In the Name of God

Journal of
Information Systems & Telecommunication
Vol. 4, No. 1, January-March 2016, Serial Number 13

Research Institute for Information and Communication Technology
Iranian Association of Information and Communication Technology

Affiliated to: Academic Center for Education, Culture and Research (ACECR)

Print ISSN: 2322-1437
Online ISSN: 2345-2773
Publication License: 91/13216

Editorial Office Address: No.5, Saeedi Alley, Kalej Intersection., Enghelab Ave., Tehran, Iran,
P.O.Box: 13145-799 Tel: (+9821) 88930150 Fax: (+9821) 88930157
Email: info@jist.ir
URL: www.jist.ir

Indexed in:

- Index Copernicus International www.indexcopernicus.com
- Journal of Information Systems and Telecommunication www.jist.ir
- Islamic World Science Citation Center (ISC) www.isc.gov.ir
- Scientific Information Database (SID) www.sid.ir
- Regional Information Center for Science and Technology (RICeST) www.ricest.ac.ir
- Magiran www.magiran.com

Publisher:
Regional Information Center for Science and Technology **(RICeST)**
Islamic World Science Citation Center **(ISC)**

This Journal is published under scientific support of
Advanced Information Systems (AIS) Research Group and
Digital & Signal Processing Research Group, ICTRC

Acknowledgement

JIST Editorial-Board would like to gratefully appreciate the following distinguished referees for spending their valuable time and expertise in reviewing the manuscripts and their constructive suggestions, which had a great impact on the enhancement of this issue of the JIST Journal.

(A-Z)

- Abdolvand Neda, Alzahra University, Tehran, Iran
- Abdollahi Mohammad, Shahrood university of Technology, Shahrood, Iran
- Amiri Parviz, Shahid Rajaee Teacher Training University, Tehran, Iran
- Andami Hooman, Islamic Azad University, Malayer Branch, Malayer, Iran
- Borna Keivan, Kharazmi University, Tehran, Iran
- Darmani Yousef, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Derakhshan Barjoei Pouya, Islamic Azad University, Naein Branch, Naein, Iran
- Falahati Abolfazl, Iran University of Science and Technology, Tehran, Iran
- Ghazvini Mahdieh, Shahid Bahonar University of Kerman, Kerman, Iran
- Hasanian Esfahani Roya, Academic Center for Education Culture and Research (ACECR), Tehran, Iran
- Kargar Mohammad Javad, University of Science and Culture, Tehran, Iran
- Kazerooni Morteza, Malek Ashtar University of Technology, Tehran, Iran
- Khalkhali Vahid, Islamic Azad University, Science and Research Branch, Tehran, Iran
- Mashoufi Behboud, Urmia University, Urmia, Iran
- Mirroshandel Seyed Abolghasem, University of Guilan, Rasht, Iran
- Mohammad Khanli Leili, University of Tabriz, Tabriz, Iran
- Mosallanejad Ali, Shahid Beheshti University, Tehran, Iran
- Nasersharif Babak, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Nooralizadeh Hamid, Islamic Azad University, Eslamshahr Branch, Tehran, Iran
- Rajabi Ruzbeh, University of Qom, Qom, Iran
- Shaker Gholam, Islamic Azad University, Science and Research Branch, Tehran, Iran
- Shirazi Jalil, Islamic Azad University, Gonabad Branch, Gonabad, Iran
- Torkamani Razieh, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Zahedi Mortaza, Shahrood university of Technology, Shahrood, Iran

Table of Contents

A New Switched-beam Setup for Adaptive Antenna Array Beamforming

Shahriar Shirvani Moghaddam*
Department of Electrical Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran
sh_shirvani@srttu.edu
Farida Akbari
Department of Electrical Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran
akbarifrd@yahoo.com

Received: 25/Sep/2014 Revised: 17/Dec/2015 Accepted: 02/Jan/2016

Abstract

In this paper, a new spatio-temporal based approach is proposed which improves the speed and performance of temporal-based algorithms, conventional Least Mean Square (LMS), Normalized LMS (NLMS) and Variable Step-size LMS (VSLMS), by using the switched beam technique. In the proposed algorithm, first, DOA of the signal source is estimated by MUltiple SIgnal Classification (MUSIC) algorithm. In the second step, depending on the desired user's location, the closest beam of the switched beam system is selected and its predetermined weights are chosen as the initial values for the weight vector. Finally, LMS/NLMS/VSLMS algorithm is applied to initial weights and final weights are calculated. Simulation results show improved convergence and tracking speed and also a higher efficiency in data transmission through increasing the Signal to Interference plus Noise Ratio (SINR) as well as decreasing the Bit Error Rate (BER) and Mean Square Error (MSE), in a joint state. Moreover, Error Vector Magnitude (EVM) as a measure for distortion introduced by the proposed adaptive scheme on the received signal is evaluated for all LMS-based proposed algorithms which are approximately the same as that for conventional ones. In order to investigate the tracking capability of the proposed method, the system is assumed to be time varying and the desired signal location is considered once in the centre of the initial beam and once in the edge of the fixed beam. As depicted in simulation results, the proposed DOA-based methods offer beamforming with higher performance in both cases of the initial beam, centre as the best case and edge as the worst case, with respect to conventional ones. The MSE diagrams for this time varying system show an ideal response for DOA-based methods in the best case. Also, in the worst case, initial height of MSE is reduced and consequently the required iteration to converge is less than the conventional LMS/NLMS/VSLMS.

Keywords: Beamforming; Switched Beam; LMS; NLMS; DOA; MUSIC.

1. Introduction

Rapid growth of requests for wireless and mobile communication and increasing operators' demands has stimulated considerable studies for development of advanced techniques and technologies. The main goal is to manage the limited radio resources such as frequency spectrum and power and also attaining fully integrated, low cost, reliable, high quality and high speed wireless communications. Smart antennas or adaptive array antennas are effective solutions for these growing demands. They can make significant performance improvement for wireless systems, with respect to other methods such as low noise amplifiers, wideband techniques and sectorization. Smart antennas utilize adaptive beamforming techniques to provide optimum performance for the system. Higher capacity and data rate, lower power consumption, extended coverage range, reduction in noise and fading effects and suppression of interferers are advantages of antenna beamforming. Various methods exist for adaptive beamforming which can be implemented using high speed digital signal processing units [1]-[6].

The least Mean Square (LMS) algorithm is a popular algorithm for adaptive beamforming and other filtering applications such as interference cancellation, echo cancellation, noise cancellation, channel estimation and channel equalization. The LMS algorithm is a training-based method that employs a temporal training reference signal and recursively updates the weight vector to minimize Mean Square Error (MSE) between a reference signal and the array output. The advantages of LMS are simplicity, high stability, robustness and low computational cost which make LMS a widely-used algorithm. However, LMS suffers from low convergence and tracking speed. Therefore, a vast number of works in the literature are accomplished to enhance the performance of LMS as well as Normalized LMS (NLMS) which is a common version of LMS, and various types of LMS are proposed that attempt to boost the speed and efficiency of LMS/NLMS [7]-[10]. Different Variable Step-size LMS (VSLMS) algorithms are presented that attempt to reduce the convergence speed of the LMS [11], [12].

The combination of the LMS and other beamforming algorithms such as Sample Matrix Inversion (SMI) is also examined and has shown remarkable improvements in the

* Corresponding Author

convergence rate and interference suppression [13]. Also, a combination of Discrete Fourier Transform (DFT) and LMS algorithm is investigated for array antenna beamforming in [14]. This algorithm has improved the performance of both time-updating algorithm and spatial estimating. In [15] a new algorithm, called LLMS, is presented which combines the use of two successive LMS sections. Simulation results have shown better convergence performance of LLMS relative to the conventional LMS as well as Constrained Stability LMS (CSLMS), and Modified Robust Variable Step Size LMS (MRVSS) algorithms. The combination of Recursive Least Square (RLS) and LMS algorithms known as RLMS is also examined and RLMS algorithm has shown better performance than other earlier algorithms such as LMS, RLS, MRVSS and CSLMS [16]. A new adaptive beamforming algorithm called Turbo-LMS is presented in [17]. The proposed Turbo-LMS algorithm has given rapid convergence with respect to the LMS. All of the above mentioned algorithms are based on temporal information and they do not use the spatial information.

DOA estimation methods are effective techniques used for detection and tracking of the signal source location. These algorithms play a key role in blind beamforming techniques. Training-based algorithms, such as LMS/NLMS, generally do not need DOA estimation. However, DOA of the desired signals can be exploited to enhance LMS/NLMS beamforming performance and some investigations are accomplished in this regard.

In [18], a detection guided NLMS algorithm is proposed which incorporates DOA detection that leads to a reduction in the number of NLMS adapted parameters. This scheme significantly improved convergence and tracking speed, as well as the performance of nulling multi-user interference signals. In another research, DOA information which is generated through the Minimum Variance Distortionless Response (MVDR) method is used to build the constrained matrix needed for different constrained beamforming techniques. These algorithms have shown a good performance in terms of stability, convergence and accuracy [19]. Some other investigations have used the LMS and Multiple SIgnal Classification (MUSIC) algorithms simultaneously to cover desired DOAs and to null unwanted signals using both algorithms [4], [20]. In [21], [22] new LMS-based/CM-based adaptive weighting algorithm is proposed which relies on the idea of predicting the next location of source, and determining the array weights before arriving to the new location. For the next time associated to the new sampling point, evaluated weights will be used. Simulation results show lower on-line required processing with respect to conventional LMS. It can also predict the next location of the target in mobile applications.

Above mentioned proposed algorithms may not be extended easily for other weighting algorithms and special array geometry is needed. Also, it is desired to extend the proposed algorithm to multiuser detection applications. Therefore, the aim of this paper is to present a new approach which exploits the DOA information in training-based algorithms that consider temporal sequence as reference, especially, LMS and NLMS algorithms. Proposed method employs the switched beam technique to enhance the LMS and NLMS speed. This new combination can be applied for different array geometries and all types of training based weighting algorithms. Also, it can be used for multibeam scenarios.

Simulation results show a better performance as well as lower convergence time for the proposed algorithm in comparison with the standard LMS, NLMS and VSLMS algorithms. Finally, the proposed DOA-based temporal-reference beamforming algorithm offers increased convergence and tracking speed, in a joint state.

The rest of this paper is organized as follows. In Section 2, an overview of switched beam technique, adaptive arrays and adaptive beamforming approaches is presented. LMS, NLMS and VSLMS algorithms are described in Section 3. Also, popular DOA estimation algorithm, MUSIC, is described, briefly. Section 4 states the proposed beamforming technique and shows the different steps of the new algorithm. In Section 5, simulation results are given and new DOA-based LMS, NLMS and VSLMS algorithms are compared with conventional LMS, NLMS and VSLMS ones, respectively. Finally, conclusion remarks are given in section 6.

2. Smart Antenna Systems and Signal Model

Smart antenna systems are generally classified into two main categories: switched beam antennas and adaptive arrays.

In switched beam systems, multiple fixed beams are defined in several predetermined directions. In fact, the switched beam technique is an extension of the cellular sectorization scheme in which each 120° macro-sector is divided into several smaller micro-sectors. Each predetermined fixed beam belongs to one micro-sector. When a mobile user moves through the cell, the system selects an appropriate fixed beam which has the strongest received signal power. The system monitors the signal strength and switches between predetermined fixed beams if required. This scheme approaches a high performance in terms of array gain and coverage range of the base station of terrestrial wireless systems and also multi-beam satellites. However, since beams are fixed and each of them has the maximum gain in the centre of the beam, the system gain decreases in the edge of the predetermined beams. Therefore, the signal strength and quality of service degrade if the user approaches the edges of the main beam or if an interferer approaches the centre of the beam. Also, the switched beam systems cannot distinguish between desired and undesired signals such as interferers. Different hardware and software designs are available for implementation of switched beam systems [1], [3].

Adaptive array antenna systems continually monitor RF environment and adjust the antenna pattern dynamically to optimize the system performance as locations of users and

interferers change. Adaptive array systems consist of an antenna array and a signal processing unit. This unit includes complex weights, a unit that combines the weighted received signals and a signal processor which computes optimum weights via adaptive beamforming algorithms. Beamforming algorithms optimize the antenna system efficiency with respect to the signal environment by focusing energy in the desired user's direction and cancelling interferer signals by creating appropriate nulls in the radiation pattern. Adaptive arrays have more capabilities than switched-beam systems and they provide more degrees of freedom [1]-[5].

Adaptive beamforming techniques may be either training-based or blind. Training-based or non-blind approaches use a reference signal to adapt weights. The reference signal is known in both the transmitter and receiver and is used for updating the weight vector at the receiver. LMS, NLMS, SMI and RLS are examples of training-based beamforming algorithms. Blind algorithms don't use the reference signal and are usually based on the known properties of the desired signal. Algorithms such as Constant Modulus (CM), Decision Directed (DD), Least Squares (LS) and Conjugate Gradient (CG) are categorized as blind algorithms [1], [5]. In this research, two well-known training-based algorithms, LMS and NLMS, are utilized for weight adaptation.

In the following, we assume that the array is composed of N sensors configured in a Uniform Linear Array (ULA) and M narrowband signals are received at the array. Received signals can be expressed as a linear combination of incident signals and zero mean Gaussian noise with variance σ_n^2. The incident signals are assumed to be direct Line Of Sight (LOS) and uncorrelated with the noise. The input signal vector can be written as follows:

$$x(t) = \sum_{m=1}^{M} a(\theta_m)\, s_m(t) + n(t) = A.S(t) + n(t) \qquad (1)$$

$s_m(t)$ is a $M \times 1$ vector concerning to the m-th source located at direction θ_m from the array boresight. $a(\theta_m)$ is a $N \times 1$ steering vector or response vector of the array for direction θ_m, and is written as:

$$a(\theta_m) = \left[1,\ e^{-jk.d\sin\theta_m}, \dots, e^{-j(N-1)k.d\sin\theta_m}\right]^T \qquad (2)$$

where d is the inter-element spacing value and $k = 2\pi/\lambda$ is the wavenumber.

Furthermore, A is a $N \times M$ matrix of steering vectors, which is named manifold matrix and expressed as:

$$A = [a(\theta_1)a(\theta_2) \quad \dots\ a(\theta_M)] \qquad (3)$$

The array output signal is given by:

$$y(n) = w^H x(n) \qquad (4)$$

where w is the N dimensional weight vector, H denotes the Hermitian transpose, and x is the received signal vector as defined in (1).

3. Beamforming Using LMS, NLMS and VSLMS Algorithms

The LMS algorithm is a popular adaptation technique based on the steepest descent method. This algorithm updates the weights recursively by estimating gradient of the error surface and changing the weights in the direction opposite to the gradient to minimise the MSE [2]. The error signal is computed by the following expression:

$$e(n) = d(n) - y(n) \qquad (5)$$

The final recursive equation for updating the weight vector is:

$$w(n+1) = w(n) + \mu x(n) e^*(n) \qquad (6)$$

where $e(n)$ is the error signal, i.e., the difference between desired signal $d(n)$ and output $y(n)$, μ is the step size parameter which controls the convergence speed of the algorithm, and $x(n)$ is the received array. The small step size causes a slow convergence but high stability around the optimum value. On the other hand, large step size results in a rapid convergence and a lower stability. Hence, step size is a major parameter that makes a trade-off between the convergence speed and the stability. The convergence rate of the LMS also depends on the eigen-value spread of the input correlation matrix. Variable step size algorithms are used to increase the convergence rate [10],[11].

In (6), the correction term $\mu x(n)\, e^*(n)$ is applied to the weight vector during LMS algorithm, is proportional to the input vector $x(n)$. Therefore, a gradient noise amplification problem occurs in the standard LMS algorithm. This problem can be solved by normalized LMS, in which a data dependent step size is used for adaption. NLMS normalizes the weight vector correction term with respect to the squared Euclidean norm of the input vector $x(n)$ at time instant n. So the updating equation in NLMS can be written as follows:

$$w(n+1) = w(n) + \frac{\mu}{\|x(n)\|^2} x(n) e^*(n) \qquad (7)$$

The NLMS algorithm shows a better performance than the LMS algorithm in terms of convergence and stability [2]. To avoid the convergence problem due to division by a small number, a positive constant ε may be added to the Euclidean norm of the input vector in (7), so the weight vector is computed through:

$$w(n+1) = w(n) + \frac{\mu}{\varepsilon + \|x(n)\|^2} x(n) e^*(n) \qquad (8)$$

This algorithm is known as ε-NLMS and results in a more reliable implementation [7]. Since LMS and NLMS algorithms are temporal-based, their slow convergence may cause tracking problems in cellular mobile systems. The main mechanism used to control the convergence rate of LMS algorithm is changing the step-size of algorithm. Variable Step-size LMS (VSLMS) algorithm can enhance the performance of beamforming including convergence rate and steady state performance. Different types of VSLMS algorithm have been proposed in the literature [7], [8]. In VSLMS algorithm usually the step-size μ is

limited between μ_{min} and μ_{max}. μ is calculated during each iteration using the gradient function of error in the previous iteration. In this paper, the VSLMS algorithm introduced by Kwong and Johnston is used for simulation [23]. The step-size is computed through equation (9).

$$\mu(n) = \alpha\mu(n-1) + \gamma e^2(n) \quad (9)$$

where $0 < \alpha < 1$ and $\gamma > 0$ and then the comparison of $\mu(n)$ is done as equation (10).

$$\mu(n) = \begin{cases} \mu_{max} & \text{if } \mu(n) > \mu_{max} \\ \mu_{min} & \text{if } \mu(n) < \mu_{min} \\ \mu(n) & \text{otherwise} \end{cases} \quad (10)$$

In the following section, a new approach is proposed which uses spatial DOA information beside adaptive weighting algorithms to achieve a better convergence speed in LMS, NLMS and VSLMS algorithms. Other LMS type algorithms also will be improved using this approach.

4. Music Algorithm

There are different methods to estimate the DOA that are divided into three basic categories, classical, subspace-based, and maximum likelihood (ML)-based. These techniques differ in modelling approach, computational complexity, resolution threshold and accuracy. Spectral-based methods, first and second categories, which rely on calculating the spatial spectrum of the received signal and finding the DOAs as the location of peaks in the spectrum, are easy to apply and need less computation than parametric or ML-based methods that directly estimate the DOAs without first calculating the spectrum [24,25].

Among them, MUSIC is a subspace-based DOA estimation algorithm referred to super-resolution technique which offers a good tradeoff between resolution and computational complexity. It should be noted that MVDR is not able to introduce high resolution peaks, Root-MUSIC is an appropriate algorithm only for ULA and the Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT) algorithm needs special array geometry, but the MUSIC algorithm offers high resolution peaks which can be used for different array geometries [25].

Eigenvectors of the covariance matrix belong to either two orthogonal signal or noise subspaces. If M signals arrive on the array, M Eigenvectors associated with the M largest eigenvalues of the covariance matrix span the signal subspace and the $N - M$ Eigenvectors corresponding to the $N - M$ smallest eigenvalues of the covariance matrix span the noise subspace. The M steering vectors that form the manifold matrix $A(\theta)$ are orthogonal to the noise subspace and so the steering vectors lie in the signal subspace.

The MUSIC algorithm estimates the noise subspace using Eigen-decomposition of the sample covariance matrix and then estimations of DOAs are taken as those θ that give the smallest value of $A^H(\theta)V_n$, where V_n denotes the matrix of Eigen-vectors corresponding to the noise subspace. These values of θ result in a steering vector farthest away from the noise subspace and orthogonal to the noise subspace as much as possible [26], [27]. This is done by finding M peaks in the MUSIC spectrum defined by:

$$P_{MUSIC}(\theta) = \frac{1}{A^H(\theta)V_n{V_n}^H A(\theta)} \quad (11)$$

Several parameters such as the number of samples (snapshots), the number of elements and also SNR affect the resolution threshold of the DOA estimation algorithms.

5. Description of the Proposed Method

The LMS algorithm and its variants are temporal training-based algorithms and they don't need DOA estimation of the signal sources for weight updating computations. To utilize DOA information in the LMS algorithm, the effect of signal source location is imposed on the initial value of the weight vector. This leads to a higher convergence rate. The new DOA-based algorithm utilizes switched beam scheme to choose an appropriate initial value for the weight vector. The angular space is divided into a plenty number of sections and a particular beam is determined for each section. An example of space division can be seen in Fig. 1. Each beam has a pre-determined weight vector that are previously, computed by the LMS, NLMS or VSLMS algorithms.

The proposed method in details is as follows.

5.1 The Preliminary Adjustment Phase

This step is carried out once before initiation of the beamforming and does not repeat in next beamforming processes. According to the required resolution and the array configuration, the angular space is divided into multiple sections of the same width and a particular beam is specified to each section as depicted in Fig. 1.

5.2 The Beamforming Phase

A set of weight vectors is calculated for the defined beams using the LMS/NLMS algorithm. The obtained weights will be saved in the memory of the system. In this phase, the initial adjustment of the system is accomplished and a set of predefined beams and their related weights are prepared for the DOA-based beamforming which is performed as the following steps.

Step 1: DOA of the signal source is estimated using a proper simple DOA estimation method such as MUSIC. In the case of multiple correlated signal sources, spatial smoothing methods same as Forward-Backward Spatial Smoothing (FBSS) can be used to help signal source detection [23].

Step 2: Depending on the desired user's location, the closest beam defined in the preliminary step is selected and its predetermined weights are chosen as the initial value for the weight vector.

Step 3: The LMS or NLMS algorithm is applied to initial weights and final weights will be calculated with a higher rate.

This step results in more accuracy in target tracking and nulling of the undesired signals. In the case of enough number of predetermined beams and absence of undesired signals, initial weights obtained in the second step can be satisfactory.

The preliminary adjustment and beamforming phases are depicted as flowcharts in Fig. 2 and Fig. 3, respectively.

6. Simulation Results

We consider a ULA with 8 elements and half wavelength inter-element spacing. The performance of the LMS, NLMS and VSLMS algorithms are investigated in different situations around one of the predetermined beams and compared with the conventional LMS, NLMS and VSLMS beamforming schemes. In this investigation, the NLMS is considered as ε-NLMS and VSLMS is considered as proposed by Kwong and Johnston [23]. In addition, for 8-elelment array, at least every 10 degrees of the space should be covered with a particular beam. So, the space is divided into several 10 degree sections with overlapped beams as demonstrated in Fig. 1.

The weight vector for each beam in a specific direction is computed through conventional LMS, NLMS and VSLMS considering single user application. The desired signal source direction is estimated using the MUSIC algorithm. It is expected that the proposed algorithm have the best and the worst performance at the centre and edges of the closest predetermined beam to the desired DOA, respectively.

To compare the conventional and the proposed spatial-temporal algorithms, the desired user's location is assumed to be at the centre of the beam placed at 25° of the best situation and at the edge of that beam at the worst case as shown in Fig. 4. Supposed that the Additive White Gaussian Noise (AWGN) is imposed to the transmitted signal and the

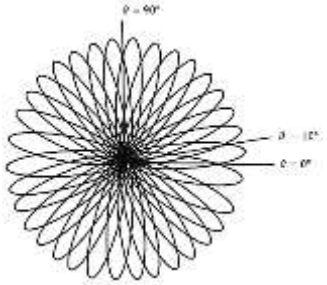


Fig. 1. Space divisions in switched beam scheme

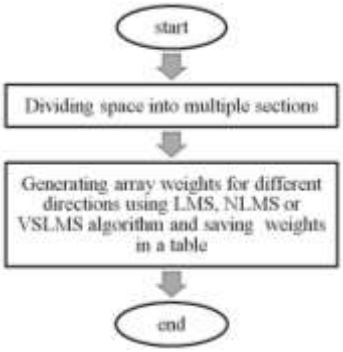


Fig. 2. The preliminary adjustment phase

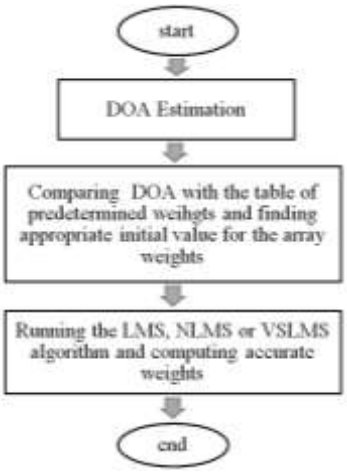


Fig. 3. The beamforming phase

Binary Phase Shift Keying (BPSK) modulation is used for data transmission. Two interference signals are also assumed to be received at $-15°$ and $70°$ with a 3 dB Signal to Interference Ratio (SIR). The convergence speed, Signal to Interference plus Noise Ratio (SINR), Bit Error Rate (BER) and MSE of the algorithms are computed via MATLAB and compared together.

Fig. 5 depicts the radiation pattern of the adaptive antenna array system when the desired signal is located at $25°$. The main lobe is placed accurately at the centre of the concerned beam when DOA is $25°$. Appropriate nulls are shaped at the direction of interferers via DOA-based techniques as well as conventional algorithms.

Fig. 6 shows the MSE changes during the training process, where DOA is placed at the centre of the considered beam ($\mathrm{DOA} = 25°$). In this case, DOA-based algorithms are converged at the beginning of the adaptation process.

For the considered beam at $25°$, signals arrive at $20°$ or $30°$ are located at the edges of the initial beam. In Fig. 6, the radiation pattern of the array is illustrated for LMS/NLMS/VSLMS and the DOA-based versions when DOA is placed at the edge of the considered beam ($\mathrm{DOA} = 30°$). Appropriate main beam and nulls are shaped in the direction of desired and interference signals, respectively. Some deviation may be seen at the location of the main lobe when DOA is $30°$. However, the array gain is equivalent or even greater at the desired signal DOA in DOA-based algorithms. By increasing the number of array elements or the number of predetermined fixed beams, the deviation will be decreased. Appropriate nulls are shaped at the direction of interferers via DOA-based techniques as well as the conventional algorithms.

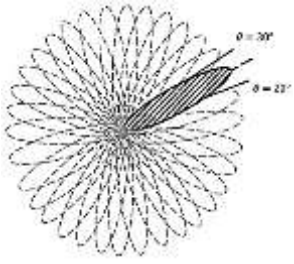


Fig. 4. Situation of the considered beam at $25°$

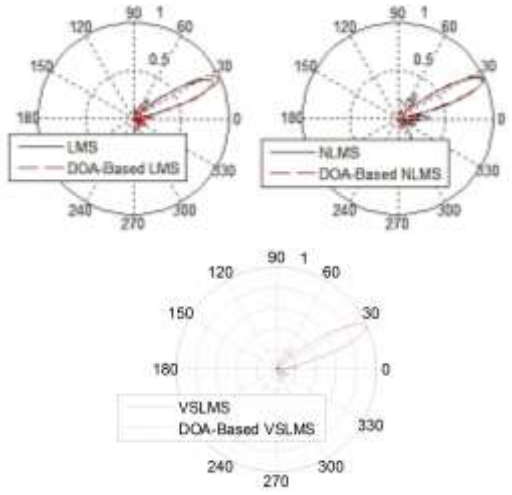


Fig. 5. Radiation Pattern of LMS/DOA-based LMS ,NLMS/DOA-based NLMS and VSLMS/DOA-based VSLMS when DOA = 25°

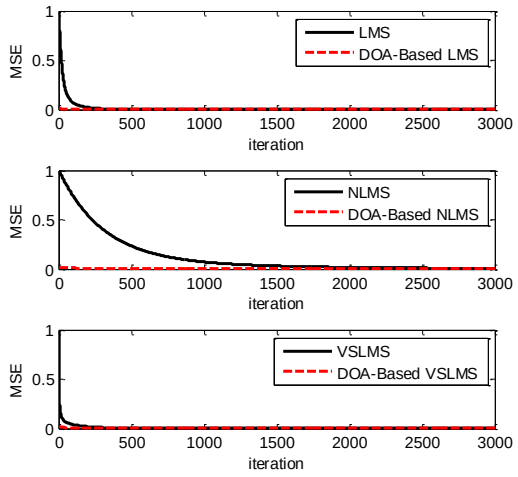


Fig. 6. MSE diagram during the training process when the desired DOA is in the centre of the initial beam

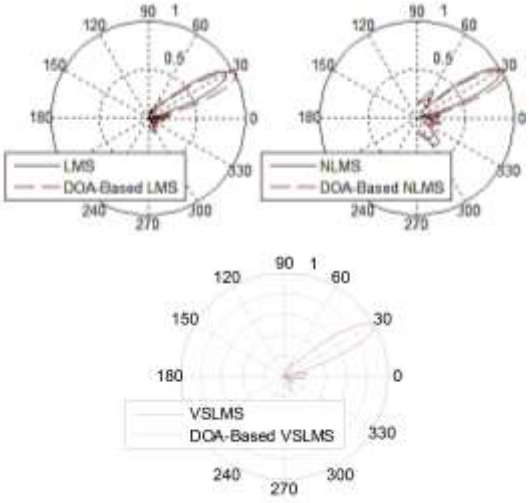


Fig. 7. Radiation Pattern of LMS/DOA-based LMS, NLMS/DOA-based NLMS and VSLMS/DOA-based VSLMS when DOA = 30°

Variations of the MSE during the training process for DOA = 30° is demonstrated in Fig. 8. The MSE diagrams also show a better convergence rate of the DOA-based algorithms than the conventional LMS/NLMS/VSLMS algorithms in this case. The lower MSE at the beginning of training process can be interpreted as better capability for adaptation with time-variant systems and moving target tracking. In other directions that are covered by this beam ($20° < \theta < 30°$), simulation trials give a better convergence speed than the edge of the initial beam.

In Fig. 9, 10, SINR versus the SNR for both the centre and edges of the initial fixed beam (DOA = 25°) are shown. As the SNR rises, SINR increases. As expected, the level of SINR in the DOA-based methods is higher than the conventional algorithms.

The BER diagram in Fig. 11 presents a lower level for the DOA-based methods in the centre of the predetermined beam (DOA = 25°). When DOA = 30° , as seen in Fig. 12, the BER levels for both of the proposed and conventional algorithms are close together. The BER in other situations is lower than BER in the edges of the initial beam placed at 25°. Therefore, it can be concluded that the BER in the DOA-based methods is the same as or lower than the conventional methods.

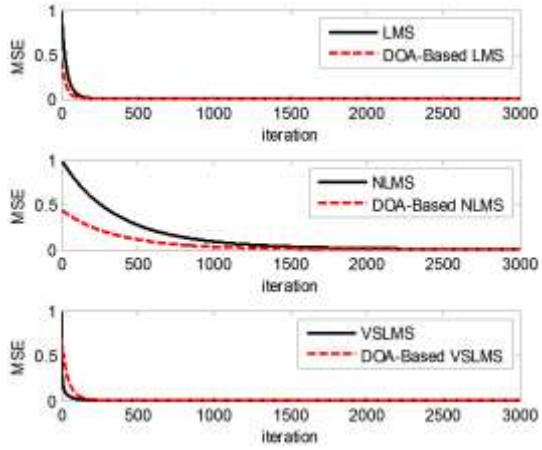


Fig. 8. MSE diagram during the training process when desired DOA is in the edge of the initial beam

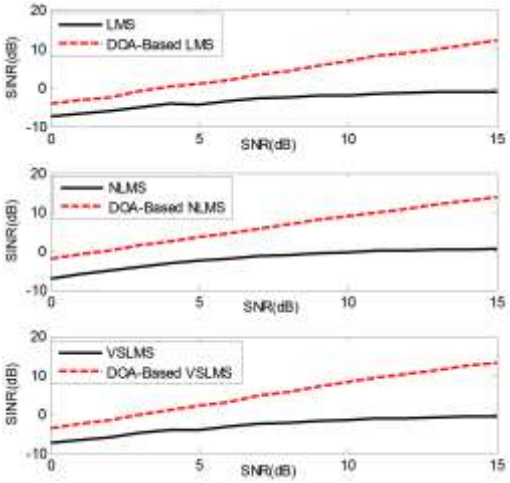


Fig. 9. SINR changes when the desired DOA is in the centre of the initial beam

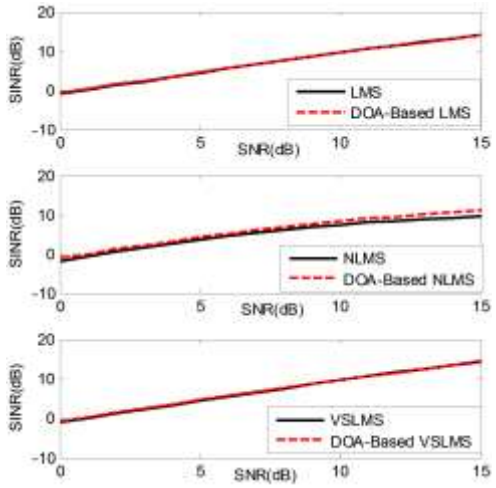


Fig. 10. SINR changes when the desired DOA is in the edge of the initial beam

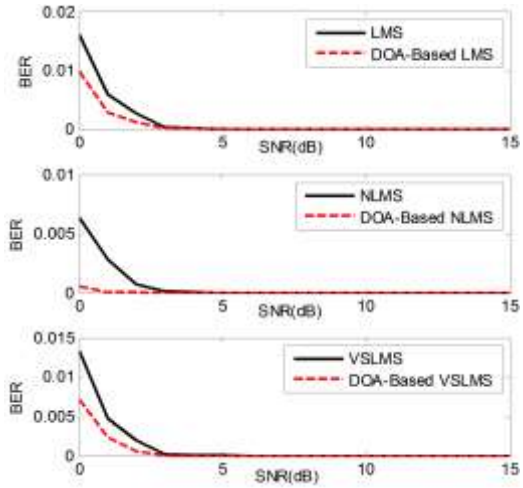


Fig. 11. BER changes when the desired DOA is in the centre of the initial beam

After the weight adjustment, data transmission begins. As depicted in Fig. 13, 14, the MSE of the transmitted data present lower levels for the DOA-based algorithms with respect to the conventional LMS, NLMS and VSLMS in all situations. Therefore, it can be concluded that convergence speed, efficiency and accuracy of the adaptive array are increased by using DOA-based LMS/NLMS/VSLMS.

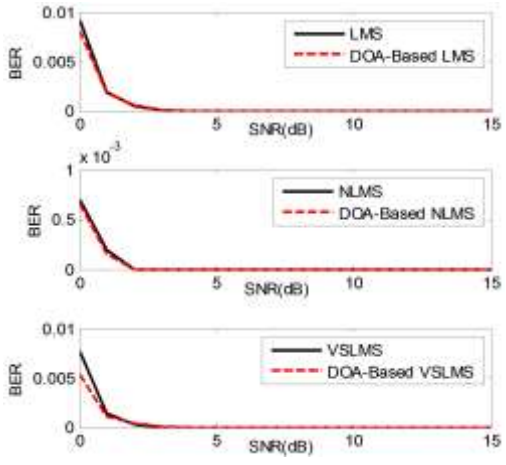


Fig. 12. BER changes when desired DOA is in the edge of the initial beam

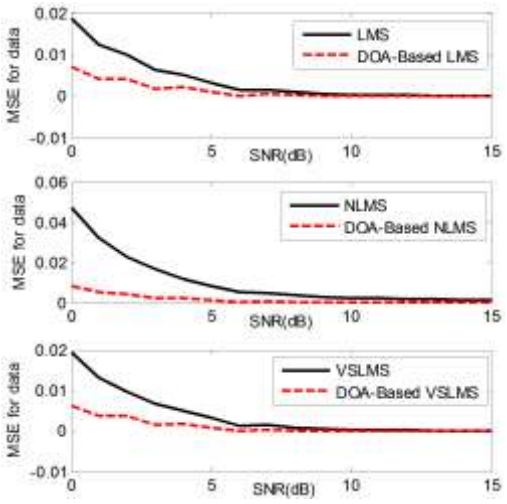


Fig. 13. MSE changes when the desired DOA is in the centre of the initial beam

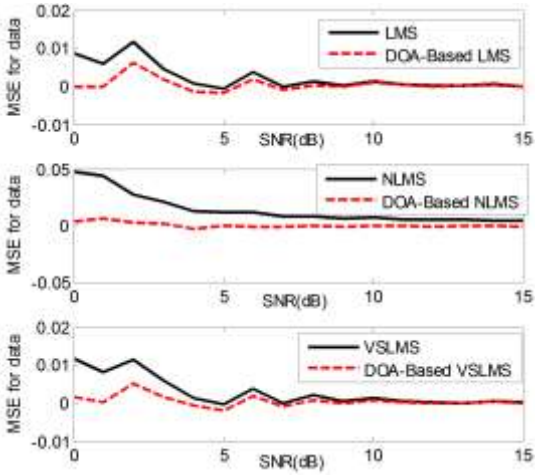


Fig. 14. MSE changes when the desired DOA is in the edge of the initial beam

In addition to the radiation pattern, convergence rate, MSE and SINR analysis, another term has been used for evaluating the performance of the proposed DOA-based LMS/NLMS/VSLMS algorithm compared to conventional one, namely Error Vector Magnitude (EVM) [15]. EVM is used for digitally modulated signals and measures the distortion introduced by the adaptive scheme on the received signal at a given SNR. EVM_{RMS} is defined as [15]:

$$EVM_{RMS} = \sqrt{\frac{1}{KP_O}\left|\sum_{j=1}^{K} S_r(j) - S_t(j)\right|^2} \quad (12)$$

where K is the number of used symbols and P_O is the average power of all symbols for the given modulation. $S_t(j)$ and $S_r(j)$ are the j^{th} transmitted symbol and the j^{th} output of the beamformer, respectively. Figures 15, 16 show the EVM_{RMS} diagrams of the above-mentioned algorithms, conventional as well as proposed, for different values of input SNR ranging from 0–25 dB for two cases of DOA. EVM diagrams show close results in the transmission of symbols using LMS/NLMS/VSLMS beamforming schemes and DOA-based versions. This is reasonable because the mechanism of the algorithm has not changed. In DOA-based LMS/NLMS/VSLMS only

the initial value of the weight vector has been changed so that the convergence rate increases.

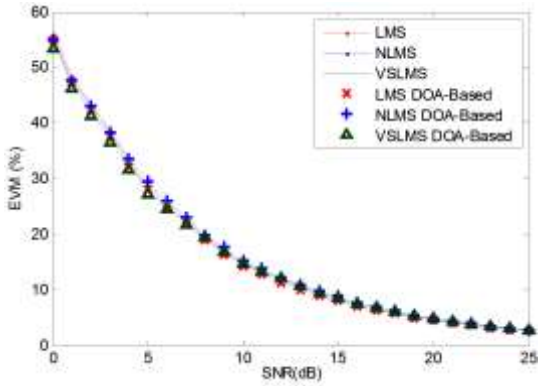


Fig. 15. EVM changes when the desired DOA is in the centre of the initial beam

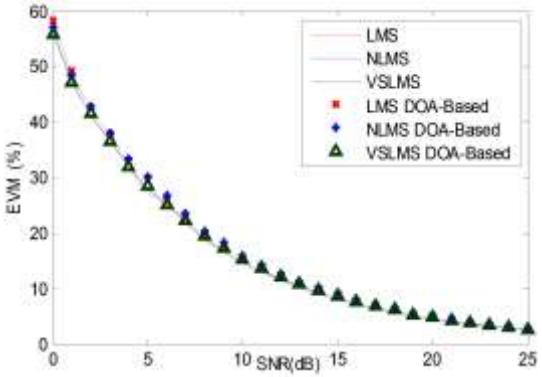


Fig. 16. EVM changes when the desired DOA is in the edge of the initial beam

Comparison of algorithms shows a larger computational cost for the DOA-based approach due to DOA estimation. This cost can be reduced by decreasing the resolution of the DOA estimation technique to a reasonable and sufficient value. In Table 1, the number of multiplications and additions for weight adjustment in the above simulated system is listed. The subspace-based DOA estimation methods such as the MUSIC also need Eigen-decomposition of the input covariance matrix for computation of the spatial spectrum. Consequently, in the DOA-based methods a better convergence speed and capability is obtained with a reasonable increase in the computational load. Utilizing a fast processor the proposed DOA-based algorithms can be converged in a short time while the LMS/NLMS/VSLMS cannot converge in that time even if a fast processor be available.

Table 1. Comparison of computational costs

Algorithm	**Type of Instruction**		
	×	+	/
LMS	1600	1600	-
NLMS	2400	2400	100
VSLMS	2600	3200	-
DOA-based LMS	6373	5448	37
DOA-based NLMS	7173	6248	137
DOA-based VSLMS	7373	7848	37

Two of the main advantages of the proposed method are lower error rate during adaption and better convergence and tracking speed. To investigate the tracking capability, in another simulation trial, the system is assumed to be time varying. The input SNR is 10 dB and the desired signal location is considered once in the centre of the initial beam at DOA = 25° and once in the edge of the fixed beam (DOA = 30°). At iteration 4000 a variation in the system response causes an interruption in the adaptation. The proposed DOA-based methods offer better response in both centre and edge of the initial beam as the best and worst situation for beamforming. The MSE diagrams of this time varying system are demonstrated in Fig. 17, 18 for DOA = 25° and DOA = 30°, respectively. An ideal response for DOA-based methods is observed in centre of the predetermined beam at 25°. Variations of the MSE is less than the conventional LMS/NLMS/VSLMS in the variation time of system when the desired signal is at 30°.

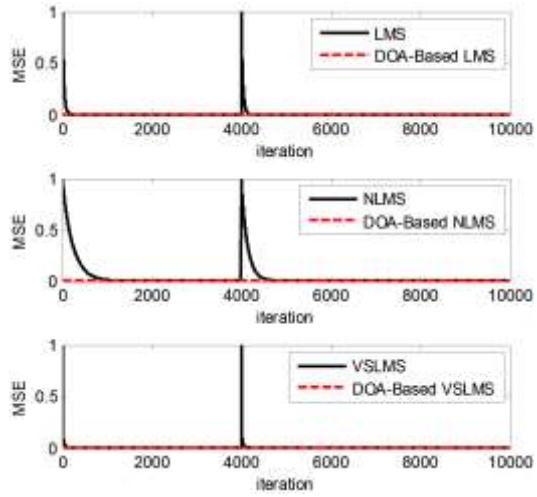


Fig. 17. MSE changes in time varying system when the desired DOA is in the centre of the initial beam

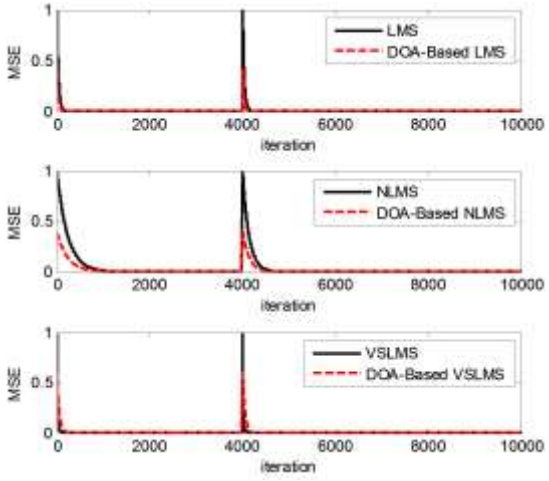


Fig. 18. MSE changes in time varying systems when the desired DOA is in the edge of the initial beam

The proposed method can be extended to the multi-beam applications. In such a condition, more initial states

for joint beams should be considered in the preliminary adjustment and for any combination of M predetermined beams, corresponding to M signal sources, one weight vector is necessary to be calculated at the preliminary adjustment of the array. So, the total number of states of the initial value of the weight vectors will be increased but the beamforming time does not change.

Increasing the number of predefined beams leads to increase in accuracy and convergence rate. So, in order to achieve more accurate beamforming system, more initial narrower beams can be supposed and larger database will be obtained.

In addition to improvement of the LMS/NLMS/VSLMS adaptive algorithms, the proposed method has the advantage of interference nulling capability and higher accuracy than the switched beam technique. Other LMS-type algorithms can be improved using this approach, too. The algorithm also does not need hardware equipment required for conventional switched beam systems.

