's feature or shape's feature stay without modification if that shape or an image submit any mixing of the following alterations

- Modification in size (Scale)
- Modification in size position (Translation)
- Modification in Orientation (Rotation)
- Reflection

The invariants of moment are extremely beneficial method for features' eliciting from two-dimensional images. Invariants of moment are characteristics of regions which are connected in binary format of images, which are invariant to scaling, translating, rotation and.

Invariants of moment are beneficial because they can be determined by simple computed set of region characteristics which can be employed for classification of the shape and recognition of the part. Suppose a two-dimensional  $F(x, y)$  in a spatial domain [Ach 05]. *Geometric moment* of order  $p + q$  is illustrated in Eq.(1)

$$m_{p,q} = \sum_x \sum_y x^p y^q F(x, y) \dots (1)$$

For  $p, q = 0, 1, 2, \dots$ . The moments' central are explained by

$$x_c = m_{1,0} / m_{0,0}$$

$$y_c = m_{0,1} / m_{0,0}$$

Where  $m_{1,0}$  in Eq. (1) and  $(x_c, y_c)$  can be called the object's region center. The normalized moment's central that represented by  $\eta_{p,q}$  can be defined by Eq.(2)

$$\eta_{p,q} = \mu_{p,q} / \mu_{0,0}^\gamma \dots (2)$$

Where

$$\gamma = p + q / 2 \dots (3)$$

A set of seven invariants can be derived from the second and third normalized central moments. This set of seven HU moment invariants (4) to (10) is invariant to translation, rotation, and scale change.

$$\phi 1 = \eta_{2,0} + \eta_{0,2} \dots (4)$$

$$\phi 2 = (\eta_{2,0} + \eta_{0,2})^2 + 4\eta_{1,1} \dots (5)$$

$$\phi 3 = (\eta_{3,0} - 3\eta_{1,2})^2 + (3\eta_{2,1} - \eta_{0,3})^2 \dots (6)$$

$$\phi 4 = (\eta_{3,0} + 3\eta_{1,2})^2 + (3\eta_{2,1} + \eta_{0,3})^2 \dots (7)$$

$$\begin{aligned} \phi 5 = & (\eta_{3,0} - 3\eta_{1,2})(\eta_{3,0} + 3\eta_{1,2})[(\eta_{3,0} + 3\eta_{1,2})^2 \\ & - 3(\eta_{2,1} + \eta_{0,3})^2] + (3\eta_{2,1} - \eta_{0,3})(\eta_{2,1} + \eta_{0,3}) \\ & [3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2] \dots (8) \end{aligned}$$

$$\phi 6 = (\eta_{2,0} + \eta_{0,2})[(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} - \eta_{0,3})^2] + 4\eta_{1,1}(\eta_{3,0} + \eta_{1,2})(\eta_{2,1} - \eta_{0,3}) \dots (9)$$

$$\begin{aligned} \phi 7 = & (3\eta_{2,1} - \eta_{0,3})(\eta_{3,0} + \eta_{1,2})[(\eta_{3,0} + \eta_{1,2})^2 \\ & - 3(\eta_{2,1} + \eta_{0,3})^2] + (3\eta_{1,2} - \eta_{3,0})(\eta_{2,1} + \eta_{0,3}) \\ & [3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} - \eta_{0,3})^2] \dots (10) \end{aligned}$$

### 3. Proposed methodology based on Bag of Visual Words (BOVW)

The suggested system in this paper for English character recognition is based on bag of visual words (BOVW). It has four

steps to build BOVW. Firstly, get the images of letters from video series. Secondly, images of letters can be transformed to frequency domain by applying

Haar transform. Thirdly, the features of interests are elicited using FAST corner detection. After eliciting interest features from all images of letters, the interest features could be described using SURF descriptor. Fourthly, the algorithm of clustering that using moving k-means is employed for described features clustering to build visual words for each image of letter. Each visual word corresponds to the center of cluster, and then builds vocabulary from all visual words. An image of letter can be entered as query

image and the numbers of features are elicited from a query. The elicited numbers of features are matched with all images of letters in vocabulary using Manhattan distance and discover the similar image of letter which is matching with query image of letter. After detected the similar image, it can recognize the name and a sequence of an image in that vocabulary. Figure 1 displays block diagram of the system for retrieve an image in a proposed method

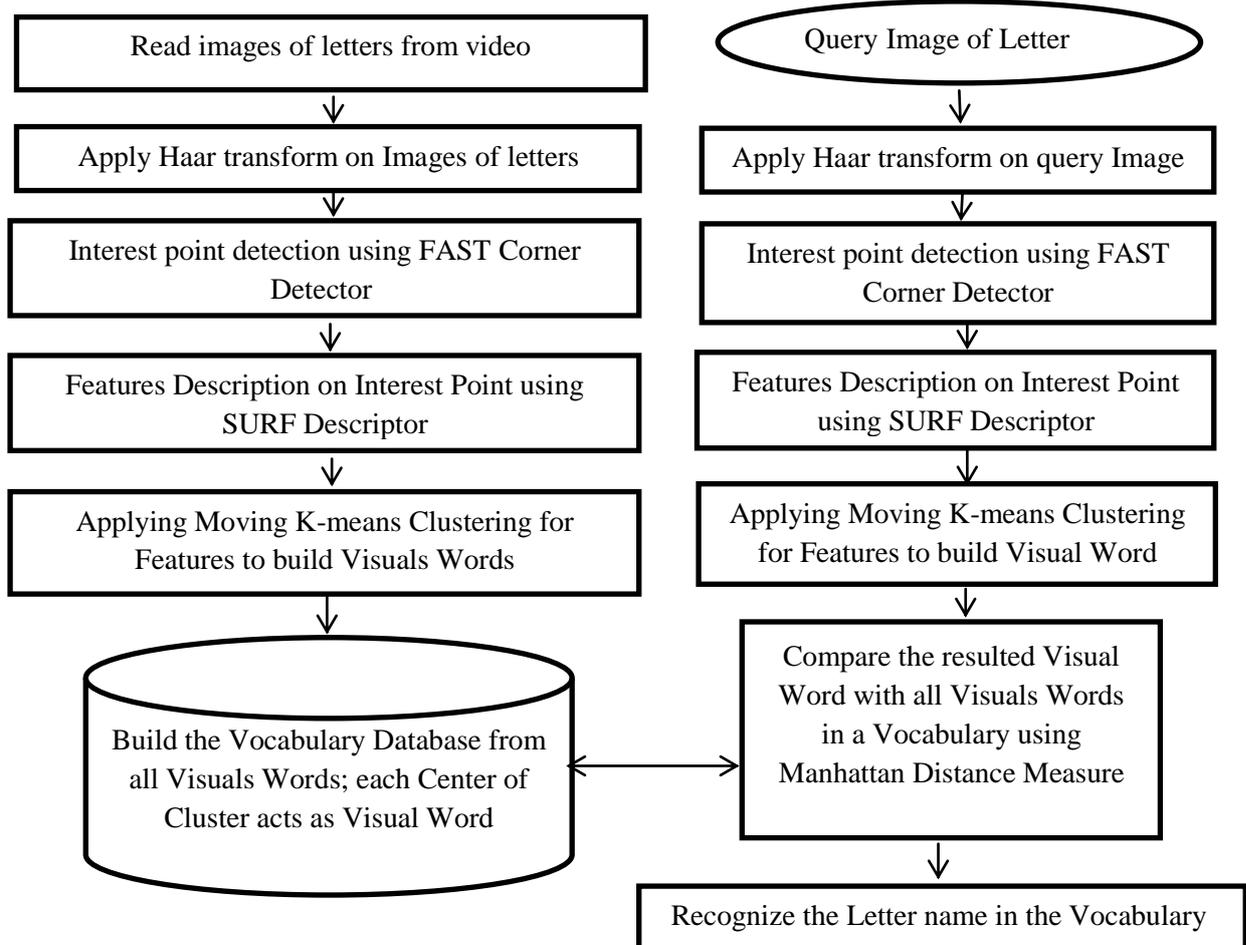


Figure (1): Block diagram of proposed system for character recognition

### 3.1 Haar Transform

Wavelet transforms based on sub-sampling low pass and high pass filters (Quadrature Mirror Filters (QMF)). By splitting the data into low pass band and high pass band with or without losing any information, matching the filters is done. Wavelet filters can be organized for applications of a broad range and numerous different sets of filters can be proposed for various applications. Wavelets are functions identified over a finite interval. The purpose from wavelet transform is to transform the data from Time-space domain to Time-frequency domain which can perform best compression results. There are a wide variety of popular wavelet algorithms,

including Daubechies wavelets, Mexican Hat wavelets and Morlet wavelets. These algorithms have the disadvantage of being more expensive to calculate than the Haar wavelets. Haar wavelet is a simplest form of wavelets; the function is defined in Eq.11. The four bands are indicates to Low-Low (LL), Low-High (LH), High-Low (HL) and High-High (HH). It can potential to implement group of wavelet filters on LL band with self-path as implemented to the main image because it contains image-like information. An image dividing operation into sub-bands can be permanent as far as wished (based on an image resolution), probably for image compression it is commonly continued only to 4 or 5 levels[5].Figure (2) shows a wavelet transform on gray scale image.

$$\varphi(x) = \begin{cases} 1 & 0 \leq x < 1/2 \\ -1 & 1/2 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \quad \dots(11)$$



Figures (2):Gray\_Scale image a) Original image b) Two Dimensional Haar Transform

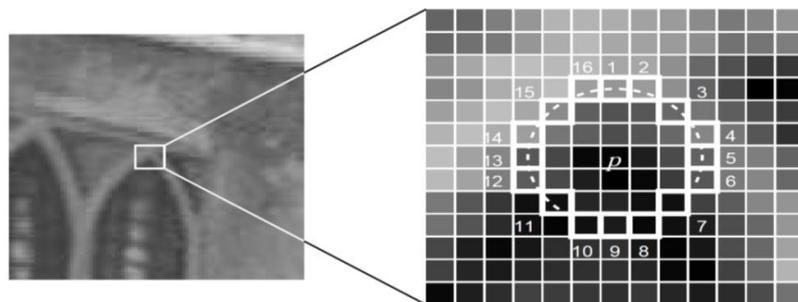
### 3.2 FAST Corner detection

Corners are important local attributes in images. These points have high drooping and lie in cross brightness of an image areas. In a diversity of image attributes, corners cannot be affected by lighting and can be have rotational constancy. Corners form 0.05% only from all image pixels. Without missing data of image, elicitation corners can reduce image's data processing [6]. Therefore, the detection of corner has factual value and it plays an important place for motion tracking, image matching, augmented reality, representation of an image and other different fields [7]. Tremendous techniques for detection the corners was suggested from multiple searchers. These techniques are divided into two groups: group of techniques focus on contour and the other group focus on intensity. Techniques focus on contour work at first to extract all contours from an image and then seek for points that have

maximum diversity over those contours [8]. Feature from an accelerated segment test (FAST) uses a Bresenham's algorithm for circle drawing with diameter of 3.4 pixels for trial mask. Trial 16 pixels compared to the nucleus's value for a complete accelerated segment. The criterion of corner should be more relaxed to block this broad trial. A pixel's criteria must be a corner based on an accelerated segment test (AST) which there must exist at least S pixels that have more brilliant circle connection or darker than a threshold. To reduce feature space of an image and increase the implementation speed of the suggested system, our algorithm was used an adaptive threshold **thr** and it can be computed using Eq. (12). Other values of 16 pixels are disregarded. So the value of S can be used to determine the detected corner at maximum angle [9].

$$thr = (Img_{max} - Img_{min}) / 2 \quad \dots (12)$$

where  $Img_{max}$  and  $Img_{min}$  are the largest and smallest gray value of whole image.



**Figure (3): Image display the point of interest under a test and the circle of 16 pixels**

### 3.2.1 Steps of FAST algorithm

The basic steps of FAST algorithm of detection the corners are illustrated bellow:-

1. From an image, chose a pixel  $p$ . IP represent pixel's intensity. This pixel can be specified as a point of interest or not. (Returning to figure (3)).
2. Get **thr** from Eq. (12) that represents the value of threshold intensity.
3. Assume periphery a pixel  $p$  represents the center of circle which has 16 pixels. (Brenham circle of radius 3.)
4. Need "N" exposure contiguous pixels out of the 16 pixels, either below or above IP by **thr** value, if the pixel wants to discover as a point of interest.
5. First match 1, 5, 9 and 13 of the circle pixels' intensity with IP to make an algorithm fast. From figure (3), at least three of these four pixels should accept the norm of the threshold for this it subsist an interest point. P is not an interest point (corner) if at least three values of - I1, I5, I9 and I13 are not below or above  $IP + \mathbf{thr}$ . For this, a pixel  $p$  can be rejected as a potential point of interest. Else if three pixels at least are up or down  $Ip + \mathbf{thr}$ , for whole 16 pixels seek and

check if 12 neighboring pixels drop in the norm.

6. A same procedure can iterate for whole image's pixels.

### 3.3 Features Description using SURF

SURF (Speeded Up Robust Features) can be widely employed for problem solving of the correspondence matching due to it is faster than SIFT (Scale Invariant Feature Transform) by briefness the showing of matching. To find candidate points, SIFT uses visual pyramids and based on the law of Gauss filters each layer with raise values of Sigma and determines differences. For image identification and matching, the proposed algorithm employs SURF descriptor for feature. Vectors of feature are elicitation by SURF which is stable to image rotation and scaling. Features can be matched using Manhattan distance measure. Local descriptors of SURF are better computational efficiency than local descriptors of SIFT because of integral images computed in SURF. At discrete locations, points of interest are chosen in the image such as corners. Every key point's neighborhood is represented by a vector of feature. The descriptor of feature has to be discriminative, robust to noise, errors' detection, deformations of geometric and photometric. Finally the vectors of SURF descriptor are matched between various images. The matching is based on Manhattan dissimilarity. To build feature space, SURF algorithm consists of

various stages. These stages are detection of interest point, for each key point, SURF descriptor must be build, and descriptor matching [8].

**3.3.1 Constructing Integral image**

$$U(x,y) = \sum_{i \leq x, j \leq y} u(i,j) \quad \dots (13)$$

where  $u(i,j)$  represent the value of pixel at the locations  $i$  and  $j$  of the original image.  $U(x,y)$  represent the value of pixel at the locations  $x$  and  $y$  of the integral image.

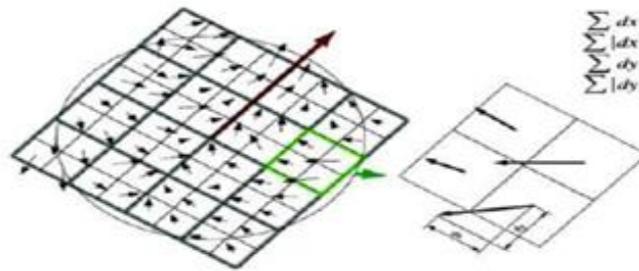
**3.3.2 Interest Point Detection**

Fast Hessian feature detector can be used in SURF .It is based on the determinant of Hessian matrix. Hessian matrix consists of partial derivatives of two dimensional functions. Our algorithm uses FAST corner for detection the interest point and can be used in applications of real time.

**3.3.3 Descriptor with SURF**

For SURF speed, integral images can be calculated. Image of integral is an intermediate representation and construct from the summation of image pixel values. It is also called as Summed Area Tables [10]. Integral image is given by Eq. (13).

Because SURF is stable to rotation, rotation can be processed by determining the direction of feature and rotating window's sampling to adjacency together with this angle. Build a quadrate area centered on the point of feature. Window volume can take about **20sX20s** from the discovered interest point, where  $s$  represents the volume. When the rotated nearness is finding, it is split into 16 sub quadrates as illustrate in figure (4). Again every sub quadrates can be divided into 4 quadrates



**Figure (4): schematic impersonation for SURF descriptor**

where  $dx$  and  $dy$  represent derivatives of  $x$  and  $y$  directions,  $|dx|$  and  $|dy|$  represent  $dx$  and  $dy$  normalization.

**3.3.4 Computation for Descriptor**

It computes Haar wavelet responses in horizontal and vertical directions for each

sub-region and summation of  $dx$ ,  $|dx|$ ,  $dy$ ,  $|dy|$  is formed and put in a vector  $V$ . For final squares, derivatives in the  $x$  and  $y$  directions are taken. The  $x$  derivatives summation over its four quadrants, similarly for  $y$  derivative is representing a descriptor for sub square. It has 4 values

for total descriptor. Normalize V to length 1 and feature's descriptor. A vector supplies the feature descriptor of SURF with aggregate 64 dimensions. Providing good discriminative to features for lower dimension with maximum computation's speed and matching [10].

### 3.4 Clustering of features using Moving k-means

Clustering means collect elements of data from a data set to clusters of several likeness norms. In Content Based Image Retrieval (CBIR) systems likeness among the database of images is not see just the likeness between database images and query image that is used for retrieval . In this trouble of imposing computation time which is existing because comparison of features of query image with all images features in database. Clustering is employed to decrease the time of computation which considers likeness among the database images. There is no requiring after clustering for query image comparison with all images in database

which decreases the time of computation and accuracy improvement. The algorithm of clustering for k-means is very straightforward for execution. The generations of clusters are not perfect in quality and time consuming for clusters generation is very imposing which is the problem in this way. An algorithm of clustering with moving k-means provides perfect time consuming and well clusters' quality to introduce cluster is less than the algorithm k-means [11]. An algorithm of clustering with moving k-means has been employed. The main steps of moving k-means clustering algorithm are illustrated bellow:-

- 1) The user has been specified the numbers of coveted clusters as input.
- 2) The set of factures that can be resulted from SURF descriptor are split randomly into number of coveted clusters based on Variance. The variance also tells something about the contrast.

$$\text{variance} = \sqrt{\sum_{g=0}^{L-1} ((g - \text{Mean})^2 * p(g))} \dots (14)$$

L :-is the total number of gray levels available , for example, for typical 8-bits image data l is 256 and range from 0 to 255.

g:-is the gray level value.

$$p(g) = N(g) / X * Y.$$

N (g):- is the number of pixels at gray level p.

P (g):- probability of each gray level value.

The midst point is treated as the cluster's centroid in every cluster. A distance among every data point to whole an initial centroid can be computed and a point of data is specified to cluster with aftermost centroid.

- 3) Through this, the cluster is enrolled to which data item can be specified and item distance of data for that cluster can be maintained. Cluster's centroids the can be recomputed.
- 4) From current after most cluster centroid ,compute a distance again for every data point if the distance is equal or lower than current nearest distance, a point of data in the selfsame cluster stays else data item distance is computed from whole the centroids.
- 5) The procedure continued until the convergence norm is not satisfied.

### 3.5 Building the Vocabulary

The total numbers of features which can be extracted from the descriptors are huge.

:

$$d = \sum_{i=0}^n |x_i - y_i| \quad \dots (15)$$

To solve this problem, the feature descriptors are clustered by applying the clustering algorithm, such as moving K-Means technique to generate a visual vocabulary. Each cluster can be treated as a distinct visual word in the vocabulary, which is represented by their respective cluster centers. The size of the vocabulary is determined using the clustering algorithm. In addition, it depends on the size and the types of the dataset [12]. The frequency of each word can be considered as the number of features in the corresponding cluster.

### 3.5 Feature Matching

Matching speed of feature is performed by a unique step of indexing for interest point which depends on the value of the Manhattan. Compute a distance that would be traveled through a Manhattan distance function to obtain with one point of data to another if a grid-like track is followed. The distance of Manhattan among two components is the summation of differences of their corresponding items [13].The distance's formula among a point  $X=(X1, X2, \text{ etc.})$  with a point  $Y=(Y1, Y2, \text{ etc.})$  is

### 3.6 Proposed algorithm

The algorithm of the proposed algorithm is illustrated as:

**Input :** video stream of images of letters, image of letter to be test and recognized

**Output :** name of letter in the tested image of letter of video frames

**Step1:** 1) Enter AVI video and Covert images of letters of video stream into frames (**Imgs**)

2)Enter the test image for matching ( **test** )

**Step 2:** Compute Haar transform for frames ((**Imgs**) and test image (**test**) and put the result in (**H\_imges**) and (**H\_test**) respectively using Eq.11.

**Step 3:** Detect the interest points for (**H\_imges**) and (**H\_test**) using FAST corner detection with adaptive threshold based on Eq.12 and put the results in (**D\_H\_Imgs**) and (**D\_H\_test**) respectively.

**Step 4:** Construct the integral images for both (**D\_H\_Imgs**) and (**D\_H\_test**) using Eq.(13).

**Step 5:** Apply the algorithm of moving k-means clustering for all features descriptor of video frames and test image by using Eq.(14).

**Step 6:** //To Create vocabulary

6.1) Each center of cluster is treated as a distinct visual word for video frames image can be called (**video\_visual\_words**) and the center of cluster for test can be called (**test\_visual\_words**).

6.2) build the vocabulary for all visual words from video (**video\_visual\_words**).

6.3) The frequency of each word can be considered as the number of features in the corresponding cluster.

**Step 7:** Track, match and recognize the letter features of the (**test\_visual\_word**) with the letters features of (**video\_visual\_words**) using Eq.(15) and recognize the name of the letter based on matching process.

### 4. Experimental results

The outcomes of suggested method are offered and discussed at this part. The suggested method is executed in C#. Three types of databases like ((Time New Roman, Arial and Calibri) video are employed for evaluation the suggested method. Database images are colored, and with size  $256 \times 256$  pixels. The suggested method consists from multiple steps:-

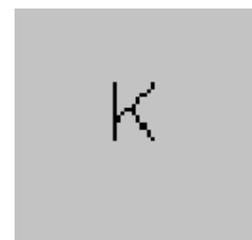
- 1) At the first step, loading the video stream and tested image as shown in the figure (5) for Time New Roman database and figure (12) for Arial database.



a)



b)



c)

**Figure (5): Time New Roman Database, a) Input Video of Images of Letters b) Extraction Frames from video c) Input Test Image of Letter from Time New Roman Database**

- 2) Second step exhibited initialization. During the initialization, it computes Haar transform on video frames and testing image as exhibited in figure (6) for Time New Roman database and figure (13) for Arial database.



**Figure(6): Haar transform on a) video frames b) test image**

- 3) The interest points are detected in the third step using FAST corner detection as exhibited in figure (7) for Time New Roman database and figure (14) for Arial database.



**Figure (7): FAST corner detection on a) video frames b) test image**

- 4) In the fourth step, complete SURF feature descriptor can be computed for the frames of video and tested image as exhibited in figure(8) for Time New Roman database and figure (15) for Arial database.



**Figure (8): SURF descriptor on a) video frames b) test image**

- 5) In the fifth step, clustering the features that are resulted from SURF descriptor, visual words can be built from each cluster's center with its histogram and then construct bag of visual words as exhibited at figure (9) for Time New Roman database and figure (16) for Arial database.

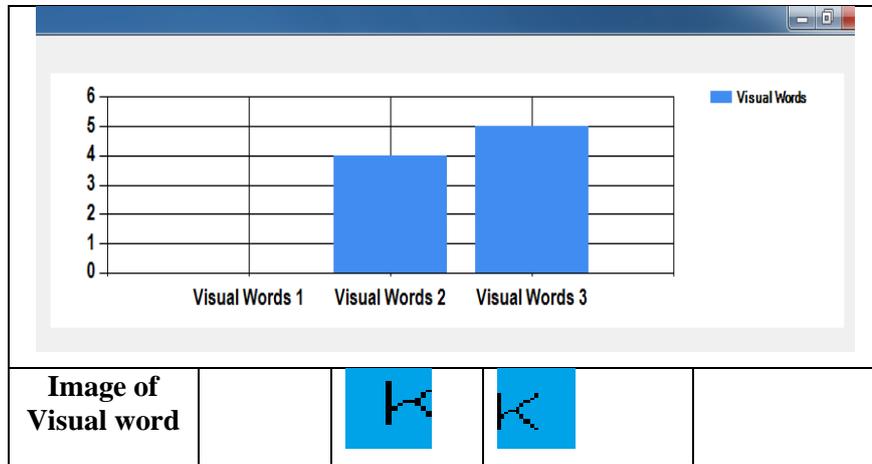


Figure (9):Histogram and Visual words for Letter K in Time New Roman Database the frequency of each word can be considered as the number of features in the corresponding cluster.

Image #	Word #
letter D	1
letter E	2
letter F	2
letter G	2
letter H	3
letter I	3
letter J	1
letter K	2,3
letter L	2,3

Figure (10):Vocabulary of all Images of Letters for Time New Roman Database and Corresponding Visual Words Numbers

- 6) In the sixth step, test image recognition with video frames to identify the letter name in vocabulary as exhibited in figure (11) for Time New Roman database and figure (18) for Arial database.

