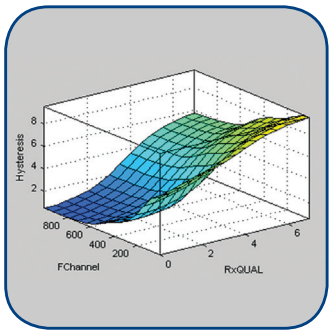
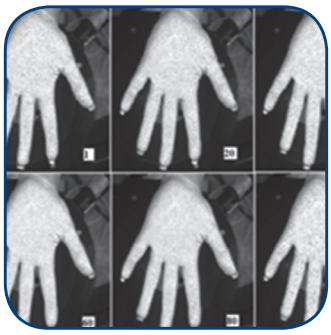
Volume 13, Number 3, 2022 ISSN 1847-6996

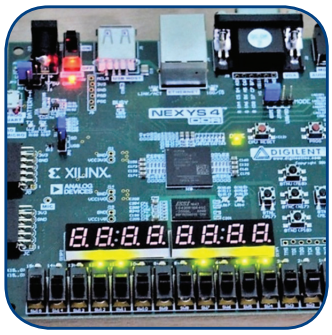
FERIT
FACULTY OF ELECTRICAL ENGINEERING, COMPUTER SCIENCE AND INFORMATION TECHNOLOGY **OSIJEK**

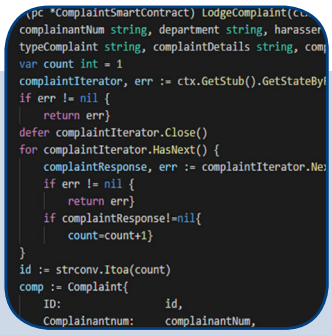
IJECES International Journal of Electrical and Computer Engineering Systems

International Journal of Electrical and Computer Engineering Systems









ISSN: 1847-6996

INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING SYSTEMS

Published by Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Josip Juraj Strossmayer University of Osijek, Croatia

Osijek, Croatia | Volume 13, Number 3, 2022 | Pages 165 - 244

The International Journal of Electrical and Computer Engineering Systems is published with the financial support of the Ministry of Science and Education of the Republic of Croatia

CONTACT

International Journal of Electrical and Computer Engineering Systems (IJECES)

Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia
Kneza Trpimira 2b, 31000 Osijek, Croatia
Phone: +38531224600, Fax: +38531224605
e-mail: ijeces@ferit.hr

Subscription Information

The annual subscription rate is 50€ for individuals, 25€ for students and 150€ for libraries.
Giro account: 2390001 - 1100016777,
Croatian Postal Bank

Journal is referred in:

- Scopus
- Web of Science Core Collection (Emerging Sources Citation Index - ESCI)
- Google Scholar
- CiteFactor
- Genamics
- Hrčak
- Ulrichweb
- Reaxys
- Embase
- Engineering Village

Bibliographic Information

Commenced in 2010.
ISSN: 1847-6996
e-ISSN: 1847-7003
Published: quarterly
Circulation: 300

IJECES online

https://ijeces.ferit.hr

TABLE OF CONTENTS

Sparse and Incomplete Signal Dictionaries for Reconstruction of MR Images

Original Scientific Paper

Deepak M Devendrappa
Department of CSE, K S School of Engineering and Management, Bengaluru.
Affiliated to Visvesvaraya Technological University, Gnana Sangama, Belagavi, India
deepak.m.d@kssem.edu.in

Karthik Pilani
Department of ECE, K S School of Engineering and Management,
Bengaluru, India.
karthik.p@kssem.edu.in

Deepak N Ananth
Department of CSE, R V Institute of Technology and Management,
Bengaluru, India
deepak.n.ananth@gmail.com

Abstract *– Compressed Sensing(CS) is a mathematical approach for data acquisition in which the signals are compressible and sparse w.r.t. to an orthonormal basis. These sparse signals are reconstructed from very less measurements. CS technique Is widely used in Magnetic Resonance Imaging (MRI) where the doctors suggest the patients to undergo MRI scans for diagnosing their body parts. During the prolonged MRI Scan, the exact slice of the MRI cannot be achieved due to the difficulties faced by the patient or irregular changes in the body position of the patient. The idea is to reduce the exposure time of the patient's body against the MRI scan by considering only fewer samples. Is it possible to Reconstruct the signal by making use of a fewer number of samples that are less than the Nyquist rate? Yes, it is possible to reconstruct the signal by making use of the Compressed Sensing or sampling Technique. Compressed sensing is a new framework for signal acquisition and representation in a compressible manner less below the Nyquist sampling rate. In this article, Sampling and reconstruction are dealt here thoroughly as part of the research activity. Compressive Sensing Matching pursuit (CoSaMP) is a novel technique for optimization. It is an iterative approximation method for sparse and incomplete signal recovery. CoSaMP method along with Different transform techniques is used for reconstruction. The FFT_CoSaMP, DCT_CoSaMP and DWT_CoSaMP are proposed methods for MR Image Reconstruction, where DWT-based CoSaMP along with different wavelet families give the best results when compared to other CS-based techniques w.r.t. PSNR, SSIM and RMSE analysis.*

Keywords*: Compressed Sensing, MRI, Nyquist Rate, CoSaMP, Dictionary learning*

1. INTRODUCTION

Compressed sensing (CS) [1] has become an interesting topic in these days for research in the field of mathematical, statistical, electrical and computer sciences, and engineering [2-5]. CS works on the basic fact that signals are represented by using a few non-zero coefficients in an appropriate dictionary or basis. As all of us know to capture a two-dimensional raw image of 256 X 256 size, each pixel will take around 8 bits of storage, the amount of storage required is 256 X 256 X 3 X 8, i.e., approximately, 1.5 Mb of memory. But as we observed, the image is stored in terms of KB. The reason is that image is stored in the compressed form only in terms of 15kb or 50 kb size approximately for example. When you try to open the image, the image is getting reconstructed by making use of some reconstruction algorithms. So that it is possible to look into the image in its original reconstructed form. The idea here is to minimize the number of sensors that are necessary to capture the raw image since the raw image is going to be compressed in its storage level. The notion behind the reduction of the number of sensors is to choose a less quantity of samples during the image acquisition. The number of samples captured violates the Nyquist theorem [11,32]. The Nyquist rate says that the sampling rate must be at least a minimum of two times the maximum frequency of the signal for the exact recovery of the signal. In contrary to this, the CS technique samples the signal at a level far lesser

than the Nyquist rate and it can be restored with high possibility. To compress the signal during the acquisition is the idea behind the compressed sensing and the compressed signal is reconstructed by making use of CS techniques.

Compression techniques which aim to get the concise representation of a signal with tolerable distortion. Some of the signal compression techniques are JPEG, MP3, JPEG2000, MPEG. In all these techniques, the whole signal is captured and duplicate signal data is removed stage by stage.

Donoho [4] challenged the traditional method of acquiring the whole signal and compressing it. He footed his first step into the area of Compressed Sensing or Compressive Sampling (CS). CS is a technique of reconstructing signals well from a fixed set of far fewer non-linear measurements bearing some incoherent characteristics [5,6,7]. Many real-time signals are characterized as sparse in some transforming domains like FFT, DCT, DWT. Since sparse signals exhibit very few non-zero coefficients, they can be represented by some linear measures. Non-linear approximations can approximate the recovery of such signals from a very less number of coefficients. Hence compressed sensing achieves a large reduction in the sampling of the signal and its recovery is achieved by using highly nonlinear methods.