7. Conclusions

LMS, NLMS algorithms are common algorithms which are exploited for adaptive antenna beamforming. Despite simplicity, high stability and low computational complexity, LMS and NLMS algorithms take long times to converge leading to cause problems in mobile user tracking or adaptation with time-variant channels. In this paper a new spatial-temporal beamforming method is presented that exploits either common LMS or NLMS weighting algorithms and switched beam scheme in a joint state to increase the convergence speed and tracking capability of the conventional LMS or NLMS algorithms. Simulation results show an improved convergence and tracking speed while obtaining a higher efficiency and accuracy in data transmission through increased SINR and decreased BER and MSE. A reasonable increase in computational load occurs during DOA-based techniques which is practicable using signal processor unit.

It is important that the weight vectors obtained by the proposed DOA-based beamforming algorithms are closer than the final desired weight vectors compared to the conventional ones. It can be seen obviously in the comparative studies which are presented in terms of BER, SINR and MSE measures. In addition to the higher performances obtained by the proposed algorithms, faster convergence and also lower deviation in tracking study than the respected conventional ones are the main features of the proposed DOA-based beamforming algorithm. Hence, the type of algorithm cannot affect this issue. However, the results of variable step-size LMS (VSLMS) algorithm are added to the new version of paper. It should be noted that the speed increase will be obtained in the cost of computation of DOA information in VSLMS as well as conventional LMS and NLMS algorithms. Therefore, this result can be obtained in the other constrained or non-constrained LMS or LS family algorithms.

Also, low complexity DOA estimation algorithms can be applied to decrease the required computational complexity.

Recently, new antenna array geometry, Shirvani-Akbari array, is proposed for both DOA estimation and antenna array beamforming in the case of angles close to array endfires [28,29]. As a new work, Shirvani-Akbari array can be combined with the new idea of the present work to find better antenna beamforming in the case of angles close to array endfires.

Acknowledgments

This work was supported by Shahid Rajaee Teacher Training University (SRTTU) under contract number 720 (20.1.1391). We would like to thank Mr. Mahyar Shirvani Moghaddam (University of Newcastle, Australia) for the great help he provided.

References

[1] C.A. Balanis, P.I. Ioannides, "Introduction to Smart Antennas," Morgan & Claypool, 2007.

[2] J.S. Blogh, L. Hanzo, "Third Generation Systems and Intelligent Wireless Networking," John Wiley & Sons, 2002.

[3] M. Chryssomallis, "Smart antennas," IEEE Antenna and Propagation Magazine, vol. 42, no. 3, pp. 129-136, 2000.

[4] R.M. Shubair, M.A. Al-Qutayri, and J.M. Samhan, "A setup for the evaluation of MUSIC and LMS algorithms for a smart antenna system," Journal of Communications, vol. 2, no. 4, pp. 71-77, 2007.

[5] S. Shirvani Moghaddam, M. Shirvani Moghaddam, "A comprehensive survey on antenna array signal processing," Journal of Trends in Applied Sciences Research, vol. 6, no. 6, pp. 507-536, 2011.

[6] H. Singh, R.M. Jha, "Trends in adaptive array processing," International Journal of Antennas and Propagation, vol. 2012, pp. 1-20, 2012.

[7] B.F. Boroujeny, "Adaptive Filters: Theory and Applications," John Wiley & Sons, 1998.

[8] A.H. Sayed, "Adaptive Filters," John Wiley & Sons, 2008.

[9] S. Shirvani Moghaddam, M. Shirvani Moghaddam, and R. Kalami Rad, "CMA-based adaptive antenna array digital beamforming with reduced complexity," In: IEEE, IET International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP2010), 2010, New Castel, England, pp. 375-379.

[10] S. Shirvani Moghaddam, M. Shirvani Moghaddam, "Speed-sensitive weighting algorithm for digital beamforming of adaptive antenna arrays," Wireless Engineering & Technology, vol. 2, no. 3, pp. 165-174, 2011.

[11] Y. Zhang, N. Li, J.A. Chambers, and Y. Hao, "New Gradient-based variable step size LMS algorithms," EURASIP Journal of Advances on Signal Processing, vol. 2008 (Article Id: 529480), pp. 1-9, 2008.

[12] J. Lee, J.W. Chen, and H.C. Huang, "Performance comparison of variable step-size NLMS algorithms," In: World Congress on Engineering and Computer Science (WCECS 2009), 2009, pp. 1-4.
[13] M. Tariqul Islam, N. Misran, "MI-NLMS adaptive beamforming algorithm with tracking ability," Journal of Applied Sciences, vol. 9, no. 12, pp. 2335-2339, 2009.
[14] K. Watanabe, R. Kohno, "Hybrid adaptive algorithm based on temporal update and spatial spectrum estimation for adaptive array antenna," In: 52nd IEEE Vehicular Technology Conference (VTC 2000), 2000, pp. 175 – 180.
[15] J.A. Srar, Kah-Seng Chung, and A. Mansour, "Adaptive Array Beamforming Using a Combined LMS-LMS Algorithm," IEEE Transactions on Antennas and Propagation, vol.58, no.11, pp.3545-3557, Nov. 2010.
[16] J.A. Srar, K.S. Chung, and A. Mansour, "Analysis of the RLMS adaptive beamforming algorithm implemented with finite precision," In: 16th Asia-Pacific Conference on Communications (APCC), 2010, pp. 231-236.
[17] H.V. Kumaraswamy, A. Nawaz, and S.A. Younus, "A new TURBO LMS beamforming for mobile communication," International Journal of Computer Applications, vol. 11, no. 2, pp. 5-9, 2010.
[18] J. Homer, P.J. Kootsookos, and V. Selvaraju, "Enhanced NLMS adaptive array via DOA detection," IET Communications, vol. 1, no. 1, pp. 19-26, 2007.
[19] R.M. de Castro, M.H.C.Dias, and J.A. Apolinario, "On the performance of constrained adaptive algorithms for combined beamforming and AOA tracking of a moving target," In: IEEE International Conference on Microwave and Optoelectronics, 2005, pp. 274-278.
[20] Ch.S. Rani, P.V. Subbaiah, and K.C. Reddy, "MUSIC and LMS algorithms for a smart antenna system," In: IET-UK International Conference of Information and Communication Technology in Electrical Sciences (IETES), 2007, pp. 965-969.
[21] S. Shirvani Moghaddam, M. Shirvani Moghaddam, and R. Kalami Rad, "A novel adaptive LMS-based algorithm considering relative velocity of source," In: 7th IEEE, IET International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP2010), 2010, New Castel, England, pp. 375-379.
[22] S. Shirvani Moghaddam, M. Shirvani Moghaddam, "A new on-line/off-line adaptive antenna array beamformer for tracking the mobile targets," International Journal of Communications, Network and System Sciences, vol. 4, no. 5, pp. 304-312, 2011.
[23] R.H. Kwong, E.W. Johnston, "A variable step size LMS algorithm," IEEE Trans. Signal Process., vol. 40, no. 7, pp. 1633-1642, 1992.
[24] S. Shirvani Moghaddam, F. Akbari, V. Tabataba Vakili, "A novel array geometry to improve DOA estimation of narrowband sources at the angles close to the array endfire," In: 19th Iranian Conference on Electrical Engineering (ICEE2011), 2011, Tehran, Iran, pp. 1-6.
[25] S. Shirvani Moghaddam, S. Almasi Monfared, "A comprehensive performance study of narrowband DOA estimation algorithms," International Journal on Communications Antenna and Propagation (IRECAP), vol. 1, no. 4, pp. 396-405, August 2011.
[26] E. Tuncer, B. Friedlander, "Classical and Modern Direction-of-Arrival Estimation," Elsevier, 2009.
[27] S. Shirvani Moghaddam, F. Akbari, "A novel ULA-based geometry for improving AOA estimation," EURASIP Journal of Advances on Signal Processing, vol. 2011 (Article Id: 39), pp. 1-9, 2011.
[28] S. Shirvani Moghaddam, F. Akbari, "Efficient narrowband direction of arrival estimation based on a combination of uniform linear / Shirvani-Akbari arrays," International Journal of Antennas and Propagation, vol. 2012 (Article Id: 280845), pp. 1-9, 2012.
[29] S. Shirvani Moghaddam, F. Akbari, "Improving LMS/NLMS-based beamforming using Shirvani-Akbari array," American Journal of Signal Processing, vol. 2, no. 4, pp. 70-75, 2012.

Shahriar Shirvani Moghaddam received the B.Sc degree from Iran University of Science and Technology (IUST), Tehran, Iran and M.Sc degree from Higher Education Faculty of Tehran, Iran, both in Electrical Engineering, in 1992 and 1995, respectively. Also, he received the Ph.D degree in Electrical Engineering from Iran University of Science and Technology (IUST), Tehran, Iran, in 2001. He has more than 100 refereed international scientific journal and conference papers, 2 text books on digital communications and one book chapter on MIMO systems. Since 2003, he has been with the Faculty of Electrical and Computer Engineering, Shahid Rajaee Teacher Training University (SRTTU), Tehran, Iran. Currently, he is an associate professor in Faculty of Electrical Engineering and DCSP research laboratory of SRTTU. His research interests include adaptive antenna array beamforming, direction of arrival estimation, D2D communications, Cognitive radio communications and Compressive Sensing.

Farida Akbari received the B.Sc degree from Amir Kabir University of Technology, Tehran, Iran and M.Sc degree from Islamic Azad University, Tehran South Branch, Tehran, Iran, both in Electrical Engineering, in 2005 and 2010, respectively. Her research interests include narrowband adaptive antenna array beamforming and direction of arrival estimation.

An Intelligent Algorithm for the Process Section of Radar Surveillance Systems

Habib Rasi*
Department of Electrical and Electronic Engineering, Shiraz University of Technology, Shiraz, Iran
h.rasi@sutech.ac.ir

Received: 03/Nov/2014 Revised: 14/Nov/2015 Accepted: 21/Dec/2015

Abstract

In this paper, an intelligent algorithm for clustering, intra-pulse modulation detection and separation and identification of overlapping radar pulse train is presented. In most cases, based only on primary features of incoming radar signals, the modern electronic intelligence system cannot recognize the different devices of the same type or class. A very important role is played by Measurement and Signature Intelligence. A radar intercept receiver passively collects incoming pulse samples from a number of unknown emitters. The information such as Pulse Repetition Interval (PRI), Angle of Arrival (AoA), Pulse Width (PW), Radio Frequencies (RF), and Doppler shifts are not usable. In the proposed algorithm, for clustering of overlapping pulses received from self-organization neural network SOFM (due to its high accuracy in comparison with other neural networks, such as CLNN and neural networks (Fuzzy ART), and for detecting intra-pulse modulation type, matrix method, and for identifying the radar type, RBF neural network have been used. The simulation results of the proposed algorithm shows that in the presence 5% noise and 5% missing pulse, the accuracy of the clustering part of the proposed algorithm is equivalent to 91/8%, intra-pulse modulation recognition accuracy is 98%, the detection accuracy is 2/99%, and the total output of the algorithm precision is 89/244%, respectively.

Keywords: Pulse Train; Matrix Multiplication Method; Radar Identification; Neural Networks.

1. Introduction

The main parts of the electronic equipment of military forces are radars, thus, identifying them is of particular importance. Figure 1 shows the general division of electronic warfare recently termed as electronic defense [1]. Most of the systems that are used to detect enemy electronic equipment systems are ELINT and ESM. The responsibility of ELINT system is strategically accurate identification of active radars in the region and the responsibility of ESM systems is immediate identification of radars deployed in the threatening equipment so that the type of threats can be revealed by them. In general, the task of ELINT and ESM systems are similar and they are only different in their duration of performance time. With successful and sustained improvements in the technology of constructing effective radars and the immense complexity of the regional combat, the effectiveness of disturbance systems and electronic deception is highly dependent on the performance of radar detection system. Thus, the performance of electronic attack sections (EA) and the electronic protection (EP) in the radar field (Figure 1) directly depend on performance of radar detection systems. Radar detection systems include sections such as antennas, receivers, processors and displayer. In these systems, processor has the task of clustering, separating and identifying radars. [2]

Figure 2 shows block diagram of the processing unit of ELINT and ESM systems. As it can be seen in Figure, first, the pulse details of word (PDW) is extracting for all of overlapping pulses received in the given time frame and then according to the extracted PDWs, clustering operation is performing on the pulses. Due to the possibility of existing of different pulse trains in clusters, the processor operates separation clusters and eventually detects the pulses on each cluster. Gained information from the identification of pulse trains and pulse PDWs is the basis for comparison with existing data in the radar database that will identify the types of threat. [3]

In the proposed algorithm, for clustering and separation of overlapping pulse strings received from the region radars, neural networks with the feature of self-organization for detection of PRI type and calculating PRI average, by using the methods of matrix multiplication for identification of the radar type, neural networks with radial basis function are used. In section 2 self-organizing neural networks and radial basis function and in section 3, the proposed algorithm will be presented. Sections 4 and 5 respectively evaluate the proposed algorithm and presents the conclusion.

* Corresponding Author

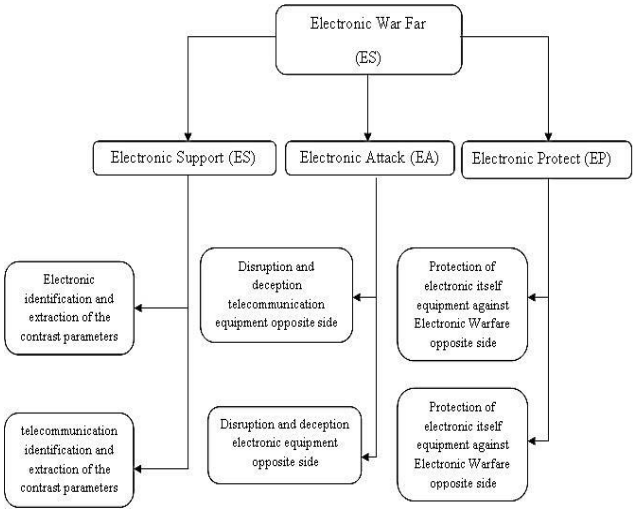


Fig. 1. The position of radar detection systems in electronic warfare

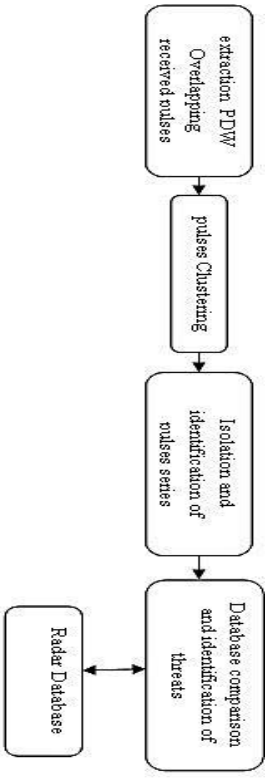


Fig. 2. General block diagram of the processing unit in radar detection systems

2. Self-organizing Neural Networks and Radial Basis Function

One of the most commonly used artificial neural networks is self-organizing neural networks (SONN). Till now several neural networks with self-organizing feature has been reported that three of the most commonly used of them, are CLNN, SOFM and Fuzzy ART networks[4]. CLNN neural networks have a two-layer and leading structure, the first layer is the feature domain encoder and the second layer is competitive layer that its neurons such generalized themselves to they can recognize the input vectors provided.

Each neuron of the second layer is connected with all neurons of the first layer by weight vectors. Each neuron in the competitive layer through a competitive process by stimulating local connections, stimulates itself and perhaps some neighboring neurons and reduces the activity of the farther neurons by inhibiting connections. In this network, after enough training, each output neuron represents a cluster and its weights represent the center of the cluster. [4,5]

SOFM neural networks are like CLNN, the only difference is that bias is not used in it. In this network, in addition to classifying input vectors, the neighboring neurons recognize adjacent parts of the input space [5]. Fuzzy-neural networks, which have been developed in recent years, use fuzzy logic gates. Fuzzy ART network is a kind of network that combines the theory of fuzzy calculation with the ART1 neural network and accepts binary and analog inputs [6]. Artificial neural networks with radial basis function (RBF) are two-layer networks with radial basis activation functions and have been proposed for various applications in signal processing [4].

Radial basis function is a multidimensional function that its output, depends on the distance of between the input vector and the center vector. In RBF networks nonlinear basis functions can take many forms like as Gaussian and polygons, etc. In practical applications, is using mostly the Gaussian function, which is known as the Gaussian RBF neural networks [7]. Two different types of RBF networks are regression networks and probabilistic networks. Regression networks are mostly used in estimating functions and probabilistic networks in classification problems.

In the probabilistic neural network, when the input vector is applied to the network, the first layer calculate the distance between input vector and the training input and thus provides a vector that its elements determines the distance between the input and training input. The second layer using the first layer output generates vector of probabilities as output of the network. Finally, the competitive transfer function in the second layer selects the maximum probabilities from vector probabilities, and produces 1for that output and 0for the rest of probabilities.

3. The Proposed Algorithm

Figure 3 shows block diagram of the proposed algorithm. In the proposed algorithm, after reception of PDW from detector and pulse analyzer, the normalization operating is performing on the input data to prepare the data to applying for clustering section. After clustering the input pulses into several clusters, types of PRI modulation in each cluster is extracted using matrix multiplication. Then according to the three parameters of pulse width, carrier frequency, and PRI, the radar types detected using PNN neural network. If the recognized

specifications does not matched with existing radars in radar data archive, as a new radar will be added to the radar data archive.

In following, details of the proposed algorithms, including algorithms of clustering section of received strings of overlapping pulses, algorithms of PRI type detection, and separation and identification of radars are presented.

3.1 Algorithms of Clustering Section of Received Strings of Overlapping Pulses:

In this section, an intelligent algorithm is designed for clustering of received pulse trains overlapping, based on self-organizing neural network. Figure 4 shows the flowchart of this part of the algorithm where in the first, three parameters of AOA, RF, and PW from PDW are selected then in the section of pre-processing and normalization, a row or column of symmetric matrix D is calculated as follows, then its elements are normalized between 1 and 0.

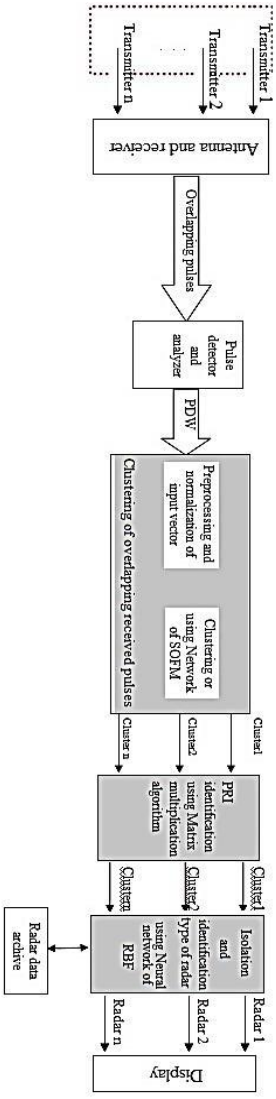


Fig. 3. Overall block diagram of the proposed algorithm

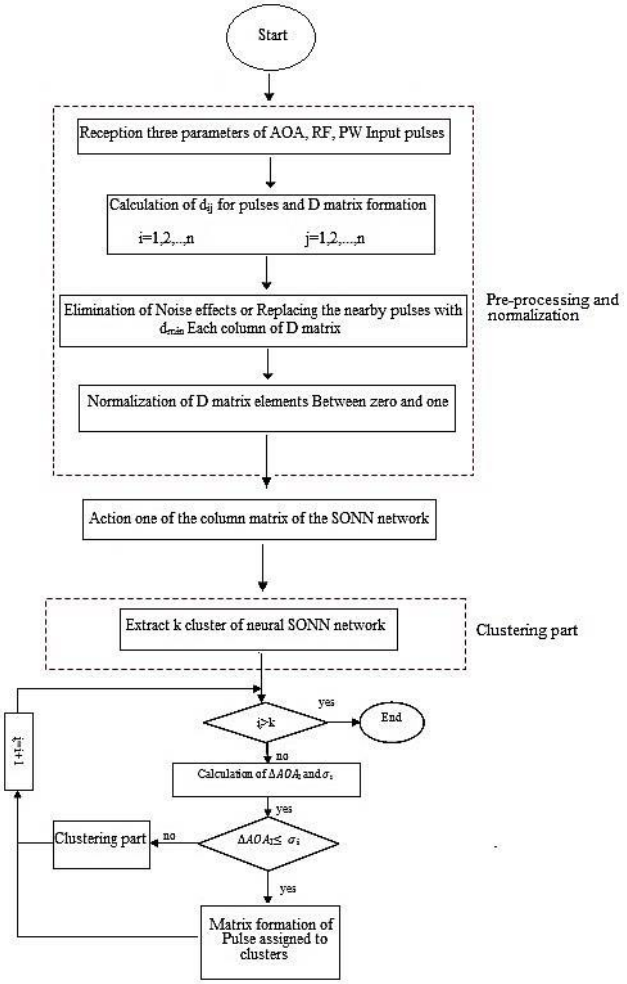


Fig. 4. Flowchart of the clustering of the proposed algorithm.

$$D = \begin{bmatrix} 0 & d_{12} & \cdots & d_{1n} \\ d_{21} & 0 & \dots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & 0 \end{bmatrix} \quad (1)$$

$$d_{ij} = \sqrt{\frac{(AOA_i - AOA_j)^2}{\sigma_{aoa}{}^2} + \frac{(PW_i - PW_j)^2}{\sigma_{pw}{}^2} + \frac{(RF_i - RF_j)^2}{\sigma_{rf}{}^2}} \quad (2)$$

The values of one of the rows or columns of the matrix D are applied to SONN neural network and with training the network, clustering pulses are performing. For each cluster, difference of the entry pulses angle are compared with the value (the value is selecting with respect to the accuracy requires to measure the entry angle radars in operation area, at this point the value has been set at 2.5° with respect to the accuracy of the existing systems), if the difference of entry angle of pulses was less than 2.5°, for that cluster matrix M calculates as equation (3), and if it is more than 2.5°, that cluster is archiving. Then for each cluster of the archive, the algorithm is executed and the process will be continued until the number of clusters in the archive is zero.

$$M = \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \\ \vdots \\ p_m \end{array} \begin{array}{c} \begin{array}{ccccc} c_1 & c_2 & c_3 & \dots & c_k \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & \cdots & 0 \end{bmatrix} \end{array} \quad (3)$$

3.2 PRI Type Recognition Section of the Proposed Algorithm

After the clustering process, the obtained clusters enter the intra-pulse modulation detection section and calculation section of PRI average. In this section, matrix multiplication technique is using the following way that its aim to detecting of techniques of fixed PRI, PRI stager, PRI Jitter and calculating mean of PRI clusters.

This method involves the following steps:

- For N pulse, harmonic matrix of cluster is formed as follows:

$$HM = \begin{bmatrix} 0 & 1 & 2 & 3 & \dots & N-1 \\ 1 & 0 & 1 & 2 & \dots & N-2 \\ 2 & 1 & 0 & 1 & \dots & N-3 \\ 3 & 2 & 1 & 0 & \dots & N-4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ N-1 & N-2 & N-3 & N-4 & \dots & 0 \end{bmatrix} \quad (4)$$

- Matrix of the difference in arrival times of pulses is calculated which is a symmetric matrix.

$$\Delta TOA(I,j) = |TOAj - TOAi|\ ,\ 1 \le I,j \le N \quad (5)$$

$$\Delta TOA = \begin{bmatrix} 0 & d_{12} & d_{13} & d_{14} & \dots & d_{1N} \\ d_{21} & 0 & d_{23} & d_{24} & \dots & d_{2N} \\ d_{31} & d_{32} & 0 & d_{34} & \dots & d_{3N} \\ d_{41} & d_{42} & d_{43} & 0 & \dots & d_{4N} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & d_{N3} & d_{N4} & \dots & 0 \end{bmatrix} \quad (6)$$

- By multiplying matrix of difference in arrival times of pulses and the inverse matrix of HM, the detection matrix of pulse train identifying (PTI) is obtaining.

$$PTI = \Delta TOA \times HM\text{-}1 \quad (7)$$

To reduce the computational complexity, we can only compute the main diagonal elements instead of calculating PTI matrix with the following equation:

$$V_{PTI}(i) = \sum_{j=1}^{N} \left(\Delta TOA_{ij} \times HM_{ji}^{-1}\right)\ ,\ i = 1,2,\dots,N \quad (8)$$

- With reviewing of VPTI vector elements, the used technique in PRI is specifying.

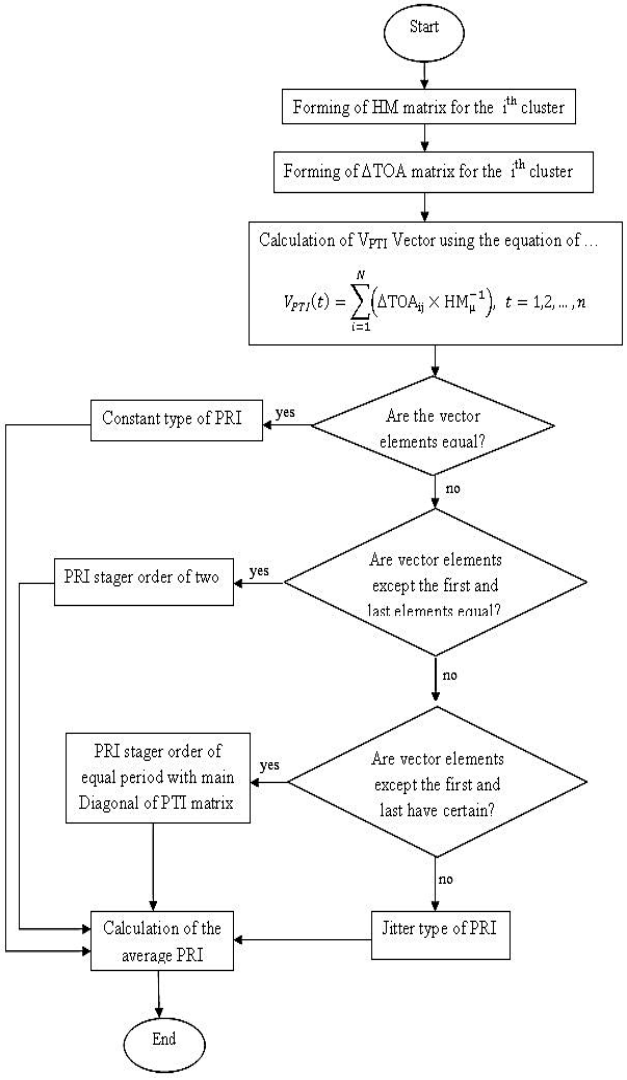


Fig. 5. Flowchart of PRI detection section of the proposed algorithm

3.3 Separation and Identification Section of the Proposed Algorithm:

The first step to identifying the type of radar by PNN neural network, formation of the pulse descriptor word vector (VPDW). To form vector VPDW from three inherent parameters of the radar such as RF, PRI and PW are used [10]. VPDW vector is formed for Ith cluster with N received pulses as follows.

$$V_{PDW}(i) = [PDW_1(i) \quad PDW_2(i) \quad \dots \quad PDW_N(i)] \quad (9)$$

In which PDW for jth pulse of ith cluster is defined as in the equation (10):

$$PDW_j(i) = \begin{matrix} RF_j(i) \\ PW_j(i) \\ PRI_j(i) \end{matrix} \quad (10)$$

Thus for m cluster, the pulse descriptor word vector is as follows:

$$V_{PDW} = [V_{PDW}(1) V_{PDW}(2) \quad \dots \quad V_{PDW}(m)] \quad (11)$$

Then mean of VPDW vector for m clusters are calculated as follows:

$$\bar{V}_{PDM}=\begin{matrix}\overline{RF1}\\ \overline{PW}\ 1\\ \overline{PRI}1\end{matrix}\ ,\ \begin{matrix}\overline{RF2}\cdots\\ \overline{PW}2\cdots\\ \overline{PRI}2\cdots\end{matrix}\ \begin{matrix}\overline{RF}m\\ \overline{PW}m\\ \overline{PRI}m\end{matrix} \quad (12)$$

To learn the above-mentioned neural network, first, the matrix of $\bar{V}_{PDW}$ of all the radars in radar data archive is calculating and training to the network. Also for the input clusters $\bar{V}_{PDW}$ matrices are calculated and applied to the network for detection. The network detects the type of radar corresponding to each cluster by comparing $\bar{V}_{PDW}$ matrix of the input clusters with what has been trained. Figure 6 shows the flowchart of this section of the proposed algorithm.

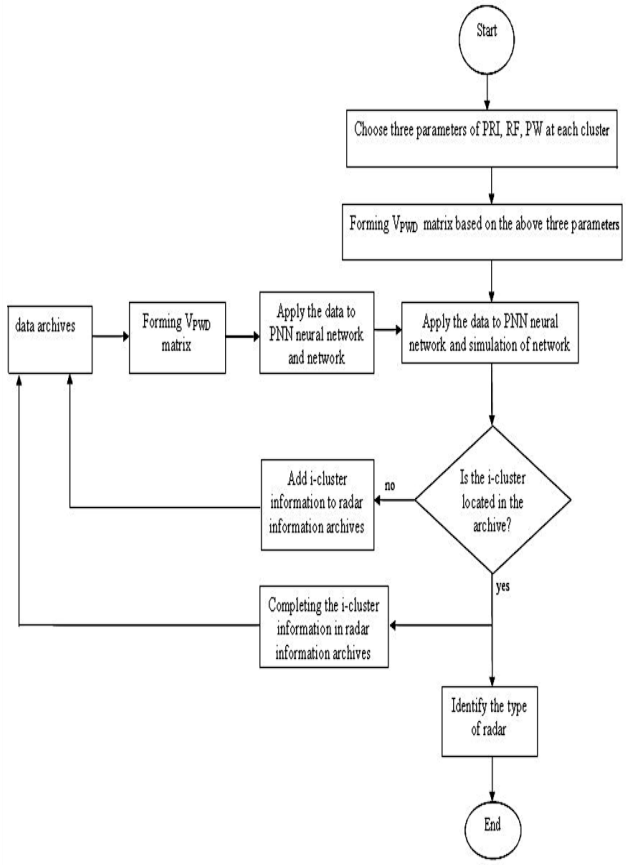


Fig. 6. Flowchart of identification section of the proposed algorithm

4. Reasons for Selecting Neural Network and Evaluating of the Proposed Algorithm

A. Selection of SOFM Neural Network and Evaluation of Clustering Section

Among the most commonly used self-organized neural networks (CLNN, SOFM and Fuzzy ART), must be selected the most suitable for clustering section of proposed algorithm. For this purpose, first, the networks are simulated using MATLAB software and then the produced data for the three parameters AOA, RF and PW from five radars, as specified in Table 1 are applied to them. These three networks have been proposed for clustering section and were compared in the terms of accuracy of clustering (error) and convergence time.

Table 1. Five radars with different capabilities

radar	AOA (deg)	RF (MHz)	PW (μs)	PRI (μs)	PA (dB)	PRI Type	RF Type
1	32	2780	3.1	2300	10	Stable	Stable
2	38	2887	2.7	2600 2800 2900	28	3 Order stagger	Stable
3	45	2670	1.3	3000	14	Jitter	Stable
4	35	2500	0.8	2700	45	Stable	Jump
5	48	2712	0.23	3100	37	Stable	Agile

Figure 7 shows the errors of the three networks for the various iterations of training. As can be seen, the networks have converged after 300 iterations. For 300 times iterations of the training, the network's error and the time required to training the network that is convergence time which represents the computational complexity, are given in Table 2 (the calculations are done by Pentium 4 computer with 2 GB of RAM).

As can be seen, the error of SOFM neural network is less, thus this network was chosen for clustering section of the proposed algorithm. According to Table 2, the clustering section of the proposed algorithm with choice of SOFM neural network and 500 times iteration training has an accuracy of about 91.8% (8.2% error).

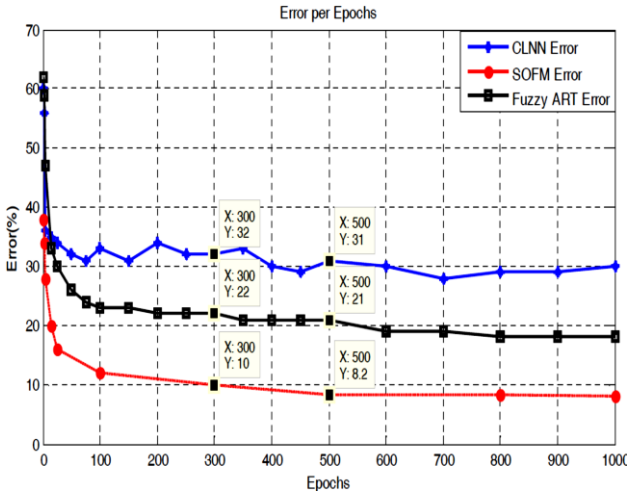


Fig. 7. Comparison of errors of CLNN, SOFM and Fuzzy ART neural network

Table 2. Output error and training time of self-organizing networks in the clustering section of the proposed algorithm.

The time required for training Repeat 500 times (s)	Error after 500times of training	The time required for training Repeat 300 times(s)	Errorafter 300 times of training	Kind of self-organizing networks
78	8.2%	45	10%	SOFM
25	21%	15	22%	Fuzzy ART
11	31%	6	32%	CLNN

B. Assessment of Recognition Section of Inter-pulse Modulation Type

For the evaluation of this part of data's produced, Table 1 was used. For data that generated from the angle of arrival and pulse width, the changes were added as Gaussian noise with a variance of 5%. For data relating to the frequency based on the technique, frequency change of variances is different. The variance of changes in fixed frequency radars is set as 5%, in radars with capability of frequency jumping is set as 10%, and in radars with capability of frequency agility changes is set as 30%. For generated data from the PRI parameter, the variance of changes in radars with fixed PRI technique was set at 5%, in Astgger PRI technique is set as 10%, and in jitter PRI technique is set as 30%, and the generated data was applied to the algorithm matrix multiplication considering 5% of missing pulses. By running the algorithm for 1000 times and calculating the average of errors, the accuracy of the algorithm was 98% obtained.

In this section, instead of calculating the pulse train identifying matrix, only calculation of the main diagonal elements is proposed (pulse train identifying vector: VPTI) in equation 8. To compare calculation time of the two equation, pulse train with PRI = 10 and N = 8, with 5% of missing pulses, and 5% error and PTI matrix and VPTI vector was calculated as follows.

$$PTI = \begin{bmatrix} 10.0470 & 0.0658 & 0.0280 & 0.0494 & 0.0124 & 0.0167 & 0.0304 & 0.0014 \\ 0.0131 & 9.9942 & 0.0280 & 0.0494 & 0.0124 & 0.0167 & 0.0304 & 0.0014 \\ 0.0131 & 0.0658 & 9.9564 & 0.0494 & 0.0124 & 0.0167 & 0.0304 & 0.0014 \\ 0.0131 & 0.0658 & 0.0280 & 10.0337 & 0.0124 & 0.0167 & 0.0304 & 0.0014 \\ 0.0131 & 0.0658 & 0.0280 & 0.0494 & 10.0708 & 0.0167 & 0.0304 & 0.0014 \\ 0.0131 & 0.0658 & 0.0280 & 0.0494 & 0.0124 & 10.0752 & 0.0304 & 0.0014 \\ 0.0131 & 0.0658 & 0.0280 & 0.0494 & 0.0124 & 0.0167 & 10.0615 & 0.0014 \\ 0.0131 & 0.0658 & 0.0280 & 0.0494 & 0.0124 & 0.0167 & 0.0304 & 10.0325 \end{bmatrix}$$

$$V_{PTI} = [10.0470 \quad 9.9942 \quad 9.9564 \quad 10.0337 \quad 10.0708 \quad 10.0752 \quad 10.0615 \quad 10.0325]$$

As can be seen, the obtained elements for the vector VPTI were the same as the main diagonal elements of the PTI matrix. However, the computation time of VPTI vector is less than the PTI matrix. Figure 8 shows the computation time for VPTI vector and PTI matrix for the different numbers of input pulses (calculated by computer, Pentium 4 with 2 GB of RAM). As seen in this figure, when the number of input pulses is high, computation time of VPTI vector is much less compared to PTI matrix.

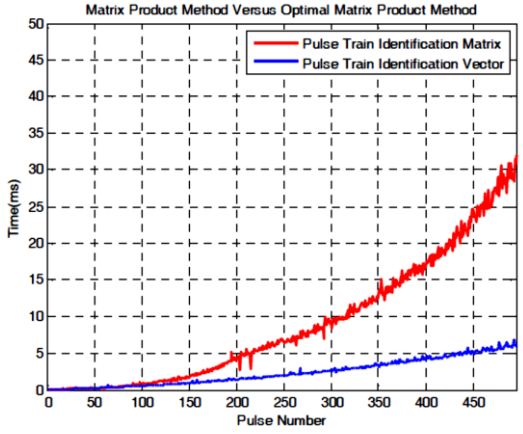


Fig. 8. Comparison of computation time of VPTI vector and PTI matrix

C. Selection of PNN Neural Network and Evaluation Section of Radar Type Identification

PNN neural network, is a type of RBF neural network which has high learning speed compared to perceptron multilayer neural networks and other monitoring networks, and is suitable for real-time processing applications. Also, with increasing of data training, it has better performance than MLP networks [10,11]. According to the mentioned features, PNN neural network was preferred to the MLP neural network for identification part of the proposed algorithm.

For the evaluation of the identification part of the algorithm, production data including RF, PRI and PW parameters of 20 practical radars with specifications as shown in Table 3 were applied to the PNN neural network and the network is training. After training the data's of 20 radars in Table 3, by introducing new input vectors we can recognize specification of related to them. For this purpose, a new input vector was applied to PNN network and the results were evaluating.

Each radar which is closer to the radars in the archives that radar will be announced as chosen, and if the difference is more than a given amount, it is considered as new radar and its specifications are added to the radar data archive.

To demonstrate this, pulse train of three radars (Table 2) was produced and applied to the PNN neural network with an error of 10% and 10% of missing pulses (Table 4). Produced data from three radars are shown in figure 9.

The results of learning (20 radars in Table 3) and applied the pulse train of three radars in Table 4 to the PNN neural network are shown in figure 10. In this figure, the circular points present the classification of 20 radars and star points related to the three radars which applied for detection. As shown in figure 10, the points resulting from the applying the three radars (the star points) are close to the learning points of the radars 3, 12 and 17. thus can be conclude that the three radars are correctly recognized. The results of Monte Carlo simulations with 1000 iterations shown that the proposed method to identifying the separated pulse trains with 5% error and 5% missing pulses, have the accuracy of about 99.2 %.

Table 3. Characteristics of the radars in the archive

Type Radar	PRF(MHz)	PW(μs)	RF(MHz)
Rdar 1	500	2	1000
Rdar 2	300	4	1500
Rdar 3	850	20	2500
Rdar 4	1500	1.2	3000
Rdar 5	800	20	3500
Rdar 6	700	1	3000
Rdar 7	900	100	2800
Rdar 8	2300	36	4000
Rdar 9	500	3.3	5000
Rdar 10	2800	1.2	5150
Rdar 11	500	3	8000
Rdar 12	6000	1.5	9000
Rdar 13	200	0.4	20000
Rdar 14	300	0.02	20000

Type Radar	PRF(MHz)	PW(μs)	RF(MHz)
Rdar 15	3000	10	18000
Rdar 16	2400	0.14	33000
Rdar 17	675	1.1	16200
Rdar 18	300	0.02	18000
Rdar 19	300	0.02	13000
Rdar 20	400	3	4300

Table 4. Characteristics of 3 radars implemented to evaluate the identification part of the proposed algorithm

Number radar in Table 3	RF(MHz)	PW(μs)	PRF(Hz)	PRF Type
3	2000-3000	20	850	Constant
12	8600-9500	1.5	4800-8100	3 Order staggered
17	16000-16400	1.1	674	Constant

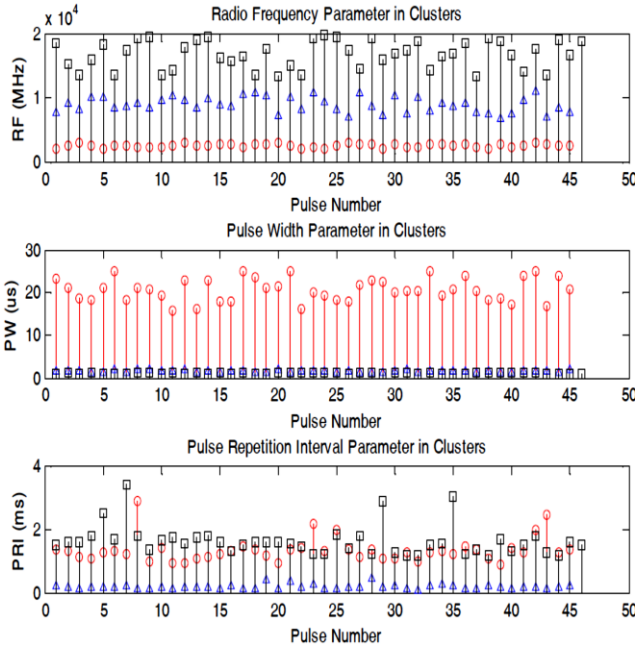


Fig. 9. Data generated from three radars for application to identification part

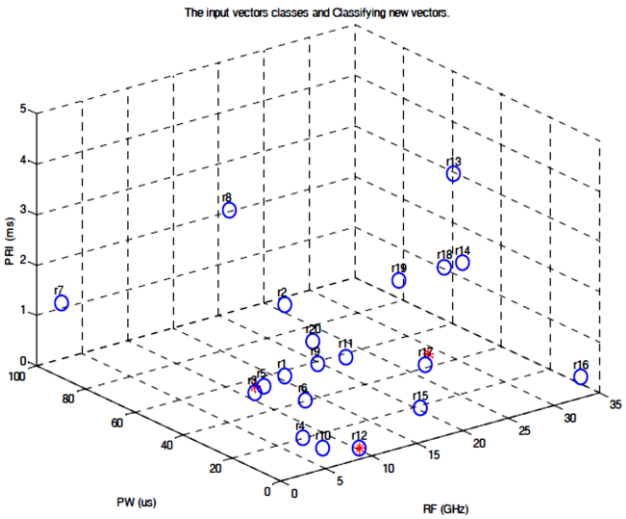


Fig. 10. The output of the RBF network after learning archive radars and detecting new radars

D. The Results of the Evaluation of the Proposed Algorithm

According to the taken evaluation, the obtained accuracy by using statistical methods for different parts of the proposed algorithm are shown in Table 5.

Table 5. Accuracy of the various parts of the proposed algorithm.

Accuracy of equivalent	Part name
91.8%	Clustering part using neural networks
98%	Part of between pulses Modulation recognition using matrix multiplication method
99.2%	Part of Identify radar using neural network

According to the table (5), to identifying radars with stagger and Jitter intra-pulse modulations, capability of the frequency jumping is 89.244 % by applying 5% error and 5% resultant noise, the accuracy of the proposed algorithm.

Table 6 shows the performance time of the proposed algorithm. As can be seen, the total performance time of the algorithm is about 4 milliseconds which it is a good time for operating equipments.

Table 6: Execution time of the proposed algorithm.

Time(ms)	Part name
2	Clustering part using neural networks
1	Part of between pulses Modulation recognition using matrix multiplication method
1	Part of Identify radar using neural network

5. Conclusions

In radar detection systems, the processor must have the least sensitivity toward deliberate changes in the pulse parameters which among TOA parameter have the most contribution. Unlike most methods, the proposed algorithm does not use this parameter in clustering and separating. In the recognition part of the intra-pulse modulation type, because using the matrix multiplication method and the possibility of its implementation using systolic array, processing speed of this section is suitable for real-time systems. In the proposed algorithm, the accuracy of clustering section is 91.8%, the accuracy of recognition part of intra-pulse modulation type for a pulse train with 5% missing pulse and 5% noise is about 98% and the accuracy of identification section for a pulse train with 5% missing pulse and 5% noise is about 99.2%. In general, the resultant accuracy of the proposed algorithm is 89.244 %.