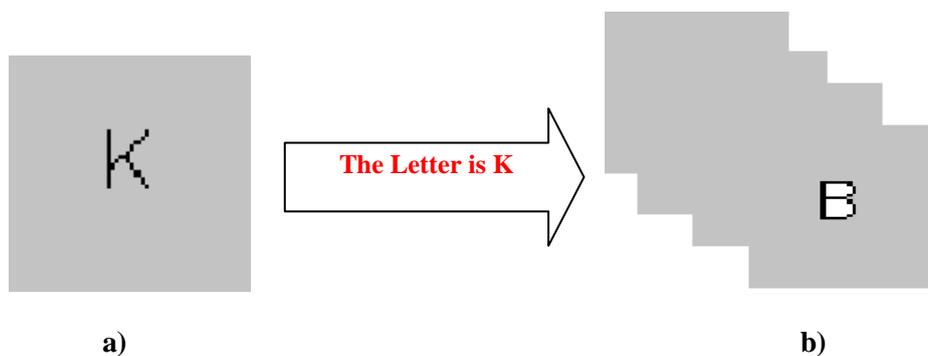


Figure (11): The matching result of the proposed system between a) test image b) video frames

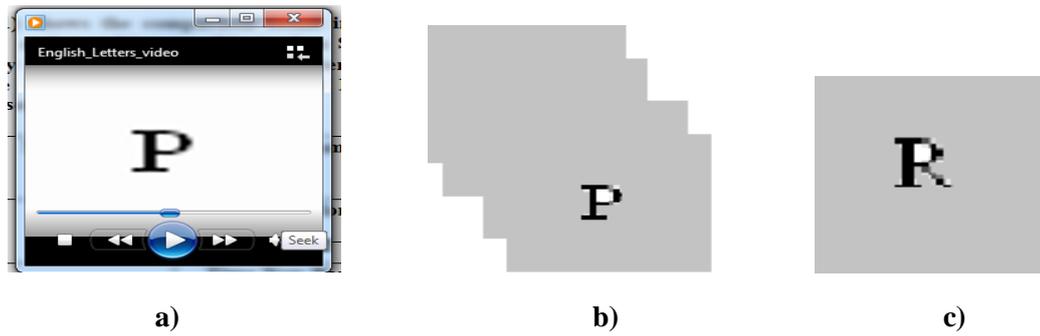


Figure (12): Arial Black Database, a) Input Video of Images of Letters b) Extraction Frames from vidro c) Input Test Image of Letter from Arial Black Database

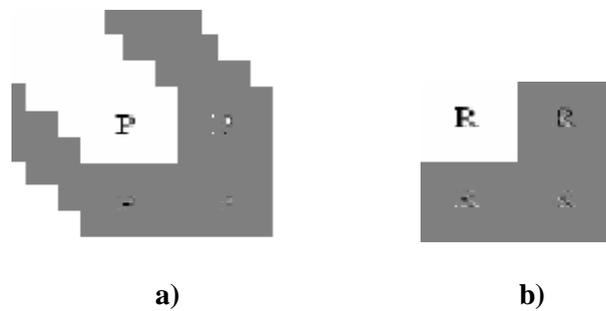


Figure (13): Haar transform on a) video frame b) test image

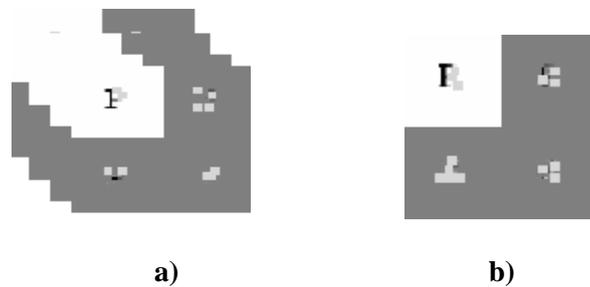


Figure (14): FAST corner detection on a) video frames b) test image

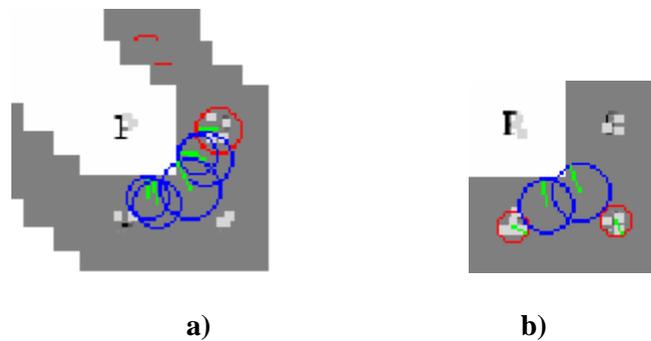


Figure (15): SURF descriptor on a) video frames b) test image

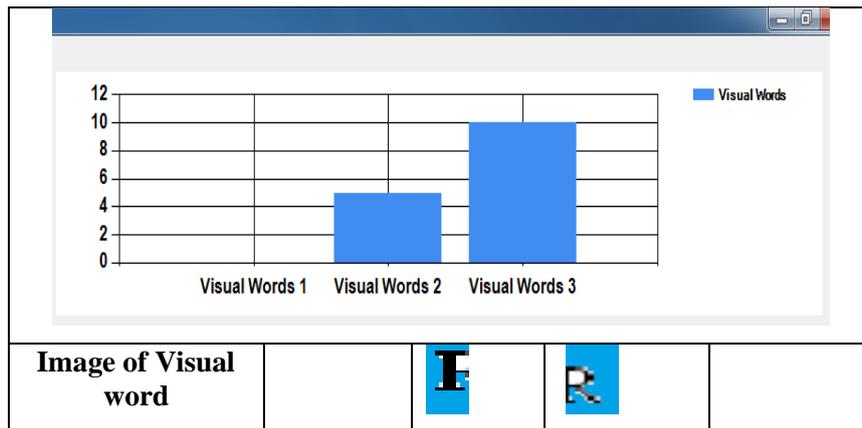


Figure (16):Histogram and Visual words for Letter R in Arial Black Database the frequency of each word can be considered as the number of features in the corresponding cluster.

Image #	Word #
letter O	2,3
letter P	2,3
letter Q	2,3
letter R	2,3
letter S	2,3
letter T	2
letter U	2,3
letter V	2,3
letter W	2,3

Figure (17):Vocabulary of all Images of Letters for Arial Black Database and Corspnding Visual Words Numbers

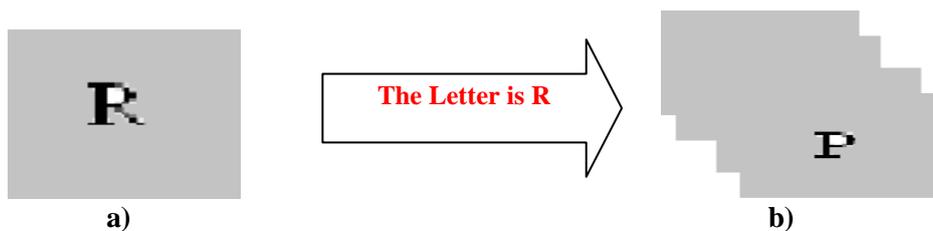


Figure (18): The matching result of the proposed system between a) test image b) video frames

### 5. Proposed methodology analysis

In this section, it can analysis the proposed methodology based on consuming time for complete the recognition process and accuracy which has been depended on computing the precision performance that can be used for the correct recognition. The accuracy can be computed using equation (16).

$$precision = \frac{True\ positivities}{True\ positivities + False\ positivities} \dots (16)$$

True Positives is the number of the images that are correctly retrieved from the image datasets, While False Positives is the number of images that are incorrectly recognized from the image database [12].

Table (1) shows the comparison result in term of time consuming and accuracy for recognition an object when Seven Moments and BOVW were employed. Accuracy recognition depends on number of correct matching between test image and database images. For example, letter K and R could be selected to display the outcome of comparison.

Method name	Object name	Time(sec) for recognition the letter	Accuracy recognition on letter
Seven Moments	Letter (K) in Time New Roman	2.991	83%
	Letter (R) in Arial Black	2.371	81%
BOVW	Letter (K) in Time New Roman	0.025	90%
	Letter (R) in Arial Black	0.011	93%

### 6. Conclusions

English letters recognition can be depended on multiple measurements. The suggested measurements method can be employed bag of visual word for recognition process. Bag of visual word is an efficient method for representation an image in the classification and recognition tasks. The suggested method deals with capital letters only and it was tested on three types of English letters font's databases (Time New Roman, Arial Black and Calibri). Experimental outcomes show that the suggested method is more efficient and fast for matching and recognizing a letter than recognition character using seven moments method. The time that BOVW could be taken for recognition a letter is less than the time that seven

moments could be taken, the BOVW accuracy depends on number of correct character recognition. BOVW have optimal accuracy in the process of recognition of letters.

### References

[1] Mandana Hamidi, Ali Borji, "Invariance analysis of modified C2 features: case study hand written digit recognition" , Machine Vision and Application, Vol.21, No.6, pp. 969-979, 2010.

[2] Nibaran Das, Ram Sarkar, Subhadip Basu, Mahantapas Kundu, Mita Nasipuri, Dipak Kumar Basu," A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application", Applied

- Soft Computing, Vol.12, No.5, pp.1592-1606, May 2012.
- [3] Nusrat Jahan, Shahrul Nizam Yaakob and Phaklen Ehkan, "**Feature Extraction for Neural Network Pattern Recognition for Bloodstain Image Analysis**", School of Computer and Communication Engineering, Malaysia, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 15 ) pp. 8583-8589 ,2016.
- [4] Muralidharan R., Dr. Chandrasekar C.," **SCALE INVARIANT FEATURE EXTRACTION FOR IDENTIFYING AN OBJECT IN THE IMAGE USING MOMENT INVARIANTS** ",Journal of Engineering Research and Studies, E-ISSN 0976-7916, JERS/Vol.II/ Issue I/ pp. 99-103, January-March 2011.
- [5] David Salomon, "**Data Compression**", the Complete Reference. Fourth Edition, Professor David Salomon (emeritus) Computer Science Department, California State University, Northridge, CA 91330 8281, USA, Springer-Verlag London Limited, 2007.
- [6] Wenjia Yang, Lihua Dou, Juan Zhang, Jinghua Lu, "**Automatic Moving Object Detection and Tracking in Video Sequences**", *SPIE Fifth International Symposium on Multispectral Image Processing and Pattern Recognition*, pp.676-712, 2007.
- [7] Asif Masood, M. Sarfraz, "**Corner detection by sliding rectangles along planar curves**", *Computers & Graphics*, Vol. 31, pp.440-448, 2007.
- [8] Jasmine J. Anitha and Deepa S.M., "**Tracking and Recognition of Objects using SURF Descriptor and Harris Corner Detection**", *ÀNehru Institute of Engineering and Technology (Anna University) ,Coimbatore, India, Vol.4, No.2,pp.1-6, 2014.*
- [9] Bay H., Ess A., Tuytelaars T., and L., "**Speeded-up robust features (SURF)**", *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
- [10] Aswini C. and Chitra D., "**Enhanced Logo Matching and Recognition using SURF Descriptor**", Department of Computer Science and Engineering, P. A. College of Engineering and Technology, Pollachi, Tamil Nadu, India, 2014.
- [11] Shefalli and Balkrishan Jindal, "**Enhancing Content-based Image Retrieval using Moving K-Means Clustering Algorithm** ", *International Journal of Computer Applications (0975 – 8887) Volume 102 – No.9, September 2014.*
- [12] Mohammed Elmogy and Hazem Elbakry," **Content-Based Image Retrieval using Local Features**

**Descriptors and Bag-of-Visual Words** ",Dept. of Information Technology, Faculty of Computers and Information, Mansoura University, Mansoura, Egypt , (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 9,pp.212-219, 2015.

- [13] Jiawei Han, and Micheline Kamber , **"Data mining, concepts and techniques"**, Third Edition book, Morgan Kaufmann Publishers is an imprint of Elsevier.225 Wyman Street, Waltham, MA 02451, USA, 2012.

## A New Multiple Blind Signatures Using El-Gamal Scheme

*Muthanna Abdulwahed Khudhair*

*Department of computer science, Dijlla University College,*

*Iraq-Baghdad*

**Tel:-009647706253593**

**E-mail:- muthanna.khudhair@duc.edu.iq**

### **Abstract**

This paper presents a new secure blind signature. In order to secure the blind signature, this scheme generates two blind signatures. Each of the blind signatures has its own blind factors. In addition, the author of the message encrypts the information sent to the signing author by an encryption key generated by El-Gamal cryptosystem, which is considered a strong cryptographic key. This key adds a secure and hidden layer to the message being signed. In addition, the author of the message binds an agreement factor for each type of message and a context that characterized the type of message signed. This method provides the most important properties for a secure blind signature.

**Keywords:** Blind Signature, Message Context, Agreement Factors, El-Gamal cryptosystem.

### **الخلاصة:-**

تقدم هذه الورقة البحثية طريقة امنية للسرية للتوقيع الاعمى . من اجل تأمين امنية اكثر للتوقيع الاعمى ، تولد هذه الطريقة توقيعين اعمى . يمتلك كل نوع من التوقيع الرقمي معاملات عمياء. اضافة الى هذا ، يقوم مؤلف (مالك) الرسالة بتشفير المعلومات المرسله الى مؤلف توقيع وذلك بواسطة مفتاح تشفيرى متولد بنظام ال-El-Gamal والذي يعتبر مفتاح تشفيرى قوي . سيضيف هذا المفتاح طبقة امنية ومخفية للرسالة المطلوب توقيعها . علاوة على ذلك يربط مؤلف الرسالة معامل اتفاق لكل نوع من الرسائل ومحتوى يميز نوع الرسالة المطلوب توقيعها . توفر هذه الطريقة معظم الخصائص المطلوبه لتوقيع اعمى امن .

## 1. Introduction

David Chaum introduced the concept of blind signatures. The aim is to generate an electronic way of money transfer such that e-coin cannot be easily traced from the center of the bank to the customer. In such a way, any two spending of the same user have no ability to linked together. Chum explained that when applying blind signatures we could get two important properties; namely intractability and unlink ability. In the blind signature the signer does not know anything about the contents of any sent message. One important and derivate of digital signatures is a blind signature. We can generate blind signature by adding some other properties to any type of digital signature. The concept of a blind signature is that it is considered a protocol for generating a signature  $s$  for a message  $m$  from a certain signer  $SG$  such that  $SG$  does not know anything about  $s$  and  $m$ . So the contents of the message are hidden from  $SG$ . The operation of this protocol is that any user  $U$  can select a random number  $k$  and merge this  $r$  into  $m$  to produce  $m'$ .  $m'$  is sent to  $SG$  who will generate a signature  $s'$

corresponding to  $m'$ .  $SG$  returns back  $s'$  to the user  $U$ . The user  $U$  will remove the blind factor to get  $s$  which is the signature of  $m$ . So according to this procedure, both  $m$  and  $s$  are hidden from  $SG$ . The most important property of blind signature is that it unforgeable. Blind signatures are used in many sensitive applications such as cash protocol to provide a strong protection for the privacy of customers. [1]

Blind signature scheme consists of three parts. The first part is key generation which is a probabilistic polynomial time. The second part is blind signature generation. This generation is an interactive protocol between the signer  $SG$  and the user  $U$ . The last part is blind signature verification which is a deterministic polynomial time algorithm. [1,2]

Blind signatures use the traditional public key cryptography in which each user has two different cryptographic keys, a private and a public key. This needs to simplify the key management, Shamir [3] invented the idea of identity-based public key cryptography. In this scheme each user must

identify himself/herself and we need a way to facilitate the users to register them at key generator centre (**KGC**).

There are different proposed blind signatures schemes [4–9]. The blind signature schemes are useful especially when we need to provide an anonymity. Such applications are sensitive, for example the online voting systems and the electronic cash systems [8].

El-Gamal is a cryptographic scheme used to provide a signature. If we compare it with symmetric algorithms in terms of encryption and decryption speed, we find that ElGamal is relatively slower [10-12]. This scheme is one type of a non-deterministic public key cryptography. This scheme has different signatures for the same message because it chooses different random factors. The generated signature allows message recovery and so it has many advantages [13-18]. One advantage is that it generates a shorter signature corresponding to shorter plaintexts. ElGamal scheme combines the plaintext with the validation of the signature [4]. An improvement to ElGamal mode had been suggested by Nyberg and

Rueppel. In this modification it is possible to receive a series of signature schemes. These new schemes can provide a verification to the signature at the time of performing a message recovery [17-19].

El-Gamal signature scheme was first introduced by Taher Elgamal in 1984. This scheme can generate a signature based on the difficulty of computing the discrete logarithms so it is considered a type of public key cryptosystem. There are two attacks that may occur against El-Gamal scheme. These attacks are low modulus attack and plaintext attack [20]. The first attack can arise when we use values of modulus that are low while plaintext attack is applicable when the enemy discovers the plaintext which is corresponding to its ciphertext and in this case it is easy to find the cryptographic key.

In order to provide security for digital data transmitted through communication channels, the best and effective way is by using cryptography system. Cryptography is fundamental for securing and protecting private information and other

applications of most organization [21].

Cryptography is defined simply as a study for mathematical techniques. The aim of cryptography is providing different services such as confidentiality, authentication, data integrity and non-repudiation [22].

The early version of ElGamal scheme depends on applying Diffe-Hellman which is used for key exchange [23,24]

## **2. Related works**

In [25] a new scheme of blind signature was proposed. This scheme depends on ElGamal method. In this scheme, when a message is signed different multiple times, the generated corresponding signature will still be the same. This property is an important modification for blind signature in that it provides anonymity to the signature. In order to achieve this goal, the proposed scheme uses both number theory and modular arithmetic techniques. The result of this scheme shows that it is faster than the compared blind signatures of RSA.

A paper presented the usage of blind signature to design an electronic voting algorithm using ElGamal signature. This research is based on XML to analyze the security of such scheme. The result of this method shows that it provides a high level of secrecy and can be implemented practically [26].

Using the properties of blind signatures especially blindness and intractability, a paper proposed two blind signatures based which are untraceable. These blind signatures are based on the difficulty of solving the factoring problem. These types of proposed blind signature are different in traditional blind signatures in that the later schemes are based on the difficulty of solving factoring problem and quadratic residues. The signatures in this paper can fully provide all properties of the standard blind signature [27].

One of the most important applications of a blind signature is the electronic voting. Most blind signatures use an elliptic curve algorithm which is characterized by the difficulty of solving such an algorithm. This paper presents a new scheme for implementing an electronic

voting method in such a way that the elliptic curve algorithm is combined with the blind factor. The aim of this procedure is to scramble the message's content and then it is signed. So, the signer of the message does not what the content of the message is. This scheme is also provide a way to let the voter to vote and authenticate himself/herself [28].

### **3. Proposed system**

This system generates a new blind signature using El-Gamal Scheme. There are two players used to cooperate for generation the blind signature. The first one is the author of the message named as **AM** and the other player is the signing authority **SA**. Each chooses some number. **AM** chooses a number **r** and **SA** chooses a number **S**. El-Gamal scheme is a public key cryptosystem which is useful to solve the problem of key exchange so it used to bypass the possibility of an intercepted key. The proposed system generates two blind signatures instead of one signature as applied in the

traditional blind signature. The players in this system must

generate two values. Each value generated by one player is sent to the other player to compute an encryption key. After that the author of the message selects two random numbers **K<sub>1</sub>** and **K<sub>2</sub>** as initial parameters for generating the two blind signatures. In this step the author of the message uses his/her public key and sends the results to the signing authority who is responsible for producing the blind signatures. The result must include the context of the message and an agreement factor for each calculated result. So, the signing authority must examine the context of each message and the corresponding agreement factor. The context (**c**) and their agreements factors (**ag**) for each user (**u**) are illustrated in table 1. These two blind signatures are sent back to the author of the message who can then remove the blinding factors to reveal the actual signatures sent from signing authority.

Users	Message Context	Agreement
$u_1$	$\{u_1c_1, u_1c_2, \dots, u_1c_{n_1}\}$	$\{u_1c_1ag_1, u_1c_2ag_2, \dots, u_1c_{n_1}ag_{n_1}\}$
$u_2$	$\{u_2c_1, u_2c_2, \dots, u_2c_{n_2}\}$	$\{u_2c_1ag_1, u_2c_2ag_2, \dots, u_2c_{n_2}ag_{n_2}\}$
• • • $u_m$	$\{u_m c_1, u_m c_2, \dots, u_m c_{n_m}\}$	$\{u_m c_1 ag_1, u_m c_2 ag_2, \dots, u_m c_{n_m} ag_{n_m}\}$

The Algorithm

Let  $m$  be the message .

Let  $a$  be the public base.

Let  $N$  be the module.

Let  $U$  be as set of users ;  $U = \{u_1, u_2, \dots, u_n\}$ .

Let  $C$  be a set of message contexts ;  $C = \{c_1, c_2, \dots, c_m\}$  .

Let  $AG$  be a set of agreement value ;  $AG = \{ag_1, ag_2, \dots, ag_k\}$  .

Let  $e$  is a public key .

Let  $d$  is the private key .

A: Encryption Key Generation:

1: **AM** chooses some number  $r$ .

2: **SA** chooses another some number  $s$ .

3: **AM** computes  $a^s \text{ mod } N$  and sends it to **SA**.

4: **SA** computes  $a^r \text{ mod } N$  and sends it to **AM**.

5: **AM** computes the encryption key ( $e_K$ ) by taking **SA's** number  $s$  as following and sends it to **SA**:

$$K = s^r \pmod{N}$$

**B:** Generating the Blind Signatures:

1: **AM** chooses two random numbers  $k_1$  and  $k_2$  .

2: **AM** generates two encrypted messages using his/her public key ( $e$ ) as following:

$$A: m_1 \equiv m(k_1)^e \pmod{N}$$

$$B: m_{11} \equiv (m_1) * e_K \pmod{N} \parallel C_i \parallel AG_i$$

$$C: m_2 \equiv m(k_2)^e \pmod{N}$$

$$D: m_{22} \equiv (m_2) * e_K \pmod{N} \parallel C_i \parallel AG_i$$

3: **AM** sends both  $m_{11}$  and  $m_{22}$  to **SA** .

**C:** **SA** performs the following steps (generating the two blind signatures  $s'_1$  and  $s'_2$  respectively):

1: **SA** takes  $m_{11}$  and  $m_{22}$  and he/she generates two blind signatures as following:

2: **SA** extracts  $C_i$  from both  $m_{11}$  and  $m_{22}$  and search for the corresponding agreement factor ( $AG_i$ ) in the table 1.

If the  $AG_i$  is found and authenticated then **SA** generates  $s'_1$  and  $s'_2$  , otherwise **AM** is considered an unauthenticated party.

3: If  $AG_i$  matches the its context  $C_i$  then :

$$S'_1 \equiv (m_{11})^d \pmod{N}$$

$$S'_2 \equiv (m_{22})^d \pmod{N}$$

4: **SA** sends both  $s'_1$  and  $s'_2$  to **AM** .

Step 4: **AM** can assure the validity of these signatures by removing the blind factors  $k_1$  and  $k_2$  as following:

$$s_1 \equiv s'_1 \cdot k_1^{-1} \pmod{N}$$

$$s_2 \equiv s'_2 \cdot k_2^{-1} \pmod{N}$$

#### 4. Results

Suppose **base** =7 , N =71 and **M**=20.

**AM** chooses **r**=9 and **SA** chooses **s**=11.

1: **AM** computes a value of **base** to the power of **r** :

$$B = \text{base}^r = 7^9 \pmod{71} = 47.$$

2: This value is sent to **SA**.

3: **SA** computes:

$$A = \text{base}^s \pmod{71} = 7^{11} \pmod{71} = 31.$$

4: This value is sent to **AM**.

5: Generating the encryption key:

A: **AM** takes the value created by **SA** as following:

$$K_e = A^s \pmod{N}.$$

$$K_e = 31^{11} \pmod{71} = 52.$$

B: Encryption is performed as following: (the ciphertext here is denoted by **C**)

$$C = (K_e * m) \pmod{N}$$

$$C = (52 * 20) \pmod{71} = 46. \text{ This } C \text{ is sent to } SA$$

In order to perform decryption, **SA** finds  $52^{-1} \pmod{71} = 56$

Then he calculates **M** as following :

$$M = 56 \times 46 \pmod{71} = 20$$

Step2 (Generating the blind signatures):

We need here to select two prime numbers **p** and **q** to generate the public key **e** and the corresponding private key , **d** using the RSA scheme.

Let  $p=11$  and  $q=7$ .

$$N_1 = p \times q = 11 \times 7 = 77.$$

$$\phi(N_1) = (p-1)(q-1) = (11-1)(7-1) = 60.$$

Let  $e=9$

$$\text{Then } d = e^{-1} \pmod{60} = 9^{-1} \pmod{60} = 50$$

**AM** selects two random numbers  $k_1=3$  and  $k_2=7$  such that  $\text{GCD}(K_1, N)=1$  and  $\text{GCD}(K_2, N)=1$ .

$$\begin{aligned} M_1 &\equiv m (k_1)^e \pmod{N} \\ &\equiv 20(3)^{13} \pmod{71} = 5 \end{aligned}$$

Let  $G_1=2$  and  $AG_1=4$

$$\begin{aligned} M_{11} &\equiv (m_1) * (e_k) \pmod{N} \parallel G_1 \parallel AG_1 \\ &\equiv 5 * 52 \pmod{71} = 47 \parallel 2 \parallel 4 \end{aligned}$$

$$\begin{aligned} M_2 &\equiv m(k_2)^e \pmod{N} \\ &\equiv 2(7)^{13} \pmod{71} = 63 \end{aligned}$$

Let  $G_2=3$  and  $AG_2=7$

$$\begin{aligned} M_{22} &\equiv m_2(e_k) \pmod{N} \parallel G_2 \parallel AG_2 \\ &\equiv 63(52) \pmod{71} = 10 \parallel 3 \parallel 7 \end{aligned}$$

Generating the two blind signatures  $S_1'$  and  $S_2'$  by the **SA** after extracting  $G_1$ ,  $G_2$ ,  $AG_1$  and  $AG_2$ :

$$\begin{aligned} S_1' &\equiv (M_{11})^d \pmod{N} \\ &\equiv 47^{37} \pmod{71} = 63 \end{aligned}$$

$$\begin{aligned} S_2' &\equiv (M_{22})^d \pmod{N} \\ &\equiv 10^{37} \pmod{71} = \end{aligned}$$

STEP 3 : Recovering the signatures by the **AM**:

$$S_1 \equiv S_1' \cdot K_1^{-1} \pmod{N}$$

$$\equiv 63. 3^{-1} \text{ mod } 71 = 21$$

$$S_2 \equiv S_2' . K_2^{-1} \text{ (mod } N)$$

$$\equiv 29 . 7^{-1} \text{ mod } 71 = 65$$

## **5. Conclusions**

This paper presents a new blind signature based on El-Gamal scheme. The importance of this scheme is that it generates two blind signatures by using two parameters for the same message. Each content generated by step the author of the message is encrypted by an encrypted key. This key is come from the ElGamal scheme. So this is the first modification that enforces the security of the message and increases the blindness of the

message. Another important support of this new scheme is that the author of the message desires to sign his/her message depending on its context. So the signer must check the context of each message sent. Also a corresponding agreement factor is attached to its context. So this method provides most properties of blind signature and it increases the degree of hiding for the message which is a desirable and strong property of perfect blind signatures.

## **References**

[1] D. Chaum, "Blind signatures for untraceable payments", *Advances in Cryptology, Crypto'82*, pp.199-203, 1982.

[2] S. Han and E. Chang, "A pairing-based blind signature scheme with message recovery", Ardil, C. (ed), *Sixth International Enformatika Conference (IEC)*, pp. 303-308, 2005.

[3] A. Shamir, "Identity-based cryptosystems and signature schemes", 1984 *International cryptology conference on advances in cryptology*, 1984. p. 47-53.

[4] C. Fan, L. Wu, and V. Huang, "Cryptanalysis on Chen-Qiu-Zheng blind signature scheme", *Appl Math Sci* 2008;2(16):787-91.

[5] N. Moldovyan, "Blind collective signature scheme based on discrete logarithm

- problem", *Int J NetwSecur* 2010;11(2):106–13.
- [6] D .Pointecheval and J.Stern Security, "Aguments for digital signatures and blind signatures", *Cryptol2000*;13:361–96.
- [7] A .Tripathy, I .Parta, and D. Jena," Proxy blind signature based on ECDLP", *Int J Comp NetwSecur* 2010;2(6):93–8.
- [8] C.Popescu, "Blind signature schemes based on the elliptic curve discrete logarithm problem", *Stud Inform Control* 2010;19(4):397–402.
- [9] N.Moldovyan, "Blind signature schemes from digital signature standards", *Int J NetwSecur* 2011;12(3):202–10.
- [10] C. Zhi-ming, "An improved encryption algorithm on ELGamal algorithm[J]", *Computer Applications and Software*, 2005, 22 (2): 82-85.
- [11] L.Wang, W.Xing, and Z.XuGuang , "ElGamal public key cryptosystem based on integral quaternions [J]", *Computer Applications*, 2008,28(5):1156-1157.
- [12] H.Lu and Y.Sun," A Public-key Cryptography Using Integral Quaternions[J]", *Journal of Tong Ji University*, 2003, 31(12).
- [13] H. Zhen-jie, W.Yu-min, and C. Ke-fei, "Generalization and improvement of Nyberg-Rueppel message recovery blind signatures[J] " *Journal on Communications*, 2005 , 26(12): 131-135.
- [14] C. Hui-yan, L. Shu-wang, and L.Zhen-hua , " Identity Based Signature Scheme with Partial Message Recovery [J]", *Chinese Journal of Computers*, 2006, 29 (9) : 1622- 1627 .
- [15] C. Tian-jie and L. Dong-dai , " Security analysis of a signature scheme with message recovery[J]", *Journal of Zhejiang University(Science Edition)* , 2006,33 (4) :396-397
- [16] K. Yuan-ping," A Signature Scheme with Message Recovery Based on Elliptic Curves[J]", *Computer engineering and science*, 2010, 32(2):58-59.
- [17] K.Yberg, and R.A Rueppel, "message recovery for signature schemes based on the discrete logarithm problem," in *EUROCRYPT*,1994, 182-193.
- [18] W. Qing-ju, K. Bao-yuan and H. Jin-guang , " Several New ELGamal Type Digital Signature

Schemes and Their Enhanced Schemes[J]", Journal of East China Jiaotong University, 2005, 22(5): 127-138 .