Let's consider a signal x which is said to be sparse if and only if much of the entries in it are zeros. The benefit of having most of the zeros in the signal is that we can ignore them since there is no loss in signal components. Signal of length N exhibiting S number of non-zero entries, then N-S entries will be zero. Sparsity of percentage is given by %sparsity = (N-S)/N*100. The sparse signal can be reconstructed from the compressive measurements with the help of the greedy technique.

This paper discusses the signal acquisition difficulties in many scenarios. We are looking for incomplete signal quantities for the recovery of sparse Magnetic Resonance Images(MRIs). Wavelet transforms (WT) will be seen much like image compression algorithms. It represents much more sparsity of the signal under multiple decomposition levels than other transformation domains. It takes O(n) computations for reversing the signal to its original form. We can re-form the image to its original level if all wavelet sub-bands are roughly sparse [40].

2. PROBLEM STATEMENT WITH COMPRESSED SENSING.

Compressed sensing is a new platform for acquiring signals. CS involves two unique things namely Sparsity and Incoherence which involve three important aspects such as sparse representation, measurement matrix and signal recovery via Compressed Sensing technique.

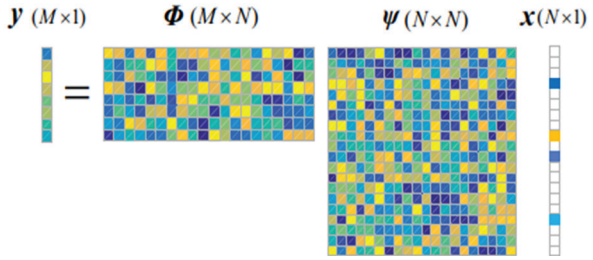


Fig. 1. Representation of Compressed Sensing model for s = 4 number of samples.

Figure 1 describes the block diagram for the CS technique where x be the input signal which is sparse in nature. The sparsity can also be achieved by making use of the transform domain. The sparsification of the image is projected onto the measurement matrix Ψ. The measurement matrix is obtained from the randn() in MATLAB 2020, where n in randn() denotes Normal Gaussian Distribution for random value generation. Consider a signal f € RN which is sparse in some domain Ψ which can be transformed by an NXN matrix Ψ = {Ψ1, Ψ2, Ψ3 - - - ΨN}. Set Ψ is the orthonormal vector bases and X be the coefficient sequence.

$$f = \sum_{i=1}^{N} \psi i x i \quad (1)$$

$$f = \Psi X \quad (2)$$

$$y = Af + e = A \Psi X + e \quad \text{Replace f by (2)} \quad (3)$$

Here y is called as measurement vector of size M, A is measurement matrix M X N and f is K-sparse signal of size N. K represent the total number of nonzero elements of signal f provided K<<M<<N. For Example, measurement matrix A can be an identity matrix, scalar matrix, random matrix or deterministic measurement. e is a noise Vector.

3. TRANSFORMATION DOMAINS

There are so many transform domains that transform the domain of the signal from one domain to another, say for example, from a frequency domain to a special domain. The Discrete Cosine Transforms (DCT) [23, 24, 26], Fast Fourier Transforms (FFT) [25] and Discrete Wavelet Transforms (DWT) [27, 28] are among the some of the transform domains. The idea behind these transformations is that we can approximate the signal in terms of its coefficients. The coefficients are in the form of low significant components and high significant components. The low significant components are the smooth areas of the image and the high significant components are the abrupt areas of the image. Low significant components hold much information about the signal/image and are highly required. These are called low-frequency components. The high significant components are less significant, less required and finally discarded. In the case of FFT, the low-frequency components are existing at 4 different corners of the image and their concertation

diminishes towards the center of the image consisting of high-frequency components. FFT is described by the phase and magnitude of the signal. FFT brings about energy compaction. In DCT, they are existing only at the upper left corner of the image and their concentration diminished towards the other three corners of the image. While in the case of DWT, it brings the transformation into 4 different sub-bands. They are High High (HH), High Low (HL), Low High (LH) & Low Low (LL). The LL components none other than our low-frequency components are exhibiting much information about the image. Wavelet transforms are used to identify both frequency and special components of the signal. Therefore, wavelet transformation is called multiresolution analysis. It is also called the next version of the Short Term Fourier Transformation (STFT). In a wavelet, the window itself is used as a wavelet. Here the window can be moved and can also be scaled. In wavelet, two types of windows are available namely wavelet window (wavelet function (ψ)) and scaling window (scaling function (Φ)).

DCT [24] and it's inverse of the equation for a 2D input signal f (x, y) is shown below by equations (4) and (5) respectively.

$$D(i,j) = 1/\sqrt{2N}(C[i]C[j] \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} p(x,y) \cos\left[\frac{(2x+1)i\pi}{2N}\right] \cos\left[\frac{(2y+1)i\pi}{2N}\right] \quad (4)$$

$$f(i,j) = {}^{2}/_{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u), C(v), F(u,v) \cos\left[\frac{(2x+1)i\pi}{2N}\right] \cos\left[\frac{(2y+1)i\pi}{2N}\right] \quad (5)$$

where C(u) = C(v) = 1/√ (2)) for u, v = 0. Otherwise, C(u) = C(v) = 1. p (x, y) is the element in the image represented by the matrix of size M X N. DCT is applied on block size N. F(u, v) is the transformed Image of f(x, y). In DCT, sine components are put in intact. The inverse of DCT is used for converting the transformed signal D(i,j) into the original signal f(i, j).

FFT [33] and its inverse are given by the following equations (6) add (7) respectively.

$$F(\omega) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx \quad (6)$$

$$f(x) = 1/2\pi \int_{-\infty}^{\infty} F(\omega) e^{i\omega x} d\omega \quad (7)$$

Where i= √(-1), e^iω=sin(ω)+i cos(ω). F(ω) is the FFT transformed signal, ω is the frequency and f(x) denotes the input image and x represents the time. The inverse of FFT is used for converting the transformed signal into the original signal.

The Wavelet transforms [34,35] is given by the following equation as follows.

$$F_{HighPass}(m,n) = \int_{-\infty}^{\infty} f(x) \Psi^{*}_{(m,n)}(t) dt \quad (8)$$

$$F_{LowPass}(m,n) = \int_{-\infty}^{\infty} f(x) \Phi^{*}_{(m,n)}(t) dt \quad (9)$$

$F_{HighPass}$ (m,n) represents the high pass filter to pass through high-frequency components and $F_{LowPass}$ (m,n) represents the low pass filter to pass through low-frequency components. * is the conjugate symbol. 'Ψ' represents the wavelet function given by equation (8) and 'Φ' represents the scaling function given by equation (9). The input signal is represented by f(x).

4. RELATED WORK

There are several reconstruction algorithms based on Compressed Sensing [20]. One among the CS Techniques is Orthogonal Matching Pursuit (OMP) which is a greedy or iterative method for solving compressed sensing problems. The greedy algorithm provides some sort of iterative estimation of the signal coefficients until a halting criterion met. The greedy algorithms provide performance guarantees when analyzed with convex optimization techniques [8]. OMP and *Iterative Thresholding* are the older methods of greedy approaches. OMP [9] starts by identifying the column of A correlated with the measurement matrix. This step is repeated in the algorithm by correlating the column with signal residuals. These signals are obtained by taking away the partial estimates of the signal from the original measurement vector. The stopping criterion is the number of iterations that limit the total number of nonzero in input signal f, such that y=Af. Where A is the measurement matrix. Iterative Thresholding [10] is a direct method. The signal is initially estimated as f'=0. The gradient descent step is iterated followed by hard thresholding till a stopping criterion happens.

The objective quality analysis for the reconstruction is calculated and verified by using SSIM (Structural Similarity index), PSNR (Peak Signal Noise Ratio), and RMSE (Root Mean Square Error).