References

[1] Stimson G. W. Introduction to Airborne Radar. Artech House, Second Edition, 1998.
[2] Wiley R. G. Electronic Intelligence: The Analysis of Radar Signals. 2nd Edition, Artech House, Inc., 1993.
[3] Hassan A., Chan F., Chan Y. T., "Joint Deinterleaving and Recognition of Radar Pulses", IEEE Radar Conference 2003, pp. 177-181.
[4] Hu, Y. Hen, Hwang, J. Neng. Handbook of Neural Network Signal Processing. Artech House, 2002.
[5] Savaria, Eric, Lavoie, Pierre .A, "Comparison of Self-Organizing Neural Networks for Fast Clustering of Radar Pulses", Puplished by Elsevier Science B.V. 1998, Signal Processing 64, pp. 249-269.
[6] Kenaya, Riydah. Eucliden ART Neural Networks, Proceedings of the World Congress on Engineering and Computer Science, WCECS 2008, San Francisco.
[7] Zhao, Chuang; Zhao, Yongjun; Lu, Jianqi, "Radar Signals Sorting with Kohonen Neural Network", 8th Signal Processing International Conference 2006, IEEE, pp. 16-20.
[8] ANDERSON, J. A.; GATELY, M. T, "Radar Signal Categorization Using a Neural Network"; International IEEE Signal Processing Conference 1990, 78(10).
[9] Mardia, H. K, "New Techniques for the Deinterleaving of Repetitive Sequences"; Radar Conference 1989, IEE Proc. 136(4), pp. 149-154.
[10] Dudczyk, J, Kawalec, A, Cyrek, J., "Applying the Distance and Similarity Functions to Radar Signals Identification", International Radar Symposium 2008, pp. 1-4

Habib Rasi received the B.Sc in Electrical and Electronic Engineering in 2009 in Sattari university, Tehran, Iran, M.Sc of radar signal processing in 2012 in Islamic azad university of Shahr-e Rey branch in Tehran province, and now Ph.D student in LPI radar signal processing in Shiraz University of Technology, Shiraz, Iran. His research interests include Radar Signal Processing, Radar Signal Design, Statistical Signal Processing, Electronic Warfare Systems

A New Method for Transformation Techniques in Secure Information Systems

Hodjatollah Hamidi*
Department of Industrial Engineering, K. N. Toosi University of Technology, Tehran, Iran
h_hamidi@kntu.ac.ir

Received: 04/Jan/2015 Revised: 03/Jan/2016 Accepted: 23/Jan/2016

Abstract

The transformation technique relies on the comparison of parity values computed in two ways. The fault detection structures are developed and they not only detected subsystem faults but also corrected faults introduced in the data processing system. Concurrent parity values techniques are very useful in detecting numerical error in the data processing operations, where a single fault can propagate to many output faults. Parity values are the most effective tools used to detect faults occurring in the code stream. In this paper, we present a methodology for redundant systems that allows to detect faults. Checkpointing is the typical technique to tolerate such faults. This paper presents a checkpointing approach to operate on encoded data. The advantage of this method is that it is able to achieve very low overhead according to the specific characteristic of an application. The numerical results of the multiple checkpointing technique confirm that the technique is more efficient and reliable by not only distributing the process of checkpointing over groups of processors. This technique has been shown to improve both the reliability of the computation and the performance of the checkpointing.

Keywords: Transformation Techniques; Information Systems; Redundancy; Checkpointing.

1. Introduction

The checkpointing of approaches considers the specific characteristic of an application and designs fault tolerance schemes according to the specific characteristic of an application [1].

Transformation techniques is a class of approaches which tolerant byzantine failures, in which failed processors continues to work but produce incorrect calculations [2]. The transformation techniques approach transforms a system that does not tolerate a specific type of fault, called the fault-intolerant system, to a system that provides a specific level of fault tolerance, namely recovery [3]. In transformation techniques, applications are modified to operate on encoded data to determine the correctness of some mathematical calculations. The Transformation techniques of approaches can mainly be applied to applications performing linear algebra computations and usually achieves a very low overhead. One of the most important characteristics of this research is that it assume a fail-continue model in which failed processors continues to work but produce incorrect calculations [4-5].

Transformation techniques can be tuned to provide the desired fault tolerance e.g., single error detection, single error correction, etc. For some computations, transformation techniques can be implemented with low overhead as shown in [6] and [7]. Transformation techniques applies error control codes to the data such that errors are detected and in some cases located and corrected. An example of transformation techniques is to encode matrices by adding checksum rows or columns as discussed in [1], [8]. Checksum encoding is used to generate what is called a "checksum matrix" from the original matrix.

Faults in numerical data processing may be detected efficiently by using parity values associated with real number codes, even when inherent round off errors are allowed in addition to failure disruptions. The basic approach for protection was discussed in [9] and has been expanded in many ways since, e.g., [10], [11], [12]. The real number convolutional codes discussed here allow data and parity values to be processed in a continuous fashion unlike block real number codes which require data and parity segmentation. There are many applications where such convolutional codes can provide protection including satellite, communication, signal processing, and large data processing systems.

Transformation techniques for arithmetic and numerical processing operations is based on linear codes. G. Bosilca et al. [13] for high-performance computing, propose a new transformation techniques method based on a parity check coding. Redinbo [14-16] presented a method to Wavelet Codes into systematic forms for Algorithm-Based Fault Tolerance applications. This method employ high-rate wavelet codes along with low-redundancy which use continuous checking attributes to detect the errors, in this paper since their descriptions are at the algorithm level can be applied in hardware or software. But, this technique is suited to image processing and data compression applications and is not a general method. Also, other constraint is on burst-error due to computational load high relatively. Moreover, there is

* Corresponding Author

onerous analytical approach to exact measures of the detection performances of the transformation techniques technique applying wavelet codes.

The paper is organized as follows: In section 2, we discuss architecture of the transformation techniques technique. In section 3, we propose the error correction system. In section 4, we discuss Factorization. In section 5, we Analysis Check pointing. In section 6, to be discussed conclusions.

2. Architecture of Transformation Techniques

To achieve fault detection and correction properties of this code in a linear process with the minimum overhead computations [15], we propose the architecture in Fig. 1.

The advantage of transformation technique is that errors which are caused by permanent or transient failures in the system can be detected and corrected by using a very low overhead and at the original throughput. Real number codes involve symbols that have real or integer values as opposed to classic binary codes. The real number convolutional codes hold great promise of protecting many data processing subsystems. There are times when the error detection capabilities of transformation techniques are not enough. Concurrent error correction at the data-level for compensating the effects of intermittent failures avoids disrupting the data flow to react to detected errors. Convolutional codes which employ real-number symbols are difficult to decode because of the size of the alphabet such codes find applications in both fault-tolerance support for signal processing subsystems and in channel coding for communication systems. In order to achieve fault detection and correction properties of convolutional code in a linear process with the minimum overhead computations, the architecture is proposed. For error correction purposes, redundancy must be inserted in some form and convolution parity codes will be employed, using the transformation technique. A systematic form of convolutional codes is especially profitable in the transformation technique detection plan because no redundant transformations are needed to achieve the processed data after the detection operations. To achieve fault detection and correction properties of convolution code in data processing with the minimum additional computations, the block diagram is proposed in. The data processing operations are combined with the parity generating function to provide one set of parity values.

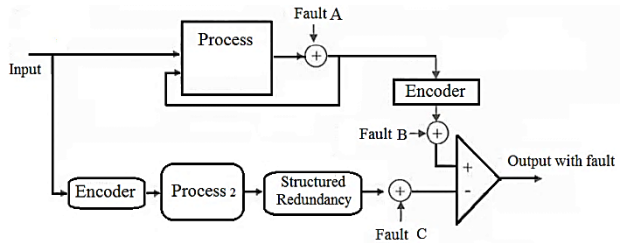


Fig. 1. Our architecture of transformation techniques

We have modeled faults in a linear process block with module fault A while the encoder and structured redundancy faults are modeled with modules B and C. Since these two last faults contribute in syndrome additively we can delete one of them without any degradation. Convolutional codes are usually used over the transmission channels, through which both information and parity bits are sent. The main architecture is similar to a normal transformation techniques scheme except of the structured redundancy and delay line in the information pass which replace the parity generator part of a systematic Convolutional encoder. The upper way is the normal Process data flow which passes through the nonlinear process block and then fed to the Convolutional encoder to make parity sequence the structured redundancy. So the syndrome sequence is a stream of zero or near zero values in normal operation [17].

2.1 Redundant Implementation

In order to avoid replication when constructing fault tolerant dynamic systems, we replace the original system with a larger, redundant system that preserves the state, evolution and properties of the original system - perhaps in some encoded form. We impose restrictions on the set of states that are allowed in the larger dynamic system, so that an external mechanism can perform error detection and correction by identifying and analyzing violations of these restrictions. The larger dynamic system is called a redundant implementation and is part of the overall.

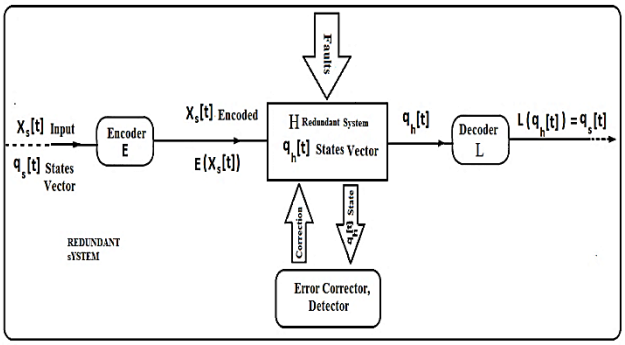


Fig. 2. Fault tolerant structure

Fault tolerant structure shown in Fig. 2,[18], the input to the redundant implementation at time step t, denoted by e(x[t]), is an encoded version of the input x [t] to the original system; furthermore, at any given time step t, the state $q_s[t]$ of the original system can be recovered from the corresponding state $q_h[t]$ of the redundant system through a decoding mapping L (i.e., $q_s[t] = L(q_h[t])$). Note that we require the error detection/correction procedure to be input-independent; so that we ensure the next-state function is *not* evaluated in the error-correcting circuit [19].

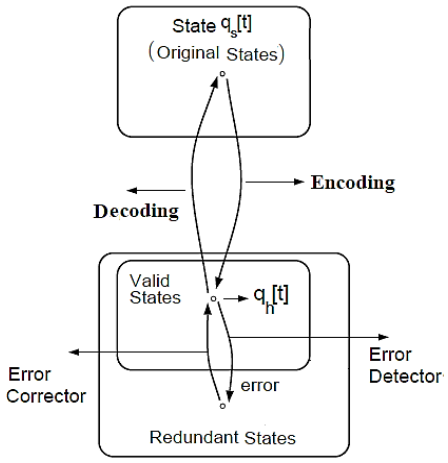


Fig. 3. Reliable state evolution using unreliable error-correction

This approach uses the scheme shown in Fig. 3 but allow failures in both the redundant implementation and the error-correcting mechanism. Clearly, since all components of this construction are allowed to fail, the system will not necessarily be in the correct state at the end of a particular time step. What we hope for, however, is for its state to be within a set of states that *correspond* to the correct one: in other words, if a fault-free error corrector/decoder was available, then we would be able to obtain the correct state from the possibly corrupted state of the redundant system. This situation is shown in Fig. 3, [20]: at the end of each time step, the system is within a set of states that could be corrected /decoded to the actual state (in which the underlying system would be, had there been no failures). Even when the decoding mechanism is not fault-free, our approach is still desirable because it guarantees that the probability of a decoding failure will not increase with time in an unacceptable fashion.

3. Error Correction System

There is no easy analytical approach to accurate measures of the detection performances of the transformation technique using convolution codes. Thus, a series of simulations provide estimates of the probability of detection and miss. Convolutional codes offer a natural protection method for lossless compression systems. A high rate Convolutional code over the ring of real corresponding to the arithmetic format can be used to dictate parity numbers that are inserted periodically in the input to the source encoder, as shown in Fig. 4, [16]. The parity values are compressed along with the normal data and the compressed stream is passed through the channel to the decoder. The decoder extracts the data and the inserted parity numbers. These parity values are compared with locally regenerated parity values. These comparisons detect error conditions and, when appropriate, error correction is engaged. A large class of burst-correcting Convolutional codes produces a single parity value for each group of data numbers. This protection method has a small impact on the overall compressing efficiency, especially for high rate codes.

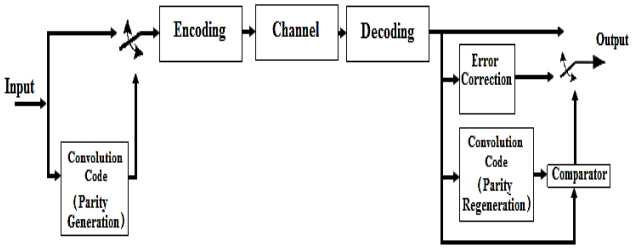


Fig. 4. Joint source-channel coding: embedding Convolutional code parity in compressed and transmitted data.

Powerful efficient Convolutional codes can be used to define arithmetic Convolutional codes that operate on the computational structures of many data processing systems. The encoding and detecting operations employ standard computational resources. When errors at the value level are detected, standard binary decoding algorithms are used in an iterative feedback manner to correct values using syndrome processing methods. The general flow of the feedback decoding method is shown in Fig. 5, [14].

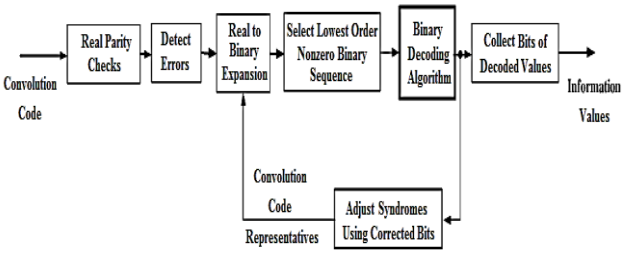


Fig. 5. Iterative syndrome decoding

Random data were encoded into convolutional code words and low level error values with variances' σ_g^2 were added. The high level error values variances' σ_b^2 and probability p were added also. The syndromes of the resulting corrupted code words were analyzed and if any syndrome's component exceeded a threshold ($r = 5\sqrt{\sigma_g^2}$). These error locations were positioned at exact integer index values. The simulations evaluated three convolutional codes, all with 7 parity positions indicating triple error correction capability, of lengths 32, 39, and 50: (32,25), (39,32), (50,43). The low-level error variances σ_g^2 were set at 10^{-6}, a very high choice so as to examine the detrimental effects on the detection and correction operations. The probability of code word error is plotted in Fig. 6 for the three selected convolutional codes. The fixed point input data used m=8 bits while the floating point input was to the precision of MatLab representation. The fixed-point encoded data always had better code word error probabilities.

This section provides typical detection performance results using the designed error detection strategies. The simulation process indicated that round off errors were on the order of 10^{-11} so the necessary thresholds were chosen well above this level. Consider the convolutional code with error injection values modeled by a Gaussian noise source. The experiment results show that when errors

were injected into the system, is on the order of 10^{-10}. The error detection performance is dependent on the error variance and the selected detection threshold. Fig. 7(a) displays three detection performance curves of the 9/7 convolutional code corresponding to single errors with three different variances: $10^{-4}\sigma_0^2$, $10^{-3}\sigma_0^2$, $10^{-2}\sigma_0^2$, where σ_0^2 is the variance of input data.

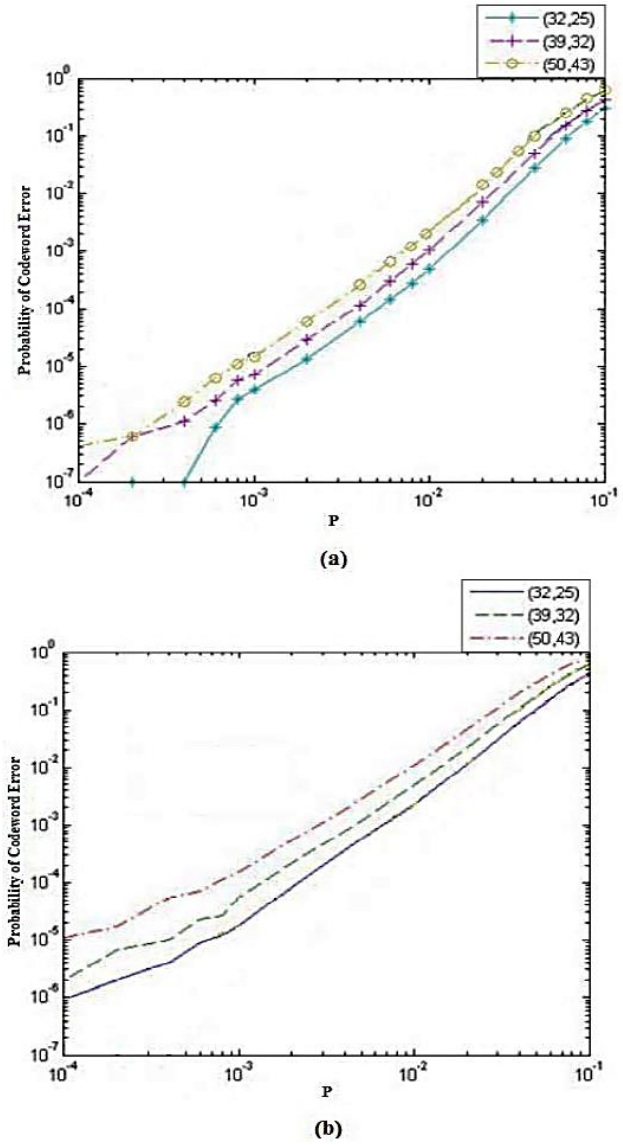


Fig. 6. Probability of Code word Error for Three Codes
(a) The fixed point input data (b) The fixed-point encoded data
($\sigma_g^2 = 10^{-6}$, $\sigma_b^2 = 50$).

The results show that the system has high detection performance when the threshold is in the range from 10^{-10} to 10^{-4} but decreases as the detection threshold increases. Fig. 7(b) shows the range of excellent performance is smaller than that of the forward transform (10^{-10} to 10^{-4}). Table 1 summarizes these simulation results. The detection performance depends on the power of the code supporting the checking capabilities.

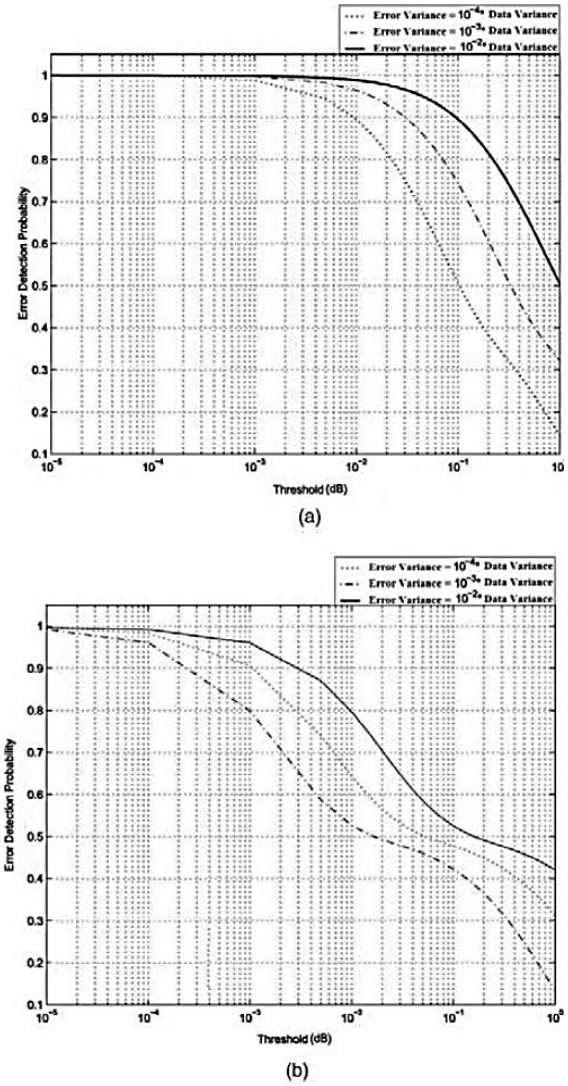


Fig. 7. Error detection performance of the proposed design model for: (a) 9/7 Convolutional code, single error (b) 9/7 Convolutional code, double error. ($10^{-4}\sigma_0^2$, $10^{-3}\sigma_0^2$, $10^{-2}\sigma_0^2$) three different variances:

Table 1. Detection Performance

Single Error			Double Error		
Injection	Detection	Percent	Injection	Detection	Percent
4096	4096	100%	4096	4096	100%
4096	4096	100%	4096	4070	99.3%
4096	4096	100%	2390	1304	54.6%

4. Factorization

4.1 LU Factorization

In LU factorization, an m×n real matrix A is factored into a lower triangular matrix L and a upper triangular matrix U, i.e. PA = LU, where P is a permutation matrix, at each iteration one column block is factored and a permutation matrix P is generated, if necessary, The LU factorization is performed in place, and P is stored as a one-dimensional array of the pivoting indices. Three variants exist for implementing LU factorization on

sequential machines. These three block algorithms of LU factorization can be constructed as follows. Suppose that we have factored A as A = LU. We write the factors in block form as follows:

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} U_{11} & U_{21} & U_{31} \\ 0 & U_{22} & U_{32} \\ 0 & 0 & U_{33} \end{bmatrix} \quad (1)$$

With these relationships, we can develop three variants by manipulating the order in which computations are formed and maintaining the final result of computations in place. These variants are called ijk variants [21] or, more specifically, right-looking, top-looking, and left-looking, respectively. They differ in which regions of data are accessed and computed during each reduction step.

4.2 Cholesky Factorization

Cholesky factorization factors an n×n real, symmetric, positive definite matrix A into a lower triangular matrix L and its transpose L^T, i.e., $A = LL^T$ or U^TU where U is upper triangular). Because of the symmetric, positive definite property of the matrix A, Cholesky factorization is also performed in place on either an upper or lower triangular matrix and involves no pivoting. Three different variants of the Cholesky factorization can be developed as above [22].

4.3 QR Factorization

Given an m×n real matrix A, QR factorization factors A such that

$$A = Q\begin{bmatrix} R \\ 0 \end{bmatrix} \quad (2)$$

Where Q is an m×m orthogonal matrix and R an n×n upper triangular matrix Q is computed by applying a sequence of householder transformations to the current column block of the form, $H_i = 1 - \tau_i v_i v_i^T$ where i =1,…., b. In one block QR algorithm Q can be applied or manipulated through the identity $Q = H_1H_2 \ldots H_b = 1 - VTV$ where V is a lower triangular matrix of "householder" vectors V_i and T is an upper triangular matrix constructed from the triangular factors V_i and τ_i of the householder transformations. When the factorization is complete, V is stored in the lower triangular part of the original matrix A, R is stored in the upper triangular part of A, and the $\tau_i' S$ are stored in the diagonal entries of A. The complete details of this algorithm are described in [23] and [24]. Both left- looking and right- looking variants can be constructed [25].

5. Analysis of Checkpointin

The basic checkpointing operation works on a panel of blocks, where each block consists of X floating- point numbers, and the processors are logically configured in a P × Q mesh (See Fig.8,[26]) .The processors take the checkpoint with a combine operation of XOR or addition. This works in a spanning- tree fashion in three parts. The checkpoint is first taken row wise, then taken column wise, and then sent to PC. The first part therefore takes [log P] steps, and the second part takes [log Q] steps. Each step consists of sending and then performing either XOR or addition on X floating- point numbers. The third part consists of sending the X numbers to PC.

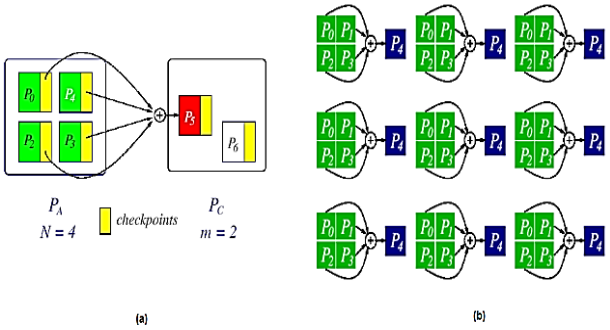


Fig. 8. (a) Single- failure recovery model: after a failure, (b) Checkpointing the matrix of (a).

We define the following terms:

γ: The time for performing a floating- point addition or XOR operation.

α: The startup time for sending a message.

β: The time to transfer one floating- point number.

The first part takes [log P] $(\alpha + X(\beta + \gamma))$, the second part takes [log Q] $(\alpha + X(\beta + \gamma))$, and the third takes $(\alpha + X\beta)$.

5.1 Implementations and Performance Evaluation

For all of the implementations, the following set of tests was performed and timed:

- Failure free algorithm without checkpointing.
- Fault tolerant implementation with single checkpointing.
- Single checkpointing implementation with one random failure.
- Fault tolerant implementation with multiple checkpointing.
- Multiple checkpointing implementations with multiple failures.

Note that the failures were forced to occur at the last iteration before the first checkpoint. The performance results of the implementations are evaluated in terms of the following parameters:

- Total elapsed wall-clock times of the algorithms in seconds (T_A, T).
- Checkpointing and recovery overheads in seconds (T_C, T_R).
- Checkpointing interval in iterations $(K, N_C = n/Kb)$.
- Average checkpointing interval in seconds $((T - T_{init})/N_C)$.
- Average checkpointing overhead in seconds $(\Delta T_C = T_C - T_{init})$.
- Total size of checkpoints in bytes (M).

- Extra memory usage in bytes (M_C).
- Checkpointing rate in bytes per second (R)

The checkpointing performed in these implementations consists of data communication and either XOR or addition of floating-point numbers. We define the checkpointing rate R as the amount of data checkpointed in bytes per second. This metric has been used to evaluate the performance of various checkpointing schemes [29-34]. In our case, the checkpointing rate is determined experimentally based on our analytic models of the fault-tolerant implementations. The total checkpointing overhead of the left-looking variant is too high compared with the right-looking variant without checkpointing (See Figures 9, 10 and 11). This checkpointing rate is used to compare the performance of the different fault tolerance techniques, Figures 12 and 13 plots the checkpointing rate for each implementation.

Parity-Based Technique: For the parity-based matrix operations, the total percentage overhead of checkpointing decreases as the problem size n increases. The total overhead of recovery is dominated by the time for taking the bitwise exclusive- or of each processor's entire data. The time it takes to recover does not depend upon the location of the failure. The multiple checkpointing implementations show performance improvement. LU factorizations benefit relatively more from the multiple checkpointing because of pivoting. Figure 10 shows the checkpointing rate experimentally determined for each implementation .This presents the overall performance of the parity-based technique for matrix operations.

	n	T_A (sec)	N_C+1	T (sec)	$\frac{T-T_{init}}{N_C}$ (sec)	T_C (sec)	T_{init} (sec)	$\frac{\Delta T_C}{N_C}$ (sec)	T_R (sec)
single checkpointing	1500	220	32	232	7.3	19	8	0.5	8
	3000	1530	66	1430	21.6	62	29	0.9	27
	4500	2610	102	2010	19.2	151	71	1.4	69
	6000	4900	144	5100	35	280	103	1.9	92
	7500	7100	180	6900	38	489	183	1.0	151
	n	T_A (sec)	N_C+1	T (sec)	$\frac{T-T_{init}}{N_C}$ (sec)	T_C (sec)	T_{init} (sec)	$\frac{\Delta T_C}{N_C}$ (sec)	T_R (sec)
multiple checkpointing	1500	220	32	220	6.9	17	5	0.5	6
	3000	1530	66	1200	18.1	46	21	0.7	20
	4500	2610	102	1902	18.3	101	55	1.0	46
	6000	4900	144	4800	33	236	79	1.6	70
	7500	7100	180	6480	35.5	409	132	0.8	111

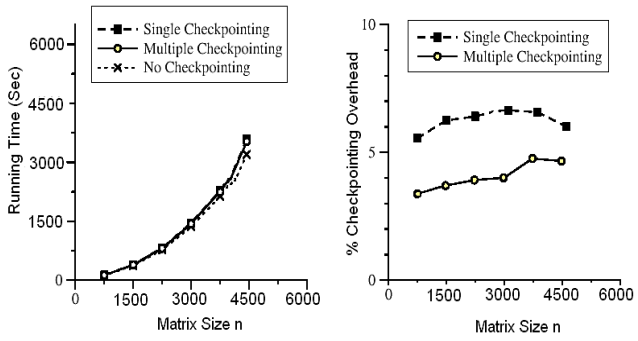


Fig. 9. Left-looking LU, timing results

	n	T_A (sec)	N_C+1	T (sec)	$\frac{T-T_{init}}{N_C}$ (sec)	T_C (sec)	T_{init} (sec)	$\frac{\Delta T_C}{N_C}$ (sec)	T_R (sec)
single checkpointing	1500	220	32	228	7.2	8	4	0.1	4
	3000	1530	66	1558	23.7	28	15	0.2	13
	4500	2610	102	2675	26.2	65	32	0.3	28
	6000	4900	144	5022	34.7	122	53	0.4	50
	7500	7100	180	7330	40.5	230	74	0.9	66
	n	T_A (sec)	N_C+1	T (sec)	$\frac{T-T_{init}}{N_C}$ (sec)	T_C (sec)	T_{init} (sec)	$\frac{\Delta T_C}{N_C}$ (sec)	T_R (sec)
multiple checkpointing	1500	220	32	226	7.2	6	2	0.1	1
	3000	1530	66	1548	23.7	18	6	0.2	5
	4500	2610	102	2659	26.1	49	20	0.3	17
	6000	4900	144	5005	34.8	105	31	0.5	27
	7500	7100	180	7291	40.4	191	55	0.8	52

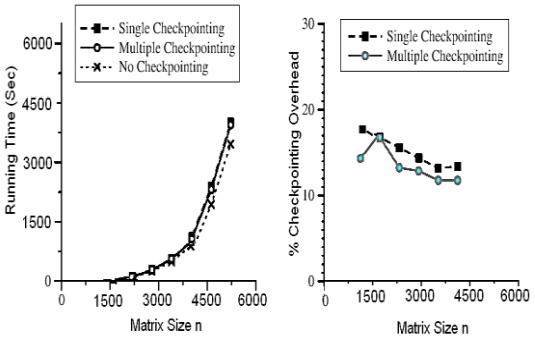


Fig. 10. Left-looking Cholesky, timing results

	n	T_A (sec)	N_C+1	T (sec)	$\frac{T-T_{init}}{N_C}$ (sec)	T_C (sec)	T_{init} (sec)	$\frac{\Delta T_C}{N_C}$ (sec)	T_R (sec)
single checkpointing	1500	220	32	244	7.6	24	8	0.5	6
	3000	1530	66	1603	24.2	73	30	0.6	27
	4500	2610	102	2750	26.7	140	54	0.8	50
	6000	4900	144	5094	38.8	194	78	0.8	71
	n	T_A (sec)	N_C+1	T (sec)	$\frac{T-T_{init}}{N_C}$ (sec)	T_C (sec)	T_{init} (sec)	$\frac{\Delta T_C}{N_C}$ (sec)	T_R (sec)
multiple checkpointing	1500	220	32	240	7.6	20	5	0.5	4
	3000	1530	66	1582	24	52	22	0.5	19
	4500	2610	102	2725	26.6	115	40	0.7	39
	6000	4900	144	5048	34.9	148	55	0.6	51

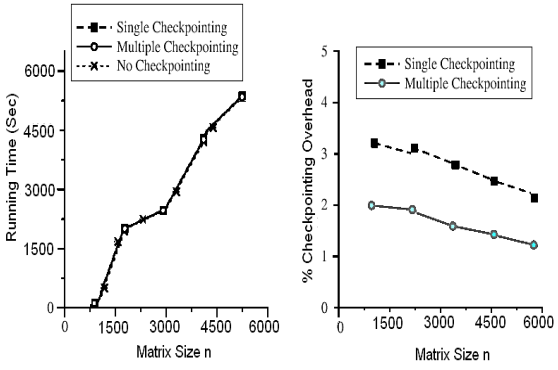


Fig. 11. Left-looking QR, timing results

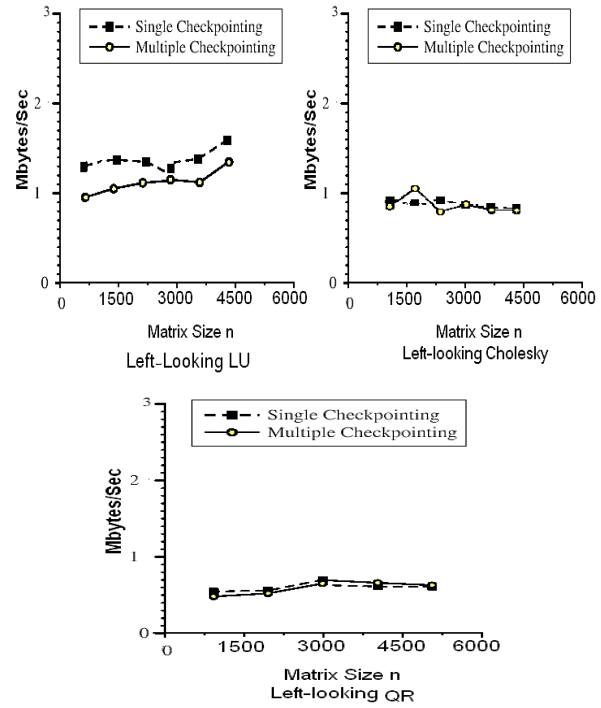


Fig. 12. Experimental checkpointing rate

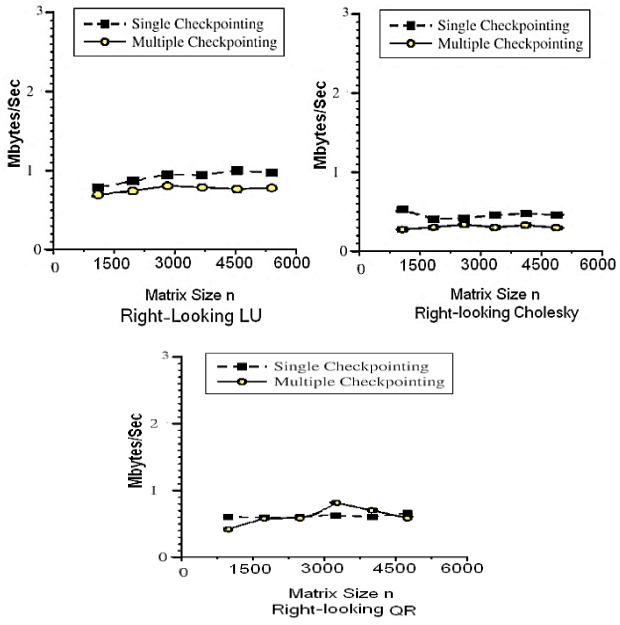


Fig. 13. Experimental checkpointing rate

Since the measured peak bandwidth of the network is 64 Mbits per second, we expect that the checkpointing rate should be somewhat lower than 8 Mbytes per second considering synchronization, copying, performing XOR, and message latency and network contention. As shown in Figures 12 and 13 the checkpointing rate determined experimentally is between 2 and 4 Mbytes per second for all the matrix operations. The right-looking variant performs the best among the failure-free variants of each factorization because it benefits from less communication and more parallelism than the others. However, for the LU and Cholesky factorizations, the left-looking variants with checkpointing perform better than the right-looking variant with checkpointing. For the QR factorization, no top-looking variant exists, and the left-looking variant performs much slower than the right-looking variant. The total checkpointing overhead of the left-looking variant is too high compared with the right-looking variant without checkpointing (Figure 9-11).

6. Conclusions

The transformation technique transforms a system that does not tolerate a specific type of faults, called the fault-intolerant system, to a system that provides a specific level of fault tolerance, namely recovery and/or safety. The advantage of transformation technique is that errors which are caused by permanent or transient failures in the system can be detected and corrected by using a very low overhead and at the original throughput.

In This paper presents a model for executing certain scientific computations on a changing distributed computing platform .The model allows a distributed computation to run on a platform where individual processors may leave due to failures, unavailability, or heavy load, and where processors may enter during the computation. The model provides an interesting way to allow reliability in computations per formed on networks of computers.

Systematic codes for data protection using the parity comparison method can be determined from general Convolutional codes by manipulating the matrix associated with such codes.

These operations involve straightforward matrix operations similar to those supporting the normal matrix forms.

The advantage of this method is that it is able to achieve very low overhead according to the specific characteristic of an application. The limitation of this method is that it is non-transparent and has to be designed according to the specific characteristic of an application.

References

[1] H. Hamidi, A. Vafaei, A.H. Monadjemi, “Algorithm based fault tolerant and check pointing for high performance computing systems,”J.Applied Sci., 9:3947-3956, 2009.

[2] H. Hamidi, A.Vafaie, S. A. Monadjemi “Analysis and design of an ABFT and parity-checking technique in high performance computing systems” Journal of Circuits,

Systems, and Computers (JCSC), JCSC Vol.21, No. 3, May 2012.
[3] H. Hamidi, A.Vafaie, S. A. Monadjemi. Analysis and evaluation of a new algorithm based fault tolerance for computing systems. International Journal of Grid and High Performance Computing, 4(1), 37–51. 2012.
[4] H. Hamidi, A.Vafaie, S. A. Monadjemi "A Framework for ABFT Techniques in the Design of Fault-Tolerant Computing Systems", EURASIP Journal on Advances in Signal Processing, Springer, Vol.2011:90, October 2011.
[5] H. Hamidi, A.Vafaie, S. A. Monadjemi, "A Framework for Fault Tolerance Techniques in the Analysis and Evaluation of Computing Systems" International Journal of Innovative Computing, Information and Control (IJICIC), Vol.8, No.7, July 2012.
[6] K.H. Huang, J.A. Abraham, "Algorithm-Based Fault Tolerance for Matrix Operations." IEEE Trans. Computers, vol. 33, pp. 518-528, 1984.
[7] Z.Chen, "Extending Algorithm-based Fault Tolerance to Tolerate Fail-stop Failures in High Performance Distributed Environments," Proceedings of the 22nd IEEE International Parallel & Distributed Processing Symposium, DPDNS'08 Workshop, Miami, FL, USA, April 14-18, 2008.
[8] C.N.Zhang, X.W. Liu, "An algorithm based mesh check-sum fault tolerant scheme for stream ciphers," International Journal of Communication Networks and Distributed Systems, Vol.3, No.3, pp.217-233, June 2009.
[9] S. Sundaram, C. N. Hadjicostis, "Fault-Tolerant Convolutional via Chinese Remainder Codes Constructed from Non-Coprime Moduli,"IEEE Transactions on Signal Processing, Vol. 56, No. 9, pp. 4244-4254, September 2008.
[10] T.Roche, M. Cunche, J.L Roch, "Algorithm-Based Fault Tolerance Applied to P2P Computing Networks," ap2ps, pp.144-149, 2009 First International Conference on Advances in P2P Systems, 2009.
[11] D. Costello, S. Lin, Error Control Coding Fundamentals and Applications, 2nd edition, Pearson Education Inc., NJ, U.S.A., 2004.
[12] Robert H. Morelos-Zaragoza, The Art of Error Correcting Coding, 2nd Edition, John Wiley & Sons, ISBN: 0470015586, 2006.
[13] G.Bosilca, R.Delmas, J. Dongarra, J. Langou, "Algorithm-based fault tolerance applied to high performance computing,"Journal of Parallel and Distributed Computing, Elsevier, Vol.69, No.4, pp.410-416, April 2009.
[14] G. Robert Redinbo, "Wavelet Codes for Algorithm-Based Fault Tolerance Applications," IEEE Transactions on Dependable and Secure Computing, vol. 7, no. 3, pp. 315-328, July-Sept. 2010.
[15] G.Robert Redinbo, "Generalized Algorithm-Based Fault Tolerance: Error Correction via Kalman Estimation", IEEE Transactions ON Computers, vol. 47, no. 6, Jun 1998.
[16] G. Robert Redinbo, "Failure-Detecting Arithmetic Convolutional Codes and an Iterative Correcting Strategy," IEEE Transactions on Computers, vol. 52, no. 11, pp. 1434-1442, Nov. 2003.
[17] S.Veeravalli, "Fault tolerance for arithmetic and logic UNIT," IEEE SOUTHEASTCON,'09, 2009, pp. 329 – 334.
[18] J. Choi, J.J. Dongarra, S. Ostrouchov, A. P. Petitet, D. W.Walker, and R. C.Whaley. "The design and implementation of the ScaLAPACK LU, QR, and Cholesky factorization routines," Scientific Programming, 1996.
[19] J. Choi, J. J. Dongarra, and D.W.Walker, "PB-BLAS: A set of parallel block basic linear algebra subprograms," Concurrency Practice and Experience, 1996.
[20] J. J. Dongarra, J, Du Croz, S. Hammarling, and I. Duff, "A set of level 3 basic linear algebra subprograms," ACM Transactions on Mathematical Software, 18(1):pp.1-17, 1990.
[21] S. Ekici, S. Yildirim, M.Poyraz, "A transmission line fault locator based on Elman recurrent networks," Applied Soft Computing, Elsevier, Volume 9, Issue 1, pp.341-347, 2009.
[22] X.She, S.Trimberger, "Scheme to minimize short effects of single-event upsets in triple-modular redundancy (TMR), IET Computers & Digital Techniques, 4(1):50, 2010.
[23] J. J. Dongarra and R. C. Whaley, A user's guide to the BLACS v1.0.LAPACK Working Note 94, Technical Report CS-95-281, University of Tennessee, 1995.
[24] A. A.Kumar, A. Makur, "Improved coding-theoretic and subspace-based decoding algorithms for a wider class of DCT and DST codes," IEEE Transactions on Signal Processing Vol. 58, Issue 2, pp. 695-708, 2010.
[25] E. N. Elnozahy, D. B. Johnson, and W. Zwaenepoel, "The performance of consistent check pointing," In 11th Symposium on Reliable Distributed Systems, 1992, pp. 39-47.
[26] R.H. Morelos-Zaragoza, "The Art of Error Correcting Coding, Second Edition," John Wiley & Sons, Ltd. ISBN: 0-470-01558-6, 2006.
[27] J. S. Plank and K. Li. Ickp, "A consistent checkpointer for multicomputer," IEEE Parallel & Distributed Technology, 2(2):62-67, 1994.
[28] J.Y. Jou and J. A.Abraham, "Fault tolerant matrix arithmetic and signal processing on highly concurrent computing structures," Proc. IEEE, vol, no.5, pp.732-741, 1986.
[29] C. N. Hadjicostis, "Coding Approaches to Fault Tolerance in Dynamic Systems," Ph.D thesis, EECS department, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1999.
[30] C. N. Hadjicostis, "Coding Approaches to Fault Tolerance in Combinational and Dynamic Systems," Boston, Massachusetts: Kluwer Academic Publishers, 2002.
[31] Y. Kim, "Fault Tolerant Matrix Operations for Parallel and Distributed Systems," Ph.D dissertation, Univ. of Tennessee, June 1996.
[32] J. S. Plank, Y. Kim, and J. Dongarra, "Algorithm-based diskless checkpointing for fault tolerant matrix operations," In 25th International Symposium on Fault-Tolerant Computing, Pasadena, CA, June 1995.
[33] H. Hamidi, An Approach to Fault Detection and Correction in Design of Fault Tolerant Computing Systems Using of Turbo codes, International Journal of Industrial Mathematics, 2016, (Accepted for Publication).

Hodjatollah Hamidi, born 1978, in shazand Arak, Iran, He got his Ph.D in computer engineering. His main research interest areas are Information Technology, Fault-Tolerant systems (fault-tolerant computing, error control in digital designs) and applications and reliable and secure distributed systems and e- commerce. Since 2013 he has been a faculty member at the IT group of K. N. Toosi University of Technology, Tehran Iran. Department of Industrial Engineering , K. N. Toosi University of Technology.