[19] H.Zhang and J.Zhang," Research and Design of an Improved ELGamal Digital Signature Scheme[J], Computer Engineering and Science, 2009, 31(12): 35-38.

[20] S.Rashmi and K.Shiv,"Elgamal's Algorithm in Cryptography", International Journal of Scientific & Engineering Research,2012, Volume 3

[21] M.Allam,' Security and Performance of ElGamal Encryption Parameter," Journal of Applied Sciences 5, Asian Network for Scientific Information, 883-886, 2005.

[22] J .Alfred and C .P. Oorschot and A .Scot, "Handbook of AppliedCryptography", CRC Press, 1997.

23. T. ElGamal," A public key cryptosystem and a signature scheme based on discrete logarithms", In G. R. Blakley and D. Chaum, editors, Advances in Cryptology – Proceedings of CRYPTO 84, volume 196 of Lecture Notes in Computer

Science, pages 10–18. Springer, 1985.

24. T.ElGamal," A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE Transactions on Information Theory, 31(4):469–472, 1985. 9.

[25] E. Mohammad, A.Emarah and K.Shennaway," A blind signature scheme based on ElGamal signature", Radio Science Conference, 2000.17th NRSC '2000. Seventeenth National.

[26] F. Song. And Z. Cui ," Electronic Voting Scheme About ElGamal Blind-signatures Based on XML", 2012 International Workshop on Information and Electronics Engineering (IWIEE) , 2012.

[27] C.Chi Lee , W.Pang Yang , and M.Shiang Hwang , " Untraceable blind signature schemes based on discrete logarithm problem ", Fundamenta Informaticae. 2002; 55(3):307-20

[28] M.Kuthe, and J.Avinash ," Implementation of Blind Digital Signature Using ECC ," Intern. J. of Computer Science and Network (IJCSN).2012; 1(5).

## Proposed Encryption and Key Generation Method Based on Geffe Generator, Genetic Algorithm, and DNA Coding

*Asst. Prof. Dr. soukaena hassan hashem*

*Computer sciences/university of technology: Soukaena.hassan@yahoo.com*

### Abstract

After the evolution in the digital world, it has become difficult to the preservation of information in the channels of communication; for this reason using of encryption methods is employed to provide security and to ensure access of information properly to the authorized persons and to ensure that the information is not manipulated and changed by others. Encryption methods are divided into two types one type is depended on using two keys (Public and Private key) while the other type is depended on using one secret key founded in both sides (Sender and Receiver). The encryption methods which depend on using two keys (Public and Private keys) sometimes are not secure because the predication of secret key is possible by using special calculations on the public key, the unauthorized person do this calculation .For this reason the methods which depend on private key are used where only authorized persons to send and receive the information have the private key. In this paper a proposed encryption method is presented based on using the Geffe Generator, Genetic algorithm, S-box and DNA coding. Since these will be used to generate special key. In the proposed method present three types of keys will be used in the encryption which are (Specified, Generated, and Static) which make it difficult to know the real key and the real secret text characters.

**Keywords:** Encryption, Geffe Generator, Genetic Algorithm, DNA coding.

**طريقه تشفير وتوليد مفتاح مقترحه تعتمد على المولد geffe والخوارزميات الجينية والترميز DNA**

### الخلاصة

بعد التطور الحاصل في العالم الرقمي اصبح من الصعب المحافظة على المعلومات في قنوات الاتصال لذلك يتم اللجوء الى استخدام طرق التشفير لتوفير الامنية لها وضمان وصول المعلومات بشكل سليم الى الاشخاص المخولين وضمان عدم التلاعب بها وتغييرها من قبل الاشخاص الغير مخولين. طرق التشفير تنقسم الى قسمين قسم يعتمد على مفاتيح (مفتاح عام ومفتاح خاص ) وقسم يعتمد على مفتاح واحد سري يكون لدى كل من الطرفين (المرسل والمستلم). ان طرق التشفير التي تعتمد على مفاتيح مفتاح عام ومفتاح خاص بعض الاحيان تكون غير امنة وذلك لانه من الممكن التتبا بالمفتاح الخاص عن طريق حسابات للمفتاح العام يقوم بها الاشخاص الغير مخولين. لذلك يفضل استخدام طرق المفتاح الخاص حيث ان المفتاح الخاص يوجد فقط لدى الاشخاص المخولين لارسال واستلام المعلومات. في هذا البحث تم اقتراح طريقة تشفير بالاعتماد على مولد ال Geffe والخوارزمية الجينية وال S-box وترميز ال DNA. وبما أن هذه سوف تستخدم لتوليد مفتاح خاص. في الطريقة المقترحة سيتم استخدام ثلاثة أنواع من المفاتيح في التشفير والتي هي (محددة، مولده، وساكنة) مما يجعل من الصعب معرفة المفتاح الحقيقي والحروف النص السري الحقيقي.

*الكلمات المفتاحية: تشفير والمولد geffe والخوارزميات الجينية و والترميز بال DNA.*

## 1. Introduction

The rapid growth of computer networks allows large files, such as text, voice and video, to be easily transmitted over the Internet, and it is important to protect confidential data from unauthorized access [1]. Information security is a very important topic in data transfer. Any loss or threat to the transfer of information can be a big loss in the process of sending information. Encryption technology plays a key role in information security systems [2]. Encryption has been used primarily to prevent the disclosure of confidential information, but it can also be used to provide credibility of the message, check the integrity of incoming data, provide a digital equivalent of handwritten signature, and nonrepudiation. Non-denial confirms that the party deals cannot deny that the transaction is taken place. Encryption is the name for the study of procedures and algorithms and methods to encrypt and decrypt information is encrypted, where, cryptanalysis is to study ways and means to defeat or encryption techniques and compromise. With some encryption methods, (and this paper used the same idea with the amendment to use the same key to encrypt and decrypt information is encrypted (multiple secret keys)), this type of encryption is known as symmetric encryption, which is also known as one of the key or secret encryption key format. Another

encryption is that it uses two keys: one key to encrypt and decrypt a different key. These systems are referred to as non-symmetric encryption, also referred to as public-private encryption, which is also necessary because one key is known to all and the other is kept secret [3].

## 2. Linear Feedback Shift Register (LFSR)

Stream cyphers are type of Symmetric Key cryptosystems that encrypts plaintext one bit/byte at a time. Pseudorandom number sequences (PNS) are sequences whose properties approximate the properties of sequences of random numbers. These are not truly random, because it is completely determined by a relatively small set of initial values. LFSRs, see figure (1), are used in a stream cypher to generate linear sequences of pseudorandom numbers. They require very less hardware and have high speed of operations. An  $n$ -stage LFSR is of maximum length if initial states repeat after every  $(2^n - 1)$  bits. The contents of the registers are moved by one position at each clock. The left-most bit fed to the register is the result of mod-2 addition of bits corresponding to the nonzero coefficients of considered primitive polynomial. The right most bit is used to form the pseudorandom number sequence. All initial states should not be "0"s because the LFSR would remain locked-up in these states [4].

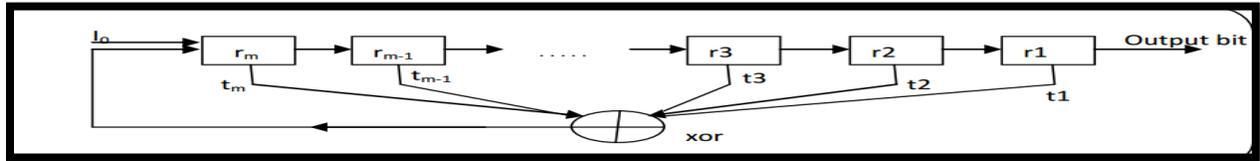


Figure (1): M-sequence Generation (LFSR) [5]

### 3. Geffe Generator

The Geffe generator [6] is defined by three maximum-length LFSRs whose lengths  $r_1, r_2, r_3$  are pair wise relatively prime, with nonlinear combining function, see equation (1).

$$F_3(x_1, x_2, x_3) = x_1 * x_2 \oplus (1 \oplus x_2) * x_3 = x_1 * x_2 \oplus x_2 * x_3 \oplus x_3 \dots\dots (1).$$

Figure (2) represents the Geffe generator.

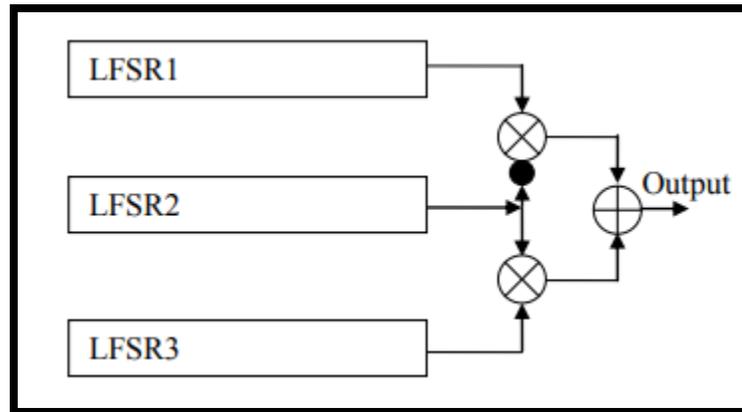


Figure (2): Geffe generator.

The keystream generated has period  $(2^{r_1} - 1)(2^{r_2} - 1)(2^{r_3} - 1)$  and linear complexity, see equation (2).

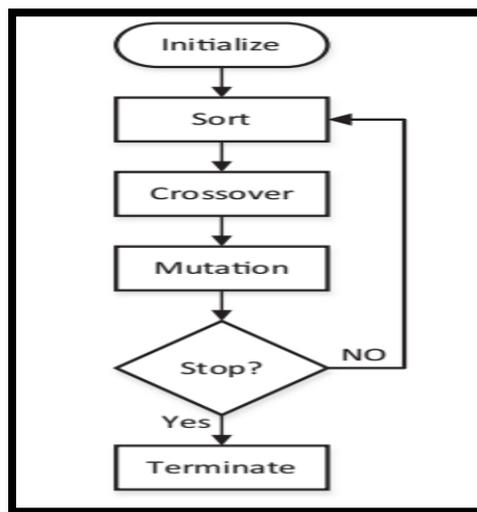
$$LC = r_1 r_2 + r_2 r_3 + r_3 \dots\dots (2).$$

The Geffe generator is cryptographically weak because information about the states of LFSR1 and LFSR3 leaks into the output sequence. Despite having high period and moderately high linear complexity, the Geffe generator succumbs to correlation attacks [6].

#### 4. Genetic algorithm

Genetic Algorithm, similar to many other meta-heuristics, is an evolutionary population based algorithm. That is to say, a population of answers will evolve through the course of the optimization to move toward the optimality of a problem. Answers or individuals in GA are presented in chromosomes, which, incidentally, are the

very strong suit of the algorithm. A chromosome is in fact an answer to the problem which is encoded to form a chromosome. The most prevalently applied chromosome is the binary chromosome. Each GA, consequently, needs to have a decoding function with the purpose of converting chromosome encoding to answers. Figure (3) presents a general flowchart of the algorithm.



**Figure (3): A general flowchart of the Genetic algorithm**

GA will initialize by randomly producing chromosomes as many as the number of populations. In the case of binary chromosome, the cells of the chromosome will be filled with 0 or 1 by the same chance. Each and every chromosome will be decoded to an answer and consequently, their fitness value will be calculated. Fitness value, by definition, is the goodness of the answer according to the problem. Next, the population will be sorted based on the individuals' fitness value. Crossover and mutation are the two very important operators of the algorithm. In both, a selection function plays an essential role.

Basically, the operators will change entering chromosomes with the hope of improving them, but choosing which chromosome to undergo the operations is by the selection function [7].

The function that is used for this purpose time and again is the roulette wheel function. This function operates in a way such that each and every member of the population has a chance to be selected, but the better the fitness value of a chromosome the more selection chance it will have. Crossover is a bi-chromosomal operator in the sense that it will work on two

chromosomes to output other(s). Two entering chromosomes will mix and produce one or two new chromosome(s) which are known by their offspring. In the case of the binary chromosome, one-cut crossover is most used. In one-cut crossover, both chromosomes will be broken from the same cell number and their parts will swap between the two, resulting in two different chromosomes that have characteristics of both entering chromosomes. Crossover is famous for being GA's optimality derive, swaying the population toward best answers. Mutation, unlike crossover, is not bi-chromosomal and does not serve the purpose of moving the population toward optimality. Its contribution to the algorithm is to keep it from local optima by radically changing the entering chromosomes. The single entering chromosome is changed by the operator harshly, without any reason, and randomly [7].

Last word about mutation is the extent that operator will change the entering chromosome. Mutation rate is the term for this behavioral factor of the algorithm. The pivotal step in the algorithm and certainly in the flowchart is deciding when to stop the evolution and be satisfied with the best answer in hand. There is no way GA can be sure of the optimal solution unless an optima is known to it in advance so there is a need for stopping strategies. In fact, there are different stoppage criteria. They can be as simple as a specific numbers of iterations or more involved by bringing the scaled improvements into equation [7].

## 5. DNA Computing

DNA is a polymer composed of monomers called nucleotides. Each nucleotide contains three components: a sugar, a phosphate group, and a base. Figure 1 illustrates the structure of a nucleotide. The sugar has five carbon atoms which are numbered from 1' to 5'. The Phosphate group is attached to the 5' carbon and the base is attached to the 1' carbon. There are four different types of bases: Adenine, Guanine, Cytosine and Thymine, abbreviated as A, G, C and T, respectively. What makes a nucleotide distinct from another is the base portion. So we can refer to every nucleotide as A, G, C or T nucleotides, depending on the type of base they have. Therefore it is possible to consider a DNA strand as a string over the alphabet {A, G, C, T}. For example ATTGCATGG is a DNA strand composed of 9 nucleotides. In every DNA computing experiment, there is a (test) tube containing lots of DNA molecules (strands). Each of the DNA molecules can be a potential solution to the problem. Also there are some biological operations available, which are performed on the DNA strands of the tube to perform computations. These operations differ according to the DNA computing model used [8].

## 6. Substitution Box (S-Box)

Substitution-boxes, or simply S-boxes, are used to increase confidentiality in substitution stage of most of cryptosystem approaches. The design of substitution box, or simply S-box, for secure cryptographic ciphers attracts a great deal of attention of most cryptographic researchers. Indeed, S-box is an important component of most

block ciphers and the good S-box ensures the nonlinearity and the confusion property [9]. In this paper S-box designing based on DNA codes is introduced in order to make the proposed text encryption method more secure and more efficient.

### 7. The Proposed Method

The encryption methods which depend on one private key to protect the information are secure but if the secret key is known by unauthorized persons this security is disappeared. In the proposed method the problem on depending on one secret key is solved by depending on multi secret keys.

In the proposed method multi secret keys are used in order to provide more security and to reduce the secret keys predication capability by the unauthorized persons. The security is achieved because every secret key is designed in special way and has special characters. Some of these keys are random based on some methods like Geffe Generator, Genetic algorithm to generate other keys and the other keys are static (tables) are found at both sides (Sending side and the Receiver side). At the beginning 3 specified seeds are used in the Geffe generator to output a key it's length is (16

bits) for a simple example, these 16 bits will be divided into two parts each part is 8 bits in order to be considered as initial population to the Genetic algorithm, then specified Genetic algorithm phases will be applied to generate a new population (children). After that the genetic algorithm result will be stored.

Then the secret text characters will be read. Then the secret text characters will be converted to numbers in the form (n1, n2) using table (1) where n1 represents the row and the n2 represents the column. Then the numbers will be converted to binary representation.

Apply merging to every two 3 bits to get 6 bits with adding specified padding (2 bits). Then the binary representation will be converted to DNA codes using table (2). Apply swapping with DNA s-box codes using table (3). Convert DNA codes to binary representation using table (4). Apply XOR with first stored Genetic algorithm children. The XOR will be applied between the previous result and the second stored Genetic algorithm children. Finally the result will be send. Algorithm (1) and algorithm (2) and the figures (4, 5) illustrate the proposed method at the both sides.

**Table (1) secret text characters will be converted to numbers in the form (n1, n2)**

	0	1	2	3	4	5
0	a	b	C	d	e	f
1	g	h	I	j	k	l
2	m	n	O	p	q	r
3	s	t	U	v	w	x
4	y	z	0	1	2	3
5	4	5	6	7	8	9

**Table (2) the binary representation will be converted to DNA codes**

Binary Codes (2 bits)	DNA Symbols
00	A
01	C
10	G
11	T

**Table (3) swapping with DNA s-box codes; take the intersection value of column and row of each one code**

Column Row	A	C	G	T
A	A	C	G	T
C	C	G	T	A
G	G	T	A	C
T	T	A	C	G

**Table (4) Convert DNA codes to binary representation**

DNA Symbols	Binary Codes (2 bits)
A	11
C	10
G	01
T	00

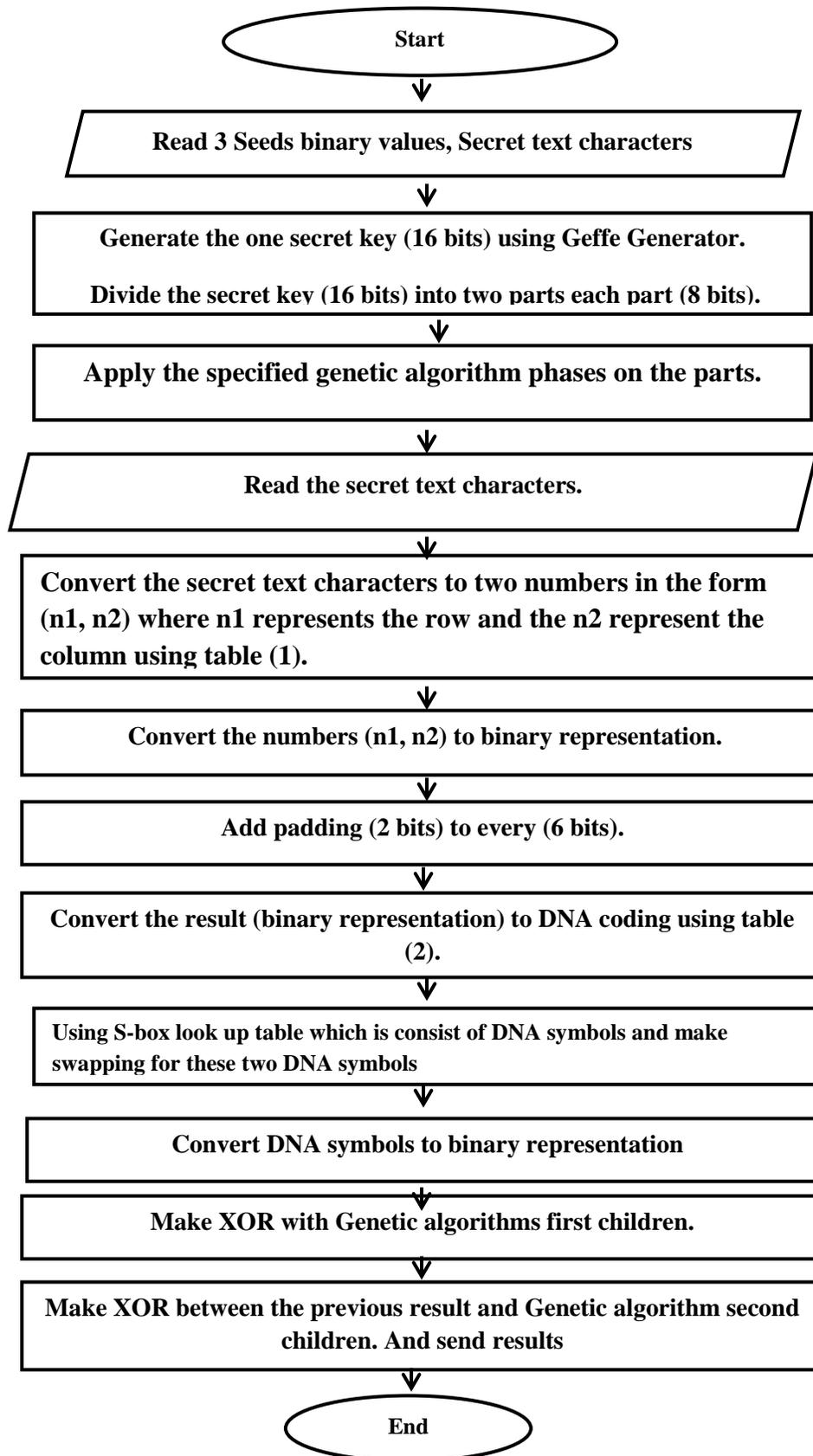


Figure (4): The proposed method at the Sender

Algorithm (1) the proposed method at the sender side
Input: Three seeds binary values, Secret text characters, the four tables.
Output: Binary representation.
<p>Step1:- Begin</p> <p>Step2:- Using the 3 specified seeds to generate the one secret key (16 bits for a simple example) using Geffe Generator.</p> <p>Step3:- Divide the secret key (16 bits) into two parts each part (8 bits).</p> <p>Step4:- Consider the keys as chromosomes in order to apply the specified genetic algorithm phases.</p> <p>Step5:- Read the secret text characters.</p> <p>Step6:- Convert the secret text characters to two numbers in the form (n1, n2) where n1 represents the row and the n2 represent the column using table (1).</p> <p>Step7:- Convert the numbers (n1, n2) to binary representation.</p> <p>Step8:- Add padding (2 bits) to every (6 bits).</p> <p>Step9:- Convert step7 result (binary representation) to DNA coding using table (2).</p> <p>Step10:- Using S-box look up table which is consist of DNA symbols and make swapping for these two DNA symbols where the first DNA symbol is used as determination to the row of the S-box lookup table and the other DNA symbol is used as determination to the column of the S-box look up table. The row determination DNA symbol is stay and the corresponding cell of that row is changed with the column determination using table (3).</p> <p>Step11:- Convert DNA symbols to binary representation using table (4).</p> <p>Step12:- Make XOR between Genetic algorithms first children.</p> <p>Step13:- Make XOR between step9 result and Genetic algorithm second children.</p> <p>Step14:- Send the step12 result.</p> <p>Step15:- End.</p>

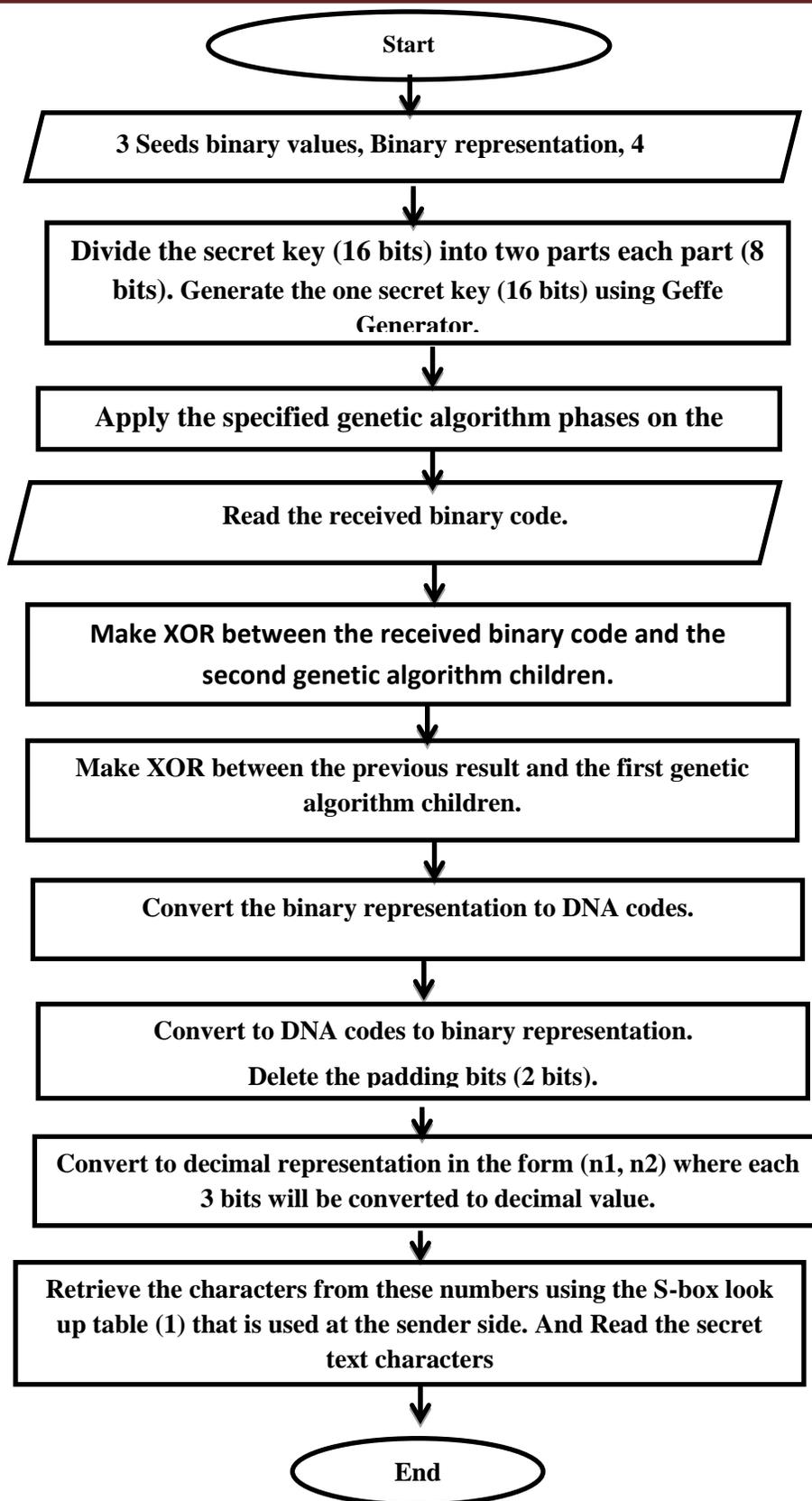


Figure (5): The proposed method at the Receiver Side

Algorithm (2) the proposed method at the receiver side
Input: Three seeds binary values, Binary representation, the four tables.
Output: Secret text characters.
<p>Step1:- Begin</p> <p>Step2:- Using the 3 specified seeds to generate the one secret key (16 bits for a simple example) using Geffe Generator.</p> <p>Step3:- Divide the secret key (16 bits) into two parts each part (8 bits).</p> <p>Step4:- Consider the keys as chromosomes in order to apply the specified genetic algorithm phases.</p> <p>Step5:- Read the received binary code.</p> <p>Step6:- Make XOR between the received binary code and the second genetic algorithm children.</p> <p>Step7:- Make XOR between the step6 result and the first genetic algorithm children.</p> <p>Step8:- Convert the binary representation to DNA codes using table (4).</p> <p>Step8:- Making inverse S-box converting where each two DNA symbols will be swapped .The first DNA symbol will be as indication to the row and the second DNA symbols as indication to cell content to retrieve the corresponding column DNA symbol with the row DNA symbol using table (3).</p> <p>Step9:- Convert to DNA codes to binary representation using table (2).</p> <p>Step10:- Delete the padding bits (2 bits).</p> <p>Step11:- Convert to decimal representation in the form (n1, n2) where each 3 bits will be converted to decimal value.</p> <p>Step12:- Retrieve the characters from these numbers using the S-box look up table (1) that is used at the sender side.</p> <p>Step13:- Read the secret text characters.</p> <p>Step14:- End.</p>

### 7.1 Example of the proposed method (16-bit for simple example)

At the beginning suppose that the three secret seeds for the Geffe generator are as following

Seed 1 65=001000001

Seed 2 70=001000110

Seed 3 135=110000111

After applying the Geffe Generator the result will be as following (1000001111011111), These 16 bits will be divided into two parts each part is (8 bits).