The PSNR [29] is given by the equation (10),

$$\text{PSNR} = 20 \log(\max(f)/ \sqrt{(\text{MSE})}) \quad (10)$$

where f is the maximum signal component of the 2-dimensional signal.

MSE [30] is nothing but the Mean Squared Error which is given equation (11),

$$\text{MSE} = (1/(m*n)) * \text{sum}(\text{sum}((f-g).\hat{}2)) \quad (11)$$

Here m and n are dimensions of the 2D signal. f is the estimated signal and g is called an actual signal.

RMSE [31] is given by equation (12),

$$\text{RMSE} = \sqrt{((\text{sum}(y`-y)2))/2)} \quad (12)$$

where y is the estimated signal and y is the actual signal.

Finally, SSIM [30] is given by equation (13),

$$\text{SSIM}(a,b) = [l(a,b)]p \,.\, [c(a,b)]q \,.\, [s(a,b)]r \quad (13)$$

Here $l(a,b)=(2\mu_a\mu_b+C_1)/(\mu_a^2+\mu_b^2+C_1)$, $c(a,b)=(2\sigma_a\sigma_b+C_2)/(\sigma_a^2+\sigma_b^2+C_2)$, $s(a,b)=(\sigma_{ab}+C_3)/(\sigma_a\sigma_b+C_3)$.

Here where μ_a, μ_b, σ_a, σ_b, and σ_{ab} are the means, standard deviations, and cross-covariance for 2 dimensional signals a, b. If p = q = r = 1 (the default for Exponents), C_1 and C_2 are constants and $C_3 = C_2/2$.

5. COSAMP

The CoSaMP [32, 35] is a Compressed Sensing Technique based on OMP. It should accept the following conditions.

1. It should proceed with a minimum number of samples.
2. It should proceed by considering the samples from all different sampling schemes.
3. It should accept all the samples that are amalgamated with the noise and should be robust.

5.1 ORTHONORMAL BASES

Let consider a set B. B is having a set of vectors i.e B = $\{v_1, v_2, \ldots\ldots v_k\}$ and each of these vectors have the length 1. $||vi||=1$for i=1,2, …k. or $||vi||^2 = 1$ or $v_i \cdot v_i = 1$ for i = 1,2,3……k. All these vectors in B have magnitude or length 1 i.e. they are normalized and are unit vectors. The dot product of any two vectors in the set is 0 i.e. $v_i \cdot v_j = 0$ for i≠j. All of these vectors are orthogonal to each other.

So $v_i . v_j = \begin{cases} 0 \; for \; i \neq j \\ 1 \; for \; i = j \end{cases}$

All of these sets have magnitude or length 1 and are orthogonal to each other. They are normalized and orthogonal. So the set B is called as an orthonormal set. Ortho means every member of the set is orthogonal and everything is normalized. If there is some subspace V,

V=span(v_1, v_2, v_3, …….v_k), then B is an orthonormal basis for V. CoSaMP has received broad attention in these days due to one of Compressed Sensing technique. The CoSaMP algorithm accepts a vector of K-sparse signal.

a. It should accept the measurement matrix.
b. It should also accept the stopping criterion as the number of iterations increases, the signal quality diminishes and this is identified as the stopping criterion.

The CoSaMP is a greedy matching pursuit algorithm. Let A be the m X N sampler matrix with constant Restricted Isometry Constant (RIC) $\delta 2_s <= c$ [3]. RIC characterizes matrices which are appearing to be orthonormal and are extensively used in the field of compressed sensing. Let y=Af + e be the observed vector of samples with arbitrary signals of K-sparse amalgamated with e noise. With precision parameter έ, CoSaMP produces k-sparse estimation that is fulfilled by the equation (14).

$$||\mathbf{f}-\mathbf{a}||_2 <= C.\max\{\acute{\varepsilon}, \mathbf{1}/\sqrt{(s)}||\mathbf{f}-\mathbf{f}_{s/2}||_1+||\mathbf{e}||_2$$

Here, $f_{s/2}$ is the $(_{s/2})$ sparse estimate to f. The running time $O(\omega .\log(||f||_2 / \acute{\varepsilon}))$ where ω bounds the cost of the matrix vector multiplied with A. $||f||_2$ is l_2 normalization. c<1 and C>1 are whole constants. CoSaMP algorithm is given below.

5.2 CoSAMP: ALGORITHM

Input: A, y, s (k-sparse signal), where y= Af+e, K non-sparse signal components and e is the noise.

Output: f`: k sparse approximation of f.

Variable initialization : let $f_0 \leftarrow 0$, $y_0 \leftarrow y$, $t \leftarrow 0$

Condition: while stopping criterion does not hold, do

1. Increment the number of repetitions: $t \leftarrow t + 1$
2. Compute an intermediates: $\delta \leftarrow A^* y0$
3. Identify the main substitutions: $B \leftarrow \sup p(\delta_{2s})$
4. Support merger:$Bt \leftarrow B \cup \sup p(f^{t-1})$
5. Signal estimation by least squares: $^{b}|B^t \leftarrow (A_T^t)+B^{tb}|B^c \leftarrow 0$
6. Prune to obtain the next estimation: $f^t \leftarrow b_k$
7. Update the samples:$y_0 \leftarrow y - Af^t$

End while

$f` \leftarrow f^t$

6. PROPOSED METHODS

The proposed CoSaMP based methods for the reconstruction of Magnetic Resonance Images (MRI) include FFT based CoSaMP (FFT_CoSaMP), DFT based CoSaMP (DCT_CoSaMP) and DWT based CoSaMP (DWT_ CoSaMP). The analysis, experimentation and comparative study of these proposed methods have been done in section 6.1 below.

6.1 EXPERIMENTATION AND ITS RESULTS

The comparative study of different compressed sensing techniques is done with the Proposed techniques. Our proposed techniques are giving good results compared to other techniques as depicted in table 1, table 2 and table 3 for PSNR, SSIM and RMSE respectively. The PSNR value of the proposed methods would be like FFT_CoSaMP of 36.01, DCT_CoSaMP of 36.33 and DWT_CoSaMP of 37.16 (db4), 36.12(coif3) and 38.5 (Sym8) for a measurement matrix of 210 X 256 of input image size of 256 X 256 as in table 1. These PSNR values are validating that our proposed DWT_CoSaMP method is the best compared to other methods. Similarly, SSIM value 0.81 of the proposed method DWT_ CoSaMP is better as compared to other CS Based techniques like FFT_CoSaMP of 0.7, DCT_CoSaMP of 0.73 and other methods of CS as in table 2. In Case of DWT_ CoSaMP, the SSIM value for Coif3 is 0.68, for db4 of 0.71 and for haar of 0.81 for a measurement matrix of 210 X 256 of the input image size of 256 X 256. In the same way, RMSE value of the proposed method DWT_CoS-

aMP 0.44 achieves good results when compared to other CS-based techniques like FFT_CoSaMP is 0.66, DCT_CoSaMP is 0.55 and other methods as shown in table 3.

The System Requirements for conduction of simulation experiment on CoSaMP based CS technique will consist of system of Windows 10 Operating System, secondary memory of 500GB, 2.2 GHz speed, RAM is 6 GB, and MATLAB software 2020.