Speech Intelligibility Improvement in Noisy Environments for Near-End Listening Enhancement

Peyman Goli*
Department of Electronic and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran
p.goli@stu.nit.ac.ir
Mohammad Reza Karami-Mollaei
Department of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran
mkarami@nit.ac.ir

Received: 01/Jul/2015 Revised: 28/Feb/2016 Accepted: 02/Mar/2016

Abstract

A new speech intelligibility improvement method for near-end listening enhancement in noisy environments is proposed. This method improves speech intelligibility by optimizing energy correlation of one-third octave bands of clean speech and enhanced noisy speech without power increasing. The energy correlation is determined as a cost function based on frequency band gains of the clean speech. Interior-point algorithm which is an iterative procedure for the nonlinear optimization is used to determine the optimal points of the cost function because of nonlinearity and complexity of the energy correlation function. Two objective intelligibility measures, speech intelligibility index and short-time objective intelligibility measure, are employed to evaluate the noisy enhanced speech intelligibility. Furthermore, the speech intelligibility scores are compared with unprocessed speech and a baseline method under various noisy conditions. The results show large intelligibility improvements with the proposed method over the unprocessed noisy speech.

Keywords: Near-end Speech Enhancement; Intelligibility Improvement; Energy Correlation; Optimization Algorithms.

1. Introduction

Mobile phones often deliver speech output to listener in noisy environments. The background noise such as traffic or babble noise reduces speech intelligibility for the near-end listener. Several preprocessing algorithms have been proposed to improve speech intelligibility for the near-end listener in noisy environment. In speech improvement methods which focus on speech intelligibility enhancement for near-end listener in background noise, the far-end speech is considered as a clean speech with good intelligibility. As the far-end speech is played for near-end listener in noisy environment, its intelligibility is degraded by background noise; thus, these methods manipulate the clean speech (i.e., the far-end speech) before it is corrupted by the background noise to improve the intelligibility of noisy speech. Therefore, the clean speech and noise are available signals in the intelligibility enhancement methods and the aim is improvement of the audibility of degraded speech, as illustrated in Figure 1. Near-end speech intelligibility improvement methods have been classified to noise-independent and noise-dependent methods.

Noise-independent modification algorithms include detecting and boosting the features of speech that have an important role in speech perception. Charturong used hidden Markov model for detecting the consonant and transient regions [1], and Raset and Motlotle applied wavelet transform for extracting these regions in clean speech [2]. Demol et al. also used non-uniform time scaling to slow down the speech and redistributed available time between the vowels and consonants to emphasis on these regions [3]. In addition, Ekramul et al. presented a speech intelligibility improvement process in which speech is modified based on an inverse Wiener filter on the vowel and consonant regions [4].

Since the speech intelligibility in noisy environments usually depends on the noise conditions, noise-dependent algorithms may be applied to speech intelligibility enhancement in application scenario where the noise-statistics are available. Noise-dependent algorithms have been carried out using estimates of the noise signal. These methods are usually based on the signal to noise ratio (SNR) modification or optimization of an objective intelligibility measure. For example, Sauert and Enzner modified the local SNR of time-frequency cells based on the global SNR [5], and Tang and Cooke presented several strategies including the time and frequency segmentation and frequency selected boost to reach the global SNR [6]. Premananda and Uma also improved the near-end speech intelligibility focusing on the selective audible speech samples by considering the threshold of hearing and auditory properties of the human ear [7].

* Corresponding Author

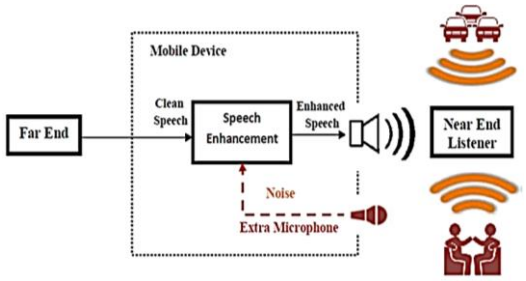


Fig. 1. Intelligibility enhancement of speech delivered in noisy environments. (Noise-independent/ Noise-dependent enhancement approach without/with extra microphone.)

Speech intelligibility enhancement based on maximization of speech intelligibility index (SII) measure [8], [9], and minimization of a perceptual distortion measure [10] are noise-dependent approaches which are considered as optimization algorithms. Tang and Cooke used a genetic algorithm-based optimization procedure, with glimpse proportion as the objective intelligibility metric to enhance the speech intelligibility in background noise [11], and Valentini-Botinhao et al. also increased the glimpse proportion measure by modifying mel-cepstral coefficients to improve the intelligibility of the synthetic speech in noise [12].

In the proposed algorithm, speech intelligibility is improved by optimizing the energy correlation between clean and noisy enhanced speeches in one-third octave frequency bands subjected to a power constraint. The fundamental idea behind the proposed method is that maximization of the cross correlation between speech degraded by noise and clean speech with good intelligibility would yield an improved speech intelligibility for the near-end listener. Hence, a cost function based on the energy correlation between the clean and noisy speeches is introduced to determine the cross correlation. Since the obtained cost function is complicated and nonlinear, the routine optimization methods (i.e., the derivative methods and Lagrange multiplier) cannot lead to an analytical solution. Therefore, an iterative algorithm is applied to optimize the cost function.

The paper is organized as follows: details of the proposed algorithm are described in three phases, namely, (1) preprocessing, (2) calculation and optimization of energy correlation function based on an iterative algorithm, and (3) estimation of statistical quantities. Finally, the objective intelligibility prediction results comparing the proposed algorithm with unprocessed speech and a baseline method are presented.

2. Proposed Speech Intelligibility Improvement Algorithm

To improve the speech intelligibility in background noise, the energy correlation function in one-third octave frequency bands within each time frame of the clean speech and speech degraded by noise, is determined and the correlation function is then optimized with a power speech constraint.

2.1 Preprocessing

Clean speech $x(n)$ and background noise $z(n)$ are available signals, and let $d(n)$ and $y(n) = d(n) + z(n)$ denote the enhanced speech and the speech degraded by noise, respectively. $x(n)$ and $z(n)$ are resampled by sample-rate of 10 kHz to capture a relevant frequency range for speech intelligibility [13]. Both signals are segmented into 50% overlapping, Hann-windowed frames with a length of 256 samples. The discrete Fourier transform (DFT) of each time frame is determined and a one-third octave band analysis is then performed by grouping DFT-bins. In total 15 one-third octave bands are used, where the lowest center frequency is set equal to 160 Hz and the highest octave band has a center-frequency equal to 4.06 kHz. Let $x(k, m)$ and $z(k, m)$ denote the DFT of the m^{th} frame of clean speech and noise in frequency index k, respectively. The energy of the j^{th} band in the m^{th} frame (i.e., $BF_{j,m}$) of clean speech, $X_{j,m}$, and noisy enhanced speech, $Y_{j,m}$, are calculated as follows,

$$X_{j,m} = \sum_{k=k_l(j)}^{k_h(j)} |x(k,m)|^2,$$

$$Y_{j,m} = \sum_{k=k_l(j)}^{k_h(j)} \left|G_{j,m}x(k,m) + z(k,m)\right|^2, \quad (1)$$

where $k_l(j)$ and $k_h(j)$ indicate the j^{th} one-third octave band edges and $|.|$ is the magnitude of the DFT. A real gain $G_{j,m} \in \mathcal{R}^+$, is applied to the $BF_{j,m}$ of the clean speech, to enhance the clean signal.

2.2 Calculation and Optimization of Correlation Function

The correlation coefficient is a statistical measure of the linear dependence between two random variables. Due to the energies $X_{j,m}$ and $Y_{j,m}$ are the random variables, the energy correlation function $\rho_{j,m}$ of the $BF_{j,m}$ of clean and noisy enhanced speeches is obtained as follows,

$$\rho_{j,m} = \frac{E(X_{j,m}Y_{j,m}) - E(X_{j,m})E(Y_{j,m})}{\sqrt{E(X_{j,m}^2) - E^2(X_{j,m})}\sqrt{E(Y_{j,m}^2) - E^2(Y_{j,m})}}, \quad (2)$$

where $E(.)$ indicates the expectation of the random variable. The numerator in Equation (2) determines the covariance of $X_{j,m}$ and $Y_{j,m}$, and the denominator is the product of theirs standard deviation. According to Equation (1) and the properties of the complex numbers, $Y_{j,m}$ would be rewritten as follows,

$$Y_{j,m} = G_{j,m}^2 X_{j,m} + G_{j,m} \sum_{k=k_l(j)}^{k_h(j)} x(k,m)z^*(k,m)$$

$$+ G_{j,m} \sum_{k=k_l(j)}^{k_h(j)} x^*(k,m)z(k,m) + Z_{j,m}, \quad (3)$$

where $.^*$ indicates the complex conjugate and $Z_{j,m}$ is the energy of the $BF_{j,m}$ of the noise. The energy correlation $\rho_{j,m}$ is obtained as a function of clean speech, noise and gain $G_{j,m}$ by substituting $Y_{j,m}$ in Equation (2).

Two obvious assumptions are considered to simplify the energy correlation: first, independency of clean speech and background noise, $E[x(n)z(n)] = E[x(n)]E[z(n)]$. Second, the zero expectation of the stochastic processes, clean speech and noise, $E[x(n)] = 0$ and $E[z(n)] = 0$.

According to Equations (2), (3) and the assumptions, the energy correlation function $\rho_{j,m}$ is simplified as follows,

$$\rho_{j,m} = \frac{\sigma_{X_{j,m}} G_{j,m}^2}{\sqrt{\sigma_{X_{j,m}}^2 G_{j,m}^4 + 4\Lambda_{j,m} G_{j,m}^2 + \sigma_{Z_{j,m}}^2}}, \quad (4)$$

Where

$$\Lambda_{j,m} = E\left\{\left[\mathrm{Re}\left(\sum_{k=k_l(j)}^{k_h(j)} x(k,m)z^*(k,m)\right)\right]^2\right\}, \quad (5)$$

$\mathrm{Re}(.)$ indicates the real part of the complex value. The statistical quantities $\sigma_{X_{j,m}}^2$ and $\sigma_{Z_{j,m}}^2$, which refer to the variance of the energies $X_{j,m}$ and $Z_{j,m}$, respectively, and $\Lambda_{j,m}$ are estimated from clean speech and noise. The energy correlation function $\rho_{j,m}$ is plotted in Figure 2 based on $G_{j,m}^2$, in the 10th frequency band for an SNR of –5 dB. The average of the energy correlation function of each frame $\sum_j \rho_{j,m}$ is maximized subjected to a power constraint to improve the speech intelligibility. The correlation function $\rho_{j,m}$ is concave in $G_{j,m}^2$, as illustrated in Figure 2. Hence, the sum of these concave functions, $\sum_j \rho_{j,m}$, is also concave. The constrained optimization problem can be formulated as follows,

$$\max_{G_{jm}} : \sum_{j=1}^{15} \frac{\sigma_{X_{j,m}} G_{j,m}^2}{\sqrt{\sigma_{X_{j,m}}^2 G_{j,m}^4 + 4\Lambda_{j,m} G_{j,m}^2 + \sigma_{Z_{j,m}}^2}};$$

$$\text{subject to} : \begin{cases} \sum_{j=1}^{15} G_{j,m}^2 X_{j,m} = P_m \\ G_{j,m}^2 \geq 0; \quad j = 1, \dots, 15 \end{cases}; \quad (6)$$

where $P_m = \sum_{j=1}^{15} X_{j,m}$ indicates the energy of the m^{th} frame of clean speech. The equality condition relates to the power constraint in the m^{th} frame, and the inequality condition satisfies the positive real gains $G_{j,m} \in \mathcal{R}^+$. The convexity is obtained by negation and the following Lagrangian cost-function characterizes the problem,

$$L = -\sum_{j=1}^{15} \frac{\sigma_{X_{j,m}} G_{j,m}^2}{\sqrt{\sigma_{X_{j,m}}^2 G_{j,m}^4 + 4\Lambda_{j,m} G_{j,m}^2 + \sigma_{Z_{j,m}}^2}} +$$

$$\lambda\left(\sum_{j=1}^{15} G_{j,m}^2 X_{j,m} - P_m\right) - \sum_{j=1}^{15} \omega_j G_{j,m}^2, \quad (7)$$

where λ and ω_j are Lagrangian multipliers related to the power constraint and inequality constraints in Equation (6), respectively. Since the objective function and constraints are differentiable, any point that satisfies the constraints in Equation (6) and the following conditions is guaranteed to optimize the problem [14].

1) $\mu_j \geq 0$; $\quad j = 1, \dots, 15,$

2) $\mu_j G_{j,m}^2 = 0; \quad j = 1, \dots, 15,$

3) $$\frac{\partial L}{\partial G_{j,m}^2} = -\frac{2\sigma_{X_{j,m}} \Lambda_{j,m} G_{j,m}^2 - \sigma_{X_{j,m}} \sigma_{Z_{j,m}}^2}{G_{j,m}^4 \sigma_{X_{j,m}}^2 + 4G_{j,m}^2 \Lambda_{j,m} + \sigma_{Z_{j,m}}^2}$$

$$+\lambda X_{j,m} - \omega_j = 0; \quad j = 1, \dots, 15, \quad (8)$$

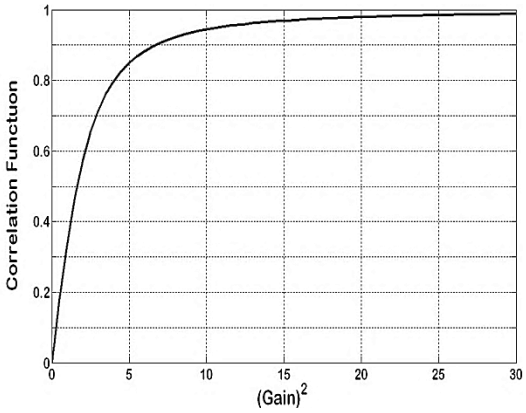


Fig. 2. Energy correlation function in the 10th band for an SNR of –5 dB based on $G_{j,m}^2$.

Because of complexity of the derivative of the Lagrangian cost-function, the optimization problem could not lead to an analytical formula. Thus, an iterative algorithm is applied to solve the optimization problem, which uses descent method to search the optimal point [14]. The Interior-point algorithm is an iterative method for non-linear optimization problem that consists of two steps. Step 1, inner loop, the Newton's method is applied to optimize the equality constrained problem. Step 2, outer loop, the Barrier method formulates the inequality constrained problem as an equality constrained problem to which Newton's method can be applied.

To formulate our constrained optimization problem as equality constrained problem, Equation (6) can be rewritten based on the Barrier logarithm as follows,

$$\min_{\breve{G}_{j,m}} : \tau f_0(\breve{G}_m) + \varphi(\breve{G}_m)$$

$$= -\tau \sum_{j=1}^{15} \frac{\sigma_{X_{j,m}} \breve{G}_{j,m}}{\sqrt{\sigma_{X_{j,m}}^2 \breve{G}_{j,m}^2 + 4\Lambda_{j,m} \breve{G}_{j,m} + \sigma_{Z_{j,m}}^2}} + \sum_{i=j}^{15} -\log(\breve{G}_{j,m}),$$

$$\text{subject to} : A\breve{G}_m = P_m, \quad (9)$$

where a simple variable change is used as $\breve{G}_{j,m} = G_{j,m}^2$. Vector $\breve{G}_m = [\breve{G}_{1,m}, \breve{G}_{2,m}, \dots, \breve{G}_{15,m}]^T$ consists of the gains of frequency bands in the m^{th} frame and vector $A = [X_{1,m}, X_{2,m}, \dots, X_{15,m}]$ refers to the energies of the clean speech bands. The equality $A\breve{G}_m = P_m$ shows the power constraint. $\varphi(\breve{G}_m) = -\sum_{j=1}^{15} \log(\breve{G}_{j,m})$ and $f_0(\breve{G}_m)$ are the Barrier logarithm and the negative energy correlation function based on $\breve{G}_m$, respectively. These functions are convex in $\breve{G}_m$. Parameter $\tau > 0$ sets the accuracy of the Barrier logarithm approximation. The optimization problem in the Equation (9) searches for the optimal vector which minimizes the cost function $f_0(\breve{G}_m)$ subjected to $A\breve{G}_m = P_m$ which satisfies power constraint and $\breve{G}_{j,m} > 0$ which is formulated in the Barrier logarithm. The optimization of the

energy correlation function using interior-point algorithm (Barrier and Newton's method) is summarized in Table I.

In interior-point algorithm shown in Table I, the inner loop minimizes the objective function $\tau f_0(\breve{G}_m) + \varphi(\breve{G}_m)$ (i.e., parameter τ is determined in outer loop) subjected to the power constraint using Newton's method. Hence, the Newton step $\Delta\breve{G}_m$ and decrement λ are calculated in the inner loop. In each iteration of the inner loop, the Newton step is added to $\breve{G}_m$ that was obtained in previous iteration, and the loop is run again until reaching inner stopping criterion, $\lambda^2/2 \le \acute{\epsilon}$ (i.e., for tolerance $\acute{\epsilon} > 0$). Backtracking line search is applied in the inner loop to determine the step size t used in gain updating at each Newton iteration. The optimal gain obtained by the inner loop is named central point, $\breve{G}_m^{*}(\tau)$, and delivered to the outer loop. In the outer loop, the central point $\breve{G}_m^{*}(\tau)$ is updated and then τ is increased by a factor $\mu > 1$. In other words, the central point $\breve{G}_m^{*}(\tau)$ is computed for a sequence of increasing values of τ until reaching the outer stopping criterion, $\tau \ge 15/\epsilon$ (i.e., for tolerance > 0), which guarantees the ϵ - suboptimal solution of the optimization problem.

Initialization of parameters of the interior-point algorithm largely affects the iterations of the inner and outer loops, and optimization accuracy. On one hand, excessive iterations of the loops lead to a large algorithmic delay; on the other hand, choosing large step sizes for reducing the iterations may results suboptimal points and reduces the accuracy. Therefore, the values of the parameters of the inner and outer loops are obtained from the experimental results, which provide the best performance of the interior-point algorithm in the optimization problem. These values yield the minimum iteration numbers and the maximum accuracy. The initialization of the parameters of the presented interior-point algorithm is shown in Table II.

Table 1. Optimization of energy correlation function using Barrier method (Newton's method).

Outer loop (Barrier Method)

Given feasible point $\breve{G}_m$, $\tau := \tau^{(0)} > 0$, $\mu > 1$, and tolerance $\epsilon > 0$.

Repeat:

1- Centering step

Inner loop (Newton's Method)

Starting point $\breve{G}_m$, tolerance $\acute{\epsilon} > 0$.

Repeat:

I. Compute the Newton step $\Delta\breve{G}_m$ and decrement λ:

$$\begin{bmatrix} \tau\nabla^2 f_0(\breve{G}_m) + \nabla^2\varphi(\breve{G}_m) & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \Delta\breve{G}_m \\ \lambda \end{bmatrix} = \begin{bmatrix} -\tau\nabla f_0(\breve{G}_m) - \nabla\varphi(\breve{G}_m) \\ 0 \end{bmatrix}$$

$$\lambda^2 = [\tau\nabla f_0(\breve{G}_m) + \nabla\varphi(\breve{G}_m)]^T[\tau\nabla^2 f_0(\breve{G}_m) + \nabla^2\varphi(\breve{G}_m)]^{-1}[\tau\nabla f_0(\breve{G}_m) + \nabla\varphi(\breve{G}_m)]$$

II. Stopping criterion. Quit and $\breve{G}_m^{*}(\tau) = \breve{G}_m$ if $\lambda^2/2 \le \acute{\epsilon}$.

III. Line search by backtracking t:

Line search

i. Given descent direction $\Delta\breve{G}_m$, and $0 < \gamma < 0.5$, $0 < \theta < 1$

i. $t := 1$

i. While $\tau f_0(\breve{G}_m + t\Delta\breve{G}_m) + \varphi(\breve{G}_m + t\Delta\breve{G}_m) > \tau f_0(\breve{G}_m) + \varphi(\breve{G}_m) + \gamma t[\tau\nabla f_0(\breve{G}_m) + \nabla\varphi(\breve{G}_m)]^T\Delta\breve{G}_m$, $t := \theta t$

IV. Update: $\breve{G}_m := \breve{G}_m + t\Delta\breve{G}_m$

2- Updating $\breve{G}_m := \breve{G}_m^{+}(\tau)$. **(Starting point for next inner iteration)**

3- Stopping criterion. Quit if $\ge 15/\epsilon$ **.**

4- Increasing τ. $\tau := \mu\tau$. **END.**

Table 2. Initialization of the parameters of the inner and outer loops in the presented interior-point algorithm

Parameter Value	Loop
Feasible point $\breve{G}_m = [1,1,..,1]^T$	Outer
$\tau^{(0)} = 0.1$	Outer
$\mu = 2$	Outer
$\epsilon = 0.002$	Outer
$\acute{\epsilon} = 0.02$	Inner
$\gamma = 0.02$	Inner
$\theta = 0.25$	Inner

2.3 Estimation of Statistical Quantities

Given that the random variables $X_{j,m}$ and $Z_{j,m}$ are short-time stationary processes over the time frames, the statistical quantities could be estimated via time frame averaging [15],

$$\sigma^2_{X_{j,m}} = \frac{1}{N-1}\sum_{r=m-N+1}^{m}\left[X_{j,r} - E(X_{j,r})\right]^2,$$

$$E(X_{j,m}) = \frac{1}{N}\sum_{r=m-N+1}^{m} X_{j,r}, \quad (10)$$

and

$$\Lambda_{j,m} = \frac{1}{N-1}\sum_{r=m-N+1}^{m}\left[\mathrm{Re}\left(\sum_{k=k_l(j)}^{k_h(j)} x(k,r)z^*(k,r)\right)\right]^2, \quad (11)$$

where N denotes the number of successive frames in which the quantities are estimated. The best results are obtained in $N = 30$. Such as Equation (10), similar estimation hold for $\sigma^2_{Z_{j,m}}$. In practice, a simple noise-tracker algorithm [16] could be applied to estimate the power of noise signal.

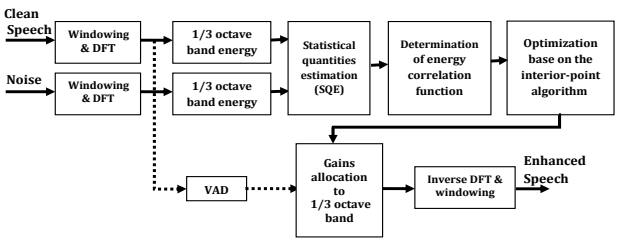


Fig. 3. Block diagram of the proposed algorithm

A simple smoother is then applied to the statistical quantities to prevent high changes which may negatively affect the estimation,

$$\hat{\sigma}^2_{X_{j,m}} = \alpha\hat{\sigma}^2_{X_{j,m-1}} + (\alpha - 1)\sigma^2_{X_{j,m}}, \quad (12)$$

where $\alpha = 0.9\,6$ leads to best results and similar smoother is applied to $\sigma^2_{Z_{j,m}}$ and $\Lambda_{j,m}$.

3. Proposed Algorithm Implementation

A simple voice activity detector (VAD) is used in the proposed algorithm, as illustrated in Figure 3. The VAD block selects the speech frames whose power is greater than $P_{max} - K$ for the process. P_{max} is the maximum power of the received frames of clean speech in dB and the constant value of K is selected 25 dB in practice. Clean speech, enhanced speech by proposed method and

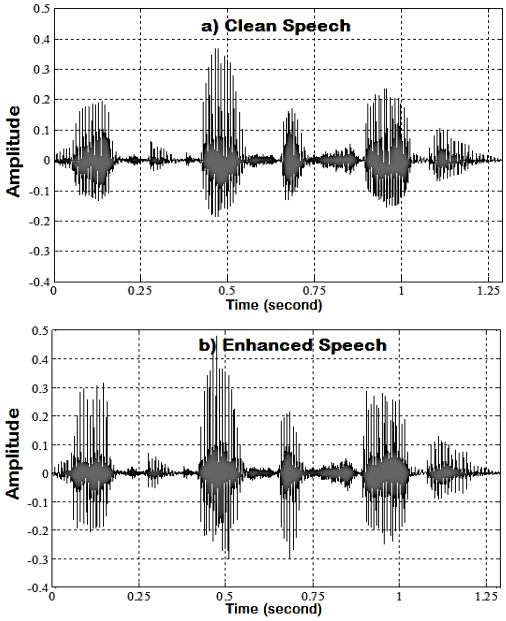
iteration number of the interior-point algorithm within each time frame in white noise for an SNR of –5 dB are shown in Figure 4. The iteration number is equal to zero in silent frames as illustrated in Figure 4, since the VAD block prevents to manipulate these frames. Also, the iteration range of the interior-point algorithm in speech-active frames is 1 to 167 with a total of 5302 iterations.

The enhanced speech is corrupted by white noise and babble noise to evaluate the performance of the proposed algorithm. The spectral power of white noise is uniformly distributed over the frequencies. This important property makes white noise appropriate to evaluate the performance of the speech enhancement methods in frequency domain; however, white noise is not an environmental noise. Babble noise, the noise of the public places, is a common environment noise that may destroy speech intelligibility in wireless communications.

The effect of the proposed algorithm on the power of one-third octave bands in clean and enhanced speech for white and babble noises with an SNR of –5 dB is shown in Figure 5. The energy is usually distributed from the frequency bands in which the clean speech power is greater than the noise power to other frequency bands, as illustrated in Figure 5. This transmission is such that the average of the energy correlation can be maximized. In other words, this method spreads the energy between the bands and makes a dense spectral power density.

4. Performance Evaluation

To evaluate the performance of the proposed algorithm, the clean and enhanced speech is degraded by white, babble, factory, and traffic noises at the SNRs of –20, –15, –10, –5 and 0 dB from the NOISEX-92 database.



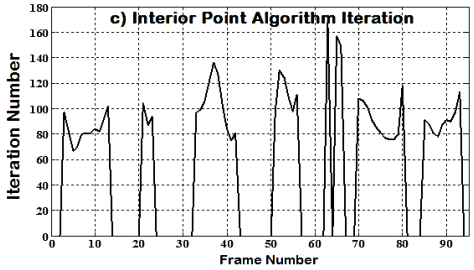


Fig. 4. The time domain plots of (a) clean speech and (b) enhanced speech with an SNR of –5 dB, and (c) iteration number of the interior-point algorithm within each frame.

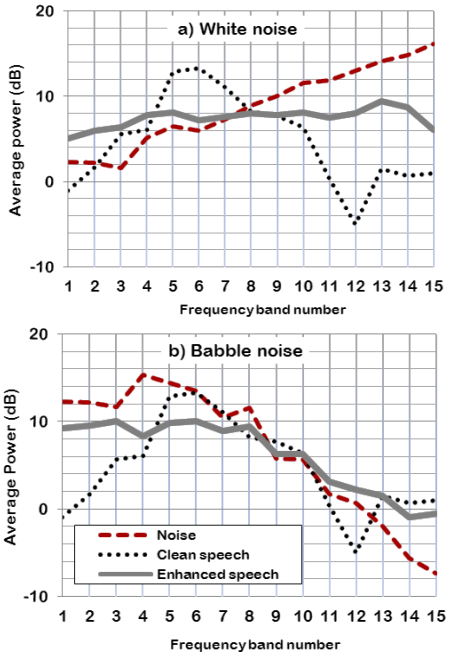


Fig. 5. Avrage power of one-third octave bands of the N=30 successive frams for noise, clean speech and enhaced speech in (a) white noise and (b) babble noise with an SNR of –5 dB.

In total, 20 random sentences from male speakers are used from the TIMIT database. The duration of each sentence is almost between 3 s and 5 s. A comparison is made with the unprocessed noisy speech and a reference method proposed by Taal *et al.* [10]. This method, which is similar to the proposed method, improves speech intelligibility by optimizing a cost function. Taal *et al.* optimally redistribute the speech energy over time-frequency cells according to a perceptual distortion measure. The baseline algorithm minimizes the mentioned objective measure by using auditory filter bank with a power constraint over all speech-active frames to improve the speech intelligibility.

The current study employs two objective intelligibility measures to predict the intelligibility of the noisy enhanced speech. The objective measures enable rapid feedback on a range of speech intelligibility enhancement methods. First measure is speech intelligibility index (SII), which is based on weighted SNRs. In the one-third octave band procedure provided by ANSI [17], the SII measure is computed by dividing the spectrum of clean speech and noise into one-third octave frequency bands and estimating the weighted average of the SNRs in each

band. The SNRs are weighted by band importance functions which differ across speech materials. The output of the SII measure is a scalar number between 0 and 1, which predicts the speech intelligibility in background noise. Second method is short-time objective intelligibility (STOI) measure, which is based on a correlation coefficient between the temporal envelopes (i.e., the root of energy of one-third octave frequency bands) of the clean and degraded speech in overlapping segments [18]. To calculate the STOI measure, the clean and degraded speech are decomposed into DFT-based. Then, short-time temporal envelope segments of the clean and degraded speech are compared by means of a correlation coefficient after normalizing and clipping. The STOI score is then obtained by averaging these short-time intelligibility measures over the speech signal. This measure provides a score in the range of 0 to 1, which refers to the speech intelligibility of degraded speech. Both measures can predict the intelligibility of noisy speech in various speech degradations.

The proposed method provides the best intelligibility scores in the STOI measure in comparison with the unprocessed noisy speech and the baseline method proposed by Taal *et al.* for all noisy conditions, as illustrated in Figure 6. The results also show that the baseline method led to a decrease in model intelligibility based on STOI at low SNRs in all noises except babble noise in which it provides a large increase in comparison with the unprocessed noisy speech in all SNRs.

Figure 7 presents the SII measure intelligibility scores. The results show intelligibility improvement with the proposed algorithm in comparison with the unprocessed noisy speech and the baseline method at low SNRs in the SII measure for white, babble, and traffic noises. However, the baseline method obtains better SII scores than our algorithm in factory noise at all SNRs. The baseline method also provides better intelligibility scores at –5 and 0 dB than the proposed method in all maskers, as illustrated in Figure 7.

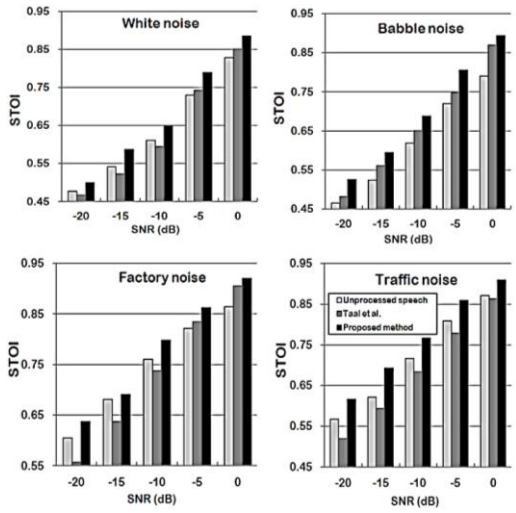


Fig. 6. STOI intelligibility predictions for the proposed method, unprocessed noisy speech, and the baseline method by Taal et al. for white, babble, factory, and traffic noises at the SNRs of –20, –15, –10, –5 and 0 dB.

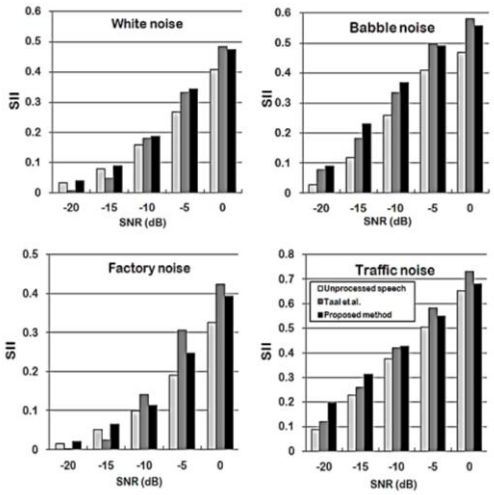


Fig. 7. SII intelligibility predictions for the proposed method, unprocessed noisy speech, and the baseline method by Taal et al. for white, babble, factory, and traffic noises at the SNRs of –20, –15, –10, –5 and 0 dB.

The significant intelligibility scores that the proposed method obtained in the STOI predictor are expected, since our method maximizes the energy correlation between clean speech and noisy enhanced speech, and also the STOI measure is based on the mean cross-correlations of the root of energy in frequency bands between these signals. Thereby, both algorithms are based on the correlation between clean speech and noisy speech. The proposed method improves speech intelligibility based on the correlation and STOI measures it to predict the speech intelligibility. Therefore, it can be stated that the proposed method maximizes an intelligibility cost function according to the STOI measure.

5. Conclusions

A new speech intelligibility improvement algorithm is proposed to enhance the speech intelligibility for the near-end listener in noisy environments without increasing the speech energy. This was performed by maximizing the energy cross correlation of the one-third octave bands between clean speech and enhanced noisy speech with a power constraint. The interior-point algorithm, an iterative algorithm for nonlinear optimization, is applied to solve the optimization problem, because of the nonlinearity and complexity of the cost function. The speech energy is redistributed over the frequency bands of clean speech according to the optimization of the energy correlation between clean and noisy speech. Two objective intelligibility predictors, the STOI and SII measures, are employed for scoring the intelligibility of the noisy enhanced speech under various noisy conditions to evaluate the performance of the proposed method. The results show significant intelligibility improvement with the proposed algorithm in comparison with the unprocessed noisy speech. As the current work is a frame-based speech enhancement method that maximizes the

cost function within each time frame, the proposed method can be appropriate for online processing. However, the iterative optimization algorithm used in the optimization problem is not appropriate for online processing due to the algorithmic delay produced with the loop iterations. Therefore, the iterative algorithm would be replaced with an alternative method with an insignificant algorithmic delay, in future works.

Acknowledgments

The research is supported by Khavaran Institute of Higher Education (KHI).

References

[1] C. Tantibundhit, J. R. Boston, C. C. Li, D. J. Durrant, S. Shaiman, K. Kovacyk, and A. El-Jaroudi. "Speech enhancement using transient speech components," in: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 2006, pp. 1–4.

[2] D. M. Rasetshwane. Enhancement of speech intelligibility using speech transients extracted by a wavelet packet-based real-time algorithm. Ph.D. Thesis, University of Pittsburgh, 2009.

[3] D. Mile, W. Verhelst, K. Struyve, and P. Verhoeve. "Efficient non-uniform time-scaling of speech with WSOLA," in: Proc. SPECOM, 2005, pp. 163–166.

[4] M. E. Hamid, S. Das, K. Hirose, and M. K. I. Molla. "Speech enhancement using EMD-based adaptive soft-thresholding EMDADT," International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 5, no. 2, pp. 1–16, 2012.

[5] B. Sauert, G. Enzner, and P. Vary. "Near-end listening enhancement with strict loudspeaker output power constraining," in: Proc. IWAENC, 2006. pp. 1–4.

[6] Y. Tang and M. Cooke. "Subjective and objective evaluation of speech intelligibility enhancement under constant energy and duration constraints," in: Proc. Interspeech, 2011, pp. 345–348.

[7] B. S. Premananda and B. V. Uma. "Selective Frequency Enhancement of Speech Signal for Intelligibility Improvement in Presence of Near-end Noise," Computer Science, Elsevier, vol. 49, pp. 244–252, 2015.

[8] B. Sauert and P. Vary. "Near-end listening enhancement optimized with respect to speech intelligibility index and audio power limitations," in: Proc. 18th European Signal Processing Conference, EUSIPCO, 2010, pp. 1919–1923.

[9] C.H. Taal, J. Jensen, and A. Leijon. "On optimal linear filtering of speech for near-end listening enhancement," Signal Processing Letters, IEEE, vol. 20, no. 3, pp. 225–228, 2013.

[10] C. H. Taal, R. C. Hendriks, and R. Heusdens. "Speech energy redistribution for intelligibility improvement in noise based on a perceptual distortion measure," Computer Speech & Language, Elsevier, vol. 28, no. 4, pp. 858–872, 2014.

[11] Y. Tang and M. Cooke. "Optimized spectral weightings for noise-dependent speech intelligibility enhancement," in: Proc. Interspeech, 2012, pp. 955–958.

[12] C. Valentini-Botinhao, J. Yamagishi, S. King, and R. Maia, "Intelligibility enhancement of HMM-generated speech in additive noise by modifying Mel cepstral coefficients to increase the glimpse proportion," Computer Speech and Language, Elsevier, vol. 28, no. 2, pp. 665–686, 2014.

[13] N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," Journal of Acoustical Society of America, vol. 19, no. 1, pp. 90–119, 1947.

[14] S. Boyd and L. Vandenberghe. Convex optimization. Cambridge university press, 2004, pp. 561–615.

[15] P. Vary and R. Martin. Digital speech transmission: Enhancement, coding and error concealment. John Wiley & Sons, 2006, pp. 143–150.

[16] T. Gerkmann and R. C. Hendriks. "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay." IEEE Transactions on Audio, Speech and Language Processing, vol. 20, no. 4, pp. 1383 – 1393, 2012.

[17] ANSI. S3.5-1997 Methods for Calculation of the Speech Intelligibility Index. New York: American National Standards Institute, 1997, pp. 90–119.

[18] C. H. Taal, R. Hendriks, R. Heusdens, and J. Jensen. "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," IEEE Transactions on Audio, Speech and Language Processing, vol. 19, no. 7, pp. 2125 – 2136, 2011.

Peyman Goli received the B.Sc and M.Sc degrees in Electronic Engineering from Noshirvani University of Technology, Babol, Iran, in 2002 and 2005, respectively. He is currently pursuing the Ph.D degree at the Electronic Engineering (Signal Processing), Noshirvani University of Technology, Babol, Iran, under the supervision of Dr. Mohammad Reza Karami-Mollaei. His main research topic is speech intelligibility enhancement in noisy environments. Other research interests include acoustic echo cancelation, speech noise cancelation and signal processing in the field of digital audio.

Mohammad Reza Karami-Mollaei received the B.Sc in Electrical and Electronic Engineering in 1992, M.Sc of signal processing in 1994, and Ph.D in 1998 in Biomedical Engineering from I.N.P.L d'Nancy of France. He is now an associate professor with the Department of Electrical and Computer Engineering, Babol University of Technology. Since 1998 his research is in signal and speech processing. He published more than 100 articles in journals and conferences. He teaches Digital Signal, Biomedical and speech processing at university. His research interests include Speech, Image and signal processing.

BER Performance Analysis of MIMO-OFDM Communication Systems Using Iterative Technique Over Indoor Power Line Channels in an Impulsive Noise Environment

Mohammad Reza Ahadiat*
Department of Electrical Engineering, Mehriz Branch, Islamic Azad University, Yazd, Iran
m.ahadiat@srbiau.ac.ir
Paeiz Azmi
Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
pazmi@modares.ac.ir
Afrooz Haghbin
Department of Electrical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
ahaghbin@gmail.com

Received: 02/Dec/2014 Revised: 16/Aug/2015 Accepted: 23/Aug/2015

Abstract

This paper addresses the performance of MIMO-OFDM communication system in environments where the interfering noise exhibits non-Gaussian behavior due to impulsive phenomena. It presents the design and simulation of an iterative technique that aims to minimize the effect of impulsive noise on the performance of the MIMO-OFDM communication system under Additive White Gaussian Noise (AWGN) channel. This is a new method to recover the signals corrupted by impulsive noise in MIMO-OFDM systems over In-home Power Line Channel. The location and amplitude Impulsive noise at the receiver using an adaptive threshold to be determined. Reduced Impulsive noise effects using the mask based on the soft decision method. By iteration, the original signal estimation can be used to improve the impulsive noise estimation. This continuous loop impulsive noise detection and mitigation a better estimate of the original signal is obtained. The Bit Error Rate (BER) performance of the MIMO-OFDM system in an impulsive noise environment was evaluated. The results show the superiority and robustness of the proposed method.

Keywords: Power Line Communication; MIMO; OFDM; Impulsive Noise; Iterative Method.

1. Introduction

The Power Line Communication (PLC) is a communication protocol that uses electrical wiring to simultaneously carry both data, and Alternating Current electric power transmission or electric power distribution. PLC networks can be divided into in-home network and out-home according to their applications. The applications of PLC in-home network have been increasing recently due to the development of Home Plug power line Alliance, which has published its standard, HomePlug specification [1]. Digital communication systems over power line channel suffer from several disturbances. Such the disturbances are composed of colored gaussian noise, impulsive noise, channel interference, frequency selective fading, attenuation, and so on[2,3].

Orthogonal Frequency Division Multiplexing (OFDM) modulation schemes is a promising technique being used for bandwidth efficient communication over the PLC, full capacity of the channel, high-speed data services broadband, reduce the complexity equalization, its capacity to minimize Intersymbol Interference (ISI), and powerful in impulsive noise environments. It is also a strong candidate as a modulation scheme the performs better than single-carrier and spread spectrum modulation methods. OFDM minimizes the effects of multipath and provides high robustness against selective fading. In PLC, the reliability of transmission is strongly influenced by the non-Gaussian impulsive noise. In OFDM, the Inverse Discrete Fourier Transform (IDFT) is used for modulating a block of N information symbols on N subcarriers[4]. The time duration of an OFDM symbol is N times larger than that of a single carrier system. This longer duration of OFDM symbol provide an advantage that the impulsive noise energy is spread among the N subcarriers due to the IDFT operation. This spreading causes less interference over all N subcarriers[5]. However, when the impulsive noise energy exceeds a certain threshold, a significant performance loss can occur due to the higher level of interference at each subcarrier[6]. It was shown in [7] that impulsive noise leads to an enormous loss in the capacity, as well as in the error rate performance.

Current high speed PLC technologies such as Hom plug AV uses as Single-Input Single-Output (SISO) communication mode and provides usable application level throughputs on typical PLC channels of about 100 Mbps. Today's in-home power line channel for 3-wire installations (phase, neutral and protective earth) basically allow more feeding and receiving possibilities, which is the precondition to apply Multiple-Input Multiple-Output

* Corresponding Author

(MIMO) and Alamouti coding methods[8]. In [9] the suitability of MIMO for in-home PLC has been investigated. It has been shown that MIMO can significantly increase the bit rate. Typically it is more than doubled compare to today's SISO systems. Alamouti scheme is considered to be the best choice of different MIMO schemes in the PLC environment. The advantages of the MIMO-PLC system mainly lies in namely increased channel capacity and reduced BER[10]. MIMO techniques in combination with orthogonal frequency-division multiplexing (MIMO-OFDM) have been identified as a promising approach for high spectral efficiency wide band systems.

Different methods based on OFDM for reducing impulsive noise over SISO in-home power line channels have been proposed. The performance of these methods, depends on the number of sub-carriers. For a large number of sub-carriers the convergence speed is fast, whereas, for a small number of sub-carriers the algorithms converge slowly or even not at all. For a small number of sub-carriers, the noise components after Discrete Fourier Transform (DFT) have an impulsive distribution. To compensate for the limitations of these systems and to be able to adopt it for all sub-carriers, a new adaptive iterative method is proposed in this work. Using this method, smaller number of iterations are needed, the impulsive noise is reduced and a better performance is achieved.

At the receiver, after ZFE, impulsive noise detection algorithm based on adaptive threshold for estimating the impulsive noise, determines the locations and amplitudes of the impulsive noise. The adaptive threshold technique utilized is optimized radar's Constant False Alarm Rate (CFAR) detector technique [11]. After noise symbol determination, impulsive noise effect on the noise symbols is reduced using the reduction factor based on the soft decision method. The reduction factor is defined as an exponential function of the absolute difference between the estimated impulsive noises of the adaptive threshold. Then, the original signal is estimated using MLD. To improve performance, the proposed method can be repeated. By iteration, the original signal estimation can be used to improve the impulsive noise estimation. As this continuous loop impulsive noise detection and mitigation to receive together with MLD, a better estimate of the original signal is obtained.

2. System Model

The general model of a PLC is depicted in Fig. 1.

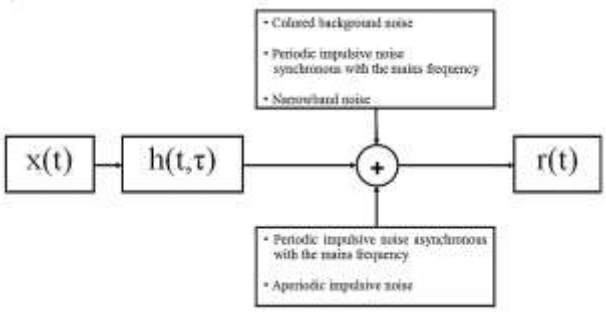


Fig. 1. Power line channel noise scenario

The $h(t,\tau)$ is channel impulse response and is represented by a channel filter and different noise components as depicted in Fig. 1 can be expressed as background noise and impulsive noise[12].