Part1 =10000011

Part2=11011111

These two parts will be considered as the initial population to the genetic algorithm. In the example we suppose that the specified genetic algorithm phases are just the crossover to generate the new children. The crossover will be applied at the middle of each part, after the crossover the result will be as following:

Part1 =10000011



Part2 =11011111



Part1 =11010011

Part2=10001111

Then the secret text characters will be read, we suppose that the secret text is "cryptography". In order to be converted to numbers in the form (n1,n2) table (1) is used the output will be as following:

c=(0,2),r=(2,5),y=(4,0),p=(2,3),t=(3,1),o=(2,2),g=(1,0),r=(2,5),a=(0,0),p=(2,3),h=(1,1),y=(4,0).

Then the numbers will be converted to binary representation as following:

c=(000,010),r=(010,101),y=(100,000),p(010,011),t=(011,001),o=(010,010),g=(001,000),r=(010,101),a=(000,000),p=(010,011),h=(001,001),y=(100,000).

Apply merging to every two 3 bits to get 6 bits with adding padding (2 bits) in order to get 8 bits which is the same length of the genetic algorithm stored keys key. Merging is:

000010 010101 100000 010011 011001 010010 001000 010101 000000 010011  
001001 100000.

After Padding:

00000100 00101010 01000000 00100110 00110010 00100100 00010000  
 00101010 00000000 00100110 00010010 01000000.  
  
 Padding bits

Then Convert to DNA representing using table (2) ,the result will be as following

00000100 00101010 01000000 00100110 00110010 00100100  
  
 AACA AGGG CAAA AGCG ATAG AGCA  
 00010000 00101010 00000000 00100110 00010010 01000000.  
 ACAA AGGG AAAA AGCG ACAG CAAA

Apply swapping with DNA S-box look up table using table (3) ,the result will be as following.

AACA AGGA CCAA AGCT ATAG AGCC ACAA AGGA  
 AAAA AGCT ACAG CCAA

Then these DNA codes will be converted to binary representation using table (4), the result will be as following:

11111010110101111010111110110001100110111011010111011111101011111111111110110  
 001110110110101111

These binary codes will be XORed with the first stored genetic algorithm children, the result will be as follow:

Binary codes =

11111010 11010111 10101111 11011000 11001101 11011010 11101111 11010111  
 11111111 11011000 11101101 10101111

Stored key=

11010011 11010011 11010011 11010011 11010011 11010011 11010011  
 11010011 11010011 11010011 11010011 11010011

First XORED result is =

```
00101001  00000100  01111100  00001011  00011110  00001001  00111100
00000100  00101100  00001011  00111110  01111100
```

Then the XOR result will be XORED with the second stored genetic algorithm children and the result will be as follow:

First XORED result is =

```
00101001  00000100  01111100  00001011  00011110  00001001  00111100
00000100  00101100  00001011  00111110  01111100
```

Second stored key =

```
10001111  10001111  10001111  10001111  10001111  10001111  10001111  10001111
10001111  10001111  10001111  10001111
```

The final result after XOR is =

```
10100110  10001011  11110011  10000100  10010001  10000110  10110011
10001011  10100011  10000100  10110001  11110011
```

This final result will be send.

## 7.2 Evaluation of the proposed method

In the proposed method there are three types of secret keys:

- 1) The specified keys (3 Primitive polynomials, 3 seeds which are used in Geffe generator and padding which can be 00, 01, 10, 11).
- 2) The Generated keys (Geffe Generator result) and the Genetic algorithm result.
- 3) The Static keys (the tables which are used like table, S-box look up table).

In the following table, table (5) the tested results are display, with randomness measures that are used in cryptography which are (Frequency test, Serial test, Poker test, Run test, Auto\_Correlation (AC) test).

Table (5) randomness test of the three keys

Example No.	LFSR polynomial	Seeds	Geffe Generator Result	Specified Keys Randomness Measures	Generated Keys	Static Keys
Example 1	$LFSR1=X^8+X^6+X^5+X^4+1$ $LFSR2=X^8+X^6+X^5+X^4+1$ $LFSR3=X^8+X^6+X^5+X^4+1$	Seed 1 =001000001 Seed 2 =001000110 Seed 3 =110000111	1000001111011111	Frequency Test=9.000 Serial Test=12.000 Poker Test= 15.200 Run Test=38.969 AC Test Move No. 1 to No. 10 all are Succeed Value average from 3.769 to 0.500.	Frequency Test=1.000 Serial Test= 3.000 Poker Test=7.200 Run Test= 8.375 AC Test Move No. 1 to No. 10 all are Succeed Value average from 3.267 to 0.000	Frequency Test=18.375 Serial Test= 19.500 Poker Test= 21.367 Run Test= 8.583 AC Test Move No. to No. 10 most of them defeat Value from 6.868 to 7.511
Example 2	$LFSR1=X^{16}+X^{14}+X^{13}+X^{11}+1$ $LFSR2=X^{19}+X^{18}+X^{17}+X^{14}+1$ $LFSR3=X^9+X^5+1$	Seed1=00101010101010101 Seed2=111100001111000 Seed3=0100110011	10010110111010111011001001100	Frequency Test= 0.125 Serial Test= 12.500 Poker Test= 5.400 Run Test= Succeed Value T0 = 8.250 AC Test AC Test Move No. 1 to No. 10 most of them succeed Value 0.043 and some defeat Value 4.172	Frequency Test=0.500 Serial Test=Succeed Value T0 = 3.000 Poker Test=Succeed Value 2.400 Run Test= Succeed Value T0 = 3.000 AC Test Move No. 1 to No. 10 all are Succeed Value from 2.793 to 0.000	Frequency Test=18.375 Serial Test= 19.500 Poker Test= 21.367 Run Test= 8.583 AC Test Move No. 1 to No. 10 some of them succeed Value 0.011 and some defeat Value 7.511

## 8. Conclusions

A deception and secure encryption method is proposed in this paper. The proposed method is better in security level since it depends on multi secret special keys and these keys are designed based on Geffe Generator, DNA, and S-box .The proposed method is fast and simple in implementation since it doesn't need any complex calculation. The private keys are kept secret since the secret keys are not used directly in encryption but will be processed using genetic algorithm. The attacker can't know the secret keys since not the original secret keys will be used to make the encryption.

## References

- [1] Ghassan M.H. ,” *Image Encryption Using Permutation and Hill Cipher* “, *Al-Rafidain University College For Sciences* ,2012.
- [2] Hasan M. Azzawi ,” *Enhancing The Encryption Process Of Advanced Encryption Standard (AES) By Using Proposed Algorithm To Generate S-Box*”, *Journal of Engineering and Development*, Vol. 18, No.2, March 2014, ISSN 1813- 7822.
- [3] Qusay M. Jafar Alsadi ,” *Proposed Method for Text Encryption Using Two Secret Keys and One Secret Mathematical Equation* “, *Al-Rafidain University College For Sciences*,2009.
- [4] Maiya Dina, Saibal K. Pal a S.K. Muttoo b, Anjali Jainc , “*Applying Cuckoo Search for analysis of LFSR based cryptosystem*”, *Perspectives in Science* (2016) 8, 435—439.
- [5] Noor Dhia Kadhm Al-Shakarchy ,” *Randomly Steganography using LFSR and NLFSR generation*”, *Journal of KerbalaUniversity* , Vol. 11 No.1 Scientific . 2013.
- [6] Hussein Ali Mohammed , “*Frequency Postulate's Theoretical Calculation for the Sequences Produced by Modified Geffe Generator*”, *Journal of Kerbala University* , Vol. 12 No.2 Scientific ,2014.
- [7] Ruholla Jafari-Marandi, Brian K. Smith “*Fluid Genetic Algorithm (FGA)*” , *Journal of Computational Design and Engineering* 4 (2017) 158–167.
- [8] Ramin Maazallahi, Aliakbar Niknafs, Paria Arabkhedri , “*A Polynomial-Time DNA Computing Solution for the N-Queens Problem*” , *Procedia - Social and Behavioral Sciences* 83 ( 2013 ) 622 – 628.
- [9] Akram Belazi , Ahmed A. Abd El-Latif b,” *A simple yet efficient S-box method based chaotic sine map*” *Optik - International Journal for Light and Electron Optics* Volume 130, February 2017, Pages 1438-1444.

## Modified Binary Particle Swarm Optimization for Solving Distribution Network Reconfiguration

*Ali Nasser Hussain*

*Department of Electrical Engineering, Electrical Engineering Technical College,  
Middle Technical University, Baghdad, Iraq*

*E-mail: alinasser1974@yahoo.com*

**Abstract:** This study utilized Binary Particle Swarm Optimization (BPSO) and Modified BPSO (MBPSO) for solving Distribution Network Reconfiguration (DNR). The search problem space for the presented algorithm is a set of lines (switches) which are normally closed or opened, this search problem may be dissimilar for different dimensions. This paper consists of two parts. First, the reconfiguration with constant load was optimized based on two algorithms BPS and MBPS. The decreasing of real power loss has been invested as an objective function; while node voltage, system radially and line current have been utilized as limits of the system. Second, the reconfiguration with variable load is optimized based on the same two algorithms BPS and MBPS. The proposed methods are applied on IEEE node 33 power system by using MATLAB software to test the effectiveness and efficiency of MBPSO algorithm. The results for the IEEE node 33 power systems indicate that MBPSO algorithm has high ability and effective in reduce power loss and voltage profile enhancing of the system compared to BPSO.

**Keywords:** BPSO, MBPSO, DNR, Reduce power loss, Voltage profile enhancing.

### أمثليه حشد جسيمات ثنائي معدلة لحل إعادة تشكيل شبكة التوزيع

علي ناصر حسين

قسم هندسة تقنيات القدرة الكهربائية، الكلية التقنية الهندسية الكهربائية، الجامعة التقنية الوسطى، بغداد، العراق

**الخلاصة:** يستعرض هذا البحث أمثليه حشد جسيمات ثنائي (BPSO) وأمثليه حشد جسيمات ثنائي معدلة (MBPSO) لحل إعادة تشكيل شبكة التوزيع (DNR). أن فضاء مشكلة البحث للخوارزميات المقدمة هو مجموعة من الخطوط (المفاتيح) التي تفتح عادةً أو تغلق، مشكلة البحث هذه قد تكون متباينة لأبعاد مختلفة. هذا البحث يتكون من جزئين. الأول، اعاده التشكيل مع الحمل الثابت التي أمثلت بالأستناد على الخوارزميتين BPSO و MBPSO. تقليل خسارة القدرة الحقيقية تم أستثمارها كدالة هدف موضوعية، بينما فولتية العقدة، نظام بشكل شعاعي و تيار الخط استعمل كحدود للنظام. الثاني اعاده التشكيل مع الحمل المتغير التي أمثلت بالأستناد على نفس الخوارزميتين BPSO و MBPSO. الطريقتان المقترحة BPSO و MBPSO طبقت على نظام قدرة كهربائي IEEE 33 حافلة باستخدام برنامج الماتلاب. النتائج لنظام قدرة كهربائي IEEE 33 حافلة حددت بان خوارزمية MBPSO لديها قدرة عالية وفعالة في تقليل خسائر القدرة وتحسين ملف الفولتية للنظام مقارنة الى BPSO.

**الكلمات المفتاحية:** أمثليه حشد جسيمات ثنائي، أمثليه حشد جسيمات ثنائي معدلة، إعادة تشكيل شبكة التوزيع، تقليل خسائر القدرة، تحسين ملف الفولتية

## 1. Introduction

Loss minimization is used to enhance the flexibility of the system. Distributed generator allocation (DG), conductor grading, capacitor placement and feeder reconfiguration are better approaches for decreasing power loss [1]. On the other hand, adding these methods into the distribution system needs much cost. DNR can be accomplished through the reconfiguration of tie switches and sectionalizing, by this method, the loss system is reduced and voltage level is enhanced by considering the operating limits devoid of costs [2]. By redistributing and arranging the loads from heavy to light, DNR can balance the feeder loads and prevents the overloading [3]. Many techniques have been described in the literature to obtain the optimal DNR. The Artificial neural network technique based on the mapping capability to decide network reconfiguration is presented in [4]. An expert system utilizing heuristic rules to reduce the search problem for decreasing the calculation time has been proposed in paper [5]. The study of load balancing and reducing power loss formulated as integer programming problem was proposed by Baran [6]. Chiang and Jumeau have been proposed a new load balancing index and they utilized it on the test power system for load balancing [7]. A new balance and unbalance load approach in distribution system for decreasing of the power loss was presented in reference [8]. Naveen was presented DNR for reducing loss via modification technique based on the

Bacterial Foraging Optimization [9]. Cuckoo Search Approach (CSA) was introduced by Nguyen and Truong; DNR have two objectives, which were to voltage level enhancement and to reduce the loss of the system [10].

In this study, BPSO and Modified BPSO MBPSO approaches are utilized in network reconfiguration to get the better solution with the objective function for decreasing line power loss and enhance voltage profile. The BPSO and MPSO algorithms are applied on 33-node IEEE system with constant loads and variable loads to find the optimal DNR. For variable load ( $\mu$  multiplied by constant load) where  $\mu$  represents the ratio value for the load variation. The range of variation for loads is linearly changed between ( $\mu = 0.75$ ) at light load up to ( $\mu = 1.250$ ) at heavy load. The results of DNR problem have been implemented for standard IEEE 33 node power system. From the results, MBPSO algorithm has high ability and effective in reduce the total real power loss and enhancing the minimum and average voltages of the system compared to BPSO and other reported papers.

## 2. Problem Formulation

### A. Load Flow

Load flow in electrical power distribution network can be defined by a number of equations that depends on the active power, reactive power and voltage at the sending end of a line to express the same quantities at receiving end of the line [11]. By

utilizing the calculation of power flow, total power loss can be obtained in

figure.1.

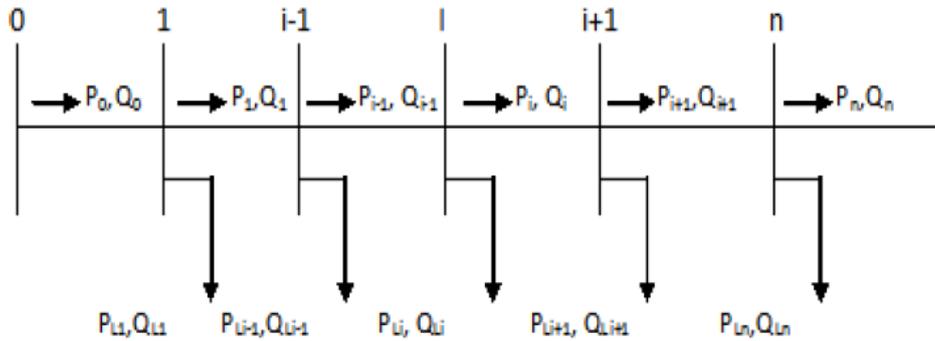


Figure 1: Simple distribution line.

The active and reactive load flow equations in the branch among  $i + 1$  and  $n - th$  nodes are:

$$P_{i+1} = [(P_i) - (P_{Li+1}) - (R_{i,i+1})] \left[ \frac{(P_i)^2 + (Q_i)^2}{|V_i|^2} \right] \quad (1)$$

$$Q_{i+1} = [(Q_i) - (Q_{Li+1}) - (X_{i,i+1})] \left[ \frac{(P_i)^2 + (Q_i)^2}{|V_i|^2} \right] \quad (2)$$

The voltage at nodes  $i$  and  $i + 1$  can be express as follows:

$$|V_{i+1}|^2 = |V_i|^2 - 2[(R_{i,i+1} \cdot P_i) + (X_{i,i+1} \cdot Q_i)] + [(R_{i,i+1}^2) + (X_{i,i+1}^2)] \left[ \frac{P_i^2 + Q_i^2}{|V_i|^2} \right] \quad (3)$$

The current equation can be express by the following equation:

$$I_T = \frac{P_i - jQ_i}{|V_i|} \quad (4)$$

The summation of real power loss can be express as shown below:

$$P_{LT} = \sum R_{i,i+1} I_T^2 \quad (5)$$

from the above equations :  $(P_i)$  and  $(Q_i)$  are the real and reactive power loss at node  $i$ ;  $(R_{i,i+1})$  and  $(X_{i,i+1})$ : are the resistance and reactance of

branch section between two nodes  $i$  and  $i + 1$ ;  $(V_i)$  is the voltage at node  $i$ ;  $(I_T)$  is the total current and  $(P_{LT})$  is the total real power losses.

**B. Objective Function**

The objective function of DNR is applied to decrease the real power loss and it is presented in equation (6):

$$f(x) = \min P_{LT} \tag{6}$$

where  $(x)$  is the control variable and  $P_{LT}$  is the total real power loss.

**C. Constrains**

In any DNR, the load flow calculation can be done by finding the node voltage, line current and active power loss of a system for every line. The necessities of the objective function are shown below:

1. Bus voltage has *min* and *max* bounds as shown in equation (7) below.

$$V_i^{min} \leq |V_i| \leq V_i^{max} ; \tag{7}$$

$$i = 1, 2, \dots, N_n$$

**D. Average Voltage Index**

This index is presented to replace the lower voltage to estimate the quality of power that is a more suitable from the

$$V_{av} = \frac{\sum_{i=1}^{N_n} V_i}{N_n} \tag{10}$$

**3. Reconfiguration Approaches**

**A. (PSO) algorithm**

From equation (6),  $V_i^{min}$  and  $V_i^{max}$  are the lower (0.9 p.u *min*) and upper (1.0 p.u *max*) voltage of node  $i$ ;  $N_n$  represent the number of nodes.

2. Line current values should not overcome constraint of each line as in equation (8).

$$|I_T| \leq I_T^{max} ; \tag{8}$$

$$T = 1, 2, \dots, N_b$$

Where  $I_T^{max}$  is the *max* bound of line current  $T$  and  $N_b$  is the total number of the lines.

3. Always save the power system in radial structure as written in (9).

$$\left. \begin{aligned} \det(A) &= 1 \text{ or } -1 && \text{(Radial System)} \\ \det(A) &= 0 && \text{(Not Radial System)} \end{aligned} \right\} \tag{9}$$

viewpoint of both sides. This index is given in equation (10).

From the above equation,  $(V_{av})$  is the average voltage for a network;  $(V_i)$  is the voltage at node  $i$  and  $(N_n)$  is the number of network nodes.

Basic idea of PSO came from the behavior of animals such as fish schooling or bird flocking to search for

food. And it is first introduced by Eberhart and Kennedy [12] in year 1995. The basic PSO algorithm is the real valued PSO, whereby each dimension in the space of the problem

can take any real valued number. The particles update their speed and position according to the following equations (11) and (12).

$$v_i^{k+1} = (w * v_i^k) + c_1 * [rand_1 * (p_{bi}^k - x_i^k)] + c_2 * [rand_2 * (g_{bi}^k - x_i^k)] \quad (11)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (12)$$

where  $(v_i^{k+1})$  is the velocity of particle at  $(k + 1)$  iteration;  $(v_i^k)$  is the velocity of particle at current iteration;  $(C_1, C_2)$  are the two positive constants within  $[0 - 2.5]$ ;  $(rand_1, rand_2)$  are the uniformly distributed positive random numbers within limit  $[0-1]$ ;  $(p_{bi}^k)$  is the local best value at  $(k)$  iteration;  $(g_{bi}^k)$  is the global best value at  $(k)$  iteration;  $(x_i^k)$  is the position at current iteration;  $(x_i^{k+1})$  is the position at  $(K + 1)$  iteration and  $(w)$ : is the inertia weight and it is reduced linearly from  $(0.9$  to  $0.4)$  at each iteration, and can be expressed as follows.

### B. (BPSO) algorithm

The first concept for BPSO algorithm has been presented by Eberhart and Kennedy in year 1997 [13]. The size of searching space is equal to number of tie switches in a system. In order to transform the exploration of PSO in a real space dimensions to binary space dimensions, sigmoid transformation is applied to the velocity element to force the velocities within a range  $[0, 1]$ , and force the component values of the locations of agents to be  $(0$  or  $1)$ . Therefore, equation (12) for changing the position is replaced by Equation (17). Also  $W$  is reduced linearly from  $(0.9$  to  $0.4)$  as shown in equation (15).

$$W = W_{max} - \left( \frac{W_{max} - W_{min}}{max_{iteration}} \right) * iter \quad (13)$$

$$v_i^{k+1} = (w * v_i^k) + [c_1 * old * rand_1 * (p_{bi}^k - x_i^k)] + [c_2 * old * rand_2 * (g_{bi}^k - x_i^k)] \quad (14)$$

$$W = W_{max} - \left( \frac{W_{max} - W_{min}}{max_{iteration}} \right) * iter \quad (15)$$

$$sigmoid(v_i^{k+1}) = \frac{1}{1 + \exp(-v_i^{k+1})} \quad (16)$$

$$\left. \begin{aligned} x_i^{k+1} &= [1, \text{if } rand < sigmoid(v_i^{k+1})] \\ x_i^{k+1} &= [0, \text{otherwise}] \end{aligned} \right\} \quad (17)$$

where old learning factors,  $c_{1old} = \text{constant}$  and  $c_{2old} = \text{constant}$ .

**C. (MBPSO) algorithm**

In the Modified BPSO (MBPSO) algorithm all agents move to be nearest to the better position based on objective function and discover the global optimum location for minimum point. It is similar to BPSO algorithm but in the MBPSO the old positive constants are modified to a random values in range between  $[0 - 1]$

instead of constant value ( $c_{1old}$  and  $c_{2old}$ ) which are given in equation (14) at BPSO. Also, the size of searching problem space is equal to number of tie switches in a network. This randomly helps to rise the ability of PSO approach in order to reach the optimal solution much faster than ( $c_{1old}$  and  $c_{2old}$ ). The equation of velocity can be written as follows in equation (18).

$$v_i^{k+1} = (w * v_i^k) + [c_{1new} * rand_1 * (p_{bi}^k - x_i^k)] + [c_{2new} * rand_2 * (g_{bi}^k - x_i^k)] \tag{18}$$

where  $c_{1new}$  and  $c_{2new}$  are given in equation (19) and (20) are the new learning factors between  $[0-1]$  instead of the old learning factors  $c_{1old}$  and  $c_{1new} = \text{Rand}$

$c_{2old}$  given in equation (14) in the BPSO algorithm.

$$\tag{19}$$

$$c_{2new} = \text{Rand} \tag{20}$$

$$\tag{20}$$

**4. Case Study**

The efficiency of MBPSO for DNR is tested on IEEE-33 node system. The data details of the network and loads for power system have been given in reference [10]. And the result of the network reconfiguration at BPSO and MPSO are obtained in two cases at constant load and at variable load.

In this case DNR is applied to the constant load ( $\mu = 1$ ) demand. IEEE 33-node is presented as test system for both the BPSO and MBPSO approaches. Table 1 describes the comparison among the proposed methods and some other methods reported in the literature [2, 9, 10, 14]. Switches status, real power loss, minimum and average voltage are given in this table.

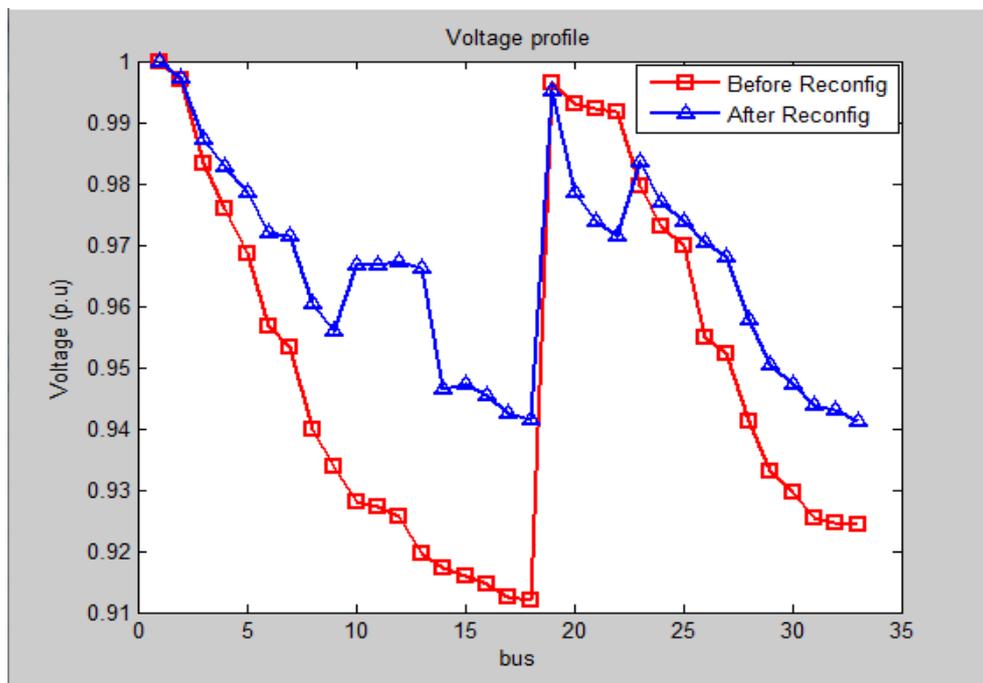
**I. Case study (1) with constant load ( $\mu = 1$ )**

**Table 1:** Result of DNR for the IEEE 33–node power system at constant load demands ( $\mu = 1$ ) while using BPSO, MBPSO and some other approaches.

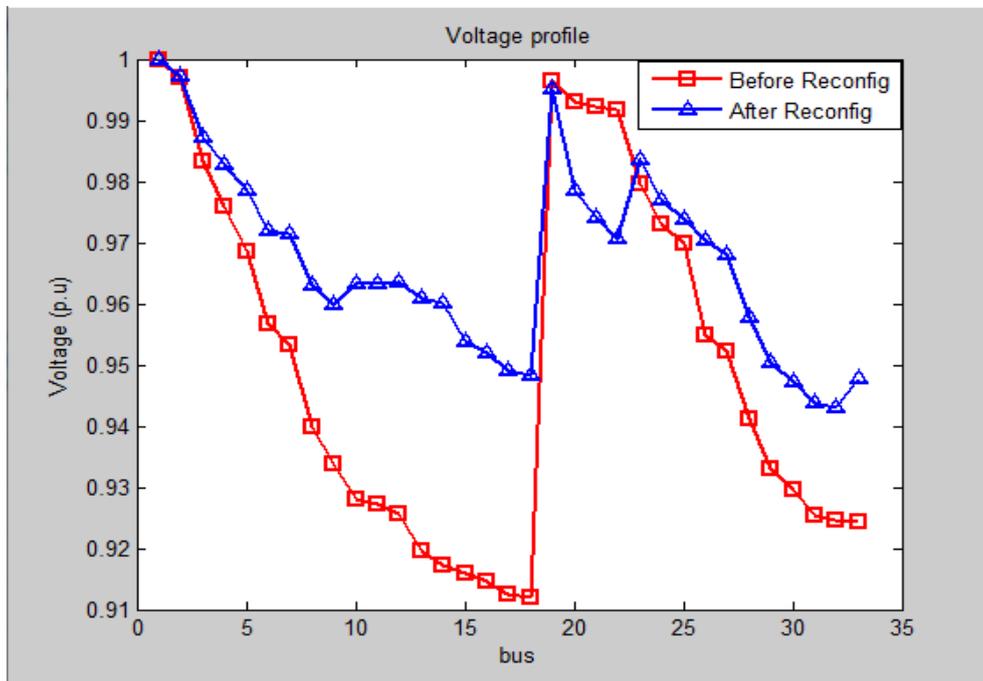
Approach	Open Switches	$P_{LT}$ (KW)	$V_{av}$ (p. u.)	$V_{min}$ (p. u.)
Initial	sw33, sw34, sw35, sw36, sw37	202.67	0.9485	0.9092
BPSO	sw7, sw9, sw13, sw32, sw37	138.61	0.9657	0.9412
MBPSO	sw7, sw9, sw14, sw32, sw37	135.17	0.9669	0.9431
FWA [2]	sw7, sw9, sw14, sw28, sw32	139.55	0.9674	0.9413
MBFOA [9]	sw7, sw9, sw13, sw32, sw37	141.91	0.9678	0.9378
ITS [10]	sw07, sw09, sw14, sw36, sw37	142.16	0.9653	0.9336
SLR [14]	sw07, sw10, sw14, sw36, sw37	142.67	0.9651	0.9336

It is seen from Table 1, the real power loss ( $P_{LT}$ ) reduces while using BPSO by 31% from 202.67 kW to 138.61kW and with MBPSO by 33% from 202.67 kW to 135.17kW. The minimum voltage ( $V_{min}$ ) while using BPSO enhances from 0.9092 p.u. to

0.9412 p.u. and with MBPSO improves from 0.9092 p.u. to 0.9412 p.u., while the average voltage enhances from 0.9485 p. u. to 0.9657p. u. while using BPSO and from 0.9485 p.u. to 0.9669 p.u. with MBPSO.