Table 1. PSNR based comparison for CoSaMP named Compressed Sensing technique

Sl. No.	References	Method	PSNR
1	Improved CoSaMP Reconstruction Algorithm Based on Residual Update[12]	COSaMP	25.93
2	A Study on Image Reconfiguration Algorithm of Compressed Sensing [13]	St.OMP	28.3
3	Adaptive gradient-based block compressive sensing with sparsity for noisy images [14]	AGbBCS SP	23..73
4	Variable atomic number MP for Image restoration [15]	CoSaMP	35.5875
5	An Improved Off-Grid Algorithm Based on CoSaMP for ISAR Imaging[[35]	CoSaMP	30.01
6	Constrained Backtracking Matching Pursuit Algorithm for Image Reconstruction in Compressed Sensing[36]	CBMP	34.04
7	Proposed method	FFT_CoSaMP	36.01
		DCT_CoSaMP	36.33
		DWT_CoSaMP	37.16 (DB4)
			38.12 (Coif3)
			38.5 (Sym8)

Table 1 depicts the comparison of PSNR values that were experimented with some research articles under different Compressed Sensing techniques. Among all the references, the proposed methods FFT_CoSaMP, DCT_CoSaMP and DWT_CoSaMP achieve better PSNR results concerning referenced results.

Similarly, Table 2 illustrates the comparison of SSIM values that were experimented with some other research articles under different compressed sensing techniques. Among all the references, the proposed method FFT_CoSaMP, DCT_CoSaMP and DWT_CoSaMP give the best SSIM results concerning referenced results.

In the same way, Table 3 demonstrates the comparison of RMSE values that were experimented with various research articles under different compressed Sensing techniques. Among all these references, the proposed method FFT_CoSaMP, DCT_CoSaMP and DWT_CoSaMP give good results concerning referenced results.

Table 2. SSIM based comparison for CoSaMP named Compressed Sensing technique

Sl. No.	References	Method	SSIM
1	Diffuse optical tomography image reconstruction via greedy algorithm [16]	CoSaMP	0.5443
2	Quantitative Comparison Of Reconstruction Methods For Compressive Sensing MRI [37]	CoSaMP	0.8
3	Recovery from compressed measurements using Sparsity Independent Regularized Pursuit, Signal Processing[38]	SIRP	0.75
4	Sparse recovery based compressive sensing algorithms for diffuse optical tomography[17]	CoSaMP	0.74
5	CS-based MRI image reconstruction via quantitative Comparison [18]	CoSaMP	0.79
6	DCT based CS recovery Strategies' in medical imaging[19]	DCT based Method	0.80
7	Proposed method	FFT_CoSaMP	0.7
		DCT_CoSaMP	0.73
		DWT_CoSaMP	0.81

Table 3. RMSE based comparison for CoSaMP named Compressed Sensing technique

Sl. No.	References	Method	RMSE
1	A novel compressive sampling method for ECG wearable measurement systems[39]	DBBD +DCT	3.8
2	Sparse Reconstruction Off-grid OFDM Time Delay Estimation Algorithm Based on Bayesian Automatic Relevance Determination[21]	FFT Expectation maximization (EM) algorithm	3.1
3	An improved statistical iterative algorithm for sparse-view and limited-angle CT image reconstruction [22]	Median and Wiener filtering algorithms	0.91
4	Proposed method	FFT_CoSaMP	0.66
		DCT_CoSaMP	0.55
		DWT_CoSaMP	0.44

Table 4 shows the analysis of various measurement matrices for the Thorax MRI image reconstruction. The measurement matrices of higher dimensions give comparatively worthy results than the lower dimensions of the measurement matrix. Here the dimensions 210X256 are the best measurement matrix for our experimental simulations. For the measurement matrix 210 x 256, DFT based CoSaMP gives PSNR of 54.95, SSIM of 0.69 and RMSE of 0.45. For the same 210 X 256 measurement matrix, the FFT-based CoSaMP gives the PSNR of 52.5, SSIM of 0.66 and RMSE of 0.45. But in the case of

DWT-based CoSaMP, for the same measurement matrix of 210 XX 256, PSNR is 57.63, SSIM is 0.71 and RMSE is 0.33. Under different wavelet families, DWT-based CoSaMP with db4, haar, sym8, and coif3 gives the PSNR values 54.16, 53.12, 55.18, and 55.12 respectively.

Table 4. Analysis of measurement matrix for MRI image reconstruction for the thorax image.

CS Technique for Thorax	Measurement matrix	PSNR	SSIM	RMSE
DCT based CoSaMP	5 X 256	27.1	0.05	11.25
	30X256	35	0.13	4.75
	55X256	37.12	0.21	3.52
	80X256	40.1	0.29	2.82
	105X256	42.23	0.37	1.6
	130X256	46.23	0.46	1.75
	155X256	47.8	0.54	1.02
	180X256	50.23	0.6	0.75
	210X256	54.95	0.69	0.45
FFT based CoSaMP	5 X 256	27.2	0.1	10
	30X256	33.62	0.15	3.52
	55X256	38.12	0.286	2.13
	80X256	42.67	0.39	1.75
	105X256	44.12	0.47	1.62
	130X256	46.2	0.52	1.52
	155X256	47.5	0.56	1.23
	180X256	48.6	0.59	0.75
	210X256	52.5	0.66	0.45
DWT based CoSaMP	5 X 256	24.5	0.02	15.75
	30X256	26.8	0.12	11.53
	55X256	32.1	0.24	6.89
	80X256	42.9	0.38	1.98
	105X256	46.23	0.48	2.05
	130X256	49.1	0.56	1.12
	155X256	51.1	0.59	0.89
	180X256	52.8	0.36	0.52
	210X256	57.63	0.71	0.33
	210 X 256	db4 54.16	-	-
		haar 53.12	-	-
		sym 55.18	-	-
		coif3 55.12	-	-

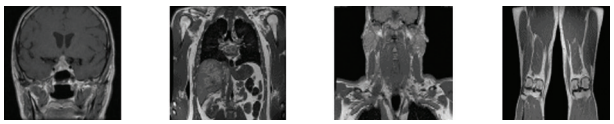


Fig. 2. MRI images of a) Brain, b) Lungs c) Thorax d) Thighs

Figure 2 shows different MRI images like a) Brain, b) Lungs c) Thorax d) Thighs The experiential results are tabulated above for the measurement matrix of 210 X 256. The tabulated values show the better reconstruction and analysis of MRI images. The average values at the bottom of the table 5 for PSNR, SSIM and RMSE give the better analysis and approximation and reconstruction of the image. And figure 3 shows the simulated graphs of PSNR, SSIM, RMSE values against different measurement matrices of Table 4.

In our experimental work, about 250 MRI images are used for experimentation by making use of a measurement matrix of 256 X 210. Among 250 images, 35 images are filtered and discussed in this research work. The below tabulation shows the statistical values of the FFT_CoSaMP, DCT_CoSaMP, and DWT_CoSaMP. The images used for experimentation are of some cancerous and non-cancerous human brain MRI images. The cancerous images are named as Yes and non-cancerous images are named as No.

Other than human brain images, we have used some of the other MRI images of human lungs, Thighs, and Thorax. The outcome for these images concerning their PSNR, PSNR, SSIM, and RMSE are tabulated in Table 5. Here the thorax MRI image is giving good results concerning PSNR, SSIM and RMSE values when compared to other MRI images. And also the average values for the same are computed, tabulated and compared as in Table 5.

Table 5. The proposed CS Techniques FFT_ CoSaMP, DCT_CoSaMP, DWT_ CoSaMP have been experimented with for 250 images and only 35 images are shown in the table images of Brain, Lungs, Thorax and Legs.