2.1 Background Noise Modeling

The background noise can create disturbances in the frequency range 0-100 MHz [13]-[14]. The background noise can be separated in colored noise and narrowband interference[12]. The background noise according to [15], can be modeled in the simple three-parameter model, where the noise is considered gaussian with the Power Spectral Density (PSD) can be expressed as:

$$R_{n_b}(f) = a + b|f|^c \quad [\frac{dBm}{Hz}] \qquad (1)$$

Where f is the frequency in MHz, and a, b and c are parameters derived from measurements. The worst and best condition in-home power line channel parameter values are characterized by $[a, b, c]$ =[-145, 53.23,-0.337] and $[a, b, c]$ = [-140, 38.75,-0.72], respectively, in this work. The PSD of these conditions is shown in Fig.2.

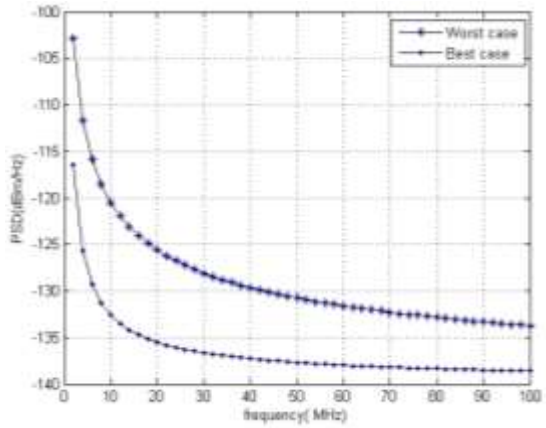


Fig. 2. The PSD of background noise for the worst and best conditions used in this work.

2.2 Impulsive Noise Modeling

The impulsive noise can be separated in periodic impulsive noise synchronous with the mains frequency, Periodic impulsive noise asynchronous with the mains frequency and aperiodic impulsive noise [12],[16]. Therefore, the overall noise can be written as:

$$\eta(t) = n_b(t) + n_{imp}(t) \qquad (2)$$

where $n_b(t)$ denotes the background noise, while $n_{imp}(t)$ denotes the impulsive noise. The impulsive noise can be modeled with an arrival process. This process is specified by the statistics of the impulse amplitude A_{imp}, interarrival time t_{iat}, and duration t_{wid}. Most of the literature focuses on the statistics of the amplitude, as the Middleton's classification of electromagnetic interference [17] and the two-term gaussian mixture [18] model. The temporal characteristics have been modeled via Markov Chains in[12], Bernoulli process in [19] and curve fitting tools from experimental measurements in [15].

A relatively simple model which incorporates background noise and impulsive noise based on the Poisson Gaussian model is known as Midlenton's class-A impulsive.

In the Middleton's model[17], the noise is categorized into three different types A, B and C. Because Midlenton's class-A impulsive noise model is more accurate and also meets all the basic requirements in modeling the real impulsive noise, this model has been used widely in performance analysis of PLC systems. This thesis also relies on Midlenton's class-A to model impulsive noise and design communications systems over power lines. Midelenton's class-A model uses the Poisson Gaussian model to represent the background noise and impulsive noise. The occurrence probability of impulsive noise is modeled by a Poisson random process with the probability of having m impulsive noise events in a time interval T given by:

$$P = \frac{(\lambda T)^m e^{-\lambda T}}{m!} \quad (3)$$

The amplitudes of both background and impulsive noise are modeled by Gaussian random processes. Let $A = \lambda T$ and call it the impulsive index. According to the Middleton's Class A noise model, the normalized complex probability density function (PDF) of the model is given by:

$$p_A(n) = \sum_{m=0}^{\infty} \frac{e^{-A} A^m}{m!} \cdot \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{n^2}{2\sigma_m^2}\right) \quad (4)$$

$$\sigma_m^2 = \left(\frac{\left(\frac{m}{A}\right) + \Gamma}{1 + \Gamma}\right) . \sigma^2$$

$$A \in [10^{-2}, 1] , \Gamma \in [10^{-6}, 1]$$

where σ_m^2 is the mth impulsive power, A is impulsive index, σ^2 is total noise power (including the powers of impulsive noise and Gaussian background noise), and $\Gamma = \frac{\sigma_G^2}{\sigma_i^2}$ is Gaussian to Impulsive noise power Ratio (GIR) with σ_G^2 and σ_i^2 are the powers of Gaussian and impulsive noise, respectively. When A is increased, the impulsiveness reduces and the noise comes closer to gaussian noise. Eq. (4) also shows that sources of impulsive noise have a Poisson distribution, and each impulsive noise source generates a characteristic Gaussian noise with a different variance. A good approximation of Eq.(4), is obtained by cutting off the cumulative sum after the third term [18]. The effects of impulsive parameters A and Γ are illustrated in Fig. 3 and Fig. 4.

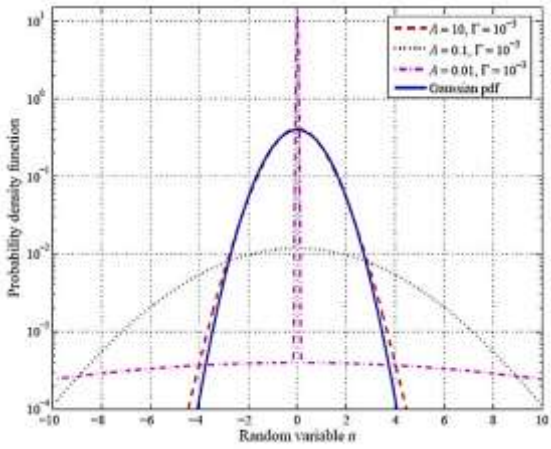


Fig. 3. Pdf of the impulsive noise with different values of impulsive index A.

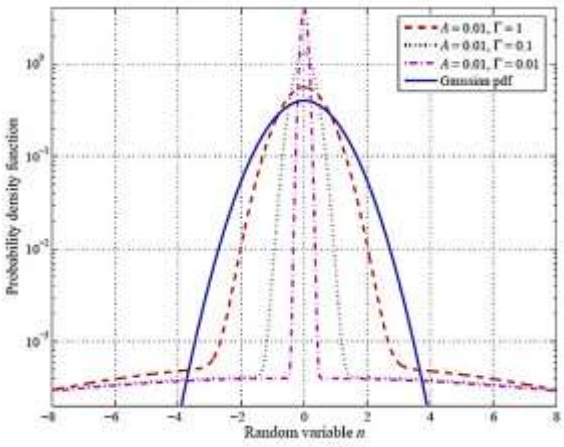


Fig. 4. Pdf of the impulsive noise with different values of Γ

Fig. 5 also illustrate the effects of the impulsive noise parameters A and Γ to the amplitude distribution of class-A impulsive noise. Again, it can be seen that if A or Γ increases, the amplitude distribution of impulsive noise comes closer to that of Gaussian noise.

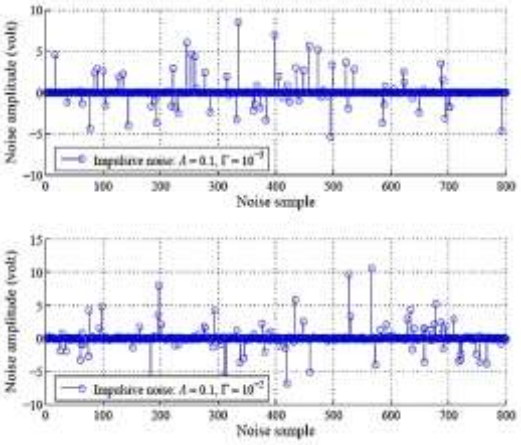


Fig. 5. Examples of impulsive noise with A=0.1 and two different values of Γ.

2.3 MIMO-PLC Channel Modeling

In this paper, MIMO-PLC system which considers the coupling effects among conductors Can be modeled by Zimmermann and Dostert model[20] where the MIMO-PLC model is considered with the Channel Transfer Function (CTF) [20]:

$$H(f) = B\sum_{p=1}^{N_p} g_p(f).\, e^{-\alpha(f)d_p}.\, e^{-j\left(\frac{2\pi f d_p}{v_d}\right)} \quad (5)$$

$$\alpha(f) = \text{Real}\{= \alpha + j\beta = \sqrt{(R' + j\omega L')(G' + j\omega C')}\}$$

where N_p, B , $|g_p| \le 1$, $\tau_p = \frac{d_p}{v_d}$, d_p , $v_d = \frac{c}{\sqrt{\varepsilon_r}}$, ε_r , γ, α and β and are the total number of fading paths, normalization term, the transmission reflection factor for path p, the pth path propagation delay, the length of the pth path, the phase velocity with c speed of light, dielectric constant , the propagation constant, the attenuation constant and the phase constant, respectively. so we are only interested in the attenuation constant. The primary cable parameters L' (inductance per unit length) and C' (capacitance per unit length) can be estimated by the geometric dimensions and the material properties. The R' (resistance per unit length) and G' (conductance per unit length) depend on frequency. The attenuation of a power line cable can be characterized by [21]:

$$A(f, d_p) = e^{-\alpha(f)d_p} = e^{(a_0 + a_1 f^k)d_p} \quad (6)$$

The parameters a_0, a_1, K are chosen to adapt the model to a specific network. In this paper According to [22] a statistical model of the channel can be derived by considering the parameters in Eq. (5) as random variables. The path lengths are assumed to be drawn from a Poisson arrival process with intensity Λ in m^{-1}. The reflection factors g_p are assumed to be real, independent, and uniformly or log-normally distributed. Finally, the parameters a_0 , a_1 and K are appropriately chosen to a fixed value[23].

All PLC available today use one transmitting and one receiving port for their communication. The signal is symmetrically fed and received between the live and neutral wire. In Iran, in-home installations consist of three wires, which offers additional feeding and receiving options. Fig. 6 shows the PLC MIMO channel. Differential signaling between any two of the three wires lead to three different feeding possibilities: P (Phase or Live) to N (Neutral), P to PE (Protective Earth), and N to PE.

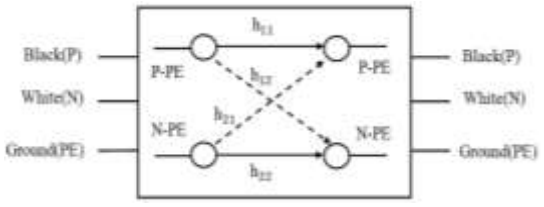


Fig. 6. In-home power line channels with two-input and two-output.

According to Kirchhoff's rule the sum of the three input signals has to be zero. Therefore only two out of the three independent input ports can be used. On receiving side all three differential reception ports are available[8]. The coupling effects between conductors in the MIMO-PLC should be considered. The CTF in MIMO system for the ith transmit to the jth receive path (where i = 1, 2 and j = 1, 2) can be written as:

$$H_{i,j}(f) = \sum_{p=1}^{N_p} B\, g_p(f).\, e^{-\alpha_{i,j}(f)d_p}.\, e^{-j2\pi f\tau_p} \quad (7)$$

Eq. (7) can be extended to the overall transfer function matrix:

$$H_{MIMO} = \begin{bmatrix} H_{1,1}(f) & H_{1,2}(f) \\ H_{2,1}(f) & H_{2,2}(f) \end{bmatrix} \quad (8)$$

where $H_{i,j}$, $i = j$ indicate co-channels and $H_{i,j}$, $i \neq j$ indicate cross channels. The attenuation constant $\alpha_{i,j}$ in Eq. (7) can be extracted from:

$$\alpha_{i,j} = \text{real}\{\left(\sqrt{(R'' + j\omega L'').*\,(G'' + j\omega C'')}\right)_{i,j}\} \quad (9)$$

where the operator $.*$ indicates the element-wise matrix multiplication. R'', L'', $C''$and G'' correspond to transmission line matrices[24], which represent the mutual interactions between conductors. The equivalent per-unit-length (p.u.l) parameter model [9] can be used to characterize the in-home transmission line, as shown in Fig. 7.

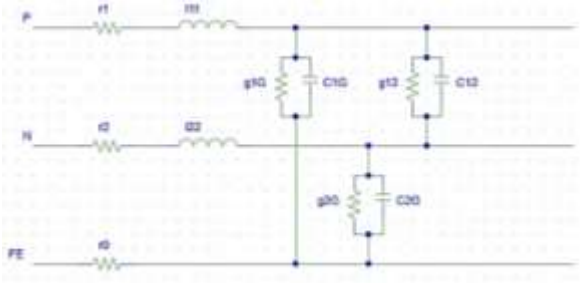


Fig. 7. The The per unit length parameter model for in-home power lines.

The resistance matrix is given as:

$$R'' = \begin{bmatrix} r_1 + r_0 & r_0 \\ r_0 & r_2 + r_0 \end{bmatrix} \quad (10)$$

where r_0 , r_1 and r_2 are the ground resistance, the resistances for line P and the resistances for line N (which indicates line 1, line 2 in Fig. 7) per unit length, respectively and computed as:

$$r_1 = r_2 = \frac{1}{2}\sqrt{\frac{\pi f \mu_c}{\sigma_c}} \quad (11)$$

where μ_c , σ_c and f and are the permeability and conductivity of conducting material, the wave frequency and, respectively. The inductance matrix is given as:

$$L'' = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} \quad (12)$$

where the l_{11} , l_{22} are the self-inductances for line P, N per unit length and the l_{12} , l_{21} are the mutual inductances. The computation for self-inductance per unit length is given as:

$$l_{11} = l_{22} = \frac{\mu_0}{2\pi}\ln\frac{GMD}{GMR_L} \quad , GMD = D_{12} \quad (13)$$

Note that GMD, μ_0 and GMR_L are geometric mean distance, the permeability of dielectric material between conductors and geometric mean radian. The anti-diagonal terms in Eq.(12) can be computed as $l_{12} = l_{21} = k\sqrt{L_{11}L_{22}}$. where k is called the coefficient of coupling ($0 \le k \le 1$). The capacitance matrix is given as:

$$C'' = \begin{bmatrix} c_{11} & -c_{12} \\ -c_{21} & c_{22} \end{bmatrix} \quad (14)$$

where c_{11} and c_{22} are the self-capacitances for line P, N per unit length and given as $c_{11} = c_{1G} + c_{12}, c_{22} = c_{2G} + c_{21}$ .and c_{PG}, c_{NG} (i.e., c_{1G}, or c_{2G}) are the capacitances between line P, N and ground (G), and given as:

$$c_{1G} = \frac{2\pi\varepsilon_0}{\ln\frac{GMD}{GMR_C}} \quad (15)$$

$$c_m = 4\pi\varepsilon_0 (= -c_{12} = -c_{21})$$

where ε_0 and GMR_C are the permittivity of dielectric material between conductors and the actual conductor radius r. The conductance matrix is given as:

$$G'' = \begin{bmatrix} g_{11} & -g_{12} \\ -g_{21} & g_{22} \end{bmatrix} \quad (16)$$

$$g_{1G} = 2\pi f c_{1G} \tan\delta \ , \ g_m (= -g_{12} = -g_{21})$$

where g_{11} and g_{22} are the self-conductances for line P, N per unit length and given as $g_{11} = g_{1G} + g_{12}, g_{22} = g_{2G} + g_{21}$ .where g_{PG}, g_{PG} (i.e., g_{1G} or g_{2G}) are the conductances between line P, N and ground (G), δ is the skin depth of the conducting material. Performing the required mathematical operation in each element of the transmission parameter matrices to solve for the attenuation factor $\alpha_{i,j}$ in Eq. (9), (i.e., $\alpha_{1,1} = \sqrt{(r_1 + r_0 + j\omega l_{11})(g_{11} + j\omega c_{11})}$, $\alpha_{1,2} = \sqrt{(r_0 + j\omega l_{12})(g_{12} - j\omega c_{12})}$), the channel matrix in Eq. (10) can be obtained. [24]-[25].

2.4 OFDM System Modeling

Fig. 8 shows the block diagram of a MIMO-OFDM system.

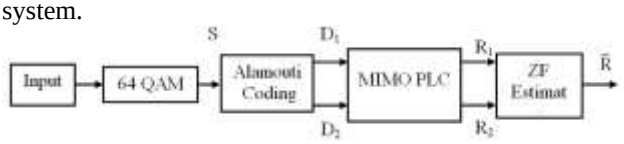


Fig. 8. Block diagram of the proposed 2*2 MIMO-OFDM System.

In the transmitter side, input signal is first converted to digital symbol. The serial data stream of the source is mapped to data symbols with employing the signal constellation scheme of 64-QAM. Then, modulated symbols are fed to the space time block encoder [26]. Complex symbols of each transmit path are processed by OFDM modulated by an Inverse Fast Fourier Transform (IFFT) and a Cyclic Prefix (CP) insertion block. In telecommunications, the term cyclic prefix refers to the prefixing of a symbol with a repetition of the end. Although the receiver is typically configured to discard the cyclic prefix samples, the cyclic prefix serves two purposes. The symbols are then transmitted to the receiver via a 2*2 MIMO in-home power line channel, as shown in Fig. 6. The receiver performs the reverse operation of the transmitter. The received signal is carried out by CP removal, FFT operation. The ZF scheme applies the inverse of the frequency response of channel to the symbol received, so that the original signal can be detected to an optimum level. ZF is the one of the best linear receiver detection method having low computational complexity, but it suffers from sudden noise enhancement.

3. The Proposed Method

The symbols are received of 2*2 MIMO-PLC by Alamouti scheme in OFDM-based systems over In-home Power Line Channels. The received symbols in the first time slot, and by assuming that the channel remains constant for the second time slot, the received symbols in the second time slot are:

$$\begin{aligned} R_{11} &= H_{11}S_1 + H_{12}S_2 + N_{11} \\ R_{12} &= H_{21}S_1 + H_{22}S_2 + N_{12} \\ R_{21} &= -H_{11}S_2^* + H_{12}S_1^* + N_{21} \\ R_{22} &= -H_{21}S_2^* + H_{22}S_1^* + N_{22} \end{aligned} \quad (17)$$

The received symbols are equalized by ZF equalization as shown in Fig. 9.The received symbols consist of original symbols and noises, which can be computed as:

$$R = HS + N, \tilde{S} = S + W.N, W = (H)^{\dagger} = (H^H H)^{-1} H^H \quad (18)$$

Where $(..)^{\dagger}$ is the pseudo-inverse and $(..)^{H}$ is the Hermitian transpose[27].

The proposed method in this paper consists of an iterative loop of three main CFAR, SOFT decision, and MLD. In this technique for impulsive noise reduction in noisy samples, of adaptive weights are used in each iteration. As depicted in Figure 6, successive detection and estimation of locations and amplitudes of impulsive noise are used to improve signal reconstruction quality [28]. For each detection and estimation step, the estimate of the noise improves using the estimated original signal from the previous step. The estimate of the impulsive noise is as follows

$$\hat{e}(n) = \hat{r}(n) - \hat{s}(n) \quad (19)$$

Where $\hat{r}(n)$ and $\hat{s}(n)$ are estimated received signal, and estimated original signal.

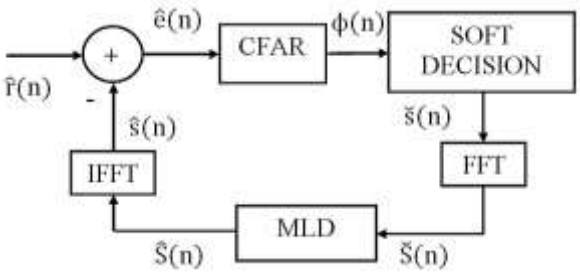


Fig. 9. Block of the proposed method.

3.1 Impulsive Noise Detection

The threshold in impulsive noise detection block can either be adaptive or non-adaptive. The CFAR is an adaptive threshold method used in radar detectors based on Neyman-Pearson criterion[28]. In this paper we use a Censored Mean Level (CML) CFAR. The reason we use CML-CFAR is to reduce the probability that impulses involved in averaging when number of them are present in adjacent cells. In the k^{th} order CML-CFAR of length $2n_c$, among the $2n_c$ adjacent cells, k of the smallest amplitudes are averaged and the other $2(n_c - k)$samples (which may contain impulsive noises) are ignored. In this method, the probability of impulsive noise presence in averaged adjacent samples is reduced. To calculate adaptive thresholds in the mth cell, $2N_c$ points are expressed as $\{n_1, \dots, n_{2N_c}\} = \{m - N_c, \dots\dots, m + N_c\}$, and the impulsive noise estimate in these points is sorted by the largest amplitudes $|\hat{e}[n_1]| \leq \cdots \leq |\hat{e}[n_{2N_c}]|$. Adaptive threshold can be computed as:

$$\eta[m] = \frac{1}{k}\sum_{l=1}^{k}|e[n_l]| \quad (20)$$

3.2 Attenuation of Impulsive Noise

The detection algorithm is not an error free process in the early detection and estimation steps, especially when the amplitudes of the impulsive noises are small compared to the samples of the signal amplitudes. When the hard decision method is employed, the modulus is either zero or one at each sample. So that some samples are erased. One of the disadvantages of the soft method is its low convergence rate even when good estimates of the impulsive noise locations are available. To overcome this problem, at each stage of the detection and estimation we gradually change the soft decision to the hard decision. Simulation results for different modes suggested that the modulus function is as follows:

$$\emptyset(e[k], \eta[k]) = \exp(-\alpha|e[k] - \eta[k]|) \quad (21)$$

where $\hat{e}$, η and α are the estimated error, the threshold generated by CFAR [29] and the softness order of the soft decision, respectively. The suggested modulus function, the value $e[k] - \eta[k]$ tends to zero, and then the α value is increased gradually through detection and estimation steps as the convergence rate increases. In each detection and estimation step, a new mask is obtained by

$$\check{s}(n) = \hat{r}(n)\emptyset(n) = \hat{r}(n)e^{-\alpha|\hat{e}(n)-\eta(n)|} \quad (22)$$

4. Simulation Results

To evaluate the proposed method, OFDM modulation with 64 sub-carriers together with *Alamouti coding* to transmit, and 2×2 In-home PLC, and CFAR, Soft Decision with MLD to receive are simulated. It is assumed OFDM with a small number of sub-carriers that all sub-carriers are used for data transmission (not be used for pilot tones) and 10^4 symbols (5×10^6 bits), the impulsive noise from Middleton's class-A model is generated, and the noise samples are independent and identically distributed (i.i.d). According to the proposed MIMO model for in-home power line channels, and power line characteristics, parameters of (10) were calculated [31].

Fig. 10 and Fig. 11 shows the BER performance of the proposed method is compared with usual method without and with 15 iteration in power line channel, respectively. The noise on the power line channel between Gaussian noise and impulsive noise is changing. This noise is the Worst mode for impulsive noise with high range and number of impulse and is in the best state for Gaussian noise. Figure 12 shows the proposed method with the usual method performance in the worst case channel noise.

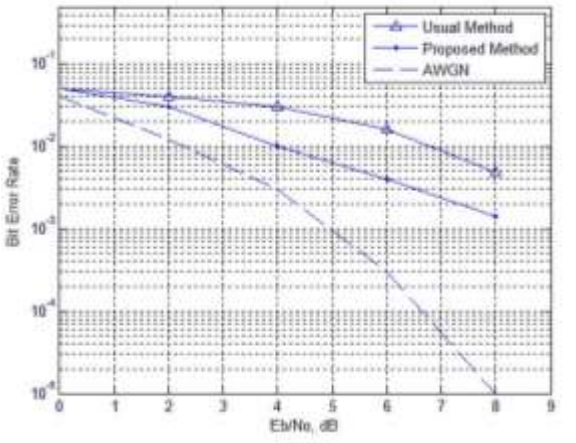


Fig. 10. Comparison of BER performance of OFDM-PLC without iteration (Impulsive Noise $\Gamma = 0.01$, $A = 0.1$).

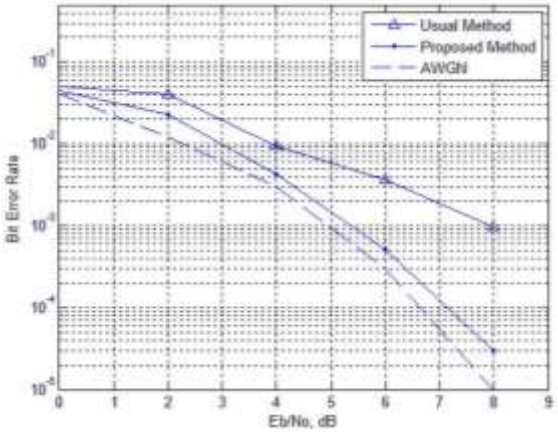


Fig. 11. Comparison of BER performance of OFDM-PLC with 15 iteration (Impulsive Noise = 0.01 , A = 0.1).

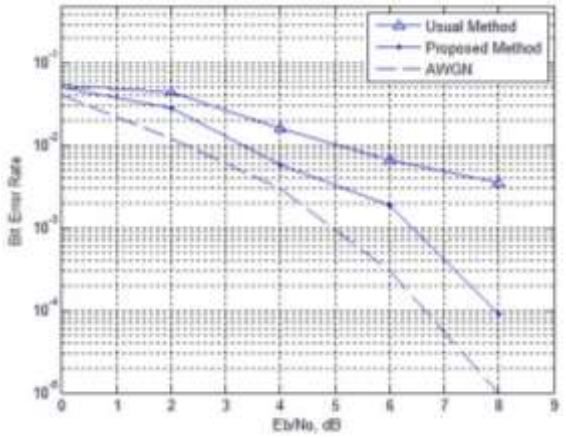


Fig. 12. Comparison of BER performance of OFDM-PLC with 15 iterations (Impulsive Noise $\Gamma = 0.0001$ and $A = 0.01$).

The best BER performance of impulsive noise reduction algorithms based on OFDM, Nonlinear Estimation and Mitigation algorithm in [32,33] and Sparse Bayesian Learning method for estimation and mitigation without training in [34] has been reported. Both of these algorithms are based on iteration. These OFDM algorithms with 64 sub-carriers are simulated. According to Fig. 13,with the same number of iterations them with proposed method are compared. Fig. 14 shows the BER performances of the proposed algorithms, for the worst case channel noise, impulsive noise with strong impulsiveness .

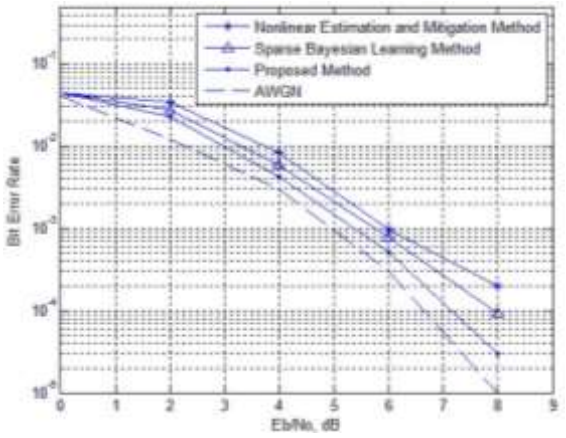


Fig. 13. The BER performance the Usual Method in comparison with the Proposed method of OFDM-PLC with 15 iterations (Impulsive Noise $\Gamma = 0.01$ and $= 0.1$).

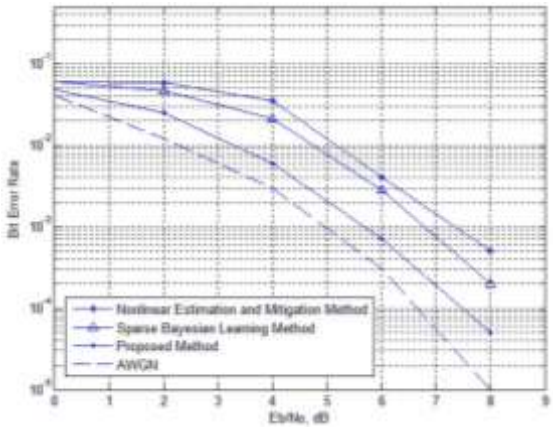


Fig. 14. The BER performance the Usual Method in comparison with the Proposed method of OFDM-PLC with 15 iterations (Impulsive Noise $\Gamma = 0.0001$ and $= 0.01$).

5. Conclusions

In this paper, we proposed an Impulsive Noise Estimation and Suppression in OFDM-MIMO Systems over In-home Power Line Channels. The proposed method in this paper consists of an iterative loop of three main CFAR, SOFT DECISION, and MLD. The location and amplitude Impulsive noise at the receiver using an adaptive threshold to be determined. Reduced Impulsive noise effects using the mask based on the soft decision method. By iteration, the original signal estimation can be used to improve the impulsive noise estimation. This continuous loop impulsive noise detection and mitigation a better estimate of the original signal is obtained. Simulation results show that the BER performance of proposed method was higher than the other impulsive noise estimation and mitigation methods.

References

[1] S. Gardner. "The HomePlug for powerline in home networking ." Proc. Int. Symp. Power Line Commun. and Its Appl, pp. 67-72, 2001.

[2] J. A. Cortes, L. Diez, F. J. Canete, J. J. Sanchez-Martinez. "Analysis of the Indoor Broadband Power-Line Noise Scenario". IEEE Trans. Electromagn. Compat, vol. 52, pp. 849-858, Nov. 2010.

[3] M. Ghosh. "Analysis of the effect of impulse noise on multicarrier and single carrier QAM systems". IEEE Trans. Commun, vol. 44, pp. 145-147, 1996.

[4] S. Dlhan, P. Farka. "Impulsive noise cancellation in systems with OFDM modulation". Journal of Electrical Engineering, vol. 59, pp. 310–316, 2008.

[5] H. Nikookar, D. Nathoeni. "Performance evaluation of OFDM transmission over impulsive noisy channels". IEEE. Int. Symp. Personal. Indoor and Mobile Radio Communications, vol. 2, pp. 550 - 554, Sep. 2002.

[6] U. Epple, D. Shutin, M. Schnell. "Mitigation of Impulsive Frequency-Selective Interference in OFDM Based

Systems". International Journal of Scientific & Engineering Research, vol. 4, pp. 2253-2256, May. 2013.
[7] Y. H. Ma, P. L. So, E. Gunawan. "Performance Analysis of OFDM Systems for Broadband Power Line Communications Under Impulsive Noise and Multipath Effects". IEEE Trans. Power Delivery, vol. 20, pp. 674-682, Apr. 2005.
[8] L. Hao, J. Guo. "A MIMO-OFDM Scheme over Coupled Multi-conductor Power-Line Communication Channel ." IEEE. Int. Symp. Power Line Commun. and Its Appl, pp. 198-203, 2007.
[9] L. Stadelmeier, D. Schneider, D. Schill, A. Schwager, J. Speidel. "MIMO for In-home Power Line Communications ." in International Conference on Source and Channel Coding (SCC).Ulm, pp. 1-6, Jan. 2008.
[10] A. Canova, N. Benvenuto, P. Bisaglia. "Receivers for MIMO-PLC channels: Throughput comparison ." IEEE. Int. Symp. Power Line Commun. and Its Appl, pp. 114-119, Mar. 2010.
[11] M. Skolnik. Introduction to Radar Systems: McGraw-Hill, 2000, pp. 126-140.
[12] M. Zimmermann, K. Dostert, "An analysis of the broadband noise scenario in powerline networks." IEEE. Int. Symp. Power Line Commun. and Its Appl, pp. 131–138, Apr. 2000.
[13] H. Philipps, "Performance measurements of power line channels at high frequencies." IEEE. Int. Symp. Power Line Commun. and Its Appl, pp. 229–237, Mar. 1998.
[14] B. Praho, M. Tlich, P. Pagani, A. Zeddam, F. Nouvel. "Cognitive detection method of radio frequencies on power line networks." IEEE. Int. Symp. Power Line Commun. and Its Appl, pp. 225-230, Mar. 2010.
[15] T. Esmailian, F. R. Kschischang, P. G. Gulak. "In-building power lines as high-speed communication channels: Channel characterization and a test channel ensemble". International Journal of Communication Systems, vol. 16, pp.381–400, 2003.
[16] A. K. S. Al-Bayati, O. M. Aloquili, J. I. AlNabulsi. "Efficient use of DS/CDMA signals for broadband communications over power lines". International Journal of Communication Systems, vol. 26, pp.212–221, Sep. 2011.
[17] D. Middleton. "Statistical-physical models of electromagnetic interference." IEEE Transactions on Electromagnetic Compatibility, vol. 19, pp.106–127, 1977.
[18] R. S. Blum, Y. Zhang, B. M. Sadler, R. J. Kozick. "Time-domain equalizer for multicarrier systems in impulsive noise". International Journal of Communication Systems, vol.25, pp.111–120, Apr. 2011.
[19] H. Dai, H. V. Poor. "Advanced signal processing for power line communications". IEEE Communications Magazine, vol. 41, pp.100–107, 2004.
[20] M. Zimmermann, K. Dostert. "A multipath model for the powerline channel". IEEE Transactions on Communications, vol. 50, pp.553–559, 2002.
[21] A. M. Tonello, S. D'Alessandro, L. Lampe. "Bit, tone and cyclic prefix allocation in OFDM with application to in-home PLC." Proc. IEEE IFIP Wireless Days Conference, pp. 23-27, Nov. 2008.
[22] A. M. Tonello. "Wideband Impulse Modulation and Receiver Algorithms for Multiuser Power Line Communications". EURASIP Journal on Advances in Signal Processing, vol. 2007, Dec. 2007
[23] A. M. Tonello, F. Versolatto. "Bottom-up statistical PLC channel modeling—Part I: Random topology model and efficient transfer function computation". Power Delivery. IEEE Transactions on, vol. 26, pp.891-898, Apr. 2011.
[24] R. Clayton, R. Paul. Analysis of Multiconductor Transmission Lines, John Wiley and Sons, Inc, 2004, pp. 46-62.
[25] J. Glover, M. Sarma, T. Overbye. Power System Analysis and Design (Fourth Edition), Thomson, 2008, pp. 155-213.
[26] C. L. Giovaneli, B. Honary, P. G. Farrell. "Space-frequency coded OFDM system for multi-wire power line communications." IEEE. Int. Symp. Power Line Commun. and Its Appl, pp. 191-195, Apr. 2005.
[27] P. Ferreira. "Stability issues in error control coding in the complex field, interpolation, and frame bounds." Signal Processing Letters, IEEE, vol. 7, pp. 57–59, 2000.
[28] J. C. Chang, F. B. Ueng. "Performance of OFDM-CDMA receivers with MIMO communications." International Journal of Communication Systems,vol. 27,pp. 732-749, May. 2014.
[29] Y. C. Tan, N. B. Ramli. "Time-domain equalizer for multicarrier systems in impulsive noise." International Journal of Communication Systems, vol. 25, pp. 111-120, Feb. 2012.
[30] S. Zahedpour, M. Ferdosizadeh, F. Marvasti, G. Mohimani, M. Babaie-Zadeh. "A novel impulsive noise cancellation based on successive approximations." in Proceedings of SampTa 2007,pp. 126-131, Jun. 2007.
[31] B. Adebisi, S. Ali, B. Honary. "Space frequency and space time frequency M3FSK for indoor multi wire communication". IEEE Trans. on Power Delivery, vol. 24, pp. 2361-2367, Oct. 2009.
[32] Y. C. Hsiao. " Iterative Interference Cancellation for OFDM Signals With Blanking Nonlinearity in Impulsive Noise Channels". IEEE Signal Processing Letters, vol. 19, pp. 147-150, Mar. 2012.
[33] R. Sukanesh, R. Sundaraguru. " Mitigation of Impulse Noise in OFDM Systems". Journal of Information & Computational Science, vol. 8, pp. 2403–2409, 2011.
[34] J. Lin, M. Nassar, L. Brian, E. Fellow. " Impulsive Noise Mitigation in Power line Communications Using Sparse Bayesian Learning". IEEE Journal on Selected Areas of Communications, pp. 1-26, Mar. 2013.

Mohammad Reza Ahadiat received the B.Sc, M.Sc, and Ph.D degrees in electrical engineering from Yazd University, Yazd, Iran, Tarbiat Modares University, Tehran, Iran, and Science and Research Branch, Islamic Azad University, Tehran, Iran, in 1999, 2003, and 2013, respectively. He is currently with the electrical and computer department of Mehriz Branch in Islamic Azad University, Yazd, Iran, as assistant professor. His areas of research include MIMO wireless communications, Power Line Communication, multicarrier modulation and estimation theory.

Paeiz Azmi was born in Tehran-Iran, on April 17, 1974. He received the B.Sc, M.Sc, and Ph.D degrees in electrical engineering from Sharif University of Technology (SUT), Tehran-Iran, in 1996, 1998, and 2002, respectively. Since September 2002, he has been with the Electrical and Computer Engineering Department of Tarbiat Modares University, Tehran-Iran, where he became an associate professor on January 2006 and he is a full professor now. Prof. Azmi is a senior member of IEEE. From 1999 to 2001, Prof. Azmi was with the Advanced Communication Science Research Laboratory, Iran Telecommunication Research Center (ITRC), Tehran, Iran. From 2002 to 2005, he was the Signal Processing Research Group at ITRC.

Afrooz Haghbin received her B.Sc degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2001. She received her M.Sc degree from Tehran University and her Ph.D degree from Tarbiat Modares University, Tehran, Iran, all in electrical engineering in 2004 and 2009, respectively. She is currently with the electrical and computer department of Science and Research Branch in Azad University, Tehran, Iran, as assistant professor. Her research interests include MIMO wireless communications, channel coding, precoding, multicarrier modulation and estimation theory.

Referral Traffic Analysis: A Case Study of the Iranian Students' News Agency (ISNA)

Roya Hassanian-Esfahani*
Research Institute for ICT, ACECR, Tehran, Iran
r.hassanian@ictrc.ac.ir
Mohammad-Javad Kargar
Department of Computer Engineering, University of Science and Culture, Tehran, Iran
kargar@usc.ac.ir

Received: 27/June/2015 Revised: 23/Dec/2015 Accepted: 31/Jan/2016

Abstract

Web traffic analysis is a well-known e-marketing activity. Today most of the news agencies have entered the web providing a variety of online services to their customers. The number of online news consumers is also increasing dramatically all over the world. A news website usually benefits from different acquisition channels including organic search services, paid search services, referral links, direct hits, links from online social media, and e-mails. This article presents the results of an empirical study of analyzing referral traffic of a news website through data mining techniques. Main methods include correlation analysis, outlier detection, clustering, and model performance evaluation. The results decline any significant relationship between the amount of referral traffic coming from a referrer website and the website's popularity state. Furthermore, the referrer websites of the study fit into three clusters applying K-means Squared Euclidean Distance clustering algorithm. Performance evaluations assure the significance of the model. Also, among detected clusters, the most populated one has labeled as "Automatic News Aggregator Websites" by the experts. The findings of the study help to have a better understanding of the different referring behaviors, which form around 15% of the overall traffic of Iranian Students' News Agency (ISNA) website. They are also helpful to develop more efficient online marketing plans, business alliances, and corporate strategies.

Keywords: Referral Traffic; Data Mining; K-means Clustering; Online News.

1. Introduction

News is a rich and message-containing context which conveys answers to some fundamental questions including what, who, how, when, where, why, for whom and so on [1-4]. It also covers a conceptual hierarchy of story, event, and topic [5]. The first electronic newspaper "News Report" was established in the University of Illinois in 1974 followed by the Columbus Dispatch starting its online service in 1980. In the late 1990s thousands of online newspapers came into existence [6].

News websites are making millions of news articles available to the public through different web services these days. Most of the professional news agencies have their own web-based services; many non-professional news agencies are also participating in the process of producing and distributing online news.

Website users come from different channels including organic search services, paid search services, referral links, direct hits, links from online social media, e-mails and so on [7]. Website owners (webmasters) can understand how to address their users' needs by paying attention to their feedbacks and analyzing their behaviors via web usage mining, which refers to activities done to understand users' behaviors by tracking the users' footprints on the web. For example, a company can realize the behavior of its visitors by tracking its referral audience, who come from website A to website B, and is called the referral audience of website B. Referral audience are acquired through referral customer acquisition, a significant marketing strategy in all industries, especially online ones, and a certain part of modern Customer Engagement (CE) Cycle indicated in Fig. 1.

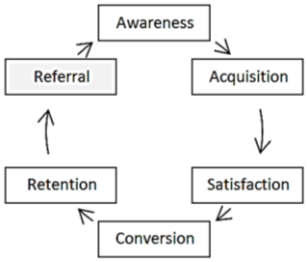


Fig. 1. Customer Engagement (CE) cycle [8]

To measure the number of acquired referral audience, one needs to measure referral traffic, a portion of website's overall traffic. In this paper we model the referral traffic of a news website via analyzing referral acquisitions and the popularity status of the referrers through data mining techniques. This approach of websites usage mining, at the best of our knowledge, has not been addressed in previous works and is used for the first time in this study.

The paper is organized as follows. In Section 2, we present a review on previous works and in Section 3

* Corresponding Author

research methodology is described. Pre-processing activities including cleaning, filtering, validation, outlier detection and normalization are described in Section 4 while Section 5 provides processing activities consisting relationship between investigation and clustering. Descriptive statics of data as well as modeling details are presented in Section 6 and conclusion and proposals for further studies are provided in Section 7.

2. Related Works

Empirical studies on online users' behavior have been performed utilizing diverse approaches ranging from descriptive social analysis and characterization to statistical analysis. Applying data mining techniques on online user-activity logs has led to valuable field literature; however, mining a website's referral traffic is a deprived category.

A common approach in web usage mining is web data clustering, which falls into sessions-based and linked-based approaches [9]. In the sessions-based approach, the one utilized in this study, the system log files of users' behaviors are retrieved, cleaned and analyzed. There are three sources of system log files including those stored on server side, on the client side and on proxy servers [10].

By investigating different time series for modeling page views, Omidvar et al. [11] propose application of regression techniques. Their research independent variables include *visitors' sources* with three options, *connection speed* with six options and *visitors type* with two options. And dependent ones include *page views of visitors' source* with three options, *page views of visitors' speed* with six options and *page views of visitors' type* with two options. The results of their study outline interesting findings. For example, search engine visitors cannot be described by single regression, while referral visitors are well-defined by linear regression. A more detailed report of their work is presented in [12].

Xu and Liu [13] cluster web users by a combination of vector analysis and K-means clustering algorithm. Their solution is independent of sessions. In another study, Xie and Phoha [14] propose a belief function based on Dempster-Shafer's theory of combining evidence to cluster web users into different profiles. Another study on clustering web usage sessions based on access patterns is performed by Fu et al. [15]. They propose a generalization based clustering method which uses an attributed-oriented induction method to reduce the dimensionality of data.

The same issue is addressed based on a-priori association rules discovery and usage-based clustering algorithm in [16]. They define two page types consisting content purpose and navigational purpose based on the amount of time spent by users. A summary of the key related works is provided in Table 1.

Table 1. Summary of key related works

Approach	Technique	Ref.
Analyzing the effectiveness of a website's analytic variables on each other	Linear regression	[11, 12]
Clustering web users	Combination of vector analysis and K-means clustering of log data	[13]
Web users profiles clustering	Belief function	[14]
Clustering web usage sessions based on access patterns	Hierarchical clustering	[15]
Web users' profile clustering	A-priori association rules	[16]

3. Research Method

The research was set to address two questions:

- Is there any significant relationship between referral traffic coming from referrer websites and their popularity status?
- How can online referral traffic be modeled through data mining techniques?

Research steps were designed according to [17] as shown in Fig. 2. It starts from data gathering and goes through data preprocessing and processing steps. In the last step the defined model is interpreted by field experts.

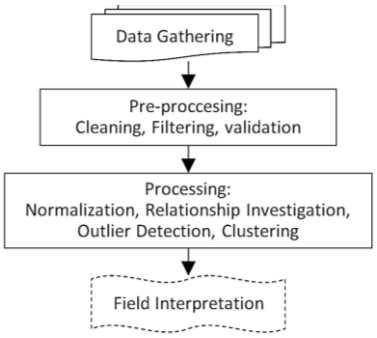


Fig. 2. The research steps

Considering the timing of the study, after running a pilot test on data of a week, the study was conducted on data of 1 month from 2015.01.17 to 2015.02.16.