**Figure 1:** Voltage profile of DNR for the IEEE 33–node power system at constant load demands ( $\mu = 1$ ) while using BPSO.



**Figure 2:** Voltage profile of DNR for the IEEE 33-node power system at constant load demands ( $\mu = 1$ ) while using MBPSO.

Figure 1 and Figure 2 show voltage profiles of the network while using BPSO and MBPSO. It is clear that the voltage at all nodes (except the nodes 19, 20, 21, 22) were improves after

## II. Case study (2) with variable load

The load demand (real and reactive) at the nodes is changes within the range ( $\mu^{min} \leq \mu \leq \mu^{max}$ ) where

$$P_{Li} = \mu P_{Li0} \tag{21}$$

$$Q_{Li} = \mu Q_{Li0} \tag{22}$$

From the above equations,  $\mu$  is the value of the load variation ratio,  $P_{Li0}$  and  $Q_{Li0}$  are the base constant real and reactive powers of the  $i$ -th load. And

reconfiguration. Finally, it is clear that from Figure 1 and Figure 2 by using MBPSO greatly improves the voltage profile compared to BPSO.

( $\mu^{min} = 0.75$ ) at light and ( $\mu^{max} = 1.25$ ) at heavy with the percent of step change ( $\Delta\mu$ ) equal to 12.5%. The load is varied by multiplying  $\mu$  with load at base case.

the results of these cases are shown below.

**At loading with variation ratio ( $\mu = 0.75$ ):** in this case DNR is applied to

the light load ( $\mu = 0.75$ ) demand. IEEE 33–node is presented as test system for both the BPSO and MBPSO approaches. Table 2 describes the comparison among the proposed

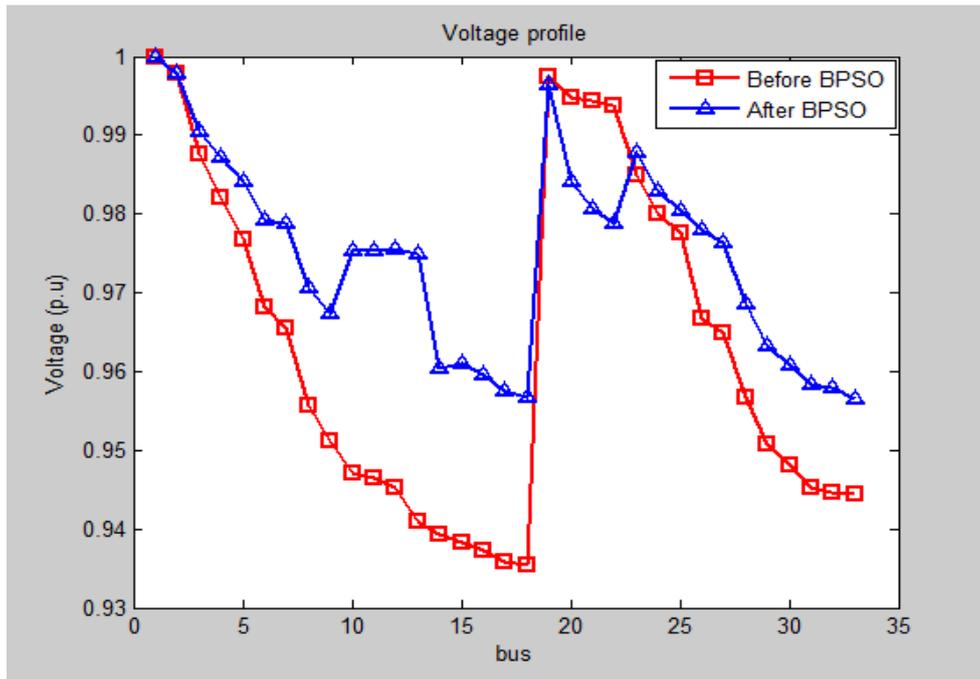
methods and some other methods reported in the literature [2,9,10,14]. Switches status, real power loss, minimum and average voltage are given in this table.

**Table 2:** Result of DNR for the IEEE 33–node power system at light load demands ( $\mu = 0.75$ ) while using BPSO, MBPSO and some other approaches.

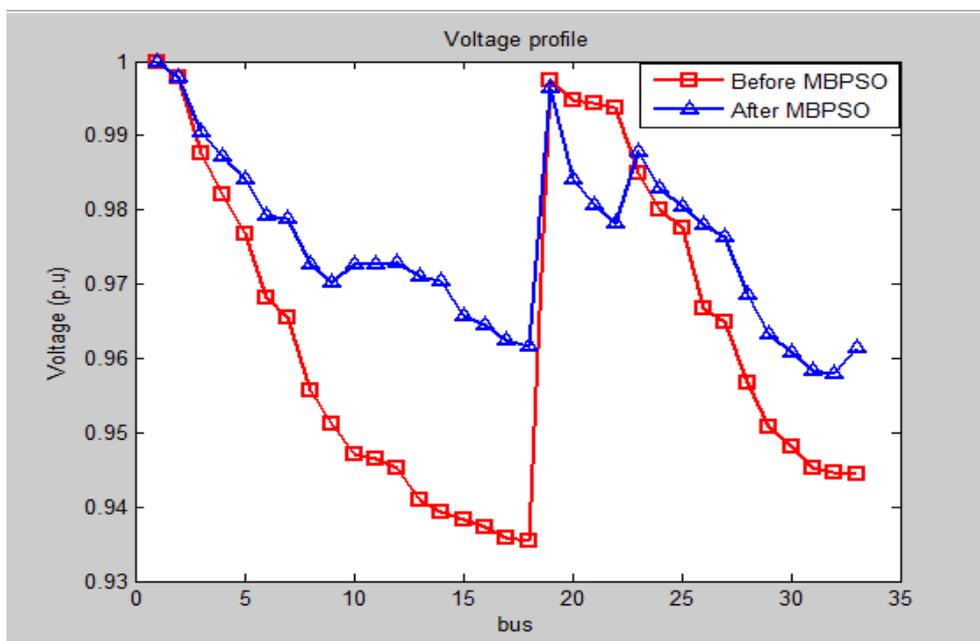
Approach	Open Switches	$P_{LT}$ (KW)	$V_{av}$ (p.u.)	$V_{min}$ (p.u.)
Initial	sw33, sw34, sw35, sw36, sw37	109.75	0.9621	0.9362
BPSO	sw7, sw9, sw13, sw32, sw37	76.16	0.9746	0.9565
MBPSO	sw7, sw9, sw14, sw32, sw37	74.33	0.9754	0.9579
FWA [2]	sw7, sw9, sw14, sw28, sw32	76.87	0.9758	0.9566
MBFOA [9]	sw7, sw9, sw13, sw32, sw37	77.88	0.9762	0.9540
ITS [10]	sw07, sw09, sw14, sw36, sw37	77.97	0.9743	0.9510
SLR [14]	sw07, sw10, sw14, sw36, sw37	78.25	0.9742	0.9510

It is seen from Table 1, the real power loss ( $P_{LT}$ ) reduces while using BPSO by 30% from 109.75 kW to 76.16 kW and with MBPSO by 32% from 109.75 kW to 74.33 kW. The minimum voltage ( $V_{min}$ ) while using BPSO enhances from 0.9362 p.u. to 0.9565 p.u. and with MBPSO

improves from 0.9362 p.u. to 0.9579 p.u., while the average voltage enhances from 0.9621 p.u. to 0.9746 p.u. while using BPSO and from 0.9621 p. u. to 0.9754 p. u. with MBPSO.



**Figure 3:** Voltage profile of DNR for the IEEE 33-node power system at light load demands ( $\mu = 0.75$ ) while using BPSO.



**Figure 4:** Voltage profile of DNR for the IEEE 33-node power system at light load demands ( $\mu = 0.75$ ) while using MBPSO.

Figure 3 and Figure 4 show voltage profiles of the network while using

BPSO and MBPSO. It is clear that the voltage at all nodes (except the nodes

19, 20, 21, 22) were improves after reconfiguration. Finally, it is clear that from Figure 3 and Figure 4 by using MBPSO greatly improves the voltage profile compared to BPSO.

**At loading with variation ratio ( $\mu = 0.875$ ):** in this case DNR is applied to the load factor ( $\mu = 0.875$ ) demand.

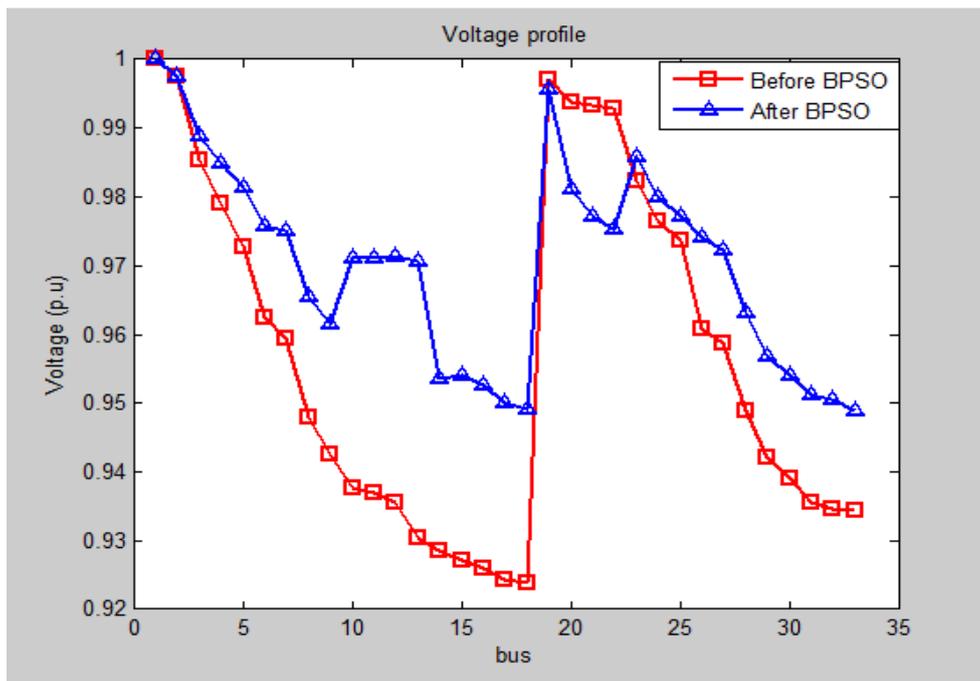
IEEE 33–node is presented as test system for both the BPSO and MBPSO approaches. Table 2 describes the comparison among the proposed methods and some other methods reported in the literature [2,9, 10, 14]. Switches status, real power loss, minimum and average voltage are given in this table.

**Table 3:** Result of DNR for the IEEE 33–node power system at load factor ( $\mu = 0.875$ ) while using BPSO, MBPSO and some other approaches.

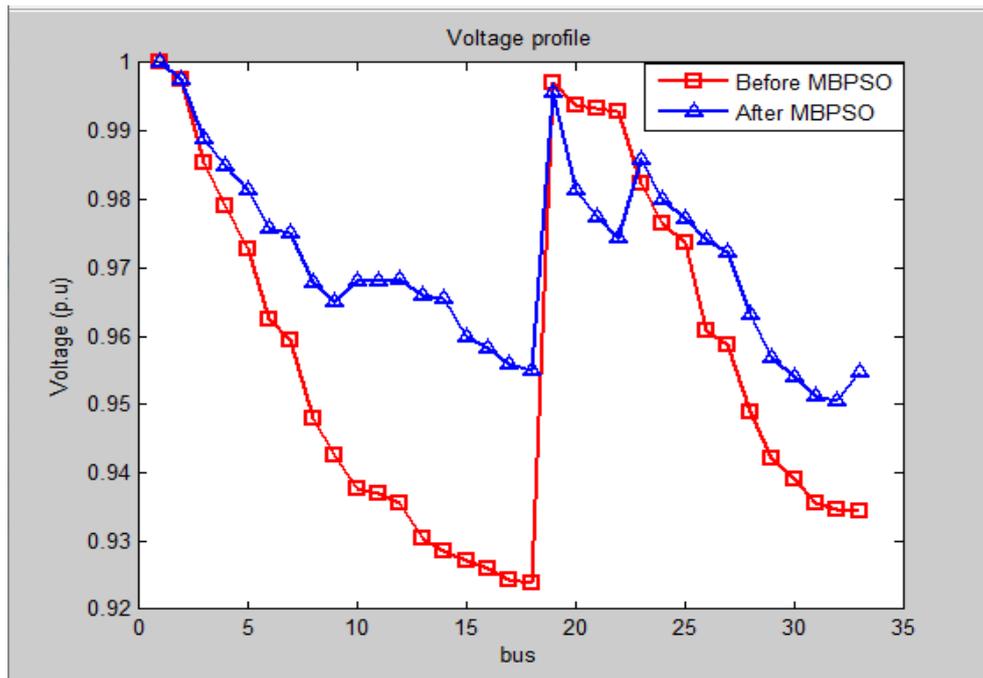
Approach	Open Switches	$P_{LT}$ (KW)	$V_{av}$ (p.u.)	$V_{min}$ (p.u.)
Initial	sw33, sw34, sw35, sw36, sw37	152.20	0.9553	0.9248
BPSO	sw7, sw9, sw13, sw32, sw37	104.87	0.9702	0.9489
MBPSO	sw7, sw9, sw14, sw32, sw37	102.31	0.9712	0.9505
FWA [2]	sw7, sw9, sw14, sw28, sw32	105.88	0.9716	0.9490
MBFOA [9]	sw7, sw9, sw13, sw32, sw37	107.31	0.9720	0.9460
ITS [10]	sw07, sw09, sw14, sw36, sw37	107.46	0.9698	0.9423
SLR [14]	sw07, sw10, sw14, sw36, sw37	107.84	0.9697	0.9423

It is seen from Table 1, the real power loss ( $P_{LT}$ ) reduces while using BPSO by 31% from 152.20 kW to 104.87kW and with MBPSO by 32% from 152.20 kW to 102.31 kW. The minimum voltage ( $V_{min}$ ) while using BPSO enhances from 0.9248 p.u. to 0.9489 p.u. and with MBPSO improves from 0.9248 p.u. to 0.9505 p.u., while the average voltage enhances from 0.9553 p.u. to 0.9702 p.u. while using BPSO and from 0.9553 p.u. to 0.9712 p.u. with MBPSO.

Figure 5 and Figure 6 show voltage profiles of the network while using BPSO and MBPSO. It is clear that the voltage at all nodes (except the nodes 19, 20, 21, 22) were improves after reconfiguration. Finally, it is clear that from Figure 5 and Figure 6 by using MBPSO greatly improves the voltage profile compared to BPSO.



**Figure 5:** Voltage profile of DNR for the IEEE 33–node power system at load factor ( $\mu = 0.875$ ) while using BPSO.



**Figure 6:** Voltage profile of DNR for the IEEE 33–node power system at load factor ( $\mu = 0.875$ ) while using MBPSO

**At loading with variation ratio ( $\mu = 1.125$ ):** in this case DNR is applied to load factor ( $\mu = 1.125$ ) demand. IEEE 33–node is presented as test system for both the BPSO and MBPSO approaches. Table 4 describes the

comparison among the proposed methods and some other methods reported in the literature [2,9,10,14]. Switches status, real power loss, minimum and average voltage are given in this table.

**Table 4:** Result of DNR for the IEEE 33–node power system at load factor ( $\mu = 1.125$ ) while using BPSO, MBPSO and some other approaches.

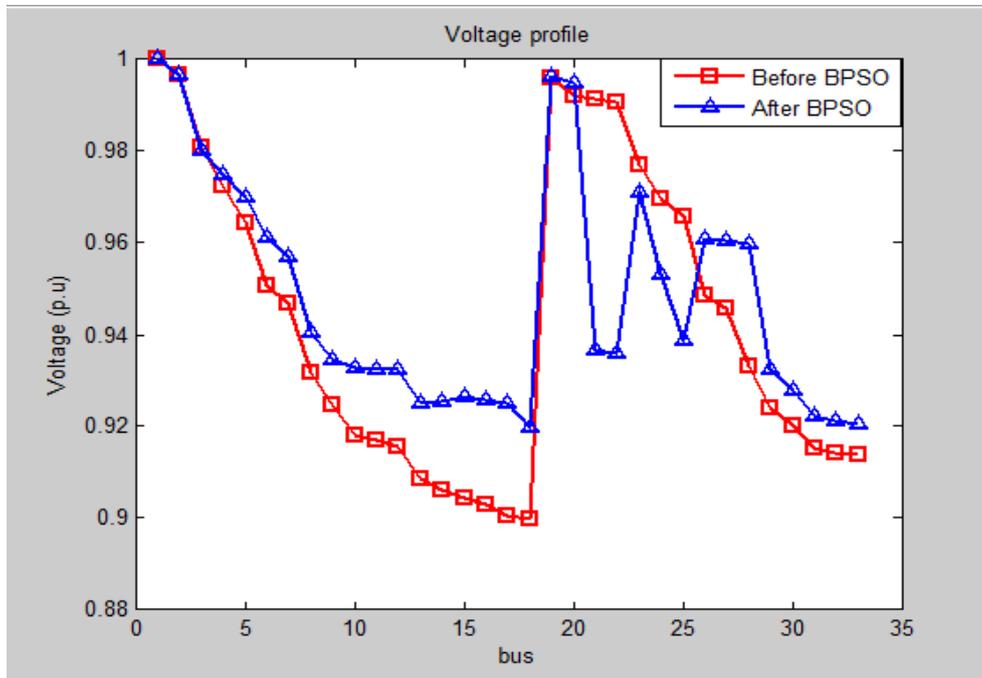
Approach	Open Switches	$P_{LT}$ (KW)	$V_{av}$ (p.u.)	$V_{min}$ (p.u.)
Initial	sw33, sw34, sw35, sw36, sw37	261.69	0.9414	0.9011
BPSO	sw7, sw9, sw13, sw32, sw37	243.63	0.9482	0.9199
MBPSO	sw7, sw9, sw14, sw32, sw37	173.08	0.9754	0.9355
FWA [2]	sw7, sw9, sw14, sw28, sw32	179.63	0.9631	0.9335
MBFOA [9]	sw7, sw9, sw13, sw32, sw37	181.91	0.9636	0.9295
ITS [10]	sw07, sw09, sw14, sw36, sw37	182.29	0.9607	0.9247
SLR [14]	sw07, sw10, sw14, sw36, sw37	182.95	0.9605	0.9247

It is seen from Table 1, the real power loss ( $P_{LT}$ ) reduces while using BPSO by 6% from 261.69 kW to 243.63 kW

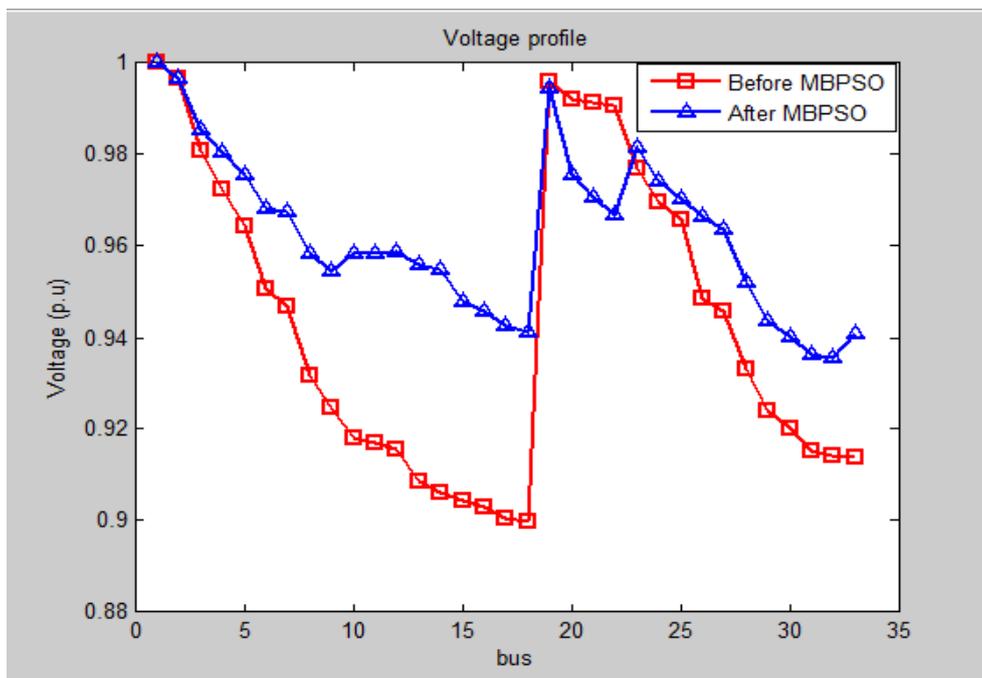
and with MBPSO by 33% from 261.69kW to 173.08 kW. The minimum voltage ( $V_{min}$ ) while using

BPSO enhances from 0.9011 p.u. to 0.9199 p.u. and with MBPSO improves from 0.9011 p.u. to 0.9355 p.u., while the average voltage

enhances from 0.9414 p.u. to 0.9482 p.u. while using BPSO and from 0.9414 p.u. to 0.9754 p.u. with MBPSO.



**Figure 7:** Voltage profile of DNR for the IEEE 33-node power system at load factor ( $\mu = 1.125$ ) while using BPSO.



**Figure 8:** Voltage profile of DNR for the IEEE 33-node power system at load factor ( $\mu = 1.125$ ) while using MBPSO.

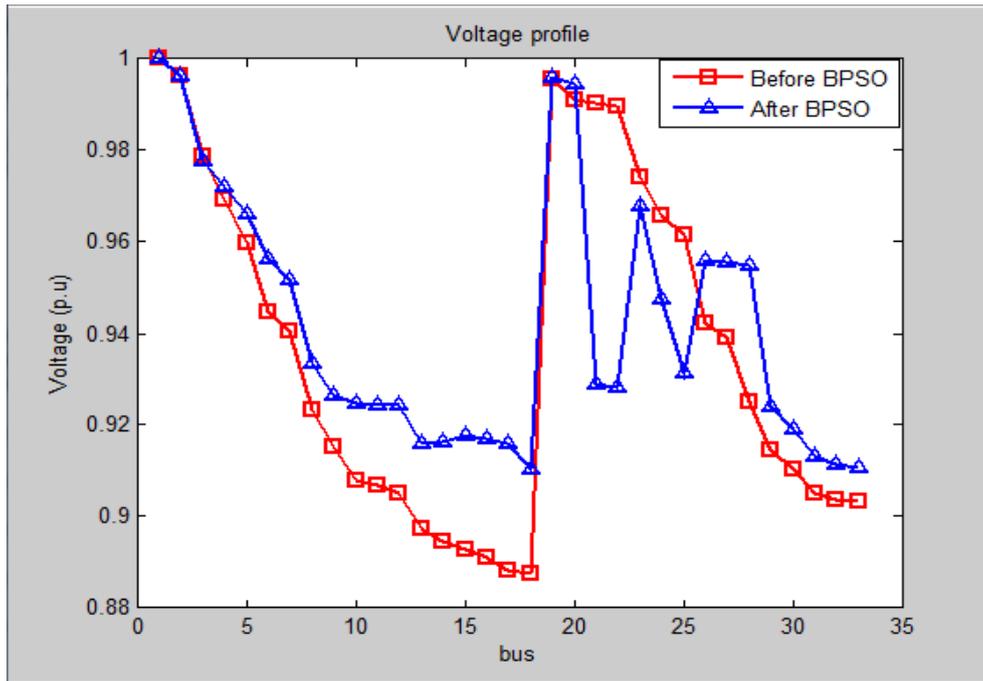
Figure 7 and Figure 8 show voltage profiles of the network while using BPSO and MBPSO. It is clear that the voltage at all nodes (except the nodes 19, 20, 21, 22) were improves after reconfiguration. Finally, it is clear that from Figure 7 and Figure 8 by using MBPSO greatly improves the voltage profile compared to BPSO.

the heavy load ( $\mu = 1.250$ ) demand. IEEE 33–node is presented as test system for both the BPSO and MBPSO approaches. Table 5 describes the comparison among the proposed methods and some other methods reported in the literature [2,9, 10, 14]. Switches status, real power loss, minimum and average voltage are given in this table.

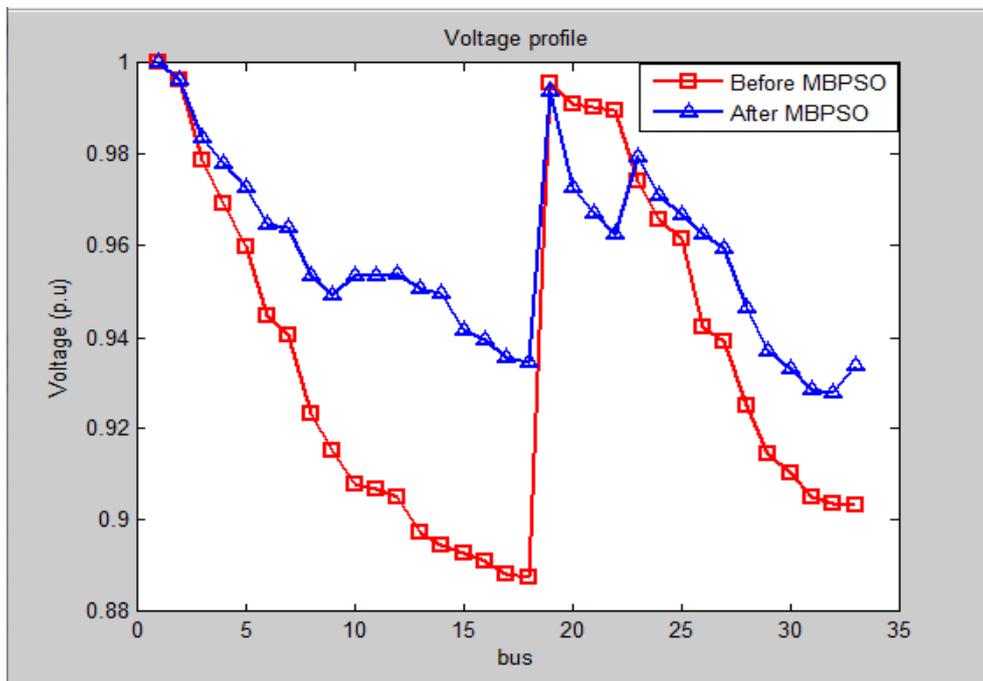
**At load with variation ratio ( $\mu = 1250$ ):** in this case DNR is applied to

**Table 5:** Result of DNR for the IEEE 33–node power system at heavy load demands ( $\mu = 1.250$ ) while using BPSO, MBPSO and some other approaches.