Sl. No.	MRI Image No- NonCancerous Yes-Cancerous	PSNR FFT_ CoSaMP	PSNR DCT_ CoSaMP	PSNR DWT_ CoSaMP
1	No 1	28.97	30.02	32.83
2	No 2	26.34	27.27	32.54
3	No 3	22.86	24.57	32.91
4	No 4	30.96	31.86	36.12
5	No 5	27.1	27.92	33.15
6	No 6	24.53	25.4	29.41
7	No 7	33.29	34.27	36.27
8	No 8	33.3	34.25	36.28
9	No 9	30.33	31.57	39.72
10	No 10	30.4	32.03	36.99
11	No 11	30.09	32.6	36.87
12	No 12	23.27	24.85	29.19
13	No 13	25.76	28.81	31.59
14	No 14	25.78	27.18	34.18
15	No 15	25.46	26.89	33.27
16	No 16	27.06	28.07	33.33
17	No17	20.28	22.54	29.47
18	No 18	27.45	28.61	32.25
19	No 19	26.05	27.21	29.36
20	No 20	23.24	24.88	28.44
21	No 21	29.16	31.03	33.89
22	No 22	23.59	24.63	28.14
23	No 23	26.02	27.18	30.68
24	No 24	25.81	27.08	34.1
25	No 25	26.41	29.1	33.04
26	No 38	24.74	25.26	38.42
27	No 94	28.2	29.07	37.49

28	No 95	28.88	31.44	39.66
29	No 96	29	30.64	38.35
30	Yes 102	28.08	28.36	37.26
31	Yes 103	28.28	29.35	32.9
32	Yes 128	32.58	34.44	37.67
33	Lungs	30.87	32.96	34.01
34	Thighs	34.71	36.86	40.1
35	Thorax	55.46	55.02	57.62
	Average	**28.98**	**30.02**	**34.72**

Sl. No.	MRI Image No- NonCancerous Yes-Cancerous	SSIM		
		FFT_ CoSaMP	DCT_ CoSaMP	DWT_ CoSaMP
1	No 1	0.5	0.5	0.67
2	No 2	0.45	0.44	0.58
3	No 3	0.34	0.35	0.51
4	No 4	0.38	0.38	0.6
5	No 5	0.42	0.41	0.51
6	No 6	0.45	0.45	0.58
7	No 7	0.6	0.6	0.65
8	No 8	0.6	0.6	0.65
9	No 9	0.25	0.25	0.53
10	No 10	0.34	0.34	0.49
11	No 11	0.35	0.37	0.6
12	No 12	0.46	0.47	0.57
13	No 13	0.62	0.65	0.71
14	No 14	0.45	0.45	0.57
15	No 15	0.38	0.39	0.47
16	No 16	0.39	0.38	0.5
17	No17	0.47	0.49	0.01
18	No 18	0.42	0.43	0.64
19	No 19	0.51	0.54	0.57
20	No 20	0.59	0.61	0.7
21	No 21	0.59	0.61	0.52
22	No 22	0.42	0.43	0.54
23	No 23	0.62	0.63	0.73
24	No 24	0.45	0.45	0.57
25	No 25	0.6	0.62	0.71
26	No 38	0.42	0.4	0.73
27	No 94	0.34	0.33	0.58
28	No 95	0.32	0.33	0.5
29	No 96	0.36	0.36	0.56
30	Yes 102	0.62	0.54	0.83
31	Yes 103	0.66	0.66	0.73
32	Yes 128	0.7	0.7	0.8
33	Lungs	0.76	0.79	0.81
34	Thighs	0.56	0.57	0.59
35	Thorax	0.66	0.69	0.71
	Average	**0.487**	**0.5**	**0.59**

Sl. No.	MRI Image No- NonCancerous Yes-Cancerous	RMSE		
		FFT_ CoSaMP	DCT_ CoSaMP	DWT_ CoSaMP
1	No 1	7.91	6.92	5.17
2	No 2	6.34	6.75	4.26
3	No 3	9.43	4.63	3.82
4	No 4	5.41	6.59	4.3
5	No 5	7.26	5.51	4.23
6	No 6	6.72	6.89	5.3
7	No 7	6.32	5.23	4.7
8	No 8	8.31	7.32	5.39
9	No 9	6.82	6.81	2.63
10	No 10	6.52	6.31	3.91
11	No 11	7.23	5.5	5.01
12	No 12	9.51	7.29	3.71
13	No 13	8.72	9.23	6.88
14	No 14	6.32	6.23	3.96
15	No 15	6.23	7.21	4.26
16	No 16	7.12	8.61	5.23
17	No17	8.56	7.69	4.28
18	No 18	7.26	6.23	5.36
19	No 19	7.12	6.23	2.35
20	No 20	5.63	5.64	4.32
21	No 21	9.23	7.25	4.5
22	No 22	7.91	6.29	5.23
23	No 23	7.23	5.23	3.85
24	No 24	9.63	7.63	6.32
25	No 25	7.52	4.23	3.25
26	No 38	6.85	5.98	4.35
27	No 94	6.35	3.54	3.54
28	No 95	7.51	6.81	2.6
29	No 96	6.32	6.23	4.65
30	Yes 102	10.71	6.6	3.5
31	Yes 103	7.98	6.25	5.91
32	Yes 128	4.91	4.81	3.21
33	Lungs	6.44	6.44	5.12
34	Thighs	2.5	2.5	2.5
35	Thorax	0.45	0.45	0.33
	Average	**7.38**	**6.31**	**4.29**

Table 6. The CoSaMP based CS technique along with Different types of dictionaries like DFT, FFT and DWT have experimented for four different images of the Brain, Lungs, Thorax and Legs.

MRI image	Proposed Method	PSNR	SSIM	RMSE
Brain	FFT_CoSaMP	31.2	0.698	4.52
	DCT_CoSaMP	33.2	0.623	6.20
	DWT_CoSaMP	36.6	0.754	4.83
Lungs	FFT_CoSaMP	30.87	0.754	6.44
	DCT_CoSaMP	32.96	0.787	6.44
	DWT_CoSaMP	34.2	0.808	5.12
Thorax	FFT_CoSaMP	52.5	0.665	0.66
	DCT_CoSaMP	54.95	0.646	0.66
	DWT_CoSaMP	57.63	0.665	0.44
Thighs	FFT_CoSaMP	34.7	0.56	3.75
	DCT_CoSaMP	36.76	0.57	3.76
	DWT_CoSaMP	40.26	0.58	2.475