3.1 Sample Population and Research Variables

The main website of Iranian Students' News Agency (ISNA) was selected for the study. ISNA is a non-governmental Iranian pure online news agency, launched in 1999. The agency provides 44 different online news services and in a month it has an average of 14,717,950 audiences, of whom around 15% are referrals.

There are two research variables in the study: *Referral Traffic* and *Referrer's Alexa Rank*.

- *Referral Traffic*

Referral traffic is defined as "the segment of traffic that arrives at a website through a link on another website" [18]. There are different traffic metrics with slight differences including "hit", "session", "page view" and "page visit". A "session" is a delimited number of a users' explicit web requests across one or more web

servers [19], and since the research does not follow click streams inside the website and only referral acquisitions are counted, the "session" was opted in this study. Therefore, referral traffic in this study is the number of visit sessions made on the main ISNA domain from referrer websites.

Referral traffic can be studied through HTTP protocol, which conveys a field named "HTTP header field". Checking it, the new webpage can understand where on the web the user's request has been sent.

Most of the times, referral traffic happens when there is a link to the second webpage on the first webpage. The user hits the link and the browser sends a request to the server of the second webpage having the HTTP header field embedded. To gather this variable, Google Analytics (GA), one of the Google's free web services used in many web mastering and monitoring tasks, was used. The validity of using GA data as input of data mining processes has been confirmed through experimental studies in [20, 21].

Although GA is one of the best and the most widely used website statistics service [22], the most source of probable inaccuracy in the research results could be originated from inaccuracies in GA statistics. It is a known problem which is discussed in several previous works. It uses users' IP to detect their locations. This may have a great impact on the validity of location-based reports in some areas. For this study, since surfing the web through proxy servers is a common manner for Iranian Internet users, to prevent any ambiguities, none of the GA's location-based information and services were used. All other scripts were coded in Python 2.7 as a strong tool for web mining and data scraping with lots of well-defined libraries including BeautifulSoup, urllib, urllib2 and lxml.

- *Referrer's Alexa Rank*

A global scoring system was needed to estimate the popularity state of referrer websites in the study. Two free public available scoring systems were Google PageRank and Alexa Rank. Google PageRank consisted of 10 score levels from 0 to 10, which did not make enough discrimination for our purpose. However, Alexa Rank assigning scores from 1 -the rank of Google.com itself- to multi-millions, was opted for this study. The other reason for choosing Alexa Rank over Google PageRank was that there were legal automatic mass extraction methods available for Alexa Rank [23], while automatic Google PageRank extracting was officially illegal and involved hacking techniques [24].

Alexa is a well-known American company in global ranking of the websites. It performs ranking based on traffic data collected from more than 30 million websites [25]. The validity of this score as a suitable general popularity metric has been proved in several previous studies such as [26, 27].

4. Pre-Processing

The frequency of total referral traffic by days is provided in Fig. 3. The chart depicts a periodic pattern during the days of the week. The most referral acquisitions happen on Sundays and the least ones on Thursdays and Fridays. This pattern is in line with the ISNA's overall page visit pattern that is fewer during Iranian weekend days.

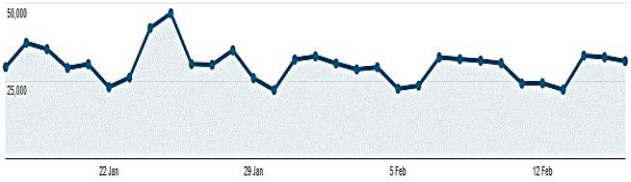


Fig. 3. The number of referral acquisitions by days (Source: GA)

After collecting data, the first step of pre-processing in any data mining project is data cleaning. In this study, to clean the data, first an integration phase was done to transform the format of collected referral URLs into the plain format needed for the next processes. The noisy data of URLs along with repeated data were then detected and removed. With regard to the research scope, sub domains were also ignored and their values were added to their related domains. A validation test was performed to omit all broken and problematic links from the list. Finally, a clean list of 6,572 referrers in a month was reached. Among them there were some nationally access blocked links which had to be recognized to be treated differently in the next steps. The statistics of collected referral URLs for one month are presented in Fig.4.

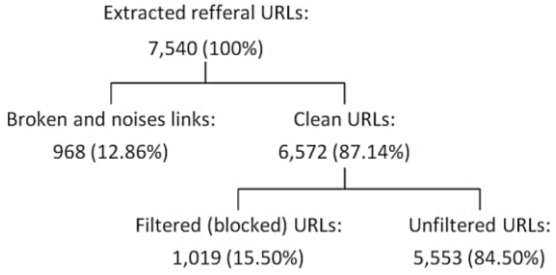


Fig. 4. Statistics of collected referral URLs

The examination shows that 1,019 sessions– around 15.50%– of the referral requests are originated from nationally access blocked domains, among which Facebook.com is the greatest. This is in line with statistics from [28], which reports Facebook.com as the number one social referral traffic source for news websites in 2013. It is a remarkable result at the same time as ISNA does not have any official account in blocked online social media.

4.1 Outlier Detection

Outliers are data that appear far away from others and diverge from the overall pattern in the dataset. The possible negative effects of outliers on data analysis make

their detection an essential part in analyzing and modeling processes. Moreover, outlier detection can give a better insight concerning the dataset.

Fig. 5 depicts variables' box plots, a graphical tool to understand the behavior of data toward its distribution.

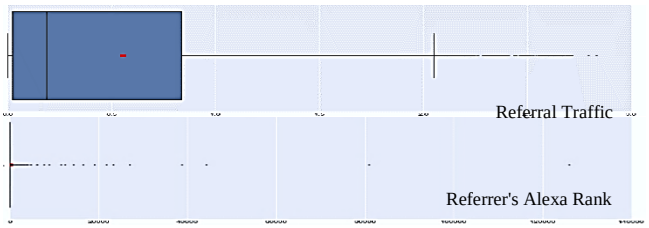


Fig. 5. Variables' box plots

As the outlier detection process strongly depends on data and research context, the Referrer's Alexa Rank outlier detection had to be rejected in the interpreting phase since being in the extremes (the least or the most) does not reflect any significant information, except the weakness or the strength of the referrer website from popularity point of view.

Referral Traffic box plot indicates four unusual behaviors shown as four diverged nodes. They are far from the others and can be interpreted as outliers. These nodes from right to left are presented in Table 2.

Table 2. Referral Traffic Outlier detection results

Website	Referral Traffic	Referrer's Alexa Rank
www.parseek.com	126039	14562
www.khabarfarsi.com	80915	2687z
www.khabarpu.com	44339	3043
www.trakhtorlink.com	38802	64690

Checking result with the field experts reveals that the first three websites are automatic news aggregator web services and the last one (trakhtorlink.com) is a sport specific news service. The surprising point which can be concluded from Table 2 is that none of the referral traffic outliers have significant status in Alexa Rank criterion.

4.2 Normalization

As many distance-based and density-based clustering algorithms are isotropic in all directions, applying them on unequal ranges of variances is equivalent to putting more weights on some variables. In this regard normalization should be considered essential in pre-processing phase to improve the accuracy of clustering model.

One of the most popular normalization methods is Min-max normalization which applies a linear transformation on the original data [29] and gives the same importance to all the variables. In this study, normalization shifted the data ranges as follows:

Referral Traffic: (1, 126,039) → *(0, 1)*

Referrer's Alexa Rank: (1, 28,335,166) → *(0, 1)*

5. Processing

Processing methods involve relationship investigation and data modeling (clustering) which are described in this section.

5.1 Relationship Investigation

The first step in investigating the relationship between two variables is the examination of their scatter plot. Scatter plot is one of the most effective graphical tools for determining if there are any relationship, patterns, or trends between two numerical attributes [30]. As Fig. 6 depicts, reviewing the overall pattern of scatter plotted data does not show any meaningful relationship between the two research variables.

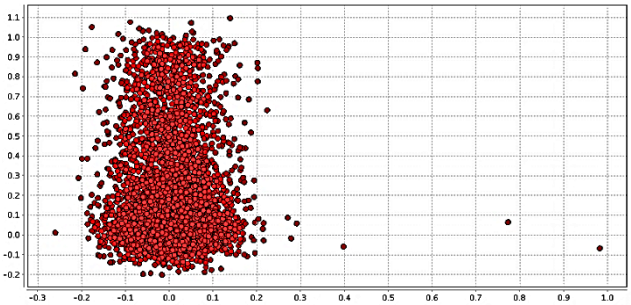


Fig. 6. Scatter plotted data with added jitter

Correlation analysis is a statistical technique to investigate whether two pairs of quantifiable variables are related. Correlation results between the two research variables are reported in Table 3.

Table 3. Correlation analysis results

	Referral Traffic	Referrer's Alexa Rank
Referral Traffic	1	-0.03479
Referrer's Alexa Rank	-0.03479	1

As it is shown in Table 3, the correlation coefficient is near to zero that confirms there is no significant relationship between the two research variables.

In other words, the relationship investigation on research data demonstrates that a website's Alexa Rank is not related to its Referring Traffic for the understudy website. This may be interpreted as the generality of Alexa ranking system. Or, the gap between popularity trends and news seeking manners.

5.2 Clustering

Clustering refers to a variety of methods aiming to group data in such a way that data in the same groups – called clusters- share more similarities (upon some factors) than to those in other groups. Clustering models fall into several approaches including connectivity-based, centroid-based, distribution-based and density-based. There is no predefined rule to choose a model, but selecting an appropriate one depends on the dataset and the aims of the study.

In this study after applying several clustering algorithms, evaluation results showed that distance-based algorithms were best fitted to the dataset. Therefore, based on experimental analysis and in line with previous

works, we opted for the K-means Squared Euclidean Distance clustering algorithm.

K-means clustering is one of the most popular unsupervised clustering algorithms. It stores K centroids and then forms clusters. Each point is considered to belong to a particular cluster if it is closer to that cluster's centroid than any others.

6. Findings

6.1 Descriptive Statistics

Descriptive statistics of the two research variables are provided in Table 4.

Table 4. Statistics report of the dataset

Item	Referral Traffic	Referrer's Alexa Rank
Mean	122.2039	5100651
Standard Error	30.3745	96454.65
Median	3	1401456
Mode	1	716577
Standard Deviation	2263.053	7186357
Sample Variance	5121409	5.16E+13
Kurtosis	1915.734	1.398512
Skewness	40.33055	1.571618
Range	123169	28335165
Minimum	1	1
Maximum	123170	28335166
Sum	678354	2.83E+10
Confidence Level (95.0%)	59.54591	189088.9

Great Standard Error (SE) and Standard Deviation (SD) values show inaccuracy of dataset mean due to their diverse distribution. Values of skewness above plus one for both variables is a sign of non-normality, which is confirmed through variables boxplots in Fig. 5, too.

In 86 detected referring Top-level Domains (TLDs), the com with 52% and ir with 41% are the most frequent referrers. Top ten most frequent referring TLDs are provided in Fig. 7.

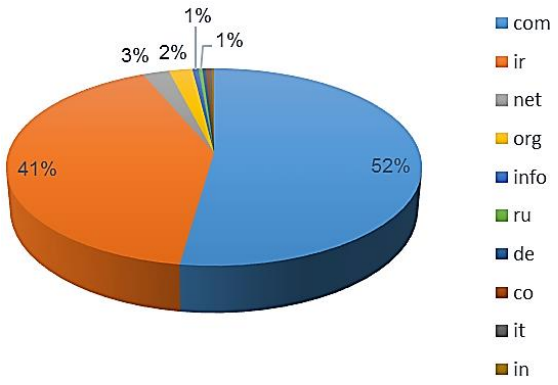


Fig. 7. Top ten most frequent referring TLDs

Since com is the most popular TLD in the world, it's the most frequent in our dataset, too. And as ISNA is an Iranian news website, its second most referral traffic comes from ir TLDs which is the national Iranian TLD code. Other national TLD codes in the top ten list are ru (Russia), de (Germany), co (Colombia), it (Italy) and in (India).

From the Referrals Traffic point of view, com and ir are again the biggest referrers to ISNA website. Top ten biggest referrer TLDs are shown in Fig. 8.

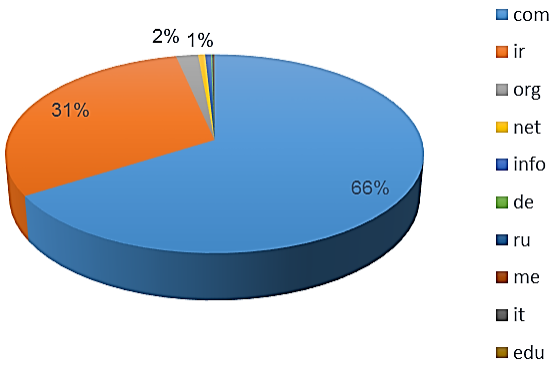


Fig. 8. Top ten biggest referrer TLDs

The two Figures depict that com TLD, which covers 52% of referrers, brings the 66% of referral traffic to the ISNA. And ir TLD, which covers 41% of referrers, brings 31% of Referral Traffic to the ISNA.

There exist some other TLDs such as me (Montenegro) which are referring a great monthly traffic to ISNA; however, they are not listed in the most frequent referrer TLDs.

As around 15% of ISNA referral traffic requests are initiated from nationally access blocked websites, the dispersion of TLDs reported in Figures 7 and 8 can be interpreted to the users equipped with VPNs or other filtering bypassing tools.

6.2 Modeling

For model selection, three main clustering algorithms were applied on data with the following overall results (Table 5).

Table 5. Applied clustering models and overall achieved results

Approach	Model	Result
Density-based	DBSCAN	The model was incapable of detecting distinct clusters and resulted in only one big cluster.
Connectivity-based	Hierarchical	The model resulted in 10019 clusters, which did not convey enough information for the purpose of our modeling. It also took a great amount of time in comparison with other algorithms.
Centroid-based	K-means	The model led to predefined number of clusters with different performances due to parameter settings.

As Table 5 shows, one of the most common clustering algorithms DBSCAN failed to detect more than one cluster. It may be because of the condense density of the dataset in two-dimension space, Although preprocessing activities proved nonuniformity distribution of variables. Hierarchical modeling also let to tremendous number of clusters in several hierarchies. Consequently, K-means

clustering was opted. The next task was to determine the right value for K, the number of clusters.

For model evaluation against different values of K, as an unsupervised technique was used, two indices of the average within centroid distance and Davies Bouldin Index (DBI) were examined. It should be mentioned that values are multiplied by -1 for the optimization propose, however it does not effect on the selection process.

Results from the average within centroid distance calculation for different number of clusters in K-means clustering model from 2 to 10 are provided in Fig. 9. It is calculated by averaging the distance between the centroid and all examples of a cluster. The average within centroid distance is a measure of clusters compactness and reduces as the clusters become more compact.

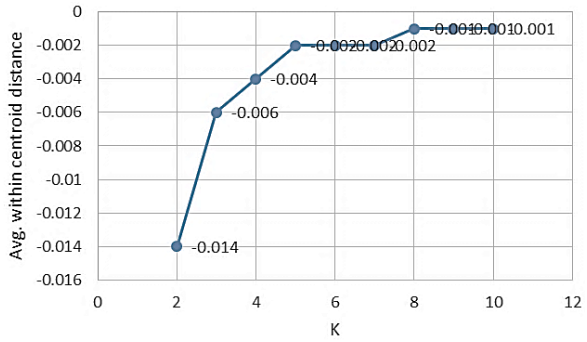


Fig. 9. The Average within centroid distance for different values of K (the number of clusters)

Fig. 9 reveals that the quality of modeling practice is in its best for Ks equal to 2, 3 and 4. It also decreases as K value increases.

DBI is another clustering quality criterion: the lower the value of DBI, the better the quality of clustering. Fig. 10 shows DBI values regarding to K variation from 2 to 10.

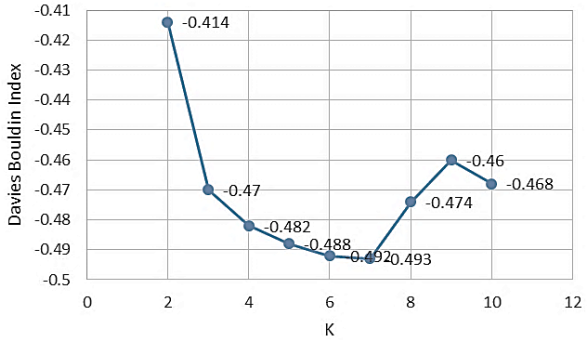


Fig. 10. DBI calculated for different values of K (the number of clusters)

The figure shows that DBI decreases dramatically when K goes from 2 to 3. It then decreases smoothly as K moves from 3 to 7. The curve rises when K goes from 7 to 9.

As a trade-off between both quality indices, the best clustering model could be selected on K=3, where the average within centroid distance and DBI are both presenting a relatively good quality simultaneously.

Finally, the performance evaluation results of clustering model for K=3 are provided in Table 6.

Table 6. Performance evaluation results for K=3

Criterion	Result
Avg. within centroid distance	-0.006
Avg. within centroid distance_cluster_0	-0.006
Avg. within centroid distance_cluster_1	-0.011
Avg. within centroid distance_cluster_2	-0.003
DBI	-0.470

Table 6 shows that the most compact cluster is cluster_2, follows by cluster_0.

Visual clustering results are provided in Figure 11. The vertical axis represents referrer's Alexa Ranks, while the horizontal axis is for the referral traffic during one month.

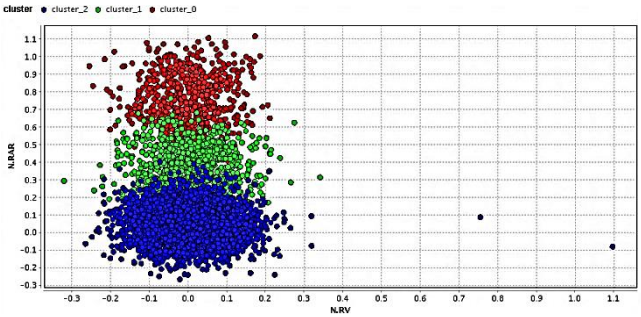


Fig. 11. Detected clusters (with added jitter) for K-means Squared Euclidean Distance algorithm (K=3)

Cluster modeling leads to three separate clusters with the characteristics mentioned in Table 7.

Table 7. Generated cluster modeling

Name of Cluster	Color	No. of Items
Cluster_0	Red	670
Cluster_1	Green	996
Cluster_2	Blue	3,885

The Interpretation of results by field experts is one of the most important steps in data mining projects. This task connects the results of a research to the real life situations. Some notable points about research results are provided below.

Referring to the dataset, it can be observed that most of the cluster_0 members are ir TLDs or com TLDs which are hosted in Iran.

Moreover, members of cluster_2 are mostly automatic news aggregator websites. Automatic news aggregators, also called feed aggregators, are types of software that collect news articles and other textual materials automatically and repeatedly from different online sources. They usually set the data in some predefined categories and deliver them to the end users in the form of comprehensive lists of news. The number of sources varies from hundreds to thousands. The different nature of such websites makes them important to detect them effectively in the process of traffic analysis. From this point of view, identifying them as a distinct cluster is a promising achievement of the clustering process.

In Alexa Rank system, a website with smaller rank has more traffic and generally seems most popular. Therefore, as the Alexa Rank increases, the popularity of the website decreases as it earns less monthly traffic. In this study, the most valuable referrer group would be cluster_2, which in

the same time is the biggest cluster. In the same vein, the less important group of referrer is one with greater Alexa Ranks, cluster_0, which is also the smallest cluster with 670 members.

7. Conclusion and Further Studies

This research was the first step to study online news content reposting status and its penetration on the web. In this study, the goal was to understand the referral audience traffic for the website. Two metric were needed. The first metric, referral traffic, was opted to investigate the number of referrals to collect the number of referral sessions. The second metric, the Alexa Rank, was chosen to estimate the extent of importance or popularity of the referrer websites.

Between several common clustering models the K-means algorithm led to better results. And by examining several values, K=3 led to the best initial value for clustering.

Some new knowledge was achieved by utilizing data mining techniques. These findings lead to a better understanding about the origin and the nature of a website's referral traffic. The findings of the study can be used for future online marketing activities, information quality improvements, and the development of better and more efficient business alliances and strategies.

Analyzing the reposting status of referrer websites and investigating the relation between reposting and referring manners are proposed as the next steps of this research.

Acknowledgements

We hereby extend our sincere appreciation to ISNA for assisting us during the process of accessing the data of the study.

References

[1] B. Kovach and T. Rosenstiel, The Elements of Journalism: What Newspeople Should Know and the Public Should Expect, Completely Updated and Revised: Random House Digital, Inc., 2007.

[2] J. B. Singer, "Five Ws and an H: Digital challenges in newspaper newsrooms and boardrooms," The International Journal on Media Management, vol. 10, pp. 122-129, 2008.

[3] D. E. Appelt, "Introduction to information extraction," Ai Communications, vol. 12, pp. 161-172, 1999.

[4] M. Karlsson and J. Strömbäck, "Freezing the flow of online news: Exploring approaches to the study of the liquidity of online news," Journalism Studies, vol. 11, pp. 2-19, 2010.

[5] X. Dai and Y. Sun, "Event identification within news topics," in IEEE International Conference onIntelligent Computing and Integrated Systems (ICISS), 2010, pp. 498-502.

[6] D. Shedden. (2014). New Media Timeline (1969-2010). Available: http://www.poynter.org/latest-news/business-news/tracker/28803/new-media-timeline/

[7] R. Watkins. (2014, 2/18/2015). Understanding Google Analytics: What is Acquisition? Available: http://southernweb.com/2014/01/understanding-google-analytics-acquisition/

[8] K. Ertell, "The Missing Links in the Customer Engagement Cycle," 2010.

[9] A. Vakali, J. Pokorný, and T. Dalamagas, "An overview of web data clustering practices," in Current Trends in Database Technology-EDBT 2004 Workshops, 2005, pp. 597-606.

[10] R. Iváncsy and I. Vajk, "Frequent pattern mining in web log data," Acta Polytechnica Hungarica, vol. 3, pp. 77-90, 2006.

[11] M. A. Omidvar, V. R. Mirabi, and N. Shokry, "Time Series modeling of visitors' type on web analytics."

[12] M. A. Omidvar, V. R. Mirabi, and N. Shokry, "Analyzing the impact of visitors on page views with Google analytics," arXiv preprint arXiv:1102.0735, 2011.

[13] l. H. Xu and H. Liu, "Web User Clustering Analysis based on K-Means Algorithm," International Conference on Information, Networking and Automation (ICINA), vol. 2, pp. 6-9, 2010.

[14] Y. Xie and V. V. Phoha, "Web user clustering from access log using belief function," in Proceedings of the 1st international conference on Knowledge capture, 2001, pp. 202-208.

[15] Y. Fu, K. Sandhu, and M.-Y. Shih, "A generalization-based approach to clustering of web usage sessions," in Web Usage Analysis and User Profiling, ed: Springer, 2000, pp. 21-38.

[16] B. Mobasher, R. Cooley, and J. Srivastava, "Creating adaptive web sites through usage-based clustering of URLs," in Knowledge and Data Engineering Exchange, 1999.(KDEX'99) Proceedings. 1999 Workshop on, 1999, pp. 19-25.

[17] M. K. Jiawei Han, "Data mining: concepts and techniques," in TheMorgan Kaufmann Series in DataManagement Systems, ed: Elsevier, 2006, pp. 1-55860. p7.

[18] B. J. Davies and N. F. Glasser, "Analysis of www. AntarcticGlaciers. org as a tool for online science communication," Journal of Glaciology, vol. 60, pp. 399-406, 2014.

[19] Google Support Page. (2/19/2015). How a session is defined in Analytics. Available: https://support.google.com/analytics/answer/2731565?hl=en

[20] T. Raudenbusch, "Can Google Analytics be a reasonable alternative to collect data for Web Usage Mining?," E-Business Applications, 2013.

[21] P. Sawant and R. Kulkarni, "A Knowledge Based Methodology To Understand The User Browsing Behavior For Quality Measurement Of The Websites Using Web Usage Mining," International Journal Of Engineering And Computer Science, vol. 2, pp. 1522-1538, 2013.

[22] W3Techs. (2/13/2015). Usage of traffic analysis tools for websites. Available: http://w3techs.com/technologies/overview/traffic_analysis/all

[23] Alexa. (2015, 5/29/2015). Alexa Internet Terms of Use Agreement. Available: http://www.alexa.com/help/terms

[24] Google. (2014, 5/29/2015). Google Terms of Service, Part:Using our Services. Available: http://www.google.com/intl/en/policies/terms/

[25] Alexa. (2/19/2015). About Us. Available: http://www.alexa.com/about
[26] A. Olteanu, S. Peshterliev, X. Liu, and K. Aberer, "Web credibility: Features exploration and credibility prediction," in Advances in Information Retrieval, ed: Springer, 2013, pp. 557-568.
[27] A. Thakur, A. Sangal, and H. Bindra, "Quantitative measurement and comparison of effects of various search engine optimization parameters on Alexa Traffic Rank," International Journal of Computer Applications, vol. 26, pp. 15-23, 2011.
[28] J. Kosur. (2013, 7/2/2015). Facebook rules social media referral traffic for mainstream media. Available: http://socialnewsdaily.com/15911/facebook-rules-social-media-referral-traffic-for-mainstream-media/
[29] N. Karthikeyani Visalakshi and K. Thangavel, "Impact of Normalization in Distributed K-Means Clustering," International Journal of Soft Computing, vol. 4, pp. 168-172, 2009.
[30] M. K. Jiawei Han, "Data mining: concepts and techniques," in TheMorgan Kaufmann Series in DataManagement Systems, ed: Elsevier, 2006, pp. 1-55860. p60.

Roya Hassanian-Esfahani received her B.Sc in Electronic Engineering and M.Sc in Information Technology Engineering from Shiraz University, Iran. As a current Ph.D Candidate in Research Institute for Information and Communication Technology, ACECR, Tehran, Iran, her focus is on news mining. Hassanian-esfahani's dominant research interests include web information quality assessments, data mining and web mining.

Mohammad Javad Kargar received B.Sc and M.Sc degree in computer engineering from IAU- University of Science and Research, and Ph.D Degree in Information Technology and Multimedia system from University Putra Malaysia in 2008. He is currently an assistant professor at the department of computer engineering in University of Science and Culture, Tehran, Iran. He is founder of the International Conference on Web Research Conference series. His research interests include web and Information Quality, Web and Data Mining and Distributed systems

A Novel Ultra-Broad Band, High Gain, and Low Noise Distributed Amplifier Using Modified Regulated Cascode Configuration (MRGC) Gain-Cell

Zainab Baharvand*
Department of Electrical and Computer Engineering, Graduate University of Advanced Technology, Kerman, Iran
z.baharvand@student.kgut.ac.ir
Ahmad Hakimi
Department of Engineering, Shahid Bahonar University of Kerman, Kerman, Iran
hakimi@uk.ac.ir

Received: 22/Jun/2015 Revised: 08/Feb/2016 Accepted: 13/Feb/2016

Abstract

In this paper, an ultra-broad bandwidth, low noise, and high gain-flatness CMOS distributed amplifier (CMOS-DA) based on a novel gain-cell is presented. The new gain-cell that enhances the output impedance as a result the gain substantially over conventional RGC is the improved version of Regulated Cascode Configuration (RGC). The new gain-cell based CMOS-DA is analyzed and simulated in the standard 0.13 μm-CMOS technology. The simulated results of the proposed CMOS-DA are included 14.2 dB average power gain with less than ± 0.5 dB fluctuations over the 3-dB bandwidth of 23 GHz while the simulated input and output return losses (S11 and S22) are less than -10 dB. The IIP3 and input referred 1-dB compression point are simulated at 15 GHz and achieved +8 dBm and -6.34 dBm, respectively. The average noise figure (NF) in the entire interest band has a low value of 3.65 dB, and the DC power dissipation is only 45.63 mW. The CMOS-DA is powered by 0.9 V supply voltage. Additionally, the effect of parameters variation on performance specifications of the proposed design is simulated by Monte Carlo simulations to ensure that the desired accuracy is yielded.

Keywords: Ultra-broad Band; CMOS Distributed Amplifier; Modified Regulated Cascode Configuration (MRGC); Low Noise.

1. Introduction

Applications of the broadband circuits in various fields such as high-rate links, high-resolution radar, imaging systems, electronic warfare, and wide band commercial or military radio systems demand a broadband amplifier as an indispensable building block at the both transmitting and receiving ends. Bandwidth increasingly becomes a controlling factor in radio frequency (RF) circuit design. Distributed amplifier (DA) is highly interested component at the high-speed amplification applications as a result of its inherent broad bandwidth, good linearity, and low sensitivity to process variations [1-4]. Distributed amplification is a method to absorb the parasitic capacitances of transistors which are main factor to restrict the bandwidth.

A simplified schematic of customary DA is shown in Fig. 1. As it can be seen, it has a pair artificial transmission lines (TL) and several active devices. The artificial gate and drain TLs essentially are constructed of series on-chip inductors in conjunction with shunt parasitic gate and drain capacitances (C_{gs},C_{ds}) of MOS transistors. The gate line is used to travel down the input signal to excite each of the active devices, in turn. Similarly, the drain line is utilized to get the desire output peak pulse through summing each of these pulses together, after amplification by active devices.

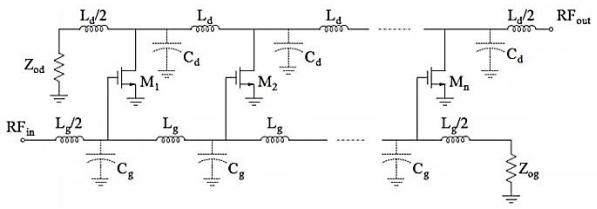


Fig. 1. The simplified circuit schematic of a conventional DA [4]

Previously reported DAs are designed into several classes, to date. DAs based on cascaded gain-cell increase the amplifier's gain while operating at low voltage and low power conditions. However, this group suffers from poor closed loop stability and lack of an ultra-broad band response, due to the incurrence of multiple poles by cascaded stages [5-8]. Another class adopts a cascode cell that is more desirable to decrease the Miller effect and to improve the reverse isolation. The voltage headroom's limitation of cascode structures makes it difficult to meet low power and high gain requirements, simultaneously. Hence, the DAs based on this group of gain-cells are unsuited to the low supply voltage applications [9]. The third class introduces two-dimensional DAs, such as cascaded single-stage distributed amplifier (CSSDA) [10], matrix DA [11], combination of the conventional

* Corresponding Author

distributed amplifier (CDA) and CSSDA [2], cascaded multi-stage distributed amplifier (CMSDA) [12, 13], and DA with cascaded gain stages [14]. This class applies the multiplicative gain mechanism to meet the high gain performance. Although this group has higher gain-bandwidth performance than other, their huge chip size and power dissipation aren't desire.

In this study, a CMOS-DA based on a new gain-cell configuration is designed to overcome the shortcomings of the previous DAs. It introduces one structure that can satisfy the combination of the most important design targets in DA designing, in terms of high power gain, broad bandwidth, low NF, and reasonable DC power consumption, simultaneously.

Additionally utilizing of Monte Carlo (MC) simulations, the parameters variation's effect on the performance specifications of the proposed design is investigated. The MC simulation results confirm that in spite of considering the tolerance effects particularly fluctuations in MOS parameters, which are unavoidable in practice, and supply voltage variations in designing the proposed DA can meet to its expected specifications. Following this introduction, Section 2 presents the analysis, design and characterization of the proposed DA. Section 3, demonstrates the simulation results. Finally, section 4 draws conclusion of the work.

2. Basic Principle of Proposed DA

Fig. 2 shows the circuit topology of the proposed CMOS-DA which is composed of three stages Modified Regulated Cascode Configuration (MRGC) gain-cell.

A summary of systematic design dividing in four steps is explained first, and these steps are then applied to design the proposed DA, as detailed in following.

Step 1) choose the optimum number of stages

Since both gate and drain artificial lines have limited quality factors, they are lossy in practice. The range of optimum number of stages reported for various DAs is between three and five. Therefore in this design to meet low power and high gain requirements, number of stages is calculated in optimum mode to be three stages.

Step 2) propose the suitable gain-cell

A new gain-cell configuration is designed to overcome the shortcomings of the previous DAs. New gain-cell based CMOS-DA alleviates the defects of the CMOS-DAs based on cascaded gain-cells, in the point of restricted bandwidth. Also, it overcomes to defects of two dimensional DAs including high power dissipation and big chip area. These advantages obtain without any limitation in the power-supply voltage and signal-swing requirements that are unavoidable in CMOS- DAs based on cascode gain-cells [6,7,9,15]. The design methodology describes in the sub-section 2-1, in detailed.

Step 3) choose the transistor aspect ratio

To avoid the bandwidth deterioration due to the pole associated to the internal node of the proposed gain-cell whose value is $P_{int,node} = \frac{g_{m2}}{(C_{i2}+C_{o1})}\left(\text{also}, P_{int,node} = \frac{g_{m3}}{(C_{i3}+C_{o4})}\right)$, the $M_2\,(\text{also}, M_3)$ transistor's transconductance value must be increased. This demands increasing the $M_2\,(\text{also}, M_3)$ transistor's width $(W_2(\text{also}, W_3))$. On the other hand, increasing $W_2(\text{also}, W_3)$ leads to increase the input capacitance $(C_{i2}(\text{also}, C_{i3}))$ of $M_2\,(\text{also}, M_3)$. The competing requirement for $W_2(\text{also}, W_3)$ in the numerator and denominator of $P_{int,node}$ implies that an optimal width for $M_2\,(\text{also}, M_3)$ exists. A bandwidth-enhancing inductor $L_n\,(\text{also}, L_m)$ is added to the source of $M_2\,(\text{also}, M_3)$ which other benefit of it is to improve the $M_2\,(\text{also}, M_3)$ transistor's transconductance. This effect arises from the fact that employing $L_n\,(\text{also}, L_m)$ enables us to raise the transistor aspect ratio and earns the objective transconductance while the similar bandwidth is obtained, at the same time. From simulation and bandwidth compensation method the optimal value for $W_2(\text{also}, W_3)$ was found to be larger than $W_1(\text{also}, W_4)$.

Step 4) choose the size L_n, L_m

For the selected transistor aspect ratio, the bandwidth-enhancing inductor $L_n\,(\text{also}, L_m)$ through simulation is adjusted to be around 0.64 nH to maximize the flat bandwidth of the propose gain-cell.

The next sub-section elaborates the effect of MRGC upon increasing the output impedance as a result increasing the gain. Also, it shows how large effect is obtained over conventional RGC.

2.1 Modified Regulated Cascode Configuration (MRGC)

Negative feedback is a known method which is greatly utilized in electronic systems design, particularly in amplification applications. There are many advantages achieved with a suitable introduction of negative feedback, in terms of: bandwidth enhancement, modified output impedance and unconditional stability [16]. This subsection presents a step-by-step approach to reach the MRGC cell that enhances RGC cell's gain.

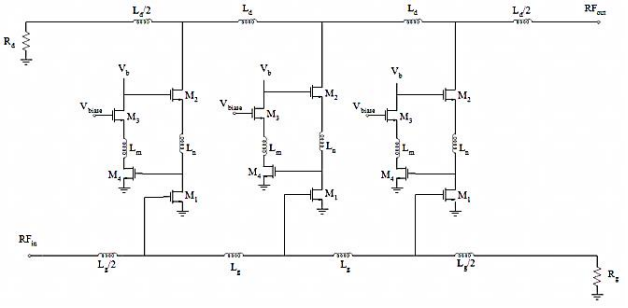


Fig. 2. The circuit topology of the proposed CMOS-DA

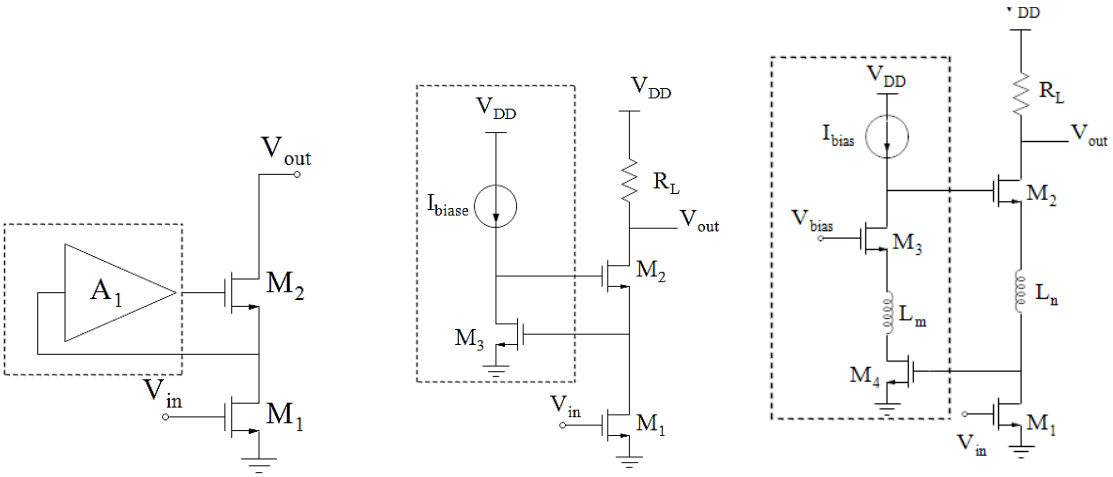


Fig. 3. a) Gain boosting in the cascode stage [17], and b) the conventional Regulated Cascode Configuration [17], and c) the Modified Regulated Cascode Configuration

Fig. 3 (a) shows the block diagram of RGC cell. In the point of calculating the R_{out} , M_1 device acts as a degeneration resistor. This device senses the output current and modulates it to a voltage signal. Now, the small signal voltage appears on the drain of M_1 device is proportional to the output current. This result suggests that the mentioned voltage can be subtracted from the gate voltage of M_2 transistor to insert this device at the current-voltage negative feedback. In fact, the A_1 amplifier forces the drain voltage of M_1 to be equal to the gate voltage of M_2. In this way, A_1 implements the negative feedback loop. The current-voltage negative feedback modifies the output

impedance as a result enhances the gain performance without stacking more cascode elements on top of M_2 transistor. Also, it increases the bandwidth and improves the stability behavior of the amplifier [16, 17]. One can prove that the voltage gain (A_v) of the RGC amplifier is given by Eq. (1) [17]:

$$A_v = -g_{m1}R_{out} \;\&\; R_{out} \approx (A_{vA1})g_{m2}r_{o1}r_{o2} \qquad (1)$$

As it shown in Fig. 3 (b), in the original case of RGC, the A_1 amplifier implements by CS cell. The output resistance

(R_{out}) and the voltage gain (A_v) of the conventional RGC amplifier are given by Eqs. (2), (3) [17]:

$$R_{out} = g_{m2}g_{m3}r_{o1}r_{o2}r_{o3} \qquad (2)$$

$$A_v = -g_{m1}g_{m2}g_{m3}r_{o1}r_{o2}r_{o3} \qquad (3)$$

Where $g_{m1,2,3}$ and $r_{o1,2,3}$ are the transconductance and the output resistances of the $M_{1,2,3}$ transistors, respectively. From the Eq. (1), it can be seen that one can increase the amount of R_{out} by increasing the value of A_{vA1} to achieve higher gain. This effect was used in designing of the proposed MRGC cell towards increasing output impedance as a result gain to that of the conventional RGC. MRGC replaces the CS cell with cascode amplifier in negative feedback loop as shown in Fig. 3(c), so it boosts the value of A_{vA1} . As a consequence, the output impedance significantly boosts over conventional RGC. It obtains in the fact that cascode cell's gain is higher than CS cell's gain. Hence, the higher gain performance of the proposed MRGC is guaranteed. Cascode cell not only illustrates higher gain to that of the CS cell, but also presents excellent reverse isolation. Also, the broader bandwidth operation can be obtained tanks to remove the Miler effect of the CS amplifier in the traditional RGC by this replacing. The output resistance (R_{out}) and the voltage gain (A_v) of the proposed MRGC amplifier can be derived to prove that the further gain can be achieved in comparing to the simple RGC:

$$R_{out} = g_{m2}g_{m3}\,g_{m4}r_{o1}r_{o2}r_{o3}r_{o4} \qquad (4)$$

$$A_v \approx g_{m1}g_{m2}g_{m3}\,g_{m4}r_{o1}r_{o2}r_{o3}r_{o4} \qquad (5)$$

As it is evident of the above equations the small-signal voltage gain of the proposed MRGC amplifier is similar to the gain of traditional quadruplet cascode amplifier while the limiting power-supply voltage and signal-swing requirements that are unavoidable in the conventional quadruplet cascode amplifier are removed. Note the parasitic capacitances in this theoretical analysis are ignored for simplicity, they are taken into account in high frequency transconductance ($G_{mt}(s)$) analysis of the proposed CMOS-DA.

To suppress the cascode cells dominant pole at higher frequencies, two bandwidth-enhancing inductors (i.e. L_m, L_n in Fig. 3(c)) are added to the proposed gain-cell. This compensation method results in high-frequency gain peaking at the drain of $M_{1,4}$ if bandwidth extension without power consumption penalty is desired [14].

2.2 High Frequency Transconductance Analysis

Fig. 4 shows MRGC gain-cell and its high frequency small-signal equivalent which is used to calculate the high frequency transconductance $G_{mt}(s)$. Calculating of $G_{mt}(s)$ is necessary because according to the CS based DA's power gain and voltage gain formulas with assuming lossless TL, $G_{mt}(s)$ is only unknown factor to

calculate the DA's power gain and voltage gain for every DA's architecture which is shaped based on a new gain cell [18]. $C'_{GS1,2}$,$C'_{DB1,2}$ are the equivalent parasitic gate-to-source, and drain-to-bulk capacitances of the $M_{1,2}$ transistors in MRGC cell, respectively. In this analysis for simplicity, it is supposed that C_{GDi} from $M_{1,2}$ devices combines with its C_{GSi} and C_{DBi} as a result of utilizing the Miller effect to shape $C'_{GS1,2}$, $C'_{DB1,2}$. Also, $g_{m1,2,3,4}$ and $r_{o1,2,3,4}$ are the transconductance and the output resistances of the MOS devices. Note that $r_{o1,2,3,4}$ are relatively large, so they can be ignored. $G_{mt}(s)$ of the MRGC cell can be derived as follows:

$$G_{mt}(s) = \frac{I_{out}}{V_{in}}(s) = \frac{I_{out}}{V_{ds1}}\frac{V_{ds1}}{V'_{ds1}}\frac{V'_{ds1}}{V_{in}} \quad (6)$$

Before of calculating the $G_{mt}(s)$, the gate-source voltage of M_2 (v_{gs2}) is determined. The gate voltage of $M_2$to ground is equal to the drain-source voltage of M_1 (v_{ds1}) which is amplified by the small signal voltage gain of A_1, that is named $(-a)$. Thus, v_{gs2} is calculated according to Eq. (7):

$$v_{gs2} = v_{g2} - v_{s2} = v_{g2} - v_{ds1} = -(a)v_{ds1} - v_{ds1} = -(a+1)v_{ds1} \approx -av_{ds1} \quad (7)$$

Also, the gate-source voltage of M_4 as evident from Fig. 4 (a) is equal to v_{ds1}. With a KCL at the output node:

$$\frac{I_{out}}{V_{ds1}} = -(ag_{m2} + g_{mb2}) \quad (8)$$

Also, from KVL at the pass of V'_{ds1}, L_n and V_{ds1}, the value of V_{ds1}/V'_{ds1} is given by Eq. (9):

$$\frac{V_{ds1}}{V'_{ds1}} = \frac{1}{SL_n(ag_{m2}+g_{mb2})+1} \quad (9)$$

Finally, With a KCL at the V_{ds1} node V'_{ds1}/V_{in} is given according to:

$$\frac{V'_{ds1}}{V_{in}} = \frac{-g_{m1}}{L_n(ag_{m2}+g_{mb2})+SC'_{DB1}} \quad (10)$$

As a consequence:

$$G_{mt}(s) = \frac{I_{out}}{V_{in}}(s) = (ag_{m2} + g_{mb2}).\frac{1}{SL_n(ag_{m2}+g_{mb2})+1}.\frac{g_{m1}}{L_n(ag_{m2}+g_{mb2})+SC'_{DB1}} \quad (11)$$

Where in Eq. (11), a defines as the high frequency small signal voltage gain of A_1 amplifier that is cascode amplifier in the proposed design and calculates according to Eq. (12) [19]:

$$\frac{V'_{out}}{V'_{in}} = -\frac{g_{m4}{g^{-1}}_{m3}\left(\frac{S^2}{\omega^2_{n,z}}+\frac{2\xi_z}{\omega_{n,z}}S+1\right)}{\left(\frac{S^2}{\omega^2_{n,p}}+\frac{2\xi_p}{\omega_{n,p}}S+1\right)}$$

$$= -\frac{g_{m4}{g^{-1}}_{m3}(L_m C_{i,3}S^2+g_{m3}L_m S+1)}{({g^{-1}}_{m3}C_{i,3}S+1)L_m C_{o,4}S^2+{g^{-1}}_{m3}(C_{i,3}+C_{o,4})S+1} \quad (12)$$

Where $C_{i,3,4}$, $C_{o,3,4}$ are the equivalent input and output parasitic capacitances of the $M_{3,4}$transistors in MRGC cell, respectively. Not that, for simplicity of high frequency analysis the small-signal equivalent circuit of cascode amplifier placed in a dashed box in Fig. 4 (b).