Approach	Open Switches	$P_{LT}$ (KW)	$V_{av}$ (p.u.)	$V_{min}$ (p.u.)
Initial	sw33, sw34, sw35, sw36, sw37	329.85	0.9342	0.8889
BPSO	sw7, sw9, sw13, sw32, sw37	305.81	0.9420	0.9102
MBPSO	sw7, sw9, sw14, sw32, sw37	216.24	0.9581	0.9279
FWA [2]	sw7, sw9, sw14, sw28, sw32	224.25	0.9587	0.9256
MBFOA [9]	sw7, sw9, sw13, sw32, sw37	227.52	0.9593	0.9211
ITS [10]	sw07, sw09, sw14, sw36, sw37	228.08	0.9561	0.9156
SLR [14]	sw07, sw10, sw14, sw36, sw37	228.92	0.9558	0.9156



**Figure 9:** Voltage profile of DNR for the IEEE 33–node power system at heavy load demands ( $\mu$  1.250) while using BPSO.



**Figure 10:** Voltage profile of DNR for the IEEE 33–node power system at heavy load demands ( $\mu$  1.250) while using MBPSO.

It is seen from Table 5, the real power loss ( $P_{LT}$ ) reduces while using BPSO by 6% from 329.85 kW to 305.81 kW and with MBPSO by 34% 0.8889 p.u. to 0.9102p.u. and with MBPSO improves from 0.8889 p.u. to 0.9279 p.u., while the average voltage enhances from 0.9342 p. u. to 0.9420 p.u. while using BPSO and from 0.9342 p.u. to 0.9581 p.u. with MBPSO.

Figure 9 and Figure 10 show voltage profiles of the network while using BPSO and MBPSO. It is clear that that the voltage at all nodes (except the nodes 19,20,21,22) were improves after reconfiguration. Finally, it is clear that from Figure 9 and Figure 10 by using MBPSO greatly improves the voltage profile compared to BPSO.

## References

- [1] S. Kalambe, and G. Agnihotri, "Loss minimization techniques used in distribution network: Bibliographical survey," *Renewable Sustainable Energy Reviews*, Vol. 29, 2014, PP.184–200.
- [2] A. Mohamed Imran, and M. Kowsalya "A new power system reconfiguration scheme for power loss minimization and voltage profile enhancement using fireworks algorithm," *Electrical Power and Energy Systems*, Vol. 62, 2014, pp. 312–322.

from 329.85 kW to 216.24 kW. The minimum voltage ( $V_{min}$ ) while using BPSO enhances from

## Conclusion

In this study, (BPSO) and Modified BPSO (MBPSO) have been presented as powerful tools to find the optimal DNR. The problem here was formulated as a non-linear problem based on the decreasing of real power loss has been invested as an objective function that is subjected to a set of constraints. The results for the IEEE node 33 power systems demonstrated that MBPSO algorithm has high ability and effective in reduce power loss and voltage profile enhancement of the system compared to BPSO and other results in the papers that reported in the literature.

- [3] E. Carpaneto, G. Chicco, and J. S. Akilimali, "Branch current decomposition method for loss allocation in radial distribution systems with distributed generation," *IEEE Trans. Power Systems*, 2006; 21(3): 1170-79.
- [4] H. Kim, Y. ko, and K. H. Jung, "Artificial neural network based feeder reconfiguration for loss reduction in distribution systems," *IEEE Transactions on Power Delivery*, Vol. 8, No. 3, 1993 pp. 1356-1366.
- [5] T. Taylor, and D. Lubkeman, "Implementation of

- heuristic search strategies for distribution feeder reconfiguration,” *IEEE Transactions on Power Delivery*, Vol. 5, No. 1, 1990, pp.239-246.
- [6] M. E. Baran, and F. F. Wu, “Network reconfiguration in distribution systems for loss reduction and load balancing,” *IEEE Trans. Power Delivery*, vol. 4, No. 2, , 1989, pp. 1401-1407.
- [7] H. D. Chiang, and R. J. Jumeau, “Optimal network reconfigurations in distribution systems: Part 1: A new formulation and a solution methodology,” *IEEE Transactions on Power Delivery*, Vol. 5, No. 4, , 1990, pp. 1902-1909.
- [8] G. K. V. Raju, and P. R. Bijwe, “Efficient reconfiguration of balanced and unbalanced distribution systems for loss minimization,” *IEE Proc. Gener. Transm. Distrib.*, 2008; 2(1): 7-12.
- [9] S. Naveen, K. Sathish Kumar, and K. Rajalakshmi, “Distribution system reconfiguration for loss minimization using modified bacterial foraging optimization algorithm,” *Electrical Power and Energy Systems*, Vol. 69, 2015, 90–97.
- [10] T. T. Nguyen, A. V. Truong, “Distribution network reconfiguration for power loss minimization and voltage profile improvement using cuckoo search algorithm,” *Electrical Power and Energy Systems*, Vol. 68, 2015, pp. 233–242.
- [11] C. T. Su, and C. S. Lee, “Network reconfiguration of distribution systems using improved mixed-integer hybrid differential evolution,” *IEEE Transactions on Power Delivery*, 2003; 18(3): 1022-1027.
- [12] T. Kennedy, and R. Eberhart, "Particle swarm optimization", In *Proc. of the IEEE international conference on neural networks*, 1995; 1942-48.
- [13] R. C. Eberhart, and J. Kennedy, “A discrete binary version of the particle swarm algorithm,” *Proc. of IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 5, 1997, pp. 4104-4108.
- [14] E. Dall’Anese, and G. B. Giannakis, “Sparsity-Leveraging reconfiguration of smart distribution systems,” *IEEE Transactions on Power Delivery*. 2014; 29(3): 1417–1426.

## False alarm reduction for Network Intrusion Detection System by using Decision Tree classifier

*Sarah Mohammed Shareef*

*Computer science department of university of technology. Baghdad-Iraq*

[sarahshareef84@gmail.com](mailto:sarahshareef84@gmail.com)

*Dr. Soukaena Hassan Hashim*

*Computer science department of university of technology. Baghdad-Iraq*

[soukaena.hassan@yahoo.com](mailto:soukaena.hassan@yahoo.com)

### Abstract

Nowadays, Network security is one of the challenging issues with the rapid growth in information technology, this subject leading people to become increasingly aware of the threats to personal privacy through computer crime. Therefore, there is important to create intrusion detection system to detect malicious activities and various attacks on the internet with elevated detection rate and minimal false positive alarm. This paper proposed Network Intrusion Detection system using Decision Tree algorithm. To detect and classify attacks into four categories (DOS, Probe, R2L, U2R). The KDDcup99 dataset has been used to evaluate the activity of proposition system. The experimental results showed that the proposed system provides better results with high detection rate in experiment 1 (99.95%), experiment 2 (97.8%) and low false alarm rate in experiment 1 (0.05%), experiment 2 (2.2%).

**Keywords:** NIDS, alarm reduction, Kddcup99 dataset, Decision Tree

### تقليل الإنذار الكاذب لنظام كشف التطفل الشبكي باستخدام مصنف شجرة القرار

سكينة حسن هاشم

سارة محمد شريف

قسم علوم الحاسوب، الجامعة التكنولوجية. بغداد-العراق

#### الخلاصة

في الوقت الحاضر، مع النمو السريع في تكنولوجيا المعلومات أصبحت أمنية الشبكات واحدة من القضايا الصعبة مما جعل المستخدمين بان يكونوا على وعي متزايد من التهديدات للخصوصية الشخصية من خلال جريمة الكمبيوتر. لذلك، هناك أهمية لخلق نظام كشف تطفل للكشف عن الأنشطة الخبيثة والهجمات المختلفة على شبكة الإنترنت مع ارتفاع معدل الكشف وانخفاض إنذار إيجابي كاذب. هذا البحث اقترح نظام كشف تطفل شبكي باستخدام خوارزمية شجرة القرار. للكشف وتصنيف الهجمات إلى أربع فئات (DOS، Probe، R2L، U2R). لتقييم أداء نظام الاقتراح، تم استخدام بيانات KDD cup 99. وأظهرت النتائج التجريبية أن النظام المقترح يوفر نتائج أفضل مع معدل كشف عال ومعدل إنذار كاذب منخفض.

**الكلمات المفتاحية:** نظام كشف التطفل الشبكي، تقليل الإنذار، مجموعة بيانات kddcup99، شجرة القرار.

## 1. Introduction

Today it is so serious to supply elevated level security to preserve highly critical and specific information. Intrusion Detection System is a fundamental technology in Network Security. These days many researchers have concerned on intrusion detection system utilizing Data mining mechanisms as an artificial proficiency [1].

Intrusion Detection System (IDS): is an appliance or software which checks network or device liveliness about bad activities and generates reports to an administration Station [2]. The techniques of IDS are divided in two categories: first one is Anomaly established on intrusion detection system was an equipment which detecting device malicious based on the "normal user profile" for utilized as a baseline and classifying it like each normal and abnormal. Second one is Misuse established on intrusion detection system was known as signing up -based detection because alerts have been created based on definite attack signing up [3].

Feature selection is the most significant preprocessing of data mining manners that utilized to recognize the irrelevant and abundant information and removing them as much as possible. Features can be defined as discrete, continuous or nominal. In general, features were identified below [4]:

**1- Relevant** : it mentions to the features that one have effectiveness on the product

and their function cannot be supposed by the remainder.

**2- irrelevant** : it mentions to the features are specified as those features not holding every effectiveness on the output, and that values are formed at random for every symbol.

**3- Redundant** : the redundancy is occurred whenever a feature can hold the function of else.

Classification is data mining mechanism which is token every case of a dataset under sight and it is a supervised machine learning technique so it can touch classified data. . A classification based intrusion detection system will assort the entire network passing into either normal or abnormal. There are different classification techniques for example decision tree [5].

**A Decision Tree (DT)** is defined as a predictive modeling technique from the subfield of machine learning within the large field of artificial intelligence. It uses divide and conquer method for splitting according to attribute values. One of the most different decision tree algorithms are described as ID3 [6].

**The ID3 algorithm** is the basic algorithm of decision tree induction, it produces decision tree by means of compulsion in detail from the top to the bottom. It is used to construct the classification rules in the form of decision tree. This is utilized Shannon's entropy (ent) like a standard of choosing the significant attribute [7] [8]:

$$Entropy(ent) (s) = \sum_{i=1}^c - p_i \log_2 p_i \quad (1)$$

Where:

$p_i$  considered the rate of patterns belong to the  $i$  th kind.

Information gain is generally utilized to determine the property for each node of generated decision tree by selecting the best feature at every step of rising a decision tree (DT). Information gain is calculates anticipated reduction within entropy occasion by awareness the amount of feature  $F_j$ . is utilized:

$$info\ gain(S, F_j) = Entropy(s) - \sum_{V_i \in V_{F_j}} \frac{|S_{V_i}|}{|S|} \cdot Entropy(S_{V_i}) \quad (2)$$

Where:

$V_{F_j}$  was represented of whole potential amounts of attribute ( $F_j$ ), ( $S_{V_i}$ ) is a subset of ( $S$ ) about that feature ( $F_j$ ) has value( $V_i$ ).

## 2. Related work

A survey has been achieved latest papers which implement training and testing of the system based on decision tree.

**Anuar N.B et al., 2008**, Design an organization to concentrate on discovery including statistical test of both anomaly and normal traffic instituted on KDDCup99 dataset, again the design involve a hybrid statistical proposition utilizes decision tree of data mining classification, The proposal proves that the decision tree for designing intrusion detection system is more suitable and accuracy than rule-based classifier [9].

**Mukund Y.R et al., 2016**, proposes the present mechanism for intrusion detection system to inform afflicted way of employing the HDFS (Hadoop Distributed File System) of machine learning algorithms, so to minimize the rate of false alarm, they were used decision tree technique and augment it in the operation with the multi-device capacity of the HDFS, therefore this approach was reduced the time taken by the DFS and improved the accuracy of the IDS [10].

**Elekar K. SH. et al., 2015**, executes various classifiers like C4.5 decision tree, Random Forest and Hoeffding Tree of intrusion detection then match the outcome

utilizing WEKA. Outcomes display that a Hoeffding Tree awards superior result between several classifiers for distinguishing anomalies by the test data [11].

**Xiang ch. et al., 2008**, suggested system Design (hybrid classifier of multi-level for intrusion detection system) using Bayesian clustering and decision tree. Detection rate can be increased by implementing a fresh multi-level intrusion hybrid classifier. A model with 4 stages of classification is utilized for the metis classifier. The first level classifies the test data at 3 divisions (Denial of service, Probe, and Others). User to Root attack and Remote to Local attack and the Normal dealings are labeled as others in this level. Second level divides others into anomaly and Normal classes, while third level dis connects the Attack class from level 2 at User to Root attack and Remote to Local. Furthermore the fourth level classifies the attacks at more particular attack kinds [12].

## 3. Dataset Description

The KDD cup 99 dataset has been the point attraction for many researchers for evaluates intrusion detection algorithms [13]. It was prepared by Stolfo et.al. (Salvatore J.S., 2000). This dataset was consisted Tcp connections; each connection has 41 features with labeled

determine the type of a connection either normal connection or type of attack connection. The artificial attacks breakdown in the following four divisions (see table 1) [14]:

A. Access or Denial of Service Attack (DoS): it mentions to an attack that an intruder creates some counting or memory resorts turn on to shaft legal demands, or dismisses legal users' incoming to computer.

B. User to Root Attack (U2R): mentions to the category of deed that the intruder set out with incoming to normal employer computation at the instrument (maybe obtained by sniffing passwords) and was

capable deeds some sensibility to earn origin incoming to the instrument.

C. Remote to Local Attack (R2L): happens when an intruder which has the capacity for transmitting packets to the computer through network but who does not have counting on which computer deeds several sensibility to acquire native access like a user of that computer.

D. Probing Attack: was a trial to collect data on the network of devices for visible intent of compassing its security dominance. Table 1 displays the four divisions and their corresponding attacks on every category.

**Table 1:** attacks Description of KDDCup99 Datasets

(4main) Attack type	Description	Attack classes
DOS	Denial of Service attacks	Neptune, apache2, back,udpstorm, teardrop.
Probe	observation and probing	Satan, ip-ssweep, saint, Mscan, port-sweep.
R2L	Forbidden access from remote instrument to local instrument	Named, Xlock, send-mail,warezclient.
U2R	Forbidden access to local user priviledges by a local unpiviledge user	perl, spy, worm, Xterm, Http-tunnel.

#### 4. Design of the proposal process

The suggested system is Network intrusion Detection System model depend on ID3 classifier to detect four types of attacks that threatened the machines security, and to reduce false alarm rates in IDS by using well-Known dataset KDD Cup 99 datasets. Fig. (1) Describe the general structure of the NIDS Model.

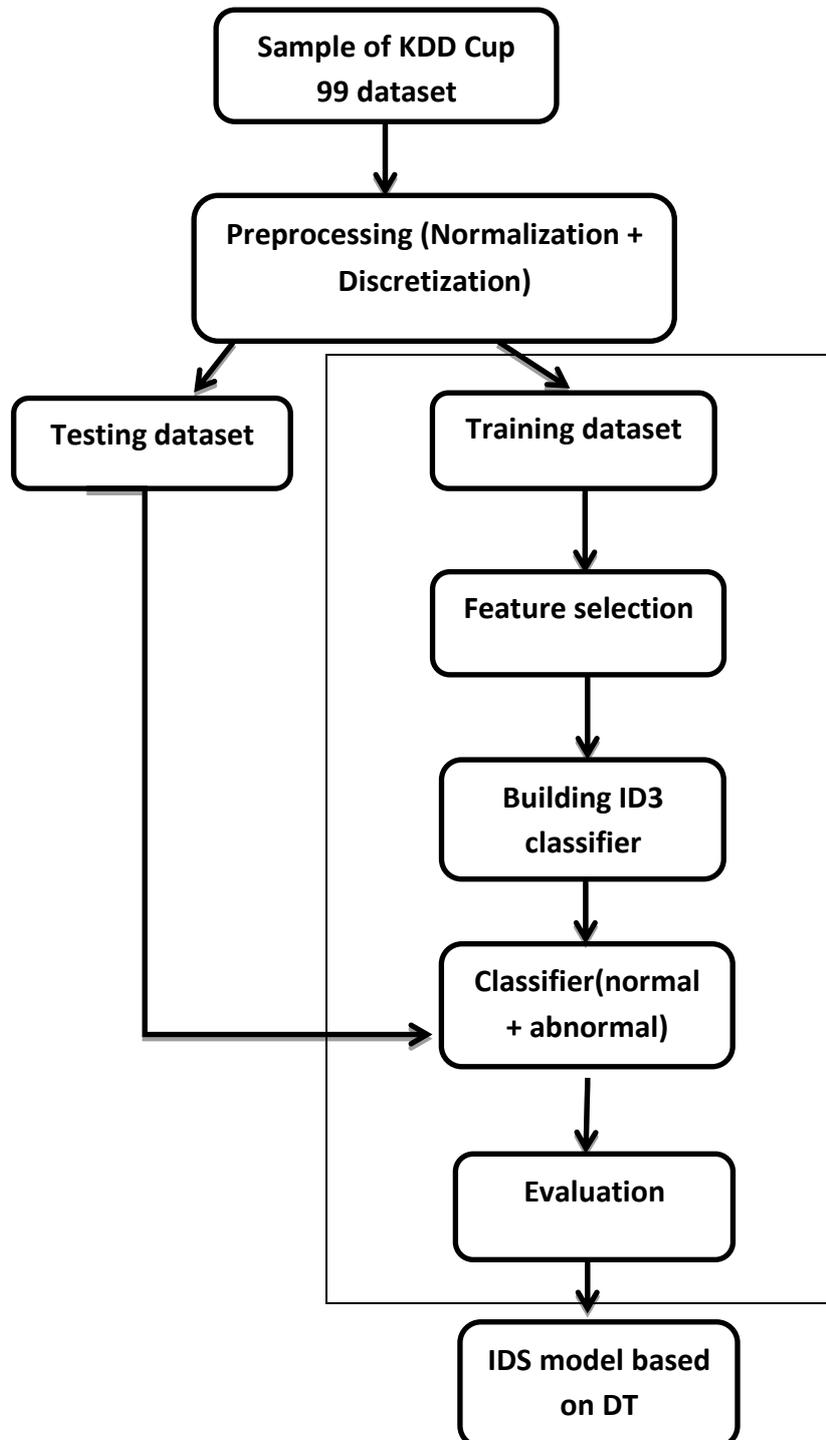


Figure 1. Block diagram for the proposed system

The proposed system illustrates in the following steps as shown in algorithm (1):

<p><b>Algorithm (1) : The proposed system</b></p> <p><b>Input:</b> Training dataset</p> <p><b>Output:</b> classify the type of attack samples from normal behavior</p> <hr/> <p>Begin</p> <p>Steps:</p> <ol style="list-style-type: none"> <li>1- Preprocessing (Normalization and Discretization) of the selected subsets of samples.</li> <li>2- Select subsets of samples for training and testing phases from Kdd99 dataset.</li> <li>3- Feature selection method for reduction the redundant and irrelevant information.</li> <li>4- Building ID3 classifier and training it by using the samples of training set.</li> <li>5- Test the ID3 classifier using the testing samples.</li> <li>6- Evaluate the performance of trained module.</li> </ol> <p>End.</p>
---

#### 4.1 Dataset preprocessing

Data preprocessing is a major and essential stage to obtain final datasets that can be considered beneficial for further data mining algorithms. In this proposal there are two types of data preprocessing are display as follows:

**a) Normalize Dataset**

First point after gaining the dataset from internet traffic, normalization process was applied to upgrade the action and efficiency of the system by scaling the accounts of feature during a small specified range [0 to 1].this proposal applied the normalization operation. Vision algorithm (2):

<p><b>Algorithm (2) preprocessing of normalized dataset</b></p> <p><b>Input:</b> all Datasets of Continuous feature</p> <p><b>Output:</b> values of dataset between 0 and 1</p> <hr/> <p><b>Steps:</b></p> <ol style="list-style-type: none"> <li>1. For every attribute in dataset</li> </ol>
--

```

compute the maximum value (max)
compute the minimum value (min)
for each value v in attribute

$$V' = \frac{V - \min_A}{\max_A - \min_A}$$


```

```

End For
End For

```

**b) Discretization dataset**

Data Discretization was one of the basic reduction techniques adversary data preprocessing. In KDD cup 99 datasets contains continuous and discrete feature so it is serious to transform the continuous feature to discrete ones for guarantee the activity of the system. Discretization techniques are classified into: supervised and unsupervised discretization based on how it is performed, if the class information is utilized by the discretization operation then supervised is said. Otherwise it is unsupervised discretization.

To improve the effectiveness of the system reducing the consuming time must be used Feature selection technique for recognizing the irrelevant and redundant feature and removing them as much as possible. Feature selection techniques such as information gain, relief, gain ratio and the proposed system will be used entropy as feature selection.

**4.2 Feature Selection Methods**

**4.3 The ID3 algorithm**

ID3 algorithm is the basic algorithm of decision tree induction; it is used to construct the classification rules in the form of decision tree. Vision algorithm (3).

**Algorithm(3) : ID3 classifier****Input:** number of samples selected from KDD99 dataset (training dataset)**Output:** set of classification rules**Steps:**

- 1- For every class  $c$  in training sample
  - Calculate  $p(c)$  from training sample
  - Compute the entropy to all training dataset using Eq.1
 end for.

- 2- For every attribute  $F$  in training sample using Eq.1

- Calculate the entropy

$$Entropy(s) = \sum_{i=1}^c - p_i \log_2 p_i$$

- Compute the Info gain using Eq.2

$$info\ gain(S, F_j) = Entropy(s) - \sum_{v_i \in V_{F_j}} \frac{|S_{v_i}|}{|S|} \cdot Entropy(S_{v_i})$$

- Find the largest info gain  
Repeat until all entry values are empty.

- 3- If all classes are the same, then stop: decision tree has one node  
Else goto on step 2

- 4- For every record in testing data:

- 1- Max=0, Class=""
  - 2- For every Rule in training rule do steps 3,4
  - 3- calculate Match that is a number of Rule conditions which is matched by record
  - 4- If Match > Max  
Then Max=Match, Class=class label of Rule
  - 5- class of record is allocate by class label of Rule
- End for  
End for

End if  
End for

#### 4.4 Training and Testing of the proposed system:

In the learning stage the system used ID3 classifier on 4000 records for training operation by choose 1000 DOS, 700 probes, 200 R2L, 100 U2R and 2000 normal in KDDcup99 dataset.

In test stage 2000 samples are utilized to evaluate the work in KDDcup99 Dataset to establish the activity of the system, where the numbers of samples selected for each class demonstrate in (Table 2).

Table (2).dataset description

Number of dataset	Total number of records					
	records	normal	dos	probe	U2R	R2L
Train dataset	4000	2000 50%	1000 25%	700 17.5%	100 2.5%	200 5%
Test1 dataset	1500	337 22.4%	562 37.4%	254 17%	95 6.3%	252 16.8%
Test2 dataset	500	138 27.6%	113 22.6%	152 30.4%	19 3.8%	78 15.6%

### 5. Experiments and Performance Evaluation

The performance of the classifier used can be paralleled according to sure metrics such as accuracy, detection rate, and error rate, the confusion matrix is explained by four values which are TP, FN, FP and TN that shown in Table (3). The parameters are argued below.

True positive (TP): It mentions the number of attack which is detected as attack correctly.

True negative (TN): It mentions that number of normal which is predicted as normal correctly.

False negative (FN): It mentions the number of attack which is detected as normal correctly.

False positive (FP): It mentions that number of normal which is predicted as attack correctly.

$$(ACC) = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$(DR) = \frac{TP}{TP+FN} \quad (4)$$

$$(FAR) = \frac{FP}{TN+FP} \quad (5)$$

**Table (3): The confusion matrix**

Actual class	Class Predicted	
	Normal class	Abnormal class
normal	(TN) True negative	(FP) False positive
abnormal	(FN) False negative	(TP) True positive

The KDDcup99 dataset used for training and testing ID3 algorithm, the suggested ID3 algorithm applied to classify dataset into five classes, (Normal, Dos, Probe, R21, U2r). In the training phase (4000) records are elected randomly from whole dataset and used for training the algorithm. This subset of records contain normal and all other types of attack. To evaluate effectiveness of the proposed algorithm will conduct two experiments. In experimental 1 the trained model tested with (1500) subset of data of records contains both normal behavior and the four types of attacks. In experimental 2 subset consist of (500) record contains both normal and attack samples used to evaluate the proposed module. The subsets of data used in this work illustrated in table

(2). Various performance measures used to evaluate the proposed module such as detection rate (DR), false alarm rate (FAR) and accuracy (ACC). The result obtain from testing phase show the high capability of proposed algorithm to distinguish normal activities from attack activities where the result from experiment 1 show effectiveness of the module to detect the attack behavior with detection rate reach to (99.95%), low false alarm rate reach to (0.05%) and accuracy of system is (98%). In experiment 1 the time for building model is (0.46) second and for testing model is (0.18) while In experiment 2 the time for building model is (0.35) second and for testing model is (0.13). The result from the two experiments shows in table (4) and figure 2.

**Table (4) the experimental result**

Performance measure	Exp1	Exp2
DR	99.95%	97.8%
FAR	0.05%	2.2%
Accuracy	98%	98%
True positive TP	99%	97%
False positive FP	0.4	0.5

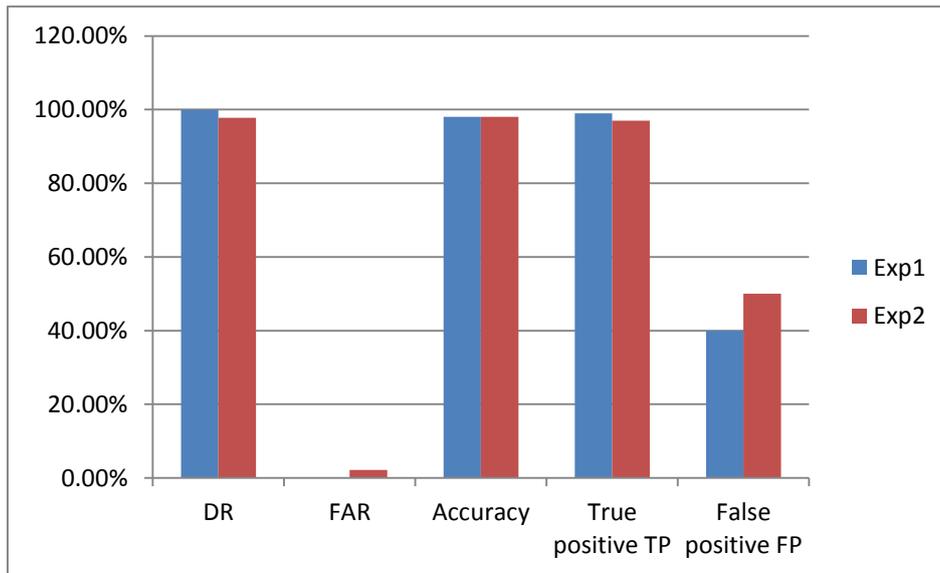


Figure .2: The results of performance measure

## 6. Conclusion

With the fast development in field of computer security, data mining mechanisms are also one of ways which security can be supplied to the network. In this paper network intrusion detection system based ID3 classifier algorithm proposed to detect more specific attack types from normal behavior. Data mining (DM) can assist to upgrade intrusion detection by processing every mentioned problem. In this work ID3 classifier algorithm used to classify the dataset into five classes. One for normal and another for attack behavior .The proposed module consist of two phase; Training phase where the module trained with dataset to be capable to distinguish normal behavior from attack event; Testing phase where new unseen samples utilized to evaluate the performance of module. The KDD99 dataset used to train and test the proposed module.experimental result display the effectiveness of system in detects attack and define normal behavior with 98% accuracy, high detection rate and low false alarm as shows in table (4).