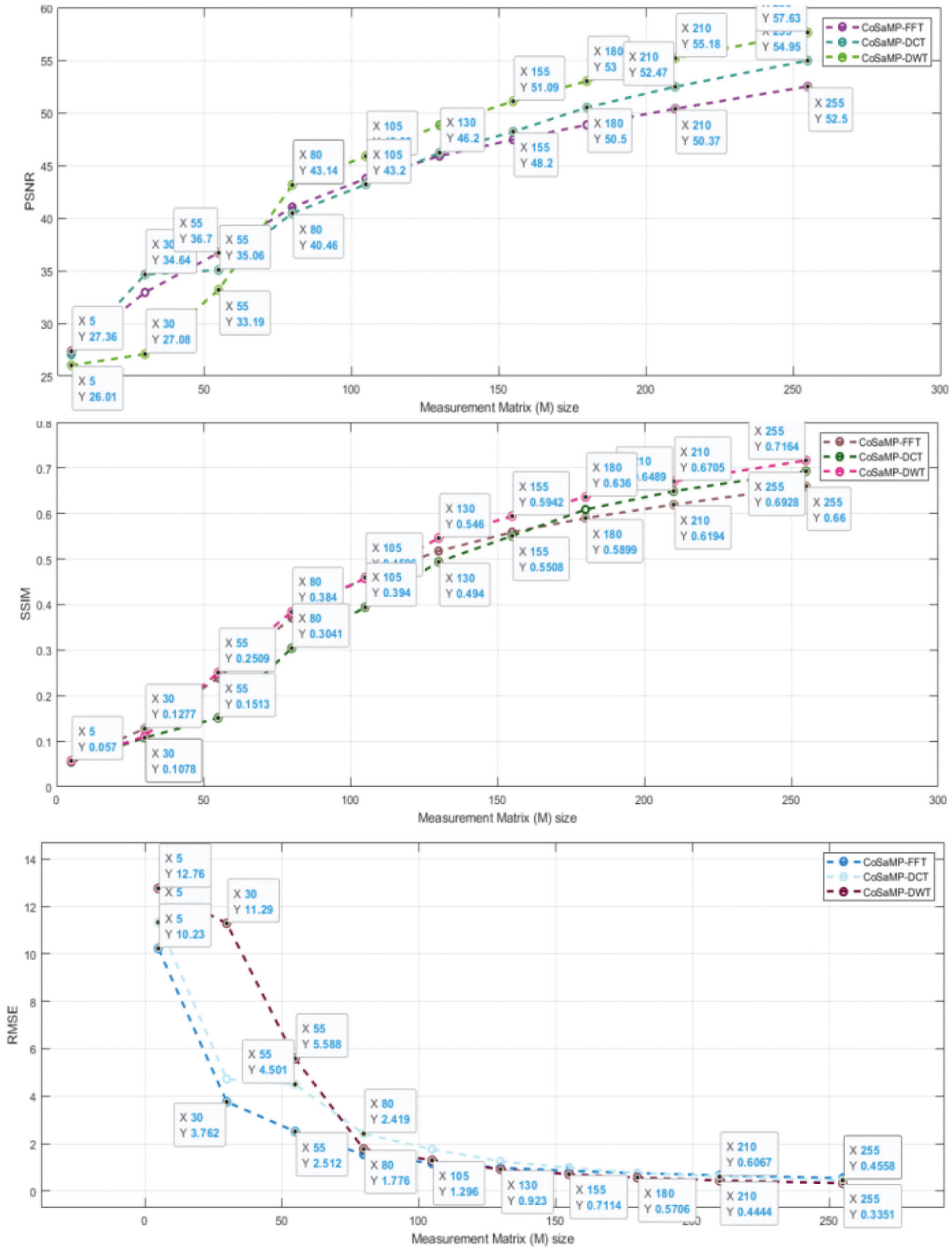


Fig. 3. PSNR, SSIM and RMSE based Simulated Graphs for the reconstruction of Thorax Image.

7. CONCLUSION AND FUTURE ENHANCEMENT

For restoration of MRI images, based on different traditional methods and proposed Compressed Sensing methods such as FFT_CoSaMP, DCT_CoSaMP and DWT_CoSaMP, DWT_CoSaMP is giving best results concerning SSIM, RMSE and PSNR statistical values. When DWT_CoSaMP is passed with different dictionaries of Wavelets families like coif3, db4 haar and sym8, the result is said to be likely good for all the families of wavelet transforms. For Further future scope, CS-based methods of reconstruction will be applied for the investigation of music and speech signals by making use of different types of sparse dictionary parameters.

8. ACKNOWLEDGMENT

I would like to thank Visvesvaraya Technological University (VTU), Gnana Sangama, Belagavi - 590018, Karnataka, India for the monetary support extended to my research work.

9. REFERENCES:

[1] D. L. Donoho, "Compressed sensing", IEEE Transactions on Information Theory, Vol. 52, No. 4, 2006, pp. 1289-1306.

[2] G. Yang et al. "DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," IEEE Transactions on Medical Imaging, Vol. 37, No. 6, 2018, pp. 1310-1321.

[3] Y. Zhang B. P. Wang Y. Fang, Z. X. Song, "SAR Observation Error Estimation Based on Maximum Relative Projection Matching", International Journal of Antennas and Propagation, 2020.

[4] D. L. Donoho, "Compressed sensing", IEEE Transactions on Information Theory, Vol. 52, 2006, pp. 1289-1306.

[5] M.B Cand'es, E. J. Wakin, "An Introduction to Compressive Sampling", Signal Processing Magazine IEEE, Vol.2, No. 5, 2008.

[6] E. Candes, J. Romberg, "Sparsity and incoherence in compressive sampling", Inverse Problems, Vol.23, 2007, p. 969.

[7] E. Candes, J. Romberg, T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information", IEEE Transactions on Information Theory, Vol.52, No.2, 2006, p. 489.

[8] M. Lustig, D. Donoho, John M. Pauly. "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging", Magnetic Resonance in Medicine, 2007, pp. 1182 1195.

[9] S. Mallat, Z. Zhang. "Matching pursuits with time-frequency dictionaries", IEEE Transactions on Signal Processing, Vol. 41, No. 12, 1993, pp. 3397-3415.

[10] M. Elad, B. Matalon, J. Shtok, M. Zibulevsky. "A wide-angle view at iterated shrinkage algorithms", Proceedings of the SPIE Optics Photonics: Wavelets, San Diego, CA, USA, August 2007.

[11] H. J. Landau", Sampling, data transmission, and the Nyquist rate", Proceedings of the IEEE, Vol. 55, No. 10, 1967, pp. 1701-1706.

[12] D. Lu, G. Sun, Z. Li, S. Wang, "Improved CoSaMP Reconstruction Algorithm Based on Residual Update" Journal of Computer and Communications", Scientific research publishing, Vol.7 No.6, 2019.

[13] Y. Zhang, D. Wang, L. Kan, P. Zhao, "A Study on Image Reconfiguration Algorithm of Compressed Sensing", TELKOMNIKA, Vol.15, No.1, 2017, pp. 299-305.

[14] H.-H. Zhao, P. L. Rosin, Y.Lai, J. Zheng, Yao-Nan Wang, "Adaptive gradient-based block compressive sensing with sparsity for noisy images", Multimedia Tools and Applications, Vol. 79, 2019, pp. 14825-14847.

[15] H. Zhao, C. Chen, C. Shi, "Image reconstruction algorithm based on variable atomic number matching pursuit", Journal of Algorithms & Computational Technology, Vol. 11, No. 2, 2017, pp. 103-109.

[16] B. P. V. Dileep, T. Das, K. Dutta "Greedy algorithms for diffuse optical tomography reconstruction", Optics Communications, Vol. 410, No. 1, 2018, pp. 164-173.

[17] B. P. V. Dileep, T. Das, K. Dutta, "Sparse recovery based compressive sensing algorithms for diffuse optical tomography", Optics & Laser Technology, Vol. 128, 2020.

[18] T. S. Kavita, K. S. Prasad, "Quantitative Comparison of Reconstruction Methods for Compressive Sensing MRI", International Journal Of Scientific & Technology Research Vol. 8, No. 10, 2019.

[19] Amira S, Ashour Yanhui Guo, Eman Elsaid Alaa, Hossam M.Kasem, "Discrete cosine transform based compressive sensing recovery strategies in medical imaging", Advances in Computational Techniques for Biomedical Image Analysis, Academic Press, 2020.

[20] R. Zhang, C. Meng, C. Wang, Q. Wang, "Compressed Sensing Reconstruction of Radar Echo Signal Based on Fractional Fourier Transform and Improved Fast Iterative Shrinkage-Thresholding Algorithm" Wireless Communications and Mobile Computing, 2021.

[21] P. Zhang, B. Ba, Y. Zhang, P. Han, "Sparse Reconstruction Off-grid OFDM Time Delay Estimation Algorithm Based on Bayesian Automatic Relevance Determination", Journal of Physics: Conference Series, Vol. 1237, No. 2, 2019.

[22] Z. Hu et al. "An improved statistical iterative algorithm for sparse-view and limited-angle CT image reconstruction". Scientific Report, Vol. 7, 2017.

[23] N. Ahmed, T. Natarajan, K. R. Rao, "Discrete cosine transform," IEEE Transactions on Computers, Vol. C-23, 1974, pp. 90-93.