2.3 The Frequency Response of Noise

The noise characteristic of MESFET DAs has been analyzed in [20], and the analysis has been adapted for MOSFET DAs in [4]. The intrinsic noise sources of DA can be identified as noise from the source, gate-termination, and drain-termination resistors (R_s, R_g, R_d). Also, the transistors have two noise sources including drain current noise and gate-induced noise [21]. The noise figure of an N stage DA can be given as Eq. (13) [22]:

$$F = 1 + \frac{P_{ngT}+P_{ndT}+P_{no}}{P_{ns}G_F} \quad (13)$$

Where P_{ngT}, attenuating by the reverse gain of the DA, presents the output noise power contribution of the gate-termination resistor according to Eq. (14):

$$P_{ndT} = KT\Delta f G_F \left|\frac{\sin N\theta}{N\theta}\right| \quad (14)$$

P_{ndT}, calculating as Eq. (15), is the output noise power because of the drain-termination resistor.

$$P_{ndT} = KT\Delta f \quad (15)$$

P_{ns}, amplifying by the DA's forward gain G_F, refers to output noise power of source resistor. Finally, P_{no} shows the output noise power due to the noise sources of transistors in the k-th stage of an N stage DA, and it can be calculated as Eq. (16):

$$P_{no} = 4KTG_{mt}R_d\frac{\gamma}{\alpha}\Delta f\sum_{k=1}^{N}\left|\frac{\sin\frac{\theta}{2}}{\theta}+\frac{1}{4}G_{mt}R_gF_cM(k)\right|^2 + \frac{1}{4}KT{G_{mt}}^2G_u{R_g}^2R_d\Delta f\sum_{k=1}^{N}|M(k)|^2 \quad (16)$$

Where K and T are the Boltzmann constant and the absolute temperature, respectively. α is the ratio of gain cell's transconductance to zero-bias drain conductance which it is about 0.85 in deep-sub-micrometer MOSFETs. Also, $G_{mt}(s)$ refers to the high frequency transconductance of the proposed DA. Finally, G_u, F_c , $M(k)$, are as following:

$$M(k) = \left(N - k + \frac{1}{2}\right) + \frac{\sin\left(k-\frac{1}{2}\right)\theta}{\theta}e^{-j\left(k-\frac{1}{2}\right)\theta} \quad (17)$$

$$G_u = \delta\frac{\omega^2{C_{gs}}^2}{5g_{do}}(1-|c|^2) \quad (18)$$

$$F_c = j|c|\sqrt{\frac{\delta}{5\gamma}}\frac{\omega C_{gs}}{g_{do}} \quad (19)$$

It should be noted that the channel drain noise (i_{nd}) and the induced gate noise (i_{ng}) because of their same physical origin are correlated with a correlation coefficient defined as c. Therefore, i_{ng} decomposed into 2 parts which one is correlated with i_{nd} that is $i_{ngc}(= F_c i_{nd})$ and another completely uncorrelated that is $i_{ngu}\left(=\sqrt{4KTG_u\Delta f}\right)$. Also, δ and γ are the coefficients of gate noise and channel noise which based on the measured results in [17] they are about 4.1 and 2.21, respectively.

Substituting the above Eqs into Eq. (13) the total noise power of DA yields as Eq. (20):

$$F = 1 + \left|\frac{\sin N\theta}{N\theta}\right|^2 + \frac{4}{N^2 {G_{mt}}^2 R_g {}^2 R_d {}^2} + \frac{16}{N^2 G_{mt} R_g}\frac{\gamma}{\alpha}\sum_{k=1}^{N}\left|\frac{\sin\frac{\theta}{2}}{\theta} + \frac{1}{4}G_{mt}R_g F_c M(k)\right|^2 + \frac{1}{N^2}G_u R_g \sum_{k=1}^{N}|M(k)|^2 \quad (20)$$

Eq. (20) can be further simplified by assuming large values of N. In that case

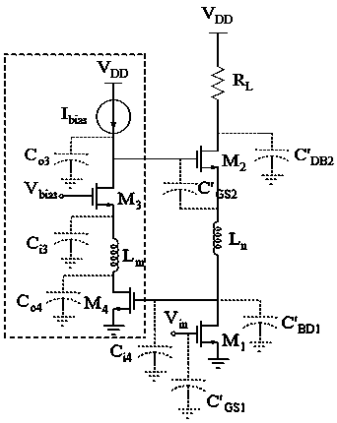
$$\sum_{k=1}^{N}|M(k)|^2 \approx \sum_{k=1}^{N}\left|\left(N - k + \frac{1}{2}\right)\right|^2 = \frac{N^3}{3} \quad (21)$$

$$\sum_{k=1}^{N}\left|\frac{\sin\frac{\theta}{2}}{\theta} + \frac{1}{4}G_{mt}R_g F_c M(k)\right|^2 \approx N\left(\frac{\sin\frac{\theta}{2}}{\theta}\right)^2 + \frac{1}{16}{G_{mt}}^2 {R_g}^2 |F_c|^2 \frac{N^3}{3} \quad (22)$$

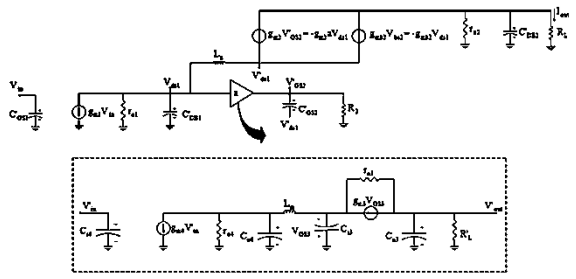
The second and third terms in Eq. (20) can be ignored for large N. Following the above supposes, the noise figure expression of DA can be given as Eq. (23):

$$F = 1 + \frac{1}{NR_g}\frac{4\gamma}{\alpha} + \frac{1}{G_{mt}}\left(\frac{\sin\frac{\theta}{2}}{\frac{\theta}{2}}\right)^2 + NR_g\frac{\alpha\delta}{3}\frac{\omega^2 {C_{gs}}^2}{5G_{mt}} \quad (23)$$

As it can be seen from Eq. (23), the noise figure expression of DA is inversely proportional to the high frequency transconductance of the proposed DA (i.e. G_{mt}), so the second and third parts of Eq. (23) can be reduced with increasing G_{mt}. Note the correlation between the gate and drain noise current sources essentially removes in Eq. (18) in the fact that $|c|^2 \ll 1$. Hence, the last term in Eq. (20) was ignored.



(a)



(b)

Fig. 4. a) Modified Regulated Cascode Configuration (MRGC) gain-cell in conjunction with parasitic capacitances, and b) the high frequency small-signal equivalent circuit of the MRGC gain-cell

3. Simulation Results

The proposed CMOS-DA is simulated using BSIM3 transistor models via Advanced Design System (ADS) simulation tool. Fig. 5 illustrates the simulated results of the power gain (S_{21}) and revers isolation (S_{12}) of the proposed CMOS-DA. As it can be seen, the flat and high gain response of 14.2 ± 0.5 dB with 3-dB bandwidth of 23 GHz is achieved. In addition, the revers isolation between input and output terminals is better than – 20 dB. The corresponding simulated results of input and output return losses (S_{11}, S_{22}) are shown in Fig. 6. Both S_{11} and S_{22} are better than −10 dB over the entire 3-dB bandwidth of 23 GHz. The simulated NF is presented in Fig. 7. As it is evident, the maximum value of NF is below 4.6 dB within the whole corresponding bandwidth while the average NF is as low as 3.65 dB.

The valuable measure of the stability is called rollett stability factor (or K-factor). If K-factor is greater than one, it implies that the amplifier is unconditionally stable [23]. As it can be seen from Fig. 8, the K-factor of the proposed CMOS-DA is bigger than unity which it shows this amplifier is unconditionally stable over range of interested frequencies from DC up to 23 GHz. The input third intercept point (IIP3) is simulated at 15 GHz. The proposed CMOS-DA achieves the good linearity performance whit IIP3 value of + 8 dBm as shown in Fig. 9. Also, simulated input referred 1-dB compression point at 15 GHz is illustrated in Fig. 10 which it achieves the value of -6.34 dBm.

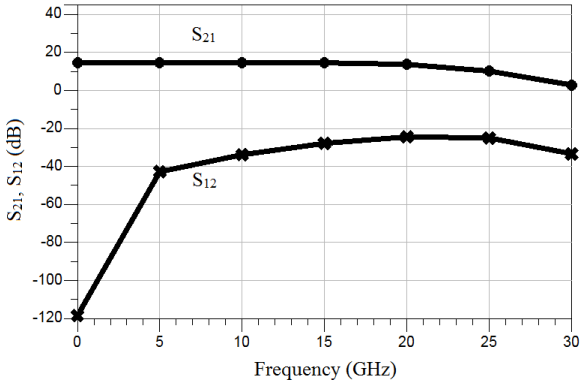


Fig. 5. Simulated results of power gain and revers isolation (S21, S12)

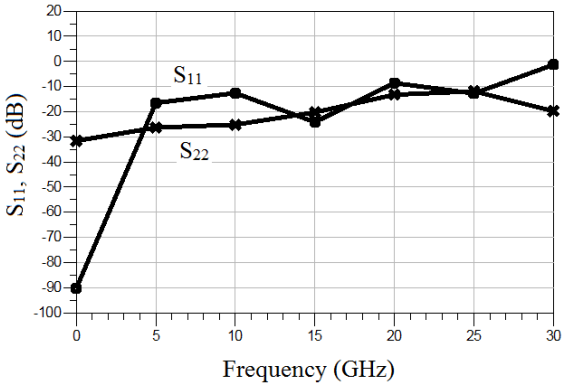


Fig. 6. Simulated results of input and output return losses (S11, S22)

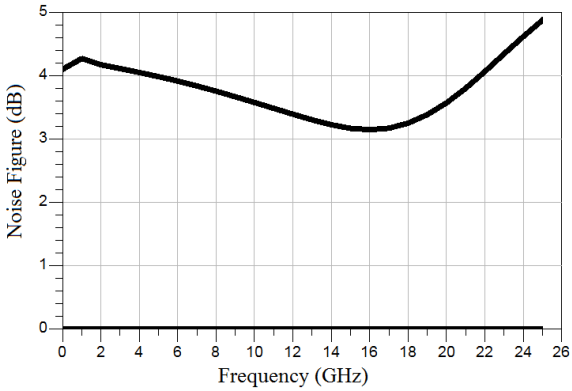


Fig. 7. Simulated noise figure (NF) of the proposed CMOS-DA

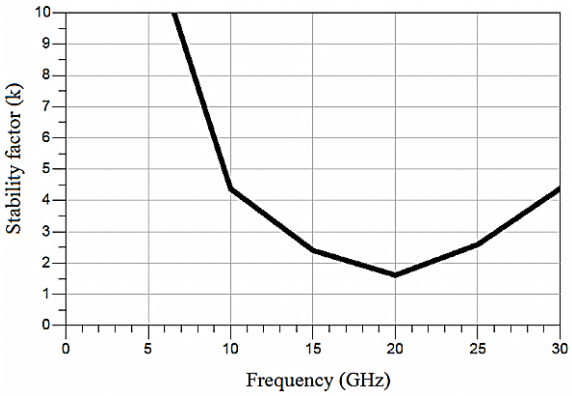


Fig. 8. Simulated stability factor (k) of the proposed CMOS-DA

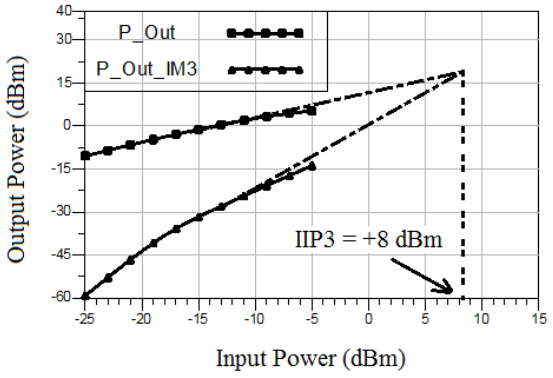


Fig. 9. Simulated fundamental and IM3 output power versus input power characteristics at 15 GHz

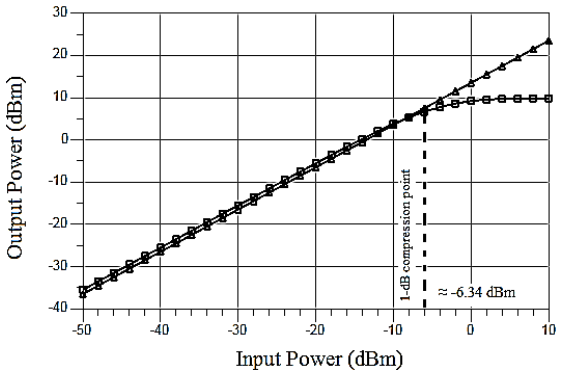


Fig. 10. Simulated input referred 1-dB compression point at 15 GHz of the proposed CMOS-DA

To date, the various methods have been presented to improve the DA design parameters. To evaluate the efficiency of each new design, a figure of merit (FoM) is needed. One FoM included the most relevant parameters such as low-power, high-gain, low-noise, and broad bandwidth can be given as Eq. (24) [8]:

$$\text{FOM} = [\text{GHz/mW}] = \frac{S_{21}[1]\ \text{BW[GHz]}}{(\text{NF}-1)[1]P_{DC}[\text{mW}]} \quad (24)$$

Where S_{21} [1] refers to the average power gain in magnitude, BW [GHz] mentions the 3-dB bandwidth in gigahertz, (NF - 1) [1] is the excess NF in magnitude and P_{DC} [mW] demonstrates power consumption in milli watts.

The comparison results of the proposed CMOS-DA with those of recently published CMOS-DAs are summarized in Table 1. The good value of the proposed DA's FoM confirms that the proposed DA well can satisfy the important design parameters which DA and LNA designs are faced with them while its ultra-broad bandwidth and high input third intercept point are highly desired. The proposed CMOS-DA comparing with the other published work achieves a good performance for ultra-broadband amplification applications.

The MC simulation is an appropriate option to take into consideration the risks associated with various input parameters which they receive little or no consideration in simulating of designs utilizing ideal components. MC simulation is a technique to understand the impact of inputs' uncertainty on the overall performance of the design. It works based on a repetitive process including a random value selection for input parameters within their specified tolerance range and getting a set of output parameters as a result of multiple trial runs [18]. Uncertainty inputs in the proposed design are device variations including fluctuations in MOS parameters such as effective gate length L_{eff}, threshold voltage V_{th}, thickness of the gate oxide T_{ox}, and the drain-source region parasitic resistance R_{dsw} [19]. Not only MOS device variations are considered, but also the effects of supply voltage and passive component tolerances are taken into account. In MC simulation, the proposed design is simulated a large number of times (e.g., 1000). For each run, all of the uncertain parameters are sampled. ADS software package can generate uniformly distributed random values of parameters, which a normal (Gaussian) distribution is used in this design. Note that the Gaussian distribution models the worst case of possible situation. The circuit is then simulated. As a result, actual observations of failures are routinely better predicted by the MC simulation results. In this way, we are beginning to understand the risk and uncertainty in the proposed design.

Figs. 11 ~ 13 show the trend of the influence of MOS device, supply voltage variation, and passive component tolerances on proposed design's performance.

Here, performance refers to S-parameters NF and stability-factor responses. The fluctuations of MOS parameters belonging to 0.13 µm CMOS technology are given according to Table 2 [24].

The result of mentioned variations on power gain of the proposed design is shown in Fig. 11 (a). As it can be seen, the density of results around of 14 dB validates the accuracy of power gain response. As it is shown in Fig. 11 (b), the revers isolation achieves very good values within total MC trial to that of ideal simulated revers isolation. Also, both input and output impedances matching within total MC trial are better than - 15 dB according to Figs. 12 (a) & 12 (b) which they imply that the good impedances matching are yielded. NF response is shown in Fig. 13 (a) that confirms the values of this factor are between 3 to 4 dB within total MC trial. Finally, Fig. 13 (b) illustrates the stability-factor response of the MC simulations. It guarantees that the proposed design is unconditionally stable under whole situation.

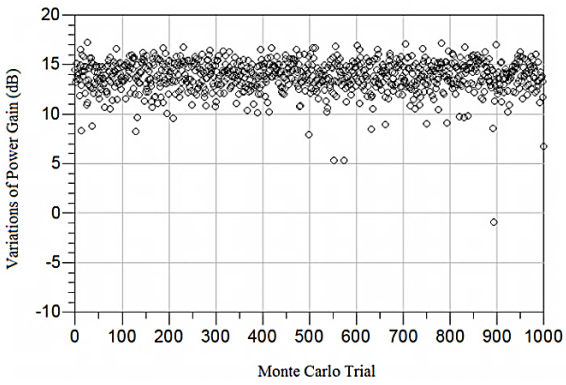
Table 1. Recently reported state-of-the-art CMOS-DAs versus the proposed CMOS-DA

References	Process	Freq (GHz)	Gain (dB)	Average NF (dB)	S11 (dB)	S22 (dB)	PDC (mw)	FOM (GHz/mW)
[9]2013a	0.18µm CMOS	32	9.5	5.85	<-15	<-10	71	0.44
[5]2011a (HG mode)	0.18µm CMOS	1.5~8.2	17.1 ± 1.5	3.52	<-11	<-10.1	46.85	0.7
[5]2011a (LG mode)	0.18µm CMOS	1.2~ 8.6	11.4 ± 1.4	3.74	<-9.4	<-10.4	9.85	1.72
[2]2011a	0.18µm CMOS	35	20.5	7.4	< -12	< -14	250	0.34
[15]2013b (HG mode)	0.13µm CMOS	0~11	20.5 ± 0.5	6.5~8	<-11	<-18	9.36	2.95
[15]2013b (LG mode)	0.13µm CMOS	0~12	15.5 ± 0.25	6~8	<-11.5	<-16.5	3.6	5
[7]2011b (HG mode)	0.13µm CMOS	0.4~10.5	20.47 ± 0.72	3.29	<-10	<-10	37.8	1.73
[7]2011b (LG mode)	0.13µm CMOS	0.7~10.9	11.03 ± 0.98	4.25	<-10.3	<-10.9	6.86	2.67
[8]2015b	0.13µm CMOS	DC ~ 13	26.5 ± 0.4	5.4	<-11.1	<-11.3	9.95	10.2
This workb	0.13µm CMOS	DC ~ 23	14.2 ± 0.5	3.65	< -10	< -10	45.63	1.96
This workc	0.13µm CMOS	DC ~ 23	14 ± 1.5	< 4	< -20	< -15	< 50	~ 1.9

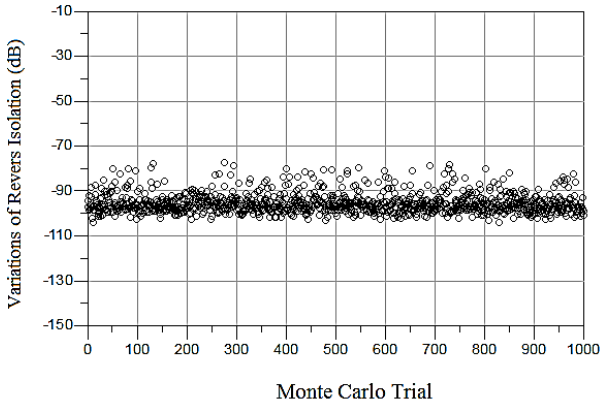
a: Based on the measurement results b: Based on the simulation results c: Based on the MC simulation results

Table 2: Parameter Values and ±3σ Variations of MOS transistor [24]

Technology	**0.13 µm**	
Parameters	**nmos**	**pmos**
$\mathbf{L_{eff}(\mu m)}$	0.09 ± 15%	0.09 ± 15%
$\mathbf{T_{ox}(A^0)}$	33 ± 4%	33 ± 4%
$\mathbf{V_{th}(V)}$	0.33 ± 15.5%	- 0.33 ± 15.5%
$\mathbf{R_{dsw}(\Omega/m)}$	200 ± 10%	400 ± 10%
$\mathbf{V_{dd}(V)}$	1.3 ± 10%	

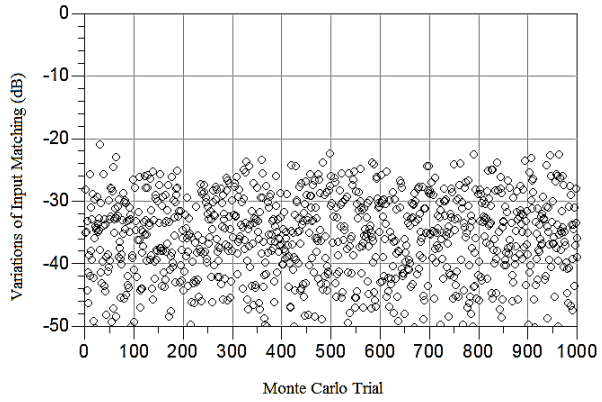


(a)

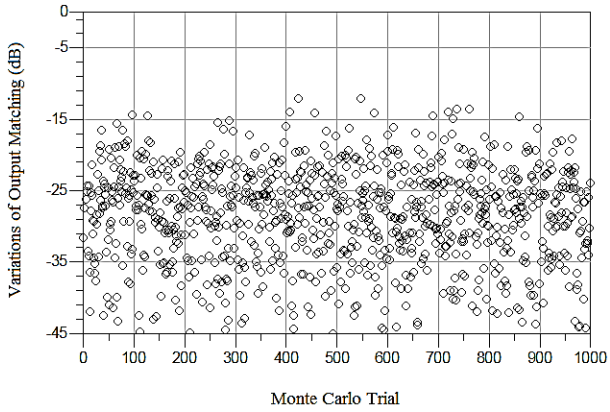


(b)

Fig. 11. Scatter plot of 1000 Monte Carlo runs for investigating a) Power Gain and b) Reverse Isolation Performances

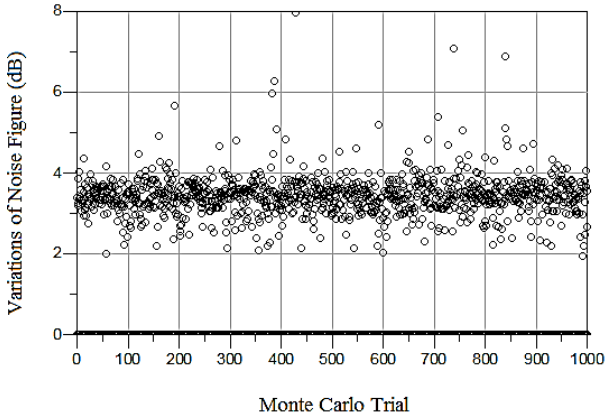


(a)

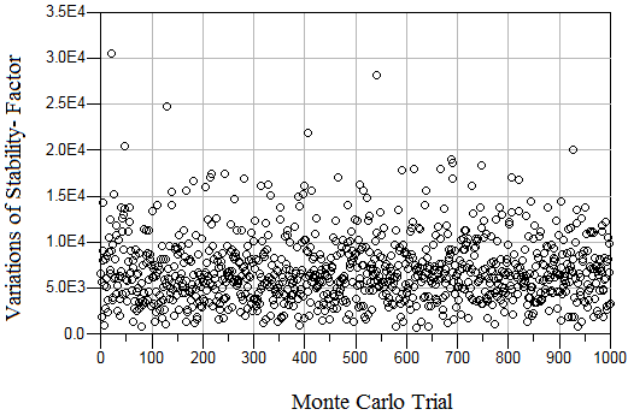


(b)

Fig. 12. Scatter plot of 1000 Monte Carlo runs for investigating a) Input Matching and b) Output Matching Performance



(a)



(b)

Fig. 13. Scatter plot of 1000 Monte Carlo runs for investigating a) Noise Figure and b) Stability-Factor Performances

4. Conclusions

In this study, a high-performance CMOS-DA construction using a new gain-cell has been reported. The new gain-cell combining the regulated cascode, and inductively coupled cascode techniques has been removed the shortcomings of both cascade gain-cells including restricted bandwidth and cascode gain-cells consisting limited signal swings to yield a significantly broadband CMOS-DA. The simulated results of the gain, input and output return losses, isolation, and NF have been illustrated the capability of utilizing this device for ultra-broadband amplification applications. It has been demonstrated promising solution over the previous reports to realize the balanced trade-off between critical challenges facing design of DAs. The effect of parameters variation on performance specifications of the proposed design has been simulated by MC simulation which it was confirmed MC and ideal simulation results are in a good agreement.

References

[1] A. Ghadiri and K. Moez, "Compact Transformer-Based Distributed Amplifier for UWB Systems," Circuits and Systems II: Express Briefs, IEEE Transactions on, vol. 58, pp. 259-263, 2011.

[2] J.-C. Kao, P. Chen, P.-C. Huang, and H. Wang, "A Novel Distributed Amplifier With High Gain, Low Noise, and High Output Power in CMOS Technology," Microwave Theory and Techniques, IEEE Transactions on, vol. 61, pp. 1533-1542, 2013.

[3] R.-C. Liu, T.-P. Wang, L.-H. Lu, H. Wang, S.-H. Wang, and C.-P. Chao, "An 80GHz travelling-wave amplifier in a 90nm CMOS technology," Power, vol. 6, p. 8dBm, 2005.

[4] F. Zhang and P. R. Kinget, "Low-power programmable gain CMOS distributed LNA," Solid-State Circuits, IEEE Journal of, vol. 41, pp. 1333-1343, 2006.

[5] J.-F. Chang and Y.-S. Lin, "A High-Performance Distributed Amplifier Using Multiple Noise Suppression Techniques," Microwave and Wireless Components Letters, IEEE, vol. 21, pp. 495-497, 2011.

[6] G. Xin and N. Cam, "Low-power-consumption and high-gain CMOS distributed amplifiers using cascade of inductively coupled common-source gain cells for UWB systems," Microwave Theory and Techniques, IEEE Transactions on, vol. 54, pp. 3278-3283, 2006.
[7] L. Yo-Sheng, C. Jin-Fa, and L. Shey-Shi, "Analysis and Design of CMOS Distributed Amplifier Using Inductively Peaking Cascaded Gain Cell for UWB Systems," Microwave Theory and Techniques, IEEE Transactions on, vol. 59, pp. 2513-2524, 2011.
[8] Z. Baharvand and A. Hakimi, "Analysis and Design of High Gain, and Low Power CMOS Distributed Amplifier Utilizing a Novel Gain-cell Based on Combining Inductively Peaking and Regulated Cascode Concepts," Amirkabir International Journal of Electrical & Electronics Engineering, pp. 35-50, 2015.
[9] H. Chih-Yin, S. Tzu-Yu, and S. S. H. Hsu, "CMOS Distributed Amplifiers Using Gate-Drain Transformer Feedback Technique," Microwave Theory and Techniques, IEEE Transactions on, vol. 61, pp. 2901-2910, 2013.
[10] B. Y. Banyamin and M. Berwick, "Analysis of the performance of four-cascaded single-stage distributed amplifiers," Microwave Theory and Techniques, IEEE Transactions on, vol. 48, pp. 2657-2663, 2000.
[11] J.-C. Chien, T.-Y. Chen, and L.-H. Lu, "A 9.5-dB 50-GHz Matrix Distributed Amplifier in 0.18-μm CMOS," in VLSI Circuits, 2006. Digest of Technical Papers. 2006 Symposium on, 2006, pp. 146-147.
[12] A. Arbabian and A. M. Niknejad, "Design of a CMOS Tapered Cascaded Multistage Distributed Amplifier," Microwave Theory and Techniques, IEEE Transactions on, vol. 57, pp. 938-947, 2009.
[13] M.-D. Tsai, H. Wang, J.-F. Kuan, and C.-S. Chang, " A 70GHz cascaded multi-stage distributed amplifier in 90nm CMOS technology, " Solid-State Circuits Conference, Digest of Technical Papers. ISSCC. 2005 IEEE International, IEEE, pp 402- 606 2005.
[14] J.-C. Chien and L.-H. Lu, "40-Gb/s high-gain distributed amplifiers with cascaded gain stages in 0.18-μm CMOS," Solid-State Circuits, IEEE Journal of, vol. 42, pp. 2715-2725, 2007.
[15] A. H. Farzamiyan and A. Hakimi, "Low-power CMOS distributed amplifier using new cascade gain cell for high and low gain modes," Analog Integrated Circuits and Signal Processing, vol. 74, pp. 453-460, 2013.
[16] B. Razavi, Fundamentals of microelectronics vol. 1: Wiley Hoboken, 2008.
[17] B. Razavi, Design of analog CMOS integrated circuits: Tata McGraw-Hill Education, 2002.
[18] D. M. Pozar, "Microwave engineering, 3rd," Danvers, MA: Wiley, 2005.
[19] P. Heydari, "Design and Analysis of a Performance-Optimized CMOS UWB Distributed LNA," Solid-State Circuits, IEEE Journal of, vol. 42, pp. 1892-1905, 2007.
[20] C. S. Aitchison, "The intrinsic noise figure of the MESFET distributed amplifier," Microwave Theory and Techniques, IEEE Transactions on, vol. 33, pp. 460-466, 1985.
[21] E. Hamidi, M. Mohammad-Taheri, and G. Moradi, "Improvements in the noise theory of the MMIC distributed amplifiers," Microwave Theory and Techniques, IEEE Transactions on, vol. 56, pp. 1797-1806, 2008.
[22] C. Mishra, "Design and implementation of low power multistage amplifiers and high frequency distributed amplifiers," Texas A&M University, 2004.
[23] J.-F. Chang and Y.-S. Lin, "DC~ 10.5 GHz complimentary metal oxide semiconductor distributed amplifier with RC gate terminal network for ultra-wideband pulse radio systems," IET microwaves, antennas & propagation, vol. 6, pp. 127-134, 2012.
[24] V. Venkatraman and W. Burleson, "Impact of process variations on multi-level signaling for on-chip interconnects," in VLSI Design, 2005. 18th International Conference on, 2005, pp. 362-367.

Zainab Baharvand received her M.Sc degree in Electrical Engineering at the Graduate University of Advanced Technology, Kerman, Iran. Now her research interests include design of RF building blocks for ultra-broadband applications. The focus of her research is on design of distributed amplifier (DA) and variable gain distributed amplifier (VGDA) in 0.13 μm CMOS technology.

Ahmad Hakimi was born in Rafsanjan, Iran, in 1961. He received the B.Sc degree in electrical engineering from Technical College of Shahid Bahonar University of Kerman, Kerman, Iran, in 1986. Using the scholarship, which was granted by the Ministry of Higher Education of Iran and Istanbul Technical University (ITU) in 1987, he has been studying for the degree of M.Sc and Ph.D in the faculty of electrical and electronic at the ITU. He received the M.Sc and Ph.D degrees from ITU in 1995 and 1996 in the field of high-frequency electronics. His research interests include the design and analysis of nonlinear RF circuits, numerical analysis and advanced engineering mathematics, analog filter design and Industrial Research Center, Kerman, Iran, and Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman, Iran.

A Unicast Tree-Based Data Gathering Protocol for Delay Tolerant Mobile Sensor Networks

Zeynab Mottaghinia
Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran
z.mottagy@gmail.com
Ali Ghaffari*
Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran
a.ghaffari@iuat.ac.ir

Received: 20/Feb/2015 Revised: 19/Aug/2015 Accepted: 22/Aug/2015

Abstract

The Delay Tolerant Mobile Sensor Networks (DTMSNs) distinguish themselves from conventional sensor networks by means of some features such as loose connectivity, node mobility, and delay tolerability. It needs to be acknowledged that traditional end-to-end routing protocols cannot be applied usefully in such challenging network conditions because of intermittent connections and/or long delays. Hence, this research is intended to propose a Unicast Tree-based Data Gathering protocol (UTDG) to resolve this problem. A UTDG includes 3 phases: tree formation phase, data collection and data transmission phase, and finally the updating phase. The proposed protocol constructs a tree in each community on the basis of transmission ranking, contact probability and the link expiration time. The selection of the next-hop node is based on the tree structure rather than forwarding the message to the neighbor node directly. Each node unicasts the data to its parent in the related community, and the root of the tree successively sends the data to the sink node. The authors contend, based on the simulation results of the study, that the proposed protocol can gain significantly higher message delivery rates with lower transmission overhead and also lower delay in data delivery than the other existing DTMSNs routing protocols in some applications.

Keywords: Delay-Tolerant Mobile Sensor Networks (DTMSNs); Unicast Tree-based Data Gathering protocol (UTDG); Transmission Ranking; Contact Probability; Link Expiration Time.

1. Introduction

Many routing protocols have been proposed for wireless sensor networks (WSNs) in the related literature in recent years. Although traditional routing methods are appropriate for many sensor applications, they cannot be applied to those scenarios with intermittent and low connections because of sensor nodes mobility, sparse network density, and energy limitation [1-3]. Two practical instances of this scenario are monitoring pervasive air quality and tracking flu virus. In order to obtain the most precise and effective measurement in these examples, wearable sensors that adapt themselves to human activities have been bound. As a result, the connection among the mobile sensors is poor; therefore, it can be concluded that establishing a well-connected mesh network to transfer data through end-to-end connections between sensor nodes and the sinks is hard.

Delay Tolerant Mobile Sensor Networks (DTMSNs) have been proposed in order to solve this problem. DTMSNs are considered as the subset of Delay Tolerant Networks (DTNs) which have many features such as node mobility, delay tolerance, frequent and prolonged communication interruption between nodes, and resource limitations. DTN is a subject that absorbs lots regards and studies have examined many DTN application domains [4]. At the earliest, it was offered to resolve the problem of interplanetary Internet communications through establishing a new network model in space system. This new model could encounter data transmission and other communication needs on the business in the space communications [5, 6]. The considered DTMSN in this paper contains two types of nodes, the wearable sensor nodes and the sink nodes. The former are attached to people (or other mobile objects) which collect information and establish a loosely connected mobile sensor network for information delivery. The second type of nodes are the high-end nodes (e.g., personal digital assistants with sensor interfaces or mobile phones), which are used as the sink nodes to receive data from wearable sensors. Sink nodes are employed are deployed at strategic positions with a high visiting probability or they are carried by people.

In this paper, the authors have a Unicast Tree-based Data Gathering protocol (UTDG) for Delay Tolerant Mobile sensor networks. The proposed UTDG has 3 phases: tree formation phase, data collection and data transmission phase, and the updating phase. In the tree formation phase, the UTDG builds a tree for each community based on the location of nodes, contact probability, transmission rankings and link expiration time; all routing decisions are made according to the formed tree. In the second phase, the next hop node is selected based on the tree structure and also the data

* Corresponding Author

which is unicast to the parent node in the tree. In the final phase, the tree in each community is updated at each time slot. UTDG can improve the network performance with reduce the cost of transmission overhead by means of elaborately selecting the next hop node to forward the data messages.

The rest of the paper is organized as follow: section 2 describes the related work of the literature. Section 3 describes the system architecture and also introduces the proposed Unicast Tree based Data Gathering algorithm. Section 4 reports the simulation results and reports the discussion of the results and the findings. Finally, Section 5 concludes and sums up the whole paper.

2. Related Work

The routing protocols for traditional networks [7-11] would fail in the DTN scenario because these protocols were intended to be deployed in a network with end-to-end connectivity. So many significant research studies have been conducted with regard to DTN architecture in the related literature and there seems to be a consensus on the general DTN architecture [12]. DTN routing algorithms can be referred to as an example of the consensus on DTN architecture. There has been wide research on routing in delay-tolerant networks. The purpose of these research studies was to gain high data delivery rate with low transmission overhead and respectably short delivery delays. A simple and basic routing protocol named direct transmission has been presented in [13]. In this protocol, whenever the sensor meets a sink, it transmits the data messages in its queue to the sink. A sensor does not receive or transmit data messages of the other sensors. This protocol has lower transmission overhead but undesirably longer delivery delay. Moreover, since it depends on the contacts of the sensor nodes and the sink node, and if there would be few sink nodes or the network is very sparse, as a result, it will have very low delivery rate. Epidemic routing [14] protocol has been proposed to increase the data delivery rate in partially connected networks. According to this protocol, a node copies its message to all those nodes with which it has contact. On the other hand, undoubtedly flooding the network with messages will consume network resources such as node energy, bandwidth, buffer, etc. If the resources are scarce and limited, the performance level might decrease [15]. Several methods have been proposed to control the flooding [16-21]. An alternative to epidemic routing is to spread copies of a message to a limited number of nodes. Spray-and-Wait is an approach that “sprays” a number of copies into the network, and then “waits” until 1 of these nodes meets the destination [22]. Moreover, spray-and-Focus [23] is very similar to Spray-and-Wait. This scheme distributes a small number of copies to a few nodes. However, instead of waiting to deliver a message to its destination by itself, each relay node can forward its copy to more nodes using a scheme based on the single-copy utility.

Other endeavors aiming to improve the performance of the DTMSN routing include [24,25]. In [24], an efficient replication-based data delivery (RED) protocol is presented on the basis of erasure coding technology. RED consists of two key components for data transmission and message management. The first component makes the decision on when (time) and where (location) to transmit data messages according to the delivery probability. The second component makes the decisions about the optimal erasure coding parameters based on its current delivery probability. The second component makes these decisions in order to achieve the desired data delivery rate and also to minimize the overhead. This history-based method is not effective and cannot have the actual ability that a node needs to deliver data to the sink nodes. In [25], the authors proposed the Message Fault Tolerance-Based Adaptive Data Delivery Scheme (FAD) to increase the data delivery rate in DTMSN. The FAD approach employed the fault tolerance feature of a message which indicates the importance of the messages. The decisions on message transmission and message dropping are made based on fault tolerance in order to minimize the transmission overhead. The system parameters are carefully tuned on the basis of thorough analyses to optimize the network performance. However, this protocol still has a high overhead.

Yong Feng et al. in [26] proposed a Distance-aware Replica Adaptive Data Gathering protocol (DRADG). This protocol uses a self-adapting algorithm to cut down the number of redundant replicas of the messages based on the distance between sensor nodes and the sink node and uses the delivery probabilities of the mobile sensors as the main routing metrics.

Also, traditional tree based routing algorithms [27-47] are not much robust in the challenged networks which suffering from frequent disruption, sparse network density and limited device. Several tree based approaches have been adopted in such challenged networks.

Two tree based algorithms are designed in [48], which are Static Tree Based Routing (STBR), Dynamic Tree Based Routing (DTBR). Regarding DTBR and STBR approaches, the message is forwarded along a time varying based end-to-end path from the source node to destination and replicated at the branch nodes which have more than one sub branches. STBR is based on the shortest path between the source node and destination and uses the link state information adopted in [49]. However, STBR cannot be dynamically adaptive to the large variation of network topology in DTNs, since the message would be constantly kept by its carrier until the connectivity is available, even if the message carrier is within the group membership of destinations. Motivated by this shortcoming, DTBR updates the path towards destination on receiving the message from previous hop.

In [50], the authors proposed an on-demand situation-aware multicast (OS-multicast) approach. Initially, a

source-rooted tree is constructed in the similar way as STBR [48]. When a node receives a bundle, it will dynamically rebuild the tree rooted at itself to all the destinations based on the current network conditions. Their simulation results showed that OS-multicast can achieve smaller delays and better message delivery ratios than DTBR [48].

These tree based approaches are multicasting algorithms. With respect to multicasting in DTNs, the large variation of network topology limits the scalability of the tree based approaches, since it is difficult to maintain and update the multicast tree using partially historical information. Note that the destinations of the multicast message are a set of nodes using Unicast Based Routing (UBR). In contrast, there is only a unique destination for unicast message performed by the unicast algorithms instead, UBR attracts more research attention by borrowing from the research activities of existing unicast algorithms in DTNs, of which to distribute the multicast destinations [51] is interesting. Then, we proposed the Unicast Tree-based Data Gathering Protocol for Delay Tolerant Mobile Sensor Networks which is described in the next section.

3. Proposed Protocol

Our protocol is a route forwarding protocol designed for delay tolerant mobile sensor networks. The major contribution of this protocol is to build a tree in each community and to select the next-hop based on the tree structure. The proposed protocol is intended to guarantee high performance. In this section, the network model is first described and then 3 important protocol parameters are introduced; after that, the proposed protocol is explained in detail.

3.1 Network Model

Initially, the authors assumed that all the *N* sensor nodes are randomly deployed in a square area of *A*. All the sensor nodes are homogeneous and have a unique ID number. A node, e.g., node *i*, maintains the table as its local information. As shown in Fig. 1, the table consists of 8 fields.

Node_ID	Sink Position	Home_ID	Level of tree	Parent Node_ID	Contact Probability	Link Expiration Time	Transmission ranking

Fig. 1. Format of node header

Each entry in the table is for a node ever met by the Node *i*. The maximum transmission range of all the sensor nodes is fixed to *R*.

As shown in Fig. 2, the mobility of all the sensor nodes is assumed to follow the community-based Mobility model depicted in [52,53] where the whole area is divided into several non-overlapped cells, a public gathering place (G) (e.g., in reality it could be a buffet in a university), and communities (C) (e.g. faculty departments of a university where each node belongs to a home community and most of the time it stays there, for example this node could be a student that belongs to a faculty department. Each node's movement accords to the Random Way Point Model [54] in each home community. Nodes randomly choose a destination and move to their destination by the specified speed v. Upon arrival at the destination, the node pauses for a while and then chooses a new destination. The destinations are selected in a way that if a node is at home, there is a high probability that it will go to the public gathering place (but it is also possible for it to go to other places), and if it is away from home, it is very likely that it will return home. Each sensor node can compute its location by means of the GPS (Global Positioning System) [55] or other GPS-less technique. The sink node is immobile and it is located at the G location which is known to all sensor nodes.

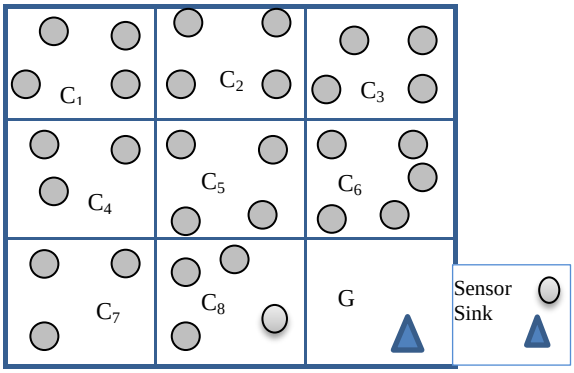


Fig. 2. Community-based mobility model [25]

3.2 Parameters of the Proposed Protocol

The proposed protocol is based on 3 important parameters, namely, contact probability, transmission ranking and Link Expiration Time which are described separately below.