## References:

- 1- Khan J.A. and Jain N., "A Survey on Intrusion Detection Systems and Classification Techniques", IJSRSET, vol 2, Issue 5, print ISSN: 2395-1990, online ISSN: 2394-4099,2016.
- 2- Kumar A., Maurya H.Ch. and Misra R., "A Research Paper on Hybrid Intrusion Detection System", International Journal of Engineering and advanced technology (IJEAT) ISSN: 2249-8958, vol-2, Issue-4, April 2013.
- 3- Gupta M., "Hybrid Intrusion Detection System: Technology and Development", International Journal of Computer applications(0975-8887), vol 115-No.9, Apr 2015.
- 4- Ladha L. and Deepa T., "FEATURE SELECTION METHODS AND ALGORITHMS", International journal on computer science and engineering (IJCSSE) , ISSN:0975-3397, vol.3, No.5, May 2011.
- 5- Patil S.S., prof Kapgate D. and prof Prasad P.S., "A Review on Detection of Web Based Attacks Using Data Mining

Techniques", International Journal of Advanced Research in computer science and software engineering, ISSN:2277 128X, vol 3, Issue 12, December 2013.

6- Rajakshmi S.Ph.D and Shanthini J.S., "DATA MINING TECHNIQUES FOR EFFICIENT INTRUSION DETECTION SYSTEM: A SURVEY", International Journal on engineering technology and sciences- IJETS, ISSN(p):2349-3968, ISSN(0): 2349-3976, vol II, Issue XI, Nov 2015.

7- Hashem S.H. and Abdulmunem I.A., "A Model to Detect Denial of Service Attack Using Data Mining Classification Algorithms", thesis for master in computer science, june 2013.

8- Essa A.S., Orman Z. and Brifcani A.M.A., "A New Feature Selection Model based on ID3 and Bees Algorithm for Intrusion Detection System", Turkish Journal of Electrical engineering and computer sciences, Doi:10.3906/e1k-1302-53, 2015.

9- Anuar N.B., Sallehudin H., Gani A. and Zakari O., "Identifying False Alarm for Network Intrusion Detection System Using Hybrid Data Mining and Decision Tree", Malaysian journal of computer science, vol.21 (2), 2008.

10- Mukund Y.R., Nayak S.S and Chandrasekaran K., "Improving False Alarm Rate in Intrusion Detection Systems using Hadoop", 2016 Intl. Conference on advance in Computing, communications and Informatics (ICACCI), India, Jaipur, sept.21-24, 2016.

11- Elekar K.Sh. and prof. Waghmare M.M., "Comparison of Tree Base Data Mining Algorithms for Network Intrusion Detection", International Journal of Engineering, Education and technology (ARDJEET), ISSN 2320-883X, vol 3, Issue 2, 01/04/2015.

12- Xiang Ch., yong P.Ch. and Meng L.S., "Design of Multiple-level Hybrid Classifier for Intrusion Detection System Using Baysian Clustering and detection Trees", Elsevier, pattern recognition letters 29(2008) 918-924, doi:10.1016/j.patrec.2008.01.008.

13- Vijayarani S. and Maria S.S., "Introduction Detection System – Astudy", International Journal of Security, Privacy and Trust Management(IJSPTM) vol 4, No 1, Feb 2015.

14- Tavallae M., Bagheri E., Lu W. and Ghorbani A.A, "A Detailed Analysis of the Kdd Cup 99 Data set", proceedings of the 2009 IEE symposium on computational intelligence in security and defense applications (CISDA 2009).

**The effect of noise on digital phase locked loop circuit of second order***Dr. Muhamed Ibrahim Shujaa**Electrical Engineering Technical College**Computer Engineering Department**dralnedawy@yahoo.com***Abstract:**

This paper present a theoretical and experimental work of noise effect on the main performance measure of the second order zero crossing digital phase locked loop (ZDPLL). The loop error probability density function (P.D.F) satisfies the Chapman-kolomogrov equation. From this and the basic equation of approximate expression for the steady state phase error p.d.f. phase error variance and we obtained the loop noise bandwidth on matlab programme. The main measurement used in this paper is value of reliability, phase error variance, probability of correct locked and maximum phase jitter.

**Keyword:** DPLL, PLL, Digital filter, DSP**تأثير الضوضاء على الدائرة الرقمية المغلقة من الدرجة الثانية**

م. د محمد ابراهيم شجاع

الكلية التقنية الهندسية الكهربائية

الخلاصة:

يتناول البحث تحليل نظري معززا بتصاميم عملية على تأثير الضوضاء على مقاييس الاداء الرئيسية لدائرة ضبط الطور الرقمية من الدرجة الثانية عابرة الصفر. ان دالة كثافة الاحتمالية للطور تتبع معادلة جابمان كولومكروف. من هذا ومن المعادلة الاساسية لعمل المنظومة تم اشتقاق معادلات تقريبية (خطية) للطور في حالة الاستقرار معدل الاهتزاز الطوري وعرض حزمة الضوضاء. ان المقاييس الرئيسية التي اعتمدت في البحث هي مقدار الاعتمادية معدل الاهتزاز الطوري، احتمالية الاقفال الصحيحة واعلى اهتزاز في الطور.

**List of symbols:**

A: single amplitude.

AWGN: Additive white Gaussian noise

C-K: Chapman-kolomogrov equation.

G1, G2, r: digital filter coefficients.

ISN: inverse signal to noise.

NO: spectral density.

P(.):probability of.

P(%): conditional probability density function.

P.D.F.: probability density function.

Q(%): transition probability density function.

ZC-DPLL: zero crossing digital phase locked loop.

DTL: digital tank lock loop

DPLL: digital phase lock loop

DCO: digital control oscillator

$\Phi$ : phase error.

$\sigma^4$ : noise variance

$\theta_0$ = phase constant.

W=input frequency

X (t): input signal

A: signal amplitude  $w_0=2\pi f_0$ .

$\theta$ t: information binary pulse.

N (t): Gaussian additive noise.

### 1. Introduction

The performance of second order ZC-DPLL in the presence of additive white Gaussian noise (AWGN) is presented in this paper. The nature of ZC-DPLL makes the statistical analysis of the phase error process obtained by the studding chapman-kolmogorov (C-K) equation [1] associated with the stochastic difference equation governing the phase error [2].In the analog case, the fokker-planch equation is derived from (C-K) equation associated with the statistic differential equation for the phase error process [1].In the following analysis, module  $2\pi$  phase error is chosen i.e. by imagining the phase error warp itself around a circle of radius one.

The C-K equation is used to solve the conditional probability density function for markov process. In the process the present value depend only on the last past value, because the phase error is a random variable following markov process will be seen later. C-K equation is applied to the probability density function (p.d.f) of the phase error.

If the signal is band limited with the bandwidth (  $B_i$ ) then noise can be approximated by sequence of independent and identically likely Gaussian random variables with zero mean and variance [3].

$\sigma^2 n = B_i N$  represented the power spectral density over the frequency range of interest.

Equation of analysing and design of second order ZC-DPLL is rewritten here as:

$$\phi(k + 1) = 2\phi(k) - \phi(k - 1) + k_1 \sin[\phi(k - 1)] + k_1 n(k - 1) - r\{k_1 \sin[\phi(k)] + k_1 n(k)\} \dots \dots \dots (1)$$

Where

A: is signal amplitude.

$K_1 = W G_1 A, k_n = W G_2 A, r = 1 + G_2 / G_1, n(k)$  is the noise component at k instant value ( $G_1, G_2$ ) are digital filter coefficient.

The C-K equation which is stated as follows is applied only for markov process.

$$P_{k+1}(\phi / \phi_0) = \int_{-\infty}^{\infty} q_k(\phi / z) P_k(z / \phi_0) dz \dots \dots \dots (2)$$

Where;

K= time index.

$\phi = \phi_0$ =initial phase error value.

$P_k(\phi / \phi_0)$ =conditional p.d.f of the  $\phi(k)$  given  $\phi_0$

$Q_k(. / z)$ =transition p.d.f of  $\phi(k + 1)$  given  $\phi(k) = z$

The phase error generated from (2) is non markovian in the present form, however by introducing an auxiliary variable [2].

$$u(k+1) = (2-1/r)\phi(k+1) - \phi(k) + u(k)/r \dots\dots\dots(3)$$

then (2) can be written as a system of two equations:

$$\phi(k+1) = -rk_1 \sin[\phi(k) + u(k) - rk_n n(k)] \dots\dots\dots(4.1)$$

$$u(k+1) = -\phi(k) - (2r-1)k_1 \sin[\phi(k)] + 2u(k) - 2r-1)k_n n_k \dots\dots\dots(4.2)$$

$$p_{k+1} \left[ \phi(k+1) = \phi, u(k+1) = \frac{u}{\phi_0 u_0} \right] = \int_{-\infty}^{\infty} qk \left[ \phi(k+1) = \phi, u(k+1) = \frac{u}{\phi} (k) = x, u(k) = Y \right] P_k \left[ \phi(k) = X, u(k) = \frac{Y}{\phi_0, u_0} \right] dXdY \dots\dots\dots(5)$$

In this form the 2- vector  $(\phi(k+1), u(k+1))$  is markovian, hence direct method of C-k equation can be applied [2].

From (4) its obvious that the kth p.d.f is free (independent) of the index k. i.e

$$\begin{aligned} & q[\phi(k+1) = \phi, u(k+1) = \frac{u}{\phi} (k) X, u(k) = Y] \\ & = q [u(k+1) = u/\phi(k+1) = \phi, \phi(k) = X, u(k) = Y]. \\ & q[\phi(k+1) \frac{\phi}{\phi} (k) = X, u(k) = Y \\ & = \delta \left\{ u - \left[ \left( 2 - \frac{1}{r} \right) \phi - X + \frac{Y}{r} \right] \right\} \dots\dots\dots(6) \end{aligned}$$

$$\frac{1}{\pi r \sqrt{2}} \int_{-\infty}^{\infty} \exp \left[ \frac{(2\sigma^2 - u - X + k_1 \sin X)^2}{2\delta^2} \right] \dots\dots\dots(7)$$

$$P_u [\phi(k) = X, u(k) = (2r-1)\phi + ru + rx] dx$$

Where the conditioning on  $\phi_0$  and  $U_0$  has been dropped for simplicity in notation.

After taking expectation and letting  $k \dots\dots > \infty$  in (4)

$$E[\sin\phi] = \phi, E[\phi] = E[u] \dots\dots\dots(8)$$

Where  $E[\dots]$  representing expectation squaring and cross multiplying (4) taking expectation and finally letting  $k \dots\dots > \infty$ , the following system of equation are obtained.

$$E[\phi^2] = E[(-rk_1 \sin\phi + u)^2] + r^2 k_n^2 2\sigma n^2 \dots\dots\dots(9.1)$$

$$E[u^2] = E[(-\phi - (2r - 1)k_1 \sin\phi + 2u)^2 + (2r - 1)^2 k_n^2 \sigma_n^2] \dots \dots \dots (9.2)$$

$$E[\phi u] = E[(-rk_1 \sin\phi + u)(-\phi - (2r - 1)k_1 \sin\phi + 2u)] + r(2r-1)k_n^2 \sigma_n^2 \dots \dots \dots (9.3)$$

When the DPLL is in the tracking mode and the phase error is small with high probability (high SNR case) then  $\sin\phi \approx \phi$ .if the approximation is used therefore.

$$E[\phi^2] = \frac{(r^2+1)(2-k_1)-2r(2-rk_1)}{C} k_n^2 \sigma_n^2$$

$$E[u^2] = \frac{(5r^2-4r+1)(2-k_1)-2r(2r-1)(2-rk_1)}{C} k_n^2 \sigma_n^2$$

$$E[\phi u] = \frac{2r^2(2-k_1)-(r^2+2r-1)(2-rk_1)}{C} k_n^2 \sigma_n^2 \dots \dots \dots (10)$$

Where  $C=k_1[2 - k_1]^2 - (2 - 2k_1)^2]$

Using (10) for the steady state carapace  $\sigma^2 = E(\phi^2)$ .then

$$\sigma^2 = \frac{1}{k_1^2} \left[ \frac{2}{\frac{4}{r+1} - k_1} - 1 \right] k_n^2 \sigma_n^2 \dots \dots \dots (11)$$

The linear loop noise bandwidth W after simplifying (11) is given by

$$\frac{Wl}{Bi} = \frac{2}{\frac{4}{r+1} - k_1} - 1 \dots \dots \dots (12)$$

Equation (11) can be further simplified to

$$\sigma^2 = \left[ \frac{2}{\frac{4}{r+1} - k_1} - 1 \right] \frac{\sigma n^2}{A^2} \dots \dots \dots (13)$$

Thus p.d.f can be re written as:

$$\text{p.d.f. } (\phi) = P(\phi) = \frac{1}{\sigma\sqrt{2\pi}} \int \exp\left(\frac{-\phi^2}{2\sigma^2}\right) d\phi \dots \dots \dots (14)$$

the optimum value of the filter parameters  $k_i, r$  minimum  $4/(rH)$ . At the same time  $k_i$  and  $r$  must also satisfy the constraints  $0 < k_i < 4/(rH)$ . It's difficult to minimize  $\sigma^2$  by differentiations. However, the digital computer can be easily programmed to search for the minimum mean square error as both parameters are varied within the mentioned constraints [4].

**2. Zero crossing DPLL**

DPLL type receives sinusoidal signal and sample an input of one to nearby zero. That's why it's called zero-crossing (ZC-DPLL).

There are two kind of (ZC-DPLL), one is (Z1-DPLL) which works on positive-going crossing, and the (ZC2-DPLL) which works on both positive and negative.

(Z1-DPLL) is the more important and simplest for implementation where its interpretation are ineffectual in general behaviour of any DPLL [5], hence (ZC2-DPLL) is faster but complicated.

(Z1-DPLL) has been developed in works [3,6], it present numerical solution for (Chapman-Kolomogorov equation) [7]. DPLL was improved in 1982 by advent of digital (tank lock loop) DTL [1], where (Z1-DPLL) based on phase detection.

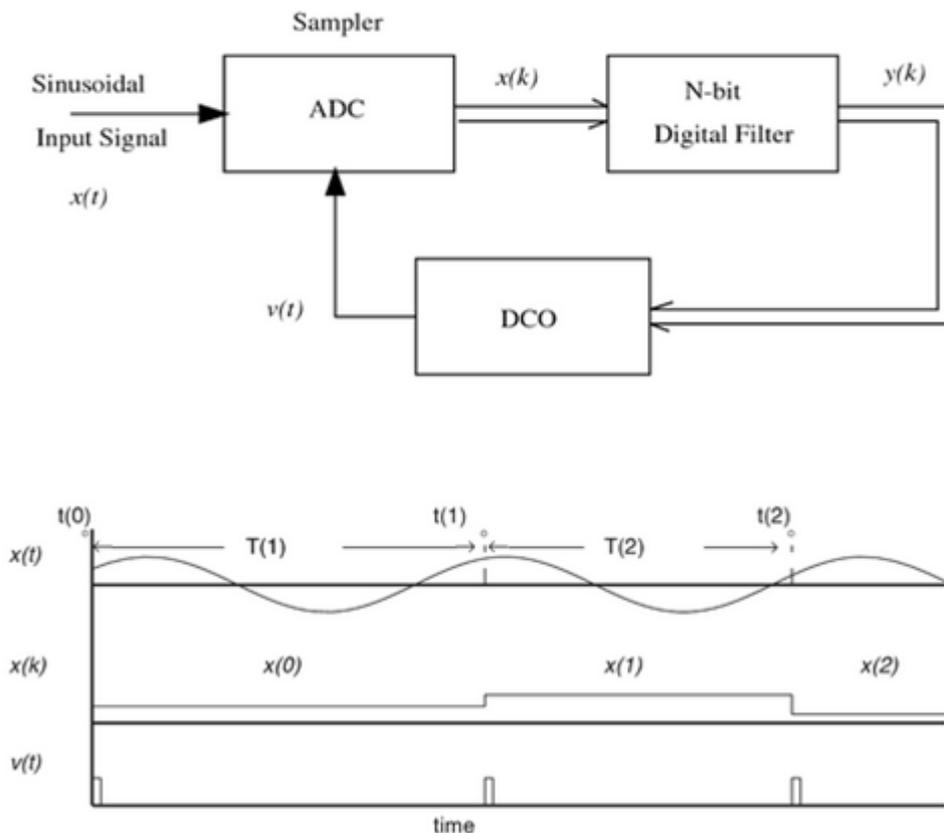


Fig 1: (ZC1-DPLL) sinusoidal with associated waveform

### 3. Digital Controlled Oscillation (DCO)

(DCO) consist of a programmable counter, binary subtract and zero detector. Subtraction is done by using second complement and full adder, while counter is decreased by only one of each clock. When it reach zero, counter gives pulse on output. Where it loads counter with binary number ( M-K), ( M-constant, K-input) M-declare (DCO) free-running frequency  $f_0$  where  $K=0$  as:

$$f_0 = f_c/m \dots \dots \dots (15)$$

$f_c$ = counter frequency clock

Period time between  $(K-1)^{th}$  and  $K^{th}$  obtained by:

$$T(K) = (M-K) T_c \dots \dots \dots (15.1)$$

$$T_c = 1/f_c$$

Pulse equation can be obtained by :

$$X(t) = A \sin \{ w_o t + \theta(t) + n(t) \} \dots \dots \dots (15.2)$$

$X(t)$  : input signal

A: signal amplitude  $w_o = 2\pi f_o$ .

$\theta t$ : information binary pulse.

$N(t)$  : Gaussian additive noise.

$$\text{Frequency steps input: } \theta(t) = (w - w_o) t + \theta_o \dots \dots \dots (15.3)$$

$\theta_o$  = phase constant.

$W$  = input frequency (first order loop.)

### 4. System Design and proposed methods:

The experimental system suggested in the work [1] is used to study the behaviour of the loop within the noise. A up system is used to store the phase error sample in its memory, and used later to plot the p.d.f. of the phase error simulated by matlab programming .the Gaussian random variants ( $n_k$ ) are generated in three steps:

1. Generating a pair of random values  $X_1, X_2$  uniformly distributed over (0, 1) [1].

2. Generating the pair:

$$Y_1 = -21n x_1 \cos(2\pi x_2)$$

$$Y_2 = -21n x_1 \sin(2\pi x_2)$$

Which are independent and Gaussian distributed with zero mean and unit variance [1].

- Scaling the variance to the appropriate value according to the S/N value. The probability density function (p.d.f.) is determined by the following equation [1].

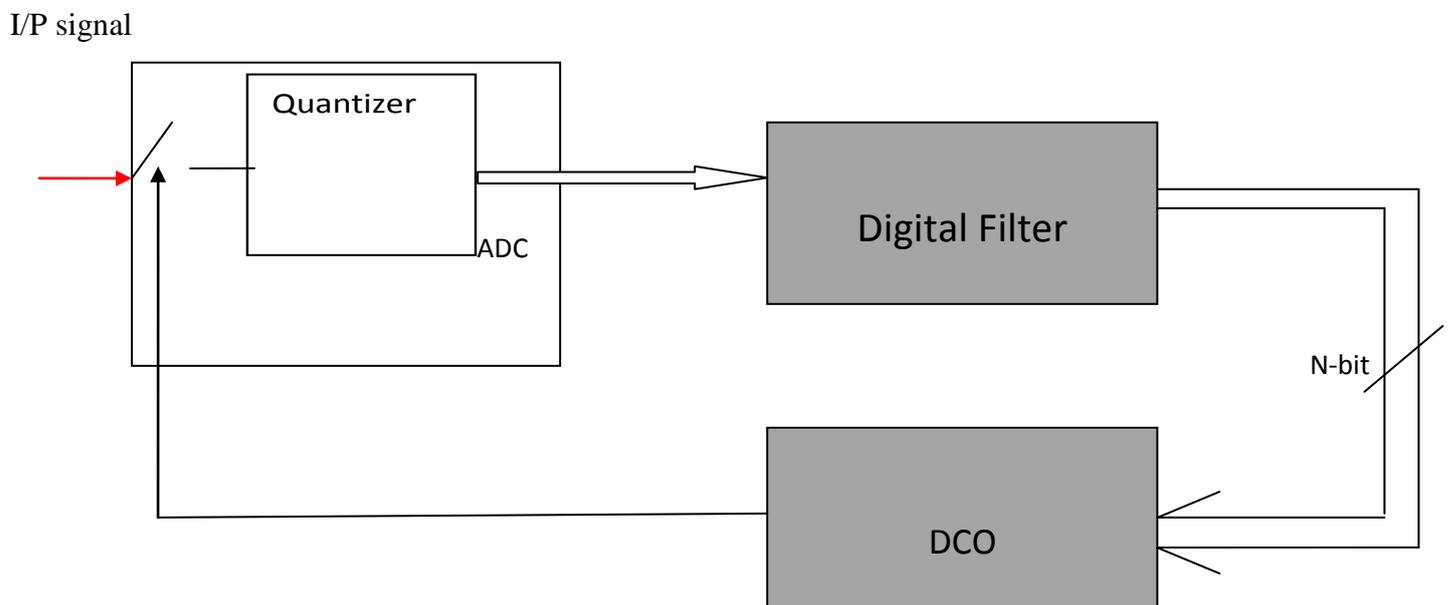
Where,

$N(x)$ : Number of phase error value which falls in range  $x \pm w$ .

$N$ : Total numbers of phase error.

$W$ : Narrow interval centred at  $x$ .

Thus p.d.f. is obtained and gain by dividing the full range of  $(x)$  into an approximate number of equal width class interval tabulating the number of data value in each class interval and dividing by the product of class interval  $W$  and sample size  $N$ .



**Figure (2): Block diagram of ZC\_DPLL.**

DCO :Digital Controller Oscillator

ADC: Analogue to Digital Converter

## 5. Results:

The main performance measures of the ZC-DPLL in the presence of the noise in the phase error cycle slipping. The p.d.f. of the phase error gives an indication about the phase error slipping. A White additive gaussian noise whose variance is equal to  $\sigma^2$  is used in the following matlab

simulation results. The actual phase error value are collected from the total simulation points and can be explained in what follows. 3000 phase error points are used in the simulation with every tenth value states as the steady states phase error are tabulated and their p.d.f. is plotted. Noise strength is measured by  $1/(\text{ISN})$  which is equal to  $\sigma_n^2/A^2$  where A is signal amplitude's/N is related to ISN as  $10 \text{ LOG}(1/\text{OSN})$ .

Fig (3) shows the p.d.f. of the phase error of  $k_1=1, r=2$  and  $S/N=10\text{dB}$  ( $\text{ISN}=0.1$ ) with the linear approximation that is derived earlier. A close matching is noticed between the simulation and the linear approximation. The figure is repeated for a positive frequency of set and plotted in fig (4). The mean value of phase error in this figure is expected when the second order ZC-DPLL is subjected to frequency step input. Fig (5) shows the p.d.f. of the phase error when a negative frequency offset is applied to the loop. The loop performance with noise is improved by decreasing the

value of (r) as shown in Fig (6a). This is slightly affected by the  $k_1$  value as shown in Fig (6b).

The effected of (r) on noise performance is shown practically with  $S/N=8\text{dB}$  and  $r=1.2$  with  $G_1=0.40\text{H}$  (hexadecimal) and  $G_2=0.20\text{H}$ , and p.d.f. of the phase error is shown in Fig(6a). With the same S/N ratio and  $r=2$  and  $G_1=0.40\text{H}$ , the p.d.f. of phase error is drawn in Fig (6b). The above two figures shows that the decrease in value of (r) increases the performance of loop. i.e. low phase error as shown in Fig(9). (Jitter and higher probability of locking). The main performance measures of the loop in the presence of noise are reliability phase error standards deviation, probability of correct looking and maximum phase jitter.

The reliability is defined by the equation  $\text{reliability \%} = (\text{total time-slip time}) / \text{total} \times 100\%$ . The probability of correct locking is probability that the loop locks at correct phase. Maximum phase jitter is the maximum spread of p.d.f. curve and represents measures of loop tendency to cycle slip. The reliability of ZC-DPLL vs input S/N, standard deviation vs S/N, probability correct locked vs S/N, and maximum phase jitter vs S/N are Plotted versus input signals /noise and shown in figures (8,9,10,11) it can be noticed that the performance of loop degraded as the S/N decreases.