[24] J. W. Cooley, P. A. W. Lewis, P. D. Welch, "The Fast Fourier Transform and Its Applications", IEEE Transactions on Education, Vol. 12, No. 1, 1969, pp. 27-34.

[25] Z. Wu; J. Sha; Z. Wang; L. Li; M. Gao, "An improved scaled DCT architecture", IEEE Transactions on Consumer Electronics, Vol. 55, No. 2, 2009, pp. 685-689.

[26] S. G. Mallat, S. Zhong, "Characterization of signals from multiscale edges", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 7, 1992, pp. 710-732.

[27] A. N. Akansu, W. A. Serdijn, I. W. Selesnick, "Wavelet Transforms in Signal Processing: A Review of Emerging Applications", Physical Communication, Vol. 3, No. 1, 2010, pp. 1-18.

[28] C. A. Chalaperumal, S. D. kumar, "An Efficient Image Quality Enhancement using Wavelet Transform", Materials Today: Proceedings, Vol. 24, Part 3, 2020.

[29] V. Bruni, D. Vitulano, "An entropy-based approach for SSIM speed up", IEEE Signal Processing, Vol. 135, 2017, pp. 198-209.

[30] J. F. Kenney, E. S. Keeping, "Root Mean Square." Mathematics of Statistics, Princeton, NJ, USA, pp. 59-60, 1962.

[31] D. Needell, J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples", Applied and Computational Harmonic Analysis, Vol. 26, 2009, pp. 301-321.

[32] E. Ayanoglu, "Data transmission when the sampling frequency exceeds the Nyquist rate", IEEE Communications Letters, Vol. 1, No. 6, 1997.

[33] P. Heckbert, "Fourier Transforms and the Fast Fourier Transform (FFT) Algorithm", Computer Graphics 2, 1998.

[34] J. D. Villasenor, B. Belzer, and J. Liao "Wavelet filter evaluation for image compression", IEEE Transactions on Image Processing, Vol. 4, 1995, pp. 1053-1060.

[35] J. Cheng "An Improved Off-Grid Algorithm Based on CoSaMP for ISAR Imaging", Proceedings of the 5th International Conference on Mechanical, Control and Computer Engineering, Harbin, China, 25-27 December 2020.

[36] X. Bi, L. Leng, C. Kim, X. Liu, Y. Du, F. Liu, "Constrained Backtracking Matching Pursuit Algorithm for Image Reconstruction in Compressed Sensing". Applied Science, Vol. 11, 2021, p. 1435.

[37] T. S. Kavita, K. S. Prasad, "Quantitative Comparison of Reconstruction Methods For Compressive Sensing MRI", International Journal Of Scientific & Technology Research, Vol. 8, No. 10, 2019.

[38] T. J. Thomas, J. S. Rani, "Recovery from compressed measurements using Sparsity Independent Regularized Pursuit", Signal Processing, Vol. 172, 2020.

[39] F. Picariello, G. Iadarola, E. Balestrieri, I. Tudosa, L. De Vito, "A novel compressive sampling method for ECG wearable measurement systems", Measurement, Vol. 167, 2021.

[40] Y. Liu, X. Chen, A. Liu, R. K. Ward, Z. J. Wang, "Recent Advances in Sparse Representation Based Medical Image Fusion," IEEE Instrumentation & Measurement Magazine, Vol. 24, No. 2, 2021, pp. 45-53.

Correlation-Based Template Tracking Of Moving Object

Original Scientific Paper

Hema Tekwani
Harcourt Butler Technical University,
Research Scholar, Department of Electronics Engineering
Kanpur, India
hematekwani19@gmail.com

Krishna Raj
Harcourt Butler Technical University,
Professor and Head, Department of Electronics Engineering
Kanpur, India
kraj_biet@yahoo.com

Abstract *– This paper presents the correlation-based motion estimation technique for the 3D displacement of objects. Two high-speed cameras are configured as a stereovision system and synchronized in real-time. Finger and hand motions are captured in form of digital images at 1500 fps and 2000 fps respectively. A complete motion acquisition system is calibrated to determine the intrinsic and extrinsic parameters which were later used in the correlation algorithm. The grayscale image frames acquired from the cameras are correlated using square templates of 10x10 pixels created from the reference image. The finger and hand motion are discussed with varying camera speed as a measure of brightness inconsistency. The observations in the correlation coefficient indicate that the proposed algorithm is efficient up to 20 and 50 templates for the finger and hand motion cases respectively. The correlation coefficient for finger motion was increased to 0.987 and 0.972 for the left and right cameras, respectively, while the correlation coefficient for hand motion was 0.924 and 0.898. The proposed algorithm is developed in MATLAB and validated by tracing the sinusoidal motion of a solid rectangular element from the image correlation technique and an accelerometer sensor mounted over the block.*

Keywords*: Grayscale Speckling, Motion Estimation, Stereo-Rig, Template, Template Tracking Coefficient, Thresholding*

1. INTRODUCTION

Flow-based motion estimation is a well-established method for tracking moving object features. The explicit application of flow-based methods in feature matching makes real-time motion estimation difficult. Since image information does not vary significantly from frame to frame, real-time motion analysis is critical in image processing. This further becomes extremely interesting to find the trajectory of moving objects with small displacements without altering the image quality. To find the best match frame, a full search technique [1] is employed in which sub-bands are used to separate the images according to their spatial and temporal orientation [2], [3]. Flow-based feature tracking has received a lot of attention, and feature-based approaches have become popular for describing distinctive visual features [4]. Computations for local motions are done in the vector field, where a motion patch [5] represents the motion, various feature attributes were included in previous motion assessment studies. The motion-tracking algorithm is developed and modified with spatial and temporal information to restore any missing feature point in tracking [6]. Another advancement was alignment in the image to find the corresponding coordinate set in the warped image [7]. To minimize intensity variations, the Lucas-Kanade algorithm also computes spatio-temporal derivatives. [8]. Researchers have attempted to measure the dynamic displacement of objects, but variations in shade and light intensity made it difficult to obtain precise results [9]. Additional algorithms are utilized in the feature tracking process, such as the warp update rule [10] in video coding motion. The illumination and lens distortion factors are also included in the iterative and hierarchical motion estimation algorithms [11]. To reduce complexity in multi-view video coding, motion and disparity estimation algorithms are used, resulting in a reduction in coding performance [12]. To retain image quality, the high-efficiency video coding standard has recently been used [13]. In another motion estimation study, Horn and Schunck developed a variational framework

that included data integrity and data matching [14]. Block matching algorithm is generally used but the hardware cost in high definition full search algorithm is a major concern [15]. Lee *et al*. implemented the searching and matching phase [16] in motion estimation by reducing the sum of absolute difference in the fast matching approach. Shen *et al*. [17] proposed the mode complexity and motion homogeneity to reduce the computational difficulty in motion and disparity estimation in multi-view coding. Most of the motion-tracking algorithms include features obtained from digital images. The digital image correlation algorithm includes subset-based correlation and element-based global correlation [18]. It was found that most of the optical flow estimation methods assume the constant brightness intensity of a pixel under the frame-wise displacement. Another approach that uses the optical flow-based framework is the gradient-based image correlation [19]. The motion estimation algorithm should be capable of countering the brightness effect and correlating the image features simultaneously with pixel-level accuracy [20]. The emergence of image correlation lies in the correlation of frames using a subset produced from a reference image [21]. The image correlation algorithm that approximates the internal and external parameters requires camera calibration [22]. Multiple square templates are obtained from the region of interest in a reference image. Monin et al. [23] modified the single-pixel algorithm and studied it with moving object frames for piece-wise 2D linear motion in single-pixel imaging. The root mean square error estimated with no regularization was 0.22. The investigation, however, was limited to simply planar motion. When using a non-cyclic basis, the technique results in a high level of reconstruction complexity. When the background is static, the local motion algorithm is limited. The root mean square error calculated in this work is 0.18 [24], and it describes a high-resolution motion measuring approach. In the present study, we propose a correlation-based three-dimensional motion estimation algorithm, which works on matching the grayscale intensity of a pattern, also called a template. Images of the moving object are captured using two high-speed cameras. The lens [25] correction in the form of both radial and tangential distortions, are calculated during camera calibration and are mentioned in the calibration Table in Section (3). The camera system is first calibrated [26] using MATLAB camera calibration application. The sets of image frames are correlated using templates to obtain template tracking coefficient or correlation coefficient (CC). We developed template-making and correlation algorithms in MATLAB. Three motion cases are investigated with the algorithm including finger and hand motion. The effect of the number of templates and speckle size on the CC is studied.

2. PROPOSED CORRELATION ALGORITHM

The templates are coupled in a correlation algorithm to estimate pixel displacement in 2D or 3D space. Generally, two cameras are preferred to estimate the three-dimensional location of the templates [27] [28].

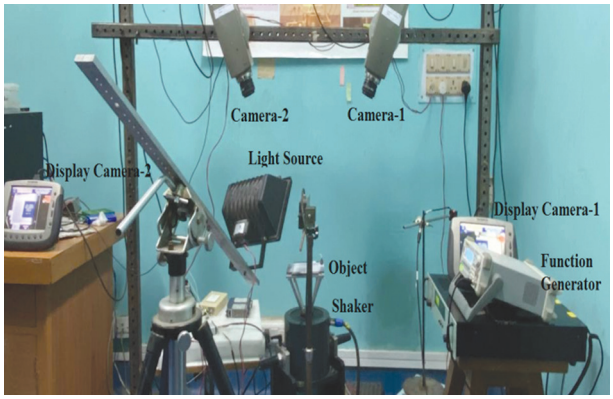


Fig. 1. Experimental Setup

The 3D motion estimation [29] demands that the object should remain in focus during the motion and the region of interest on the object is imaged on two or more cameras [19]. Since the motion of the object also depends on the camera speed, if the motion is slow then it should be correlated at a lower speed. A sufficient camera speed is able to obtain micron level displacement and time resolution. The thresholding [24] of the template and the formation of the subset also play a vital role in estimating the motion traversed. Fig. 1 shows the experimental setup and Fig. 2 shows the proposed image correlation algorithm. The algorithm involves the stereo camera calibration, image correlation, and validation of the correlation algorithm using an accelerometer sensor.

2.1. TEMPLATE TRACKING IN MOTION ESTIMATION

The template [30], [31] is correlated to a sequence of images to produce the pixel displacement. The technique is extremely helpful when motion features of the object under deliberation are not accessible. For tracking the grayscale intensity values, image correlation is performed and CCs are studied. Sutton *et al.* [27] described the CC for m observations of variables *a* and *b* as given in Equation (1). Where *a*' and *b*' are the mean values of a and b. For a perfect template matching in subsequent images, the CCs are unity [24].

$$Cor(a,b) = \frac{\sum_{j=1}^{m}(a_j - a')(b_j - b')}{\sqrt{\sum_{j=1}^{m}(a_j - a')^2 \sum_{j=1}^{m}(b_j - b')^2}} \quad (1)$$

The geometrical model for a camera requires three elementary transformations. The first transformation relates the world coordinates of a scene point to camera coordinates. The second transformation is the projection of this point onto the retinal plane. The third transforms the point into the sensor coordinate system [27]. Assuming a point in 3D space with its coordinates (X_w, Y_w, Z_w) where *w* signifies the world coordinate system and the camera [18] coordinates (*X, Y, Z*). The sen-

sor coordinates are (x_p, y_p) the focal length coordinates in pixels are (f_x, f_y) and the center of the image plane [27] is (C_x, C_y). The final customized form of the equation including all the transformations and constraints is given in Equation (2).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} R_{11}R_{12}R_{13} \\ R_{21}R_{22}R_{23} \\ R_{31}R_{32}R_{33} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \quad (2)$$

Where $[R]$ and $[T]$ are the components of rotation matrix and translation vector. The transformed and modified form of displacement in sensor coordinates is shown in Equation (3).

$$\begin{pmatrix} x_p \\ y_p \end{pmatrix} = \begin{pmatrix} C_x + f_x \frac{R_{11}X_w + R_{12}Y_w + R_{13}Z_w + T_x}{R_{31}X_w + R_{32}Y_w + R_{33}Z_w + T_z} \\ C_y + f_y \frac{R_{21}X_w + R_{22}Y_w + R_{23}Z_w + T_y}{R_{31}X_w + R_{32}Y_w + R_{33}Z_w + T_z} \end{pmatrix} \quad (3)$$

$$\begin{pmatrix} x_p \\ y_p \end{pmatrix} = \begin{pmatrix} C_x^j + f_x^j \frac{R_{11}^j X_w + R_{12}^j Y_w + R_{13}^j Z_w + T_x^j}{R_{31}^j X_w + R_{32}^j Y_w + R_{33}^j Z_w + T_z^j} \\ C_y^j + f_y^j \frac{R_{21}^j X_w + R_{22}^j Y_w + R_{23}^j Z_w + T_y^j}{R_{31}^j X_w + R_{32}^j Y_w + R_{33}^j Z_w + T_z^j} \end{pmatrix} \quad (4)$$

Two camera system [32] is employed for out-of-plane measurement of displacement as shown in Equation (4), where j ϵ N, (N = 1, 2). The parameter which describes the field of view is stereo-angle. The left camera is chosen as the reference camera and the parameters are calculated with respect to the reference camera. Lens distortion was included also in the calibration process giving rise to radial and tangential distortion. Tangential distortion is negligible as compared to the radial, hence ignored. Typically two coefficients are sufficient for radial distortion [27].

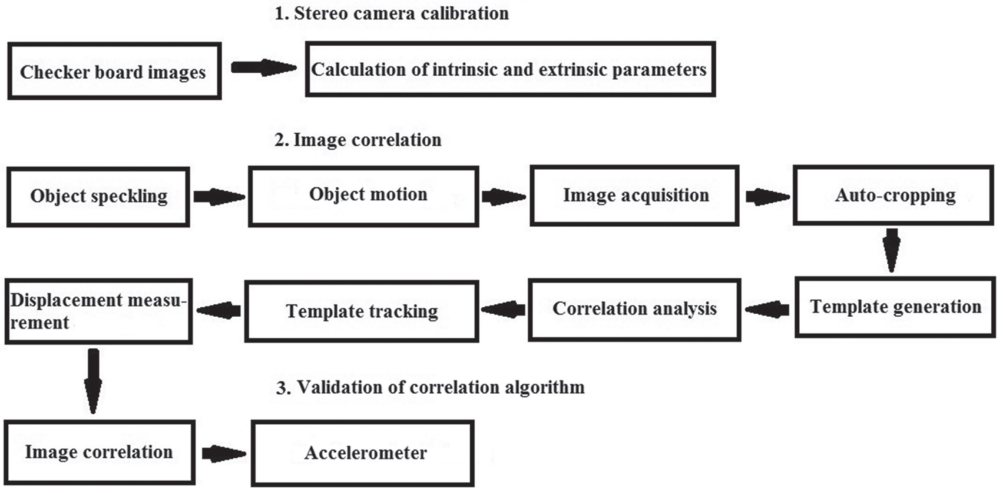


Fig. 2. Image Correlation Algorithm