3.2.1 Contact Probability

Contact probability indicates the degree of likelihood that a sensor node will communicate with its neighbors. We adopt a simple approach, namely exponentially weighted moving average (EWMA) [56] to calculate the contact probability. In this approach, Node *i* retains a list of contact probabilities ξ_{ij} for everyother node j, which it has met before. ξ_{ij} is initialized with zero. ξ_{ij} is updated in every time slot according to the following rule:

$$\xi_{ij} = \begin{cases} (1-\alpha)\left[\xi_{ij}\right] + \alpha & , \quad \text{imeetsj} \\ (1-\alpha)\left[\xi_{ij}\right] & , \quad \text{otherwise} \end{cases} \qquad (1)$$

In Eq. (1), $\alpha \in (0,1)$ is a constant parameter, and $\left[\xi_{ij}\right]$ is the old contact probability. Evidently, this is a dynamic process, and thus $\xi_{ij}$is not essentially equal to the actual contact probability p_{ij}. However, according to the following theorem p_{ij} converges to ξ_{ij}.

Theorem 1: If p_{ij} is a probability of Nodes *i* and *j* which should be met in each time slot, ξ_{ij}converges to p_{ij}.

Proof: Consider a sequence of time slots and let $\xi_{ij}(t)$ indicate ξ_{ij} in the time slot *t*. we have:

$$E(\xi_{ij}(1)) = (1-\alpha)\xi_{ij}(0)+p_{ij}\alpha$$

$$E(\xi_{ij}(2)) =\xi_{ij}(0) + p_{ij}(\alpha)[1 + (1 - \alpha)]$$

....

$$E(\xi_{ij}(t)) = (1 - \alpha)^t\xi_{ij}(0) + p_{ij}(\alpha)[1 + (1 - \alpha) + \cdots + (1 - \alpha)^{t-1}]$$

When $t \to \infty$, we arrive at:

$$\lim_{t\to\infty} E(\xi_{ij}(t))=\alpha . p_{ij}.\frac{1}{\alpha}=p_{ij},$$

Which is independent of the parameter α and the initial value $\xi_{ij}(0)$.

3.2.2 Transmission Ranking

Transmission rankings indicate the degree of likelihood that sensor nodes will communicate with the sink node. Generally, the more likely a node is to communicate with the sink node, the higher the transmission ranking attached to it. Let p_i denote the transmission ranking of sensor *i*. As depicted in [57] Due to the randomly moving characteristic of sensor nodes, p_i is a variable related to speed, current moving direction and the distance with the sink node. Based on the Random waypoint model, the process of calculating p_i of node *i* can be categorized into the following 4 cases:

If node *i* is in the transmission range of the sink node, its transmission ranking p_i is equal to 1, and hence it can communicate with the sink node directly. This is because the node *i* can directly communicate with the sink node at that time.

If the current moving path of node *i* intersects the communication range of the sink node, we let $p_i = 1$. Since node *i* is moving towards the sink node, hence, it will soon communicate directly with the sink node.

If the above-mentioned conditions cannot be held, we can calculate the current transmission ranking p_i of the node which is larger when the line is closer to the sink node.

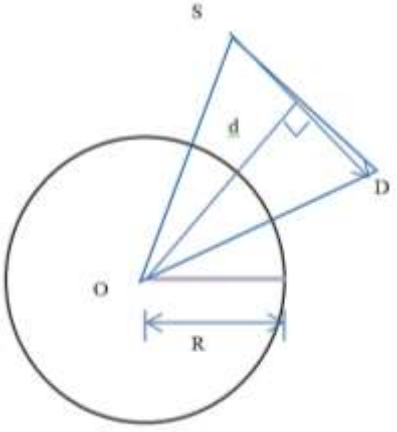


Fig. 3. The sketch map for calculating the transmission ranking [57]

As shown in Fig.3, triangle ΔOSD is composed of the sink node $O(x_o, y_o)$, the start point $S(x_s, y_s)$ and the destination point D (x_d, y_d). SD, OD and OS are denoted for sub-tenses of the angles O; S and D respectively. *d* is denoted for the distance between O and SD. Half circumference of the triangle ΔOSD is denoted by *p*, and acreage of the triangle ΔOSD is denoted by St, then the value of *p* is shown in Eq.(2) as below:

$$p = \frac{1}{2}(\|SD\| + \|OD\| + \|OS\|) = \frac{1}{2}[\sqrt{(x_s - x_d)^2 + (y_s - y_d)^2} + \sqrt{(x_o - x_d)^2 + (y_o - y_d)^2} + \sqrt{(x_o - x_s)^2 + (y_o - y_s)^2}] \quad (2)$$

And the average St is shown in Eq. (3) as follows:

$$S_t=\frac{1}{2}\big||OS|\big|.\big||SD|\big|.\text{Sin } S = \frac{1}{2}\|SD\| . d = \sqrt{p(p - \|OS\|)(p - \|OD\|)(p - \|SD\|)} \quad (3)$$

According to Eq. (3), the distance *d* is shown in Eq. (4) as follows:

$$d = 2\sqrt{p(p - \|OS\|)(p - \|OD\|)(p - \|SD\|)} \,/\, \|OS\| \quad (4)$$

We can get the value of *d* since it can be calculated by means of using p and d, as it is shown in Eq. (5):

$$d = \sqrt{\|OS\|^2 - \left(\frac{\|SD\|^2+\|OS\|^2-\|OD\|^2}{2\|SD\|}\right)^2} \quad (5)$$

In addition, if R/d < 1 and $d<\|OS\|<\|OD\|$ or $d <\|OD\|<\|OD\|$, we ought to set $p_i= R/\|OS\|$ or $p_i= R/\|OD\|$. In the end, we can get the transmission probability formula as it is shown below [57].

$$p_i = \begin{cases} 1 & , \frac{R}{d} \geq 1 \\ R \Big/ \sqrt{\|OS\|^2 - \left(\frac{\|SD\|^2+\|OS\|^2-\|OD\|^2}{2\|SD\|}\right)^2} & , \frac{R}{d} < 1 \\ \frac{R}{\|OS\|} & , \frac{R}{d} < 1 \text{ and} d < \|OS\| \leq \|OD\| \\ \frac{R}{\|OD\|} & , \frac{R}{d} < 1 \text{ and} d < \|OD\| < \|OD\| \end{cases} \quad (6)$$

3.2.3 Link Expiration Time

In this section the authors introduce the expiration time of the link that is formed between the two nodes whose locations are included in the transmission range of each other. Based on our assumption, each mobile node can learn its location by GPS or other GPS-less technique and all the sensor nodes have synchronized clocks. Thus, each mobile node can calculate its speed and direction, and hence broadcast the parameters to its neighbors by the periodic hello messages. Assume that the two nodes *i* and *j* are within the transmission range of each other at time *t*. As described in [58], we can calculate the link expiration time between the nodes *i* and *j*, which is denoted T_{ij} as follow:

$$T_{ij} = t + \frac{-(ab+cd)+\sqrt{(a^2+c^2)r^2-(ad-bc)^2}}{a^2+c^2} \quad (7)$$

In Eq. (7), the locations of the nodes *i* and *j* are represented by (x_i,y_i) and(x_j, y_j), the speeds by v_i and v_j, and the moving directions by θ_i and θ_j ($0\leq \theta_i,\theta_j \leq 2\pi$), respectively; $a= v_i \cos\theta_i - v_j \cos\theta_j$, $b= v_i \sin\theta_i - v_j \sin\theta_j$, $c=x_i - x_j$, $d=y_i - y_j$.

3.3 UTDG Algorithm

The system model, that is, a Unicast Tree-based Data Gathering protocol (UTDG) as shown in Fig. 4 has the following properties. The nodes in each community build a tree with different levels. The distance between the two levels is equal to the radio range of the sensor node.

The UDTG algorithm for delay tolerant mobile sensor networks has been proposed in this paper. This algorithm works in 3 phases: Tree formation phase, data collection and transmission phase, and finally updating phase.

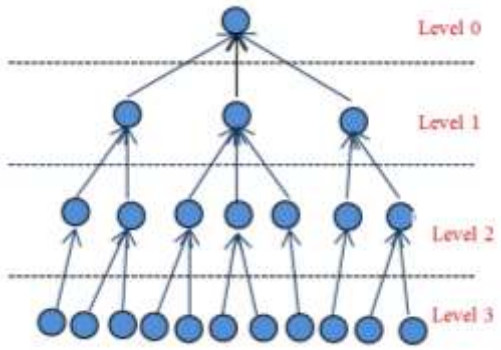


Fig. 4. System architecture

a. Tree formation phase:

All the nodes have their clock synchronized by using the NTP or the GPS clock itself. We set their clocks before deploying them in the area and at a certain time they will begin to build a tree in each home community. Our unicast tree-based protocol requires the warm-up period to construct trees in each home community. In the tree formation phase, the number of levels is calculated using the Eq. (8) as follows:

$$\text{number of levels} = \left[\frac{1}{\beta}\right] + 1 \qquad (8)$$

Where $\beta \in (0,1)$ is a constant value and should be defined based on the application, the number of levels shows the number of the levels in each community.

A node assigns a level to itself according to the value of p_i (Transmission Ranking). A node with maximum transmission ranking is selected as the trees' root. If the other nodes have the transmission rankings $(1-\beta x_{n-1}, 1-\beta x_n]$, they will assign the level n to themselves. In other words, we can assign different levels of the tree to sensor nodes in the tree with *n*+1 level according to Eq. (9) as follows:

$$\begin{cases} \text{level} = 1, p_i \in [1-\beta x_0, 1-\beta x_1] \\ \text{level} = 2, p_i \in (1-\beta x_1, 1-\beta x_2] \\ \text{level} = 3, p_i \in (1-\beta x_2, 1-\beta x_3] \\ \quad\quad . \\ \quad\quad . \\ \text{level} = n, p_i \in (1-\beta x_{n-1}, 1-\beta x_n] \end{cases} \qquad (9)$$

Where p_i is the transmission ranking of the node i; the level of the node in the tree will be as follows:

$x_j = j; j = 0, \dots, (\text{number of levels} - 1)$; $0 \leq 1-\beta x_j \leq 1$.

In fact, nodes with greater transmission ranking will be placed at higher levels of the tree. For example, if β is equal to 0.2, we have:

$$\text{number of levels} = \left[\frac{1}{\beta}\right] + 1 = 6$$

$x_0=0 \Rightarrow 1-\beta x_0=1$ — Level 1
$x_1=1 \Rightarrow 1-\beta x_1=0.8$ — Level 2
$x_2=2 \Rightarrow 1-\beta x_2=0.6$ — Level 3
$x_3=3 \Rightarrow 1-\beta x_3=0.4$ — Level 4
$x_4=4 \Rightarrow 1-\beta x_4=0.2$ — Level 5
$x_5= \Rightarrow 1-\beta x_5=0$

The node with the highest transmission ranking is placed at the zero level; if the transmission ranking of a node (p_i) is within [1, 0.8] then its level will be equal to 1. If p_i is within the interval [0.8, 0.6], then, it will be placed at level 2; if p_i is within the interval [0.6, 0.4], the level will be 3; and finally if (p_i) is within the interval (0.2, 0], the node will be at level 5

After assigning a level to the nodes, we use contact probability (ξ_{ij}) and link expiration time (T_{ij}) for connecting children to their parents. (The node that has higher T_{ij} and ξ_{ij} is selected as the parent) If the node *i* is at the level of x, the variable x will be reduced count by 1, and then will broadcast a hello message which contains the value of x-1. Each node which is located at the x-1 level and receives a hello message will send an RTR (Ready to Receive) message which contains node's ID. The node *i*, with reference to its table, calculates ξ_{ij} and T_{ij} and it gets the probability of being a parent, according to Eq. (10) as follows:

$$\text{parent}_i = \gamma\xi_{ij} + (1-\gamma)\ T_{ij} \qquad (10)$$

Where in Eq. (10), γ is weight parameter; parent_i is probability of being parent.

The node *i* becomes the child of the node that sends its RTR message earlier and has higher parent_i. it sends a message to its parent node to be aware of its presence. The parent node keeps the ID of its children in a table called children table. These steps are repeated until the whole tree has been constructed.

b. Data collection and data transmission phase

After the tree formation phase, like CSMA/CA each node asks its parent to send data by means of an RTS (Ready to Send) message. Since our network is unreliable, we will have several modes:

The child node sends the RTS message and the parent node is in the IDLE mode:

Parent nodes send CTS (Clear to Send) messages to their children nodes; these messages contain ID of the node which has sent the RTS message. By means of this message parent nodes inform their other children to go to the sleep mode and save energy until the time data transmission is finished. Then, the child will start unicasting data. If the parent node receives the message, it will send the ACK message. If the child does not get any ACK messages from

its parent, then it will send the data again (Because the network is intermittently connected). The child node sends the RTS and the parent node is in BUSY mode:

Children do not receive any messages from their parents; therefore, they must wait a random amount of time and then resend the RTS message. It is probably the case that the parent has no response to the RTS message due to the disconnection. So the child node refers to its table and sends the hello message to those nodes that have higher contact probability. Each node that receives the hello message will send an ACK message with its ID and level. Consequently, the node which has higher contact probability and lower level becomes the new parent of the child node and the child node sends its data to it. The root of each tree that has been formed in communities sends the received data to the sink node.

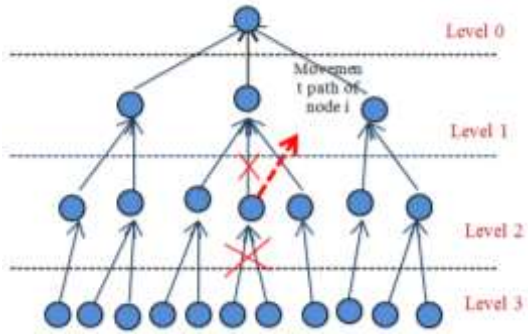


Fig. 5. Movement of the node

c. Updating phase

In each time slot, the values of ξ_{ij}, p_i and T_{ij} will be updated according to Eq.s 1, 6 and 7, respectively. When a sensor node moves from 1 location to another location in its home community, as shown in Fig. 5, there will be 2 possibilities regarding the movement of the node. The node moves either within the same level or to a higher or lower level. The node re-calculates its level according to p_i. If the level of a node does not change, then the node will check that if it is within the range of its father or not; if it is within the range of its parent node, then there is no need to re-join the tree; otherwise, the node chooses its father according to $parent_i$, as shown in Fig. 6.

When a node moves from 1 community to another community, it assigns a level to itself in the new community according to p_i.

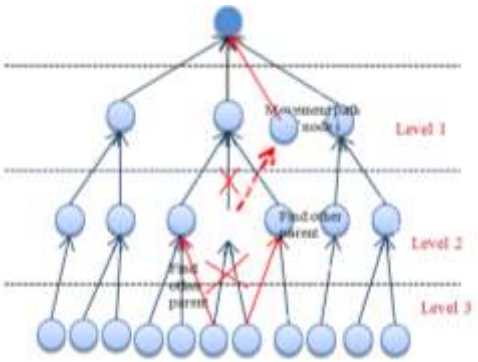


Fig. 6. New tree formation

d. Handling node invalidation

Child node invalidation: if the parent node does not get any response from any of its children in the data transmission process, hence, it will add this node in the list of invalid nodes and will wait for the maximum T_{ij} of its children. After this waiting time, the node which does not send any response will be inferred as an invalid child and as a result, the parent node will remove this node from the list of its children nodes.

Parent node invalidation: after the link expiration time between a child node and a parent node, the child node will remove the father node from the father list and will choose a new father according to a new ξ_{ij} and T_{ij}.

e. Handling energy constraints

There are 2 possible values for the energy level of a node: A node with an energy level higher than half of the original battery capacity and a node with an energy level lower than half of the original battery capacity. These energy limitation options of the nodes will be treated as follows:

If the sensor node energy level is lower than half of the original battery capacity but higher than the average energy level (Threshold value) then move the node to 1lower level and increase its level count by 1. Otherwise if the energy level of the node is lower than the threshold value, then move this node to the lowest level. If the leaf node energy level is higher than the battery capacity, then move this node to 1 higher level and decrease the level count by 1.

4. Simulation

In this section, we evaluate the performance of the schemes through MATLAB simulations. Each simulation is repeated 10 times. In our experimental environment, we assumed that the simulation area is a 1500 m × 1500 m region which was divided into non-overlapped subareas: one gathering place and eight communities. The sub-area at the left bottom is selected as the gathering place, and the sink node is positioned at the gathering place. The authors supposed that the data generation in each sensor follows a Poisson process with an average arrival interval from 10 s to 100 s. Other simulation parameters and their default values are summarized in Table 1 below.

We carried out the UTDG, OS-multicast, DRADG, FAD and epidemic routing protocols. The performance metrics used in our simulations are as follows:

Data delivery ratio, which is the ratio of the data received by the sink node to the sum of data generated by all the sensor nodes in the network.

Data delivery delay, which is defined as the duration from the very beginning of data generation time until it is received by the sink node.

Network lifespan, which is defined as the duration from the very beginning of the network operation until the time a half of all sensor nodes depletes their energy in our simulation.

Table 1: Simulation parameters.

Parameter	Value
Network size (m2)	1500 ×1500
Number of sensor nodes	150
Radio Transmission R (m)	105
Speed of sensor node v (m/s)	2~10
Maximum buffer size of sensor (message)	1000
Data packet size (bytes)	1000
Control message size (bytes)	250
Initial energy (J)	50
Packet generation ratio (packet/s)	0.01
Maximum delay tolerance value (s)	1000
Position of sink node (m)	(300, 300)
α	0.8

4.1 Impact of Message Generation Ratio

In the simulation which was conducted in this study, the authors changed the data generation rate in order to observe its impacts on the performance of the four protocols under different transmission loads. The performance of five protocols is shown in Fig. 7 and Fig. 8.

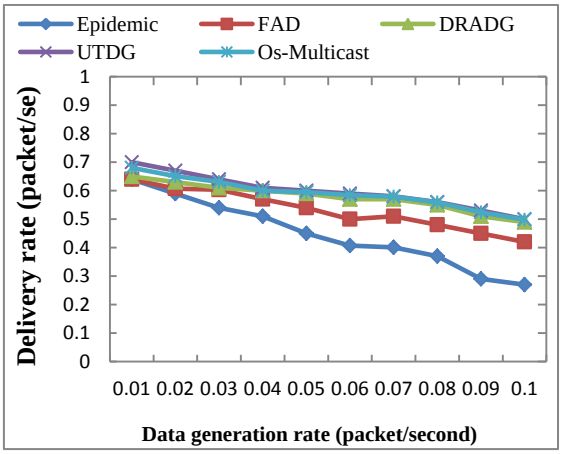


Fig. 7. Impact of data generation rate on packet delivery rate

As Fig. 7 shows, UTDG has the highest data delivery ratio in comparison with the other four protocols; the reason is that UTDG has less resource demands and can deal with high transmission loads through reducing the network traffic. Although OS-multicast tries to take advantage of the current available opportunistic links to push the data closer to the destination but the chance of network congestion also increases in this algorithm and redundant traffic has been created, so it has lower delivery rate than UTDG. The epidemic protocol will have the lowest data delivery ratio when the data generation rate is very low, but as the data generation rate increases, data delivery ratio will increase. This can be explained in view of the fact that since the epidemic protocol generates too many message copies; hence it leads to MAC layer collision and rapid exhaustion of the limited network resources. Moreover, FAD shows better performance than the epidemic routing. However, since DRADG has less resource demands than the other 2 protocols, as a result, it performs better than FAD and the epidemic protocols.

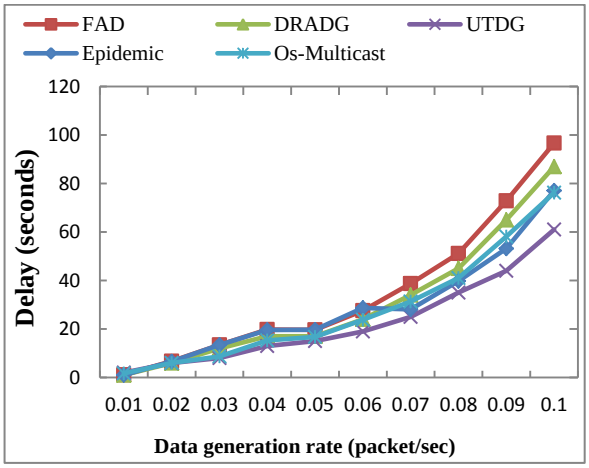


Fig. 8. Impact of data generation rate on delay

Just as Fig. 8 demonstrates, as data generation rate increases, the average delay of all protocols goes up. Obviously, the epidemic routing has the most increase in delivery delay among other protocols. It can be argued that inasmuch as UTDG efficiently cuts down the communication overhead as well as properly choosing the next hop based on nodes' delivery probabilities, hence, DRADG routing has shorter delivery delay than the epidemic routing and FAD. A lot of redundant traffic has been introduced by Os-multicast due to its nature of utilizing multiple currently available links; hence, OS-multicast has higher deliver delay than UTDG. Moreover, because in UTDG routing, sensors forward a data message for the node with the highest contact probability, and also because it can properly manage data traffic, hence, it has the shortest delivery delay among all the protocols .

4.2 Impact of Varying Sensor Node Density

The connectivity of DTMSN is related to the density of sensor nodes. The following experiments show the network performance of five protocols with different sensor node density. As shown in Fig. 9, as the density of sensor nodes goes up, the delivery rate of epidemic and FAD schemes decreases. This is logical since the number of collisions increases. The UTDG protocol almost achieves the upper bound of the data delivery rate when the node density is lower than 150 nodes, as the number of nodes becomes more than 150, the level of nodes in tree changes dramatically in UTDG routing scheme, which results in the reduction of the data delivery ratio. With the increment of number of nodes, the connectivity of the network is enhanced, and thus the performance of Os-multicast and DRADG improve. Overall, it can be concluded that UTDG has better performance than the other protocols in low node density.

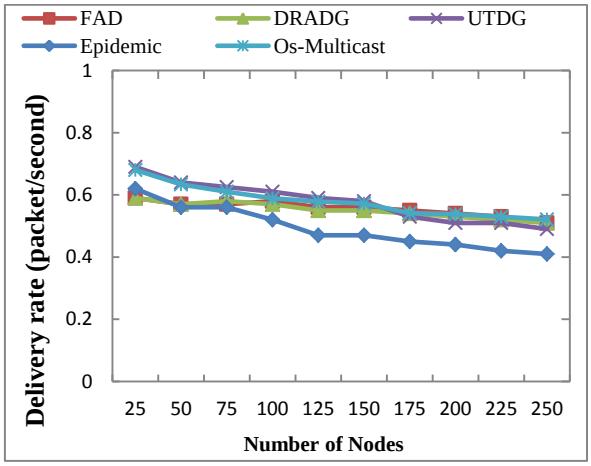


Fig. 9. Impact of varying sensor node density on average DDR

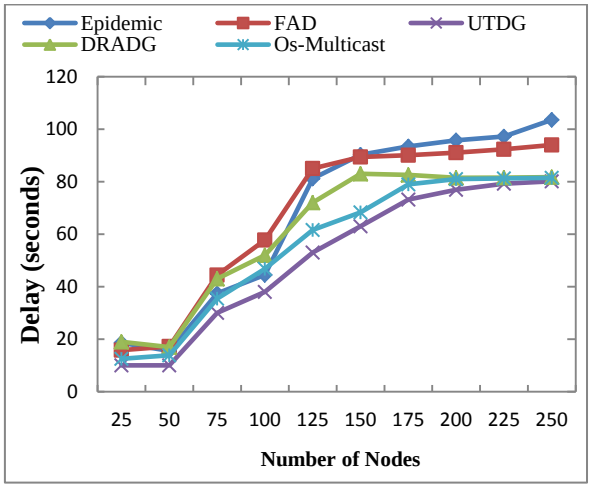


Fig. 10. Impact of varying sensor node density on average delay

4.3 Impact of Varying Buffer Size

In this section of the paper, we are evaluating the impact of buffer size on the performance (see Fig. 11 and Fig. 12). The buffer size here indicates the maximum number of messages a sensor can hold. As it is shown in Fig. 11, we can find out that as the buffer size increases, the delivery ratio also increases for all the protocols because in that case messages can stay in the memory for a longer time before they are dropped. Compared to the other evaluated protocols, the epidemic routing protocol is more sensitive to the buffer size since it generates many copies of the messages and hence needs more buffer size. It should also be noted that the UTDG protocol gains more than the other protocols with an increase in the buffer size. Fig. 12 depicts that the data delivery delay increases along with a larger buffer size; the reason for this is that when the node has larger buffer size, it exchanges more data message between the nodes; hence the chance of network congestion also increases.

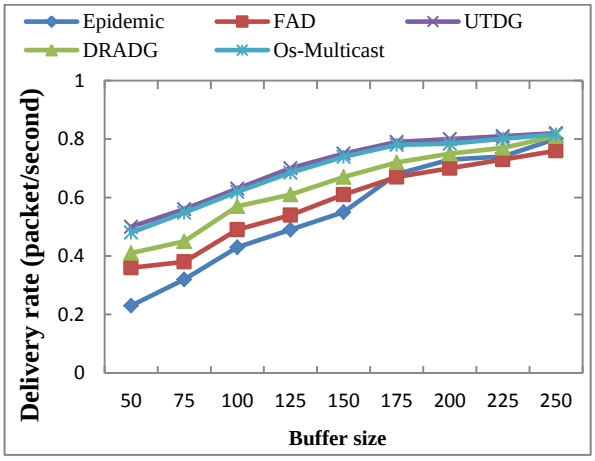


Fig. 11. Impact of varying buffer size on average DDR

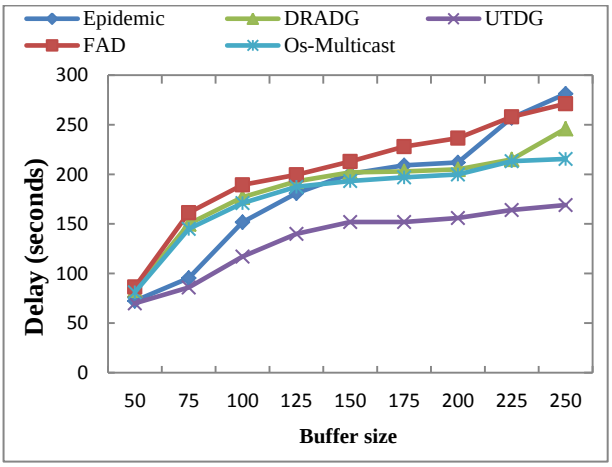


Fig. 12. Impact of varying buffer size on average delay

4.4 Analyzing Network Lifespan

Since sensor nodes are generally energy-constrained, therefore, network lifespan is considered as one of the most important metrics for DTMSNs. We can regard the network as dead when half of the sensor nodes exhaust their energy. The experiments show the network lifespan of the four protocols, and the results are shown in Fig. 13. It can be stated that because the epidemic routing sends and receives too many copies of the messages, it therefore depletes too much energy; so its network lifespan is deemed to be the shortest among the five protocols. Moreover, UTDG has much longer network lifespan than FAD, DRADG and Os-multicast. The reason is that UTDG, unlike the multiple-copy feature of FAD and DRADG, is a single-copy routing protocol, therefore it can reduce the transmission overhead. It also selects the best next-hop based on the tree structure, and also OS-multicast tries to utilize multiple paths to the receivers and multicasting data leads to more data message exchange between nodes, so it consumes more energy than UTDG. In addition, the total energy consumption in UTDG is significantly much less than Os-multicast, DRADG, FAD and the epidemic routing which is regarded as a great advantage for UTDG in terms of economizing energy.

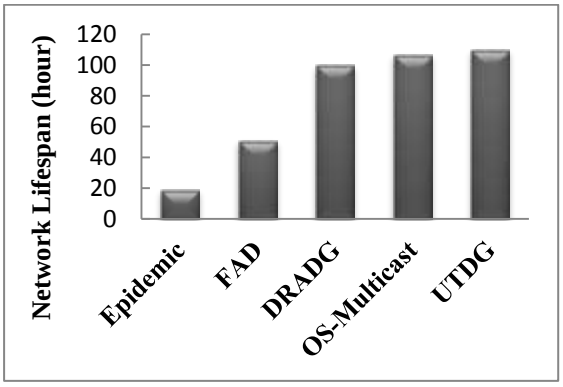


Fig. 13. Network lifespan

5. Conclusions

This paper dealt with the significant issue of efficient data transmission in the Delay-Tolerant Mobile Sensor Network (DTMSN). By taking into consideration the unique features of DTMSN such as sensor node mobility, loose connectivity, and delay tolerability which distinguish DTMSN from conventional sensor networks, we proposed a new routing approach, namely a Unicast Tree-based Data Gathering protocol (UTDG) for DTMSN. UTDG constructs a tree in each community based on 4 parameters: sensors location, transmission ranking, contact probability and link expiration time. The proposed protocol selects the next-hop to forward the data messages based on the tree structure. The simulation results showed that our proposed UTDG protocol performs significantly better than the other protocols with less traffic overhead and less energy consumption. Moreover, it has higher delivery rate than the other existing protocols in applications with 150 nodes or lower.

References

[1] Scott K and Burleigh S. (2007). "Bundle protocol specification." RFC 5050. Available: http://www.rfc-editor.org/rfc/rfc5050.txt [July 27, 2012].

[2] Y. Cao and Z. Sun. "Routing in delay/disruption tolerant networks: a taxonomy, survey and challenges," IEEE Communications and Tutorials, vol. 15, no.2, pp. 654–677. DOI: SURV.2012.04/SURV.2012.042512.00053, 2013.

[3] K. Fall . "A delay-tolerant network architecture for challenged internets," ACM SIGCOMM, New York, NY, USA, 2003.

[4] A. Voyiatzis. "A survey of delay- and disruption-tolerant networking applications," J. Internet Eng., vol. 5 (1), 2012.

[5] I.F. Akyildiz, O.B. Akan and C. Chen. "InterPlaNetary internet: state-of-the-art and research challenges," Computer Networks, vol. 43, no.2, pp. 75-112, 2003.

[6] L. Fuhong, L. Qian, Z. Xianwei and X. Ke. "Towards green for relay in InterPlaNetary Internet based on differential game model," Science China Information Sciences (CHINAF), vol. 57, no.4, pp.1-9, 2014.

[7] D.B. Johnson and D.A. Maltz. "Dynamic source routing in ad hoc wireless networks," Mobile Computing, vol. 353, pp. 153–181, 1996.

[8] C.E. Perkins and P. Bhagwat. "Highly dynamic destination sequenced distance-vector routing (DSDV) for mobile computers," In SIGCOMM'94: Proceedings of the ACM conference on computer communications, 1994, pp. 234–244.

[9] C.E. Perkins and E.M. Royer. "Ad hoc on-demand distance vector routing," In WMCSA'99: Proceedings of the 10th IEEE workshop on mobile computing systems and applications, 1999, pp. 90–100.

[10] A.V. Vasilakos, M.P. Saltouros and A.F. Atlassis . "Optimizing QoS routing in hier-archical ATM networks using computational intelligence techniques," IEEE Transactions on Systems, Man and Cybernetics; Part C, vol. 33, no.3, pp. 297–312, 2003.

[11] Y. Moustafa, M. Ibrahim, M. Abdelatif, L. Chen, and A.V. Vasilakos. "Routing metrics of cognitive radio networks: A survey," IEEE Communications Surveys and Tutorials, vol. 16, no.1, pp. 92–109, 2014.

[12] Delay Tolerant networking research group. http://www.dtnrg.org.

[13] Y. Wang and H.Y. Wu. "Delay/Fault-Tolerant Mobile Sensor Network (DFT-MSN): A new paradigm for pervasive information gathering," IEEE Transaction on Mobile Computing, vol. 6, pp.1021-1034, 2006.

[14] A. Vahdat and D. Becker. "Epidemic Routing for Partially Connected ad hoc Networks," Technical Report CS-2000-06, Duke University: Durham, NC, USA, 2000.

[15] Y. C. Tseng, S. Y. Ni,Y.S. Chen, and J. P. Sheu. "The Broadcast Storm problem in a mobile ad hoc network," Wireless Networks, vol. 8, no. 2/3, 2002.

[16] J. A. Davis, A. H. Fagg, and B. N. Levine. "Wearable computers as packet transport mechanisms in highly-partitioned ad-hoc networks," In Proc Int'l Symp. on Wearable Computers (ISWC'01), 2001, pp. 141–148.

[17] D. Marasigan and P. Rommel. "MV routing and capacity building in disruption tolerant networks," In Proc IEEE INFOCOM, 2005, pp. 398–408.

[18] T. Small and Z. J. Haas. "Resource and performance tradeoffs in delay-tolerant wireless networks," In Proc the ACM SIGCOMM Workshop on Delay-Tolerant Networking (WDTN'05), 2005, pp. 260–267.

[19] K. Tan, Q. Zhang, and W. Zhu. "Shortest path routing in partially connected ad hoc networks," In Proc. Global Telecommunications Conference (GLOBECOM), 2003, pp. 1038–1042.

[20] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine. "Maxprop: Routing for vehicle-based disruption-tolerant networking," in Proc. IEEE INFOCOM, 2006, pp. 1-11.

[21] W. Yahui, S. Deng and H. Huang. "Optimal Forwarding and Beaconing Control of Epidemic Routing in Delay Tolerant Networks," Ad Hoc & Sensor Wireless Networks (AHSWN), vol. 20, no.12, pp.71-93, 2014.

[22] T. Spyropoulos, K. Psounis and C. Raghavendra. "Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks," In Proc. ACM SIGCOMM Workshop on Delay-Tolerant Networking, 2005, pp. 252-259.

[23] T. Spyropoulos, K. Psounis and C.S. Raghavendra. "Spray and focus: Efficient Mobility-Assisted Routing for Heterogeneous and Correlated Mobility", In Proc. of IEEE ICPCC Workshop, 2007, pp. 79-85.

[24] Y. Wang and H.Y. Wu. "Replication-Based Efficient Data Delivery Scheme (RED) for Delay/Fault-Tolerant Mobile Sensor Network (DFT-MSN)," In Proc. Fourth Annual IEEE International Conference on Pervasive Computing and Communications 2006, pp. 485-489.

[25] Y. Wang and H.Y. Wu. "The Delay/Fault-Tolerant Mobile Sensor Network DFT-MSN: A New Paradigm for Pervasive Information Gathering", IEEE Transactions on Mobile Computing, vol. 6, pp. 1021-1034, 2007.

[26] Y. Feng, H. Gong, M. Fan, M. Liu and X.A. Wang. "Distance-Aware Replica Adaptive Data Gathering Protocol for Delay Tolerant Mobile Sensor Networks," Sensors, vol. 11, pp. 4104-4117, 2011.

[27] Z. Bing, Y. Takizawa, A. Hasagawa, A. Yamaguchi and S. Obana. "Tree-Based Routing Protocol for Cognitive Wireless Access Networks", Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE DOI: 10.1109/WCNC.2007.768 Publication Year: 2007, pp. 4204 – 4208

[28] M. Zeynali, L. M. Khanli, and A. Mollanejad, "TBRP: Novel Tree Based Routing Protocol in Wireless Sensor Network ", International Journal of Grid and Distributed Computing, vol. 2, pp. 35-48, 2009.

[29] L. Borsani, S. Guglielmi, A. Redondi and M. Cesana. "Tree-based routing protocol for mobile Wireless Sensor Networks," Wireless On-Demand Network Systems and Services (WONS), 2011 Eighth International Conference on DOI: 10.1109/WONS.2011.5720188 Publication Year: 2011, pp. 164 – 170.

[30] G. Bencan and J. Tingyao. "A tree-based routing protocol in wireless sensor networks," Electrical and Control Engineering (ICECE), 2011 International Conference on DOI: 10.1109/ICECENG.2011.6057953 Publication Year: 2011, pp. 5729 – 5732.

[31] A. Naderi and S.M. Mazinani. "A Differentiated Tree-Based Protocol Along with Re-routing Policy in Wireless Sensor Networks," Computer, Consumer and Control (IS3C), 2012 International Symposium on DOI: 10.1109/IS3C.2012.131 Publication Year: 2012, pp. 495 – 498.

[32] M. Caleffi, G. Ferraiuolo and L. Paura. "Augmented Tree-based Routing Protocol for Scalable Ad Hoc Networks," Mobile Adhoc and Sensor Systems, 2007. MASS 2007. IEEE International Conference on DOI: 10.1109/MOBHOC.2007.4428727 Publication Year: 2007, pp. 1 - 6 Cited by: Papers (5)

[33] S.M. Mazinani, A. Naderi and M. Jalali. "A Tree-Based Reliable Routing Protocol in Wireless Sensor Networks," Computer, Consumer and Control (IS3C), 2012 International Symposium on DOI: 10.1109/IS3C.2012.130 Publication Year: 2012, pp. 491 – 494.

[34] N.P. Karthickraja and V. Sumathy. "A study of routing protocols and a hybrid routing protocol based on Rapid Spanning Tree and Cluster Head Routing in Wireless Sensor Networks," Wireless Communication and Sensor Computing, 2010. ICWCSC 2010. International Conference on DOI: 10.1109/ICWCSC.2010.5415893 Publication Year: 2010, pp. 1 – 6.

[35] A.H. Mohajerzadeh, M.H. Yaghmaee and Z. Eskandari. "Tree based energy efficient and Congestion Aware Routing Protocol for Wireless Sensor Networks," Communication Systems, 2008. ICCS 2008. 11th IEEE Singapore International Conference on DOI: 10.1109/ICCS.2008.4737472 Publication Year: 2008, pp. 1707 – 1711.

[36] Z. Zusheng and Y. Fengqi. "Performance Analysis of Cluster-Based and Tree-Based Routing Protocols for Wireless Sensor Networks," Communications and Mobile Computing (CMC), 2010 International Conference on Volume: 1 DOI: 10.1109/CMC.2010.187 Publication Year: 2010, pp. 418 – 422.

[37] N. Mitton, T. Razafindralambo, D. Simplot-Ryl and I. Stojmenovic. "Hector is an Energy Efficient Tree-Based Optimized Routing Protocol for Wireless Networks," Mobile Ad-hoc and Sensor Networks, 2008. MSN 2008. The 4th International Conference on DOI: 10.1109/MSN.2008.24 Publication Year: 2008, pp. 31 - 38 Cited by: Papers (2).

[38] O. Tang, B. Wang and Z. Dai. "Virtual backbone tree based energy balanced hierarchical routing protocol," Power Electronics and Intelligent Transportation System (PEITS), 2009 2nd International Conference on Volume: 2 DOI: 10.1109/PEITS.2009.5406920 Publication Year: 2009, pp. 204 – 208.

[39] D. Guisen. Z. Xuejun, Z. Jun and L. Feng. "TLRP: Tree-shape Location-based Routing Protocol in Mobile Ad Hoc Network," ITS Telecommunications Proceedings, 2006 6th International Conference on DOI: 10.1109/ITST.2006.288986 Publication Year: 2006, pp. 646 – 649.

[40] N.C. Wang. "Power-aware dual-tree-based multicast routing protocol for mobile ad hoc networks," Communications, IET Vol. 6, pp. 724 – 732, 2012.

[41] H. Zhao, W. Jie, Z. Jie, L. Liefen and T. Kaiyun. "A General Self-Organized Tree-Based Energy-Balance Routing Protocol for Wireless Sensor Network.", IEEE Transactions on Nuclear Science, vol. 61, pp. 732 – 740, 2014.

[42] K. Kiwoong, H. Minkeun, K. Taehong, H.K. Seong and K. Daeyoung. "The stateless point to point routing protocol based on shortcut tree routing algorithm for IP-WSN," Internet of Things (IOT), 2012 3rd International Conference on the DOI: 10.1109/IOT.2012.6402319 Publication Year: 2012, pp. 167 – 174.

[43] K. Viswanath, K. Obraczka and G. Tsudik. "Exploring mesh and tree-based multicast. Routing protocols for MANETs.", IEEE Transactions on Mobile Computing, vol. 5, pp. 28 – 42, 2006.

[44] T. Yu, B. Changqin and F. Aihua. "A Transmission Time Based Routing Protocol for Clustered Collection Tree Wireless Sensor Networks," Computational and Information Sciences (ICCIS), 2010 International Conference on DOI: 10.1109/ICCIS.2010.13 Publication Year: 2010, pp. 21 – 24.

[45] R.S.Elhabyan, M.C.E. Yagoub. "Weighted tree based routing and clustering protocol for WSN," Electrical and Computer Engineering (CCECE), 2013 26th Annual IEEE Canadian Conference on DOI: 10.1109/CCECE.2013.6567759 Publication Year: 2013, pp. 1 - 6 IEEE CONFERENCE PUBLICATIONS

[46] N. Meghanathan, "A MANET multicast routing protocol for stable trees basedon the inverse of Link Expiration Times," Consumer Communications and Networking Conference (CCNC), 2012 IEEE DOI: 10.1109/CCNC.2012.6180965 Publication Year: 2012, pp. 947 – 951.

[47] Y. Baddi and E. Ech-Cherif . "Parallel GRASP algorithm with delay and delay variation for core selection in shared

tree based multicast routing protocols." Innovative Computing Technology (INTECH), 2013 Third International Conference on DOI: 10.1109/INTECH.2013.6653696 Publication Year: 2013, pp. 227 - 232.
[48] W. Zhao, M. Ammar, and E. Zegura, "Multicasting in delay tolerant networks: semantic models and routing algorithms," in Proc. ACM WDTN '05, 2005, pp. 268-275.
[49] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," in Proc. ACM SIGCOMM'04, 2004, pp. 145-158.
[50] Q. Ye, L. Cheng, M. C. Chuah, and B. D. Davison, "Os-multicast: On-demand situation-aware multicasting in disruption tolerant networks," in Proc. IEEE VTC, 2006, pp. 96–100.
[51] J. Wu and Y. Wang, "A non-replication multicasting scheme in delay tolerant networks," in IEEE MASS '10, San Francisco, California, USA, 2010.
[52] A. Lindgren, A. Doria and O. Schelen. "Probabilistic Routing in Intermittently Connected Networks," Lecture Notes in Computer Science, Vol. 3126, pp. 239–254, 2004.
[53] T. Spyropoulos, K. Psounis, and C. Raghavendra. "Performance Analysis of Mobility-Assisted Routing," in Proc. MobiHoc .'06: 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2006, pp. 49.–60.
[54] C. Tracy, J. Boleng, and Vanessa Davies. "A survey of mobility models for ad hoc network research," Wireless Communications & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications, vol. 2, pp. 483–502, 2002.
[55] A. Thaeler, M. Ding and X.Z. Cheng. "iTPS: An Improved Location Discovery Scheme for Sensor Networks with Long Range Beacons," Journal of Parallel. And Distributed. Computing, vol. 65, pp. 98–106, 2005.
[56] J.S. Locus, M.S. Saccucci. "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," Technimetrics session of the 33rd Annual Fall Technical Conference in Houston, Texas,1989.
[57] Z. Ke, L. Ming, Z. Jinqi, S. Jipeng and Z. Jiazhi. "Data Aggregation based Adaptive Data Reproductive Delivery Scheme for Delay Tolerant Mobile Sensor Networks," Chinese Journal of Electronics, Vol.19, pp. 175-180, 2010.
[58] W. Su, S.J. Lee, M. Gerla. "Mobility prediction and routing in ad hoc wireless networks," International Journal of Network Management, vol. 11, pp. 3-30, 2001.

Zeynab Mottaginia received the B.Sc degree in Software engineering from Payam Noor University (PNU) of Tabriz, Iran, in 2010, and her M.Sc degree in Computer Engineering from Azad University of Tabriz, Iran, in 2012. Her current scientific interests are in general area of Challenged Networks, Mobile Ad hoc Networks (MANETs), Wireless Sensor Networks (WSNs), and Delay Tolerant Networks (DTNs).

Ali Ghaffari received his B.Sc, M.Sc and Ph.D degrees in computer engineering from the University of Tehran and IAUT (Islamic Azad University), TEHRAN, IRAN in 1994, 2002 and 2011 respectively. As an assistant professor of computer engineering at Islamic Azad University, Tabriz branch, IRAN, his research interests are mainly in the field of wired and wireless networks, Wireless Sensor Networks (WSNs), Mobile Ad Hoc Networks(MANETs),Vehicular Ad Hoc Networks(VANETs), networks security and Quality of Service (QoS). He has published more than 50 international conference and reviewed journal papers.