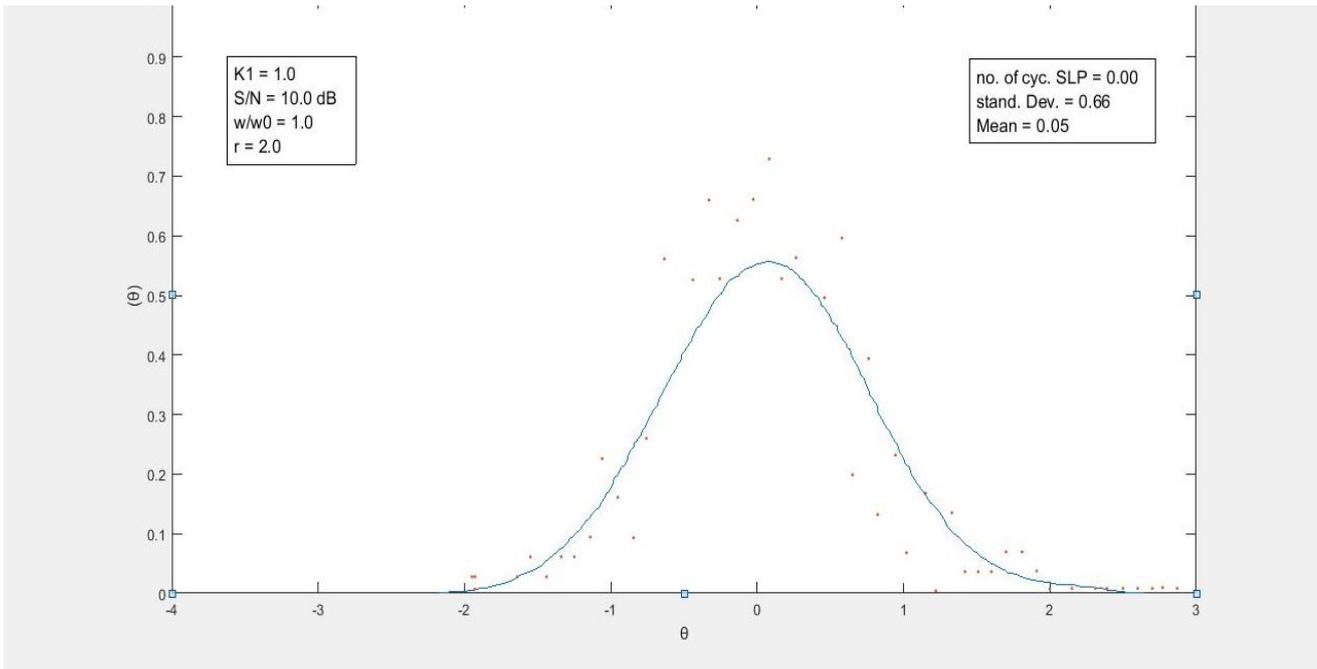


Figure (3): Steady state p.d.f for second order ZC-DPLL

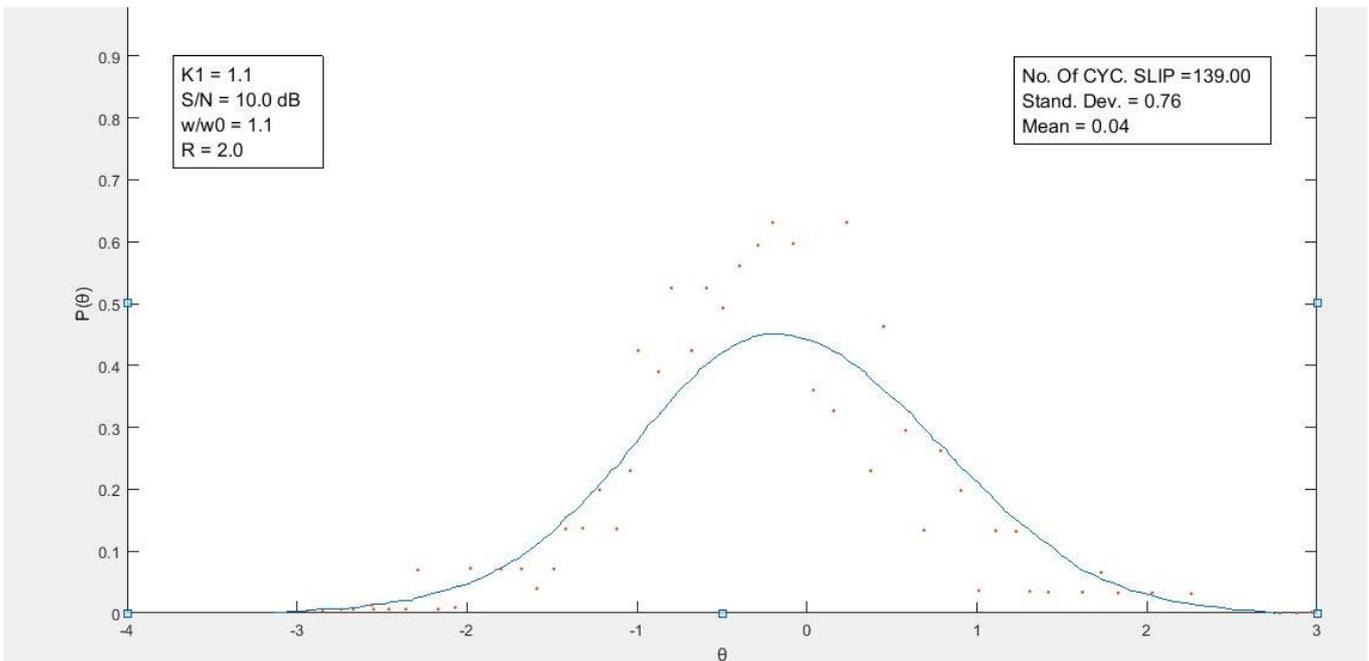


Figure (4): Steady state p.d.f for second order ZC-DPLL

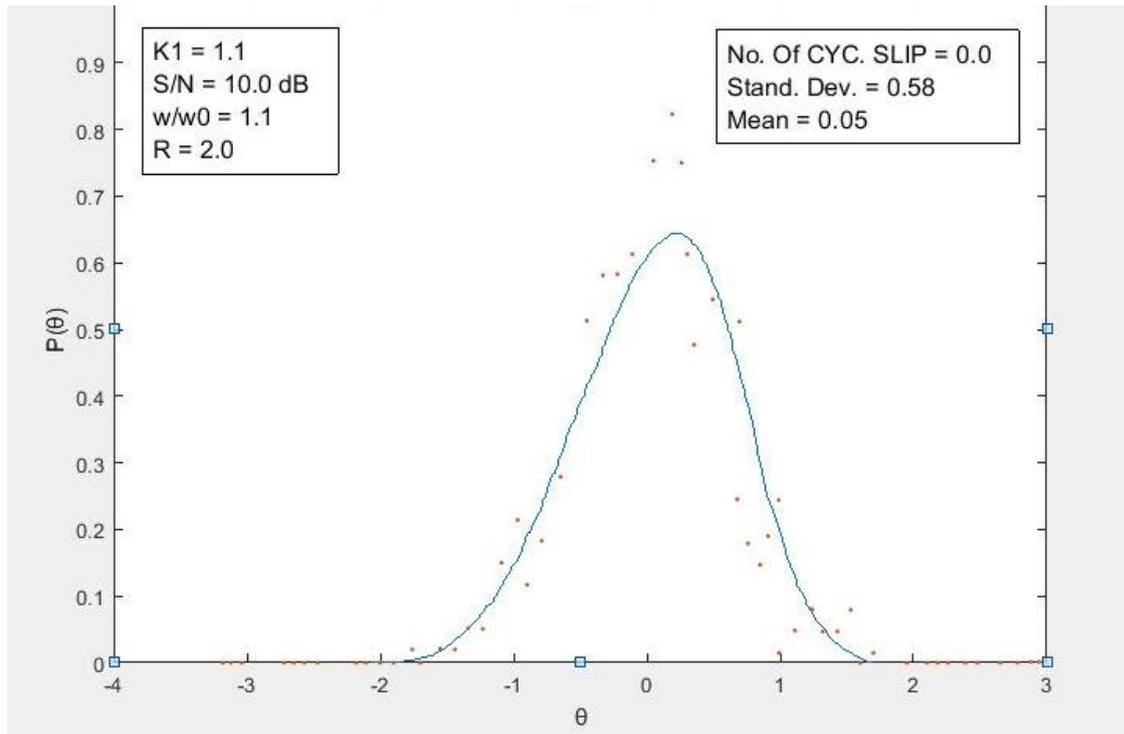
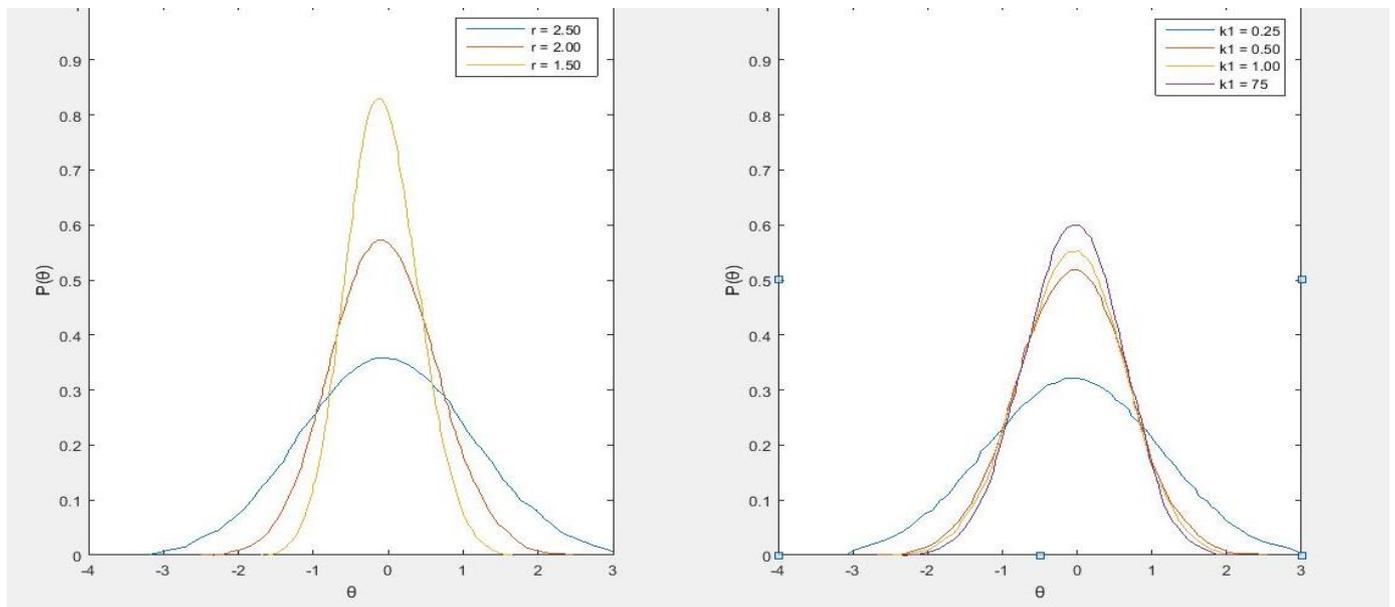


Figure (5): Steady state p.d.f for second order ZC-DPLL



(a)

(b)

Figure (6): steady state p.d.f for second order ZC-DPLL

(a) With various k value

(b) With r value

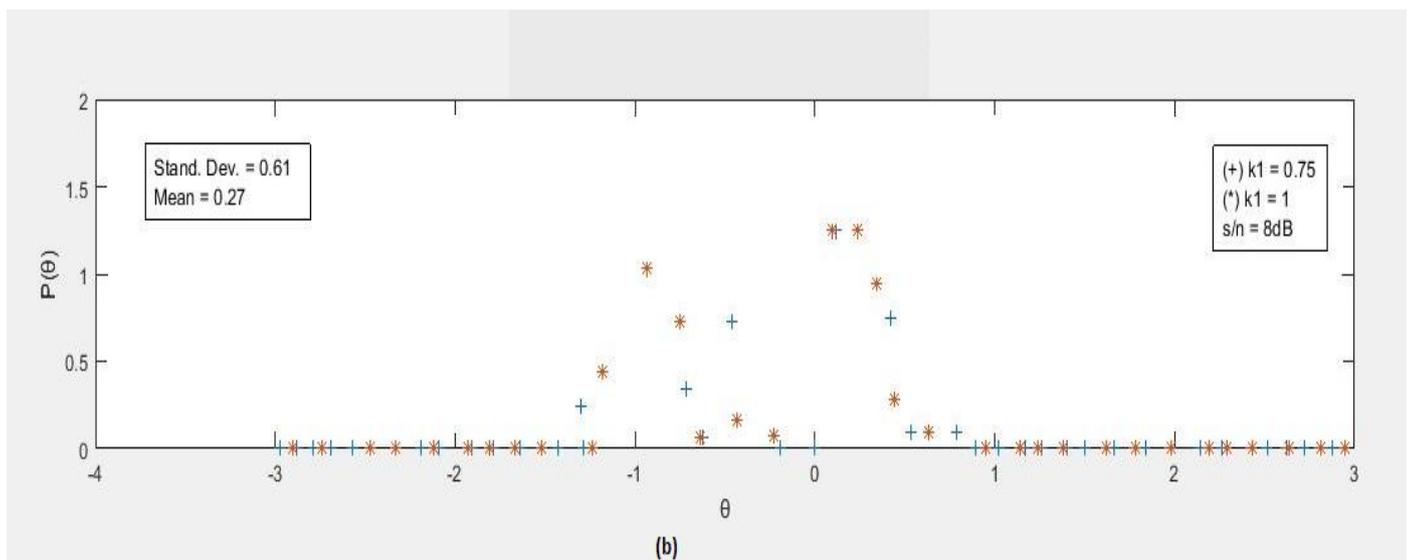
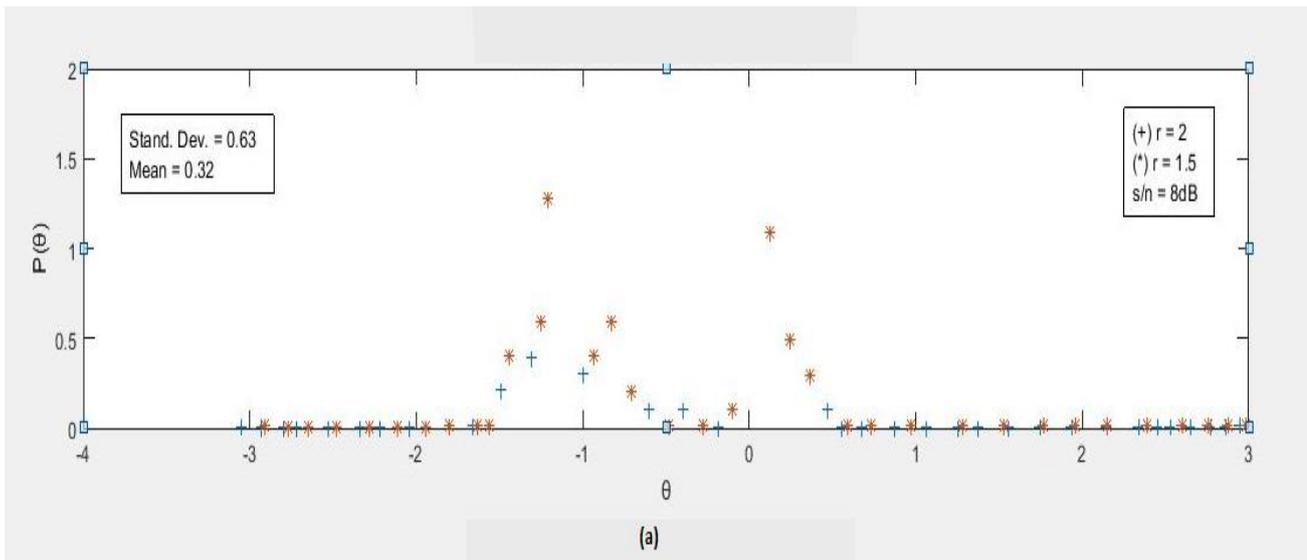


Figure (7): p.d.f for experimental second order ZC-DPLL

(a) With various r value and S/N=8dB

(b) With various K1 value and S/N=8dB

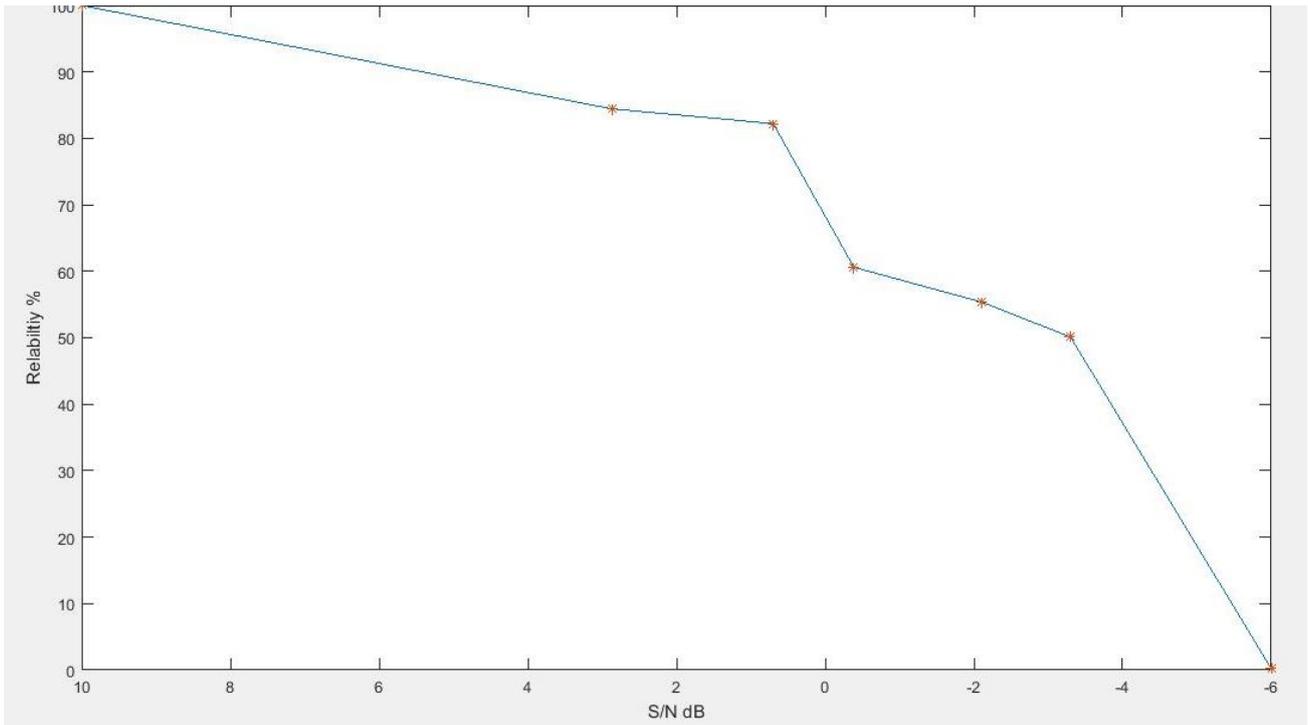


Figure (8): Reliability of second order ZC-DPLL vs. Input S/N

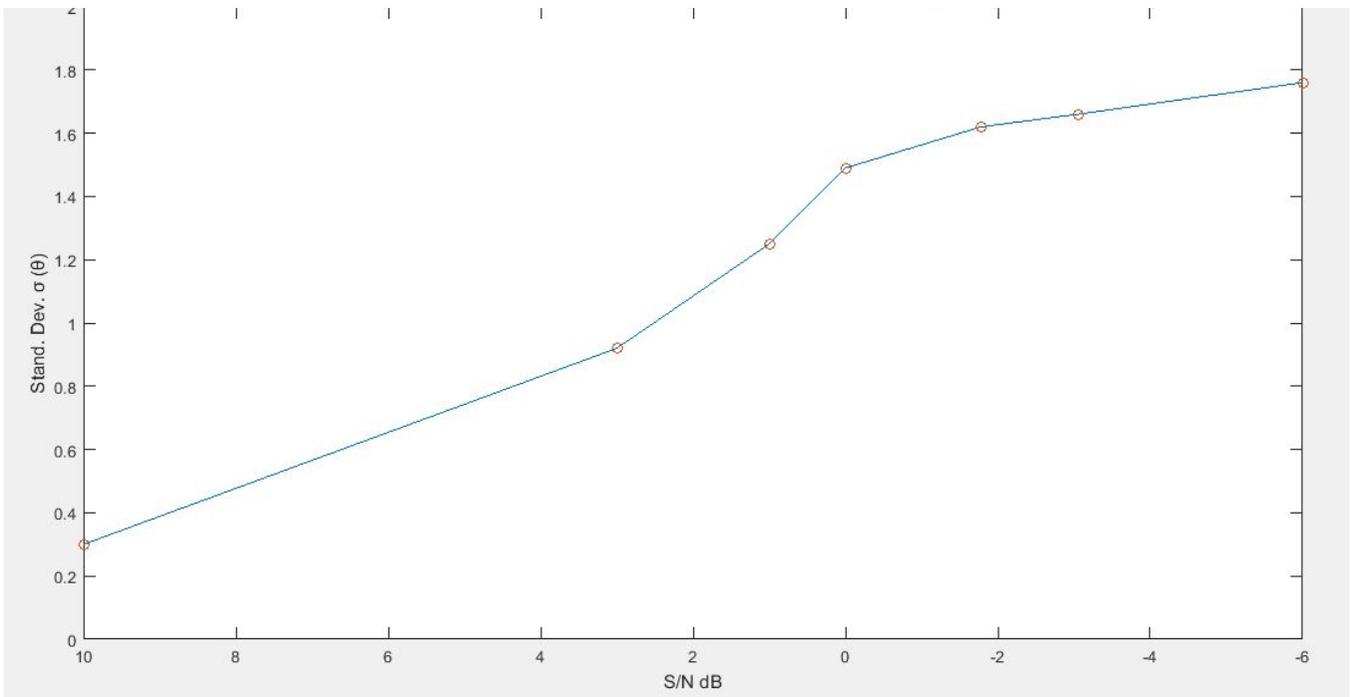
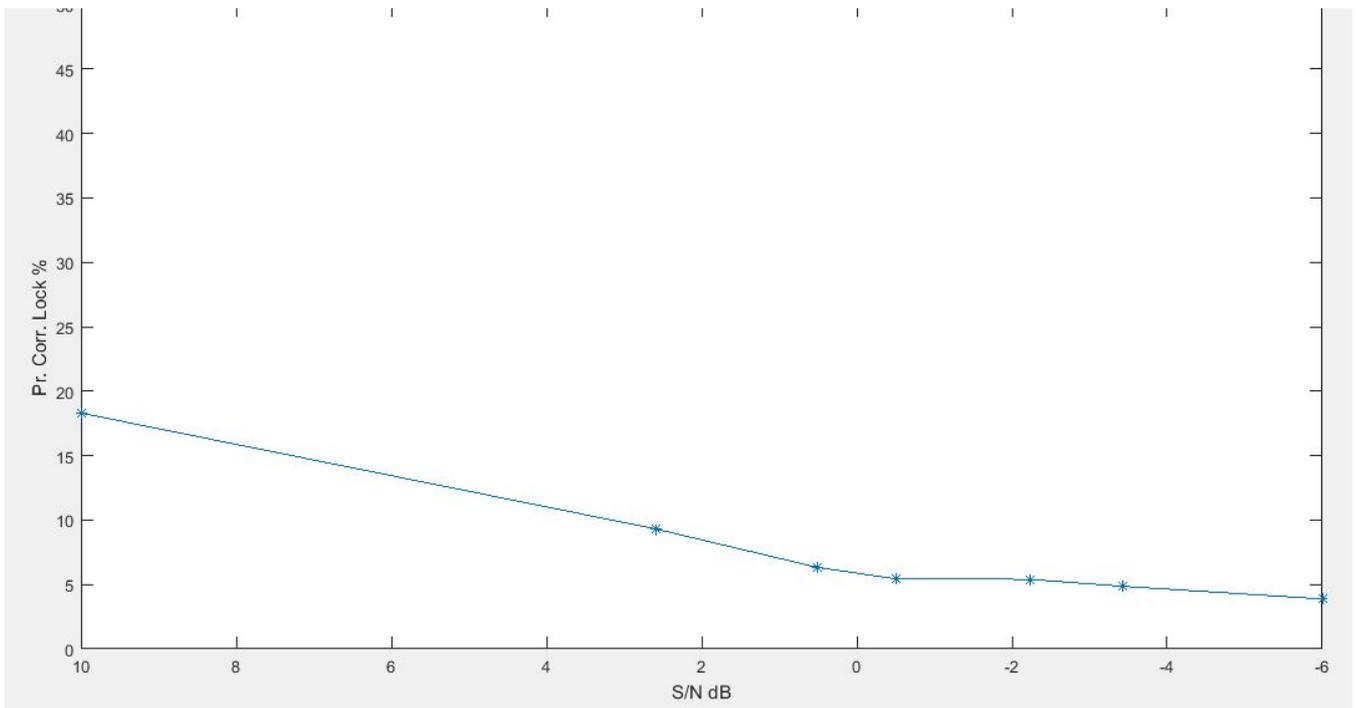
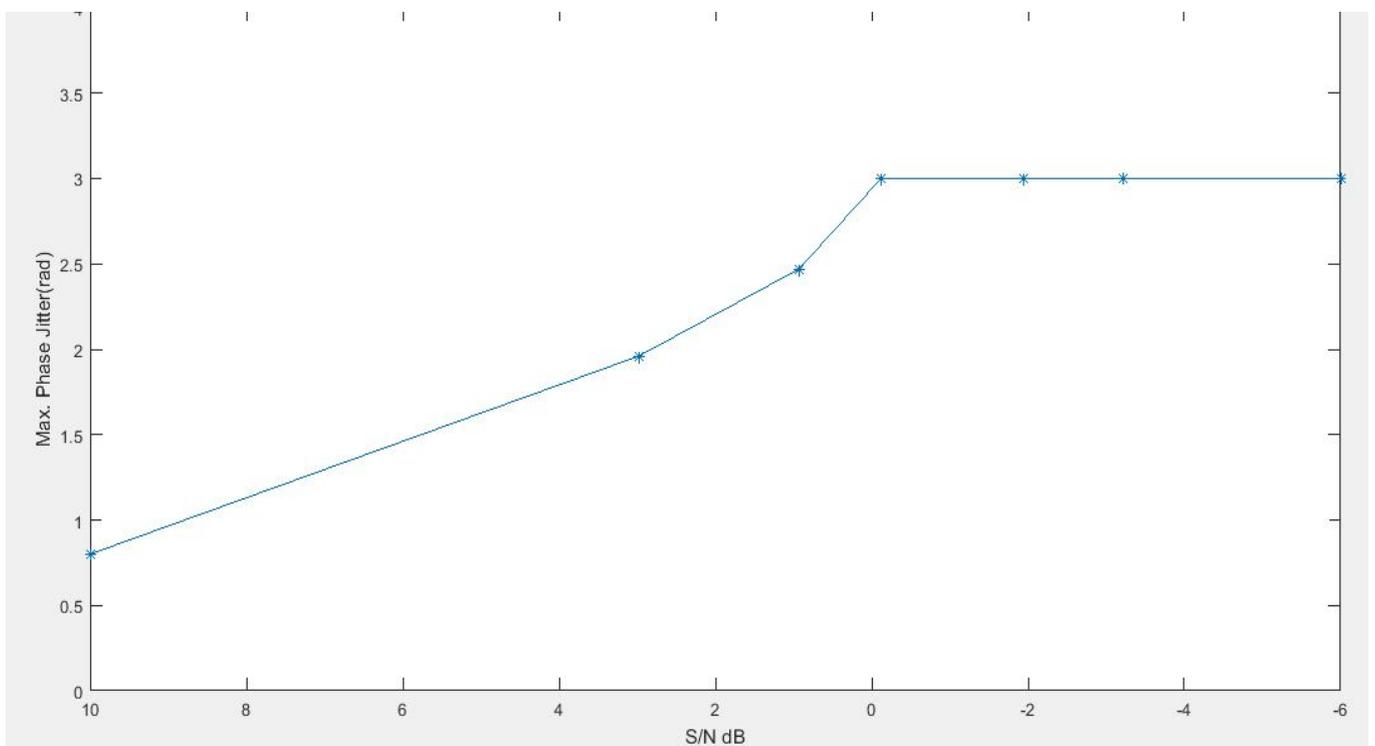


Figure (9): Phase error standard deviation of second order ZC-DPLL vs. input S/N



**Figure (10):** Probability of correct locking of second order ZC-DPLL vs. input S/N



**Figure (11):** Phase error jitter of second order ZC-DPLL vs. input S/N

## 6. Conclusions:

A second order zero crossing digital phase locked loop(ZC-DPLL) is analysed in the presence of the effect noise. An approximate expression for the steady state phase error probability density function phase error variance and noise effect loop bandwidth are obtained and given by equation (14, 13, and 12).The loop noise performance is affected by the loop filter gain  $G_1$ ,  $G_2$ .

## 7. Reference:

[1] kasim M. hamza and saleh R.AL.Araji,"analysis and design of the second order zero crossing Digital Phase locked loop in the absence of noise." J. Electronic.& computer. Res vol.1(2) /P 49-70.Electronic and computer scientific research council-Baghdad. Oct, 1987

[2] Weinberg A.and Liu B. "Discrete Time Analysis Of Non-uniform sampling first and second order-Order Digital Phase Locked Loop",IEEE Trans

Gommon.Technol.Vol COM22,pp123-137.feb 1974.

[3] Q. Nasir and S. R. Al-Araji, "Optimum Performance Zero Crossing Digital Phase Locked Loop using Multi-Sampling Technique," IEEE International Conference on Electronics Circuits and Systems, Sharjah, pp. 719-722. 14-17 December 2003,

[4] Schwartz M. "information Transmission Modulation and noise"Mc-Graw Hill,pp357. 1980

[5] Gill G.S and Gupya S.C."on higher – order discrete digital locked loop"IEEE trans AES-8,pp615-623.1972

[6] M. Nandi, 'Optimization of zero crossing digital phase locked loop performance in carrier synchronization system', Int. Journal of Electronics and Communication. Eng., vol. 9, pp.77-85, no.1, 2016

[7] Q. Nasir, "Chaos Controlled ZCDPLL for Carrier Recovery in Noisy Channels," Wireless Personal Communications, Vol. 43, pp. 1577-1582. No. 4, December 2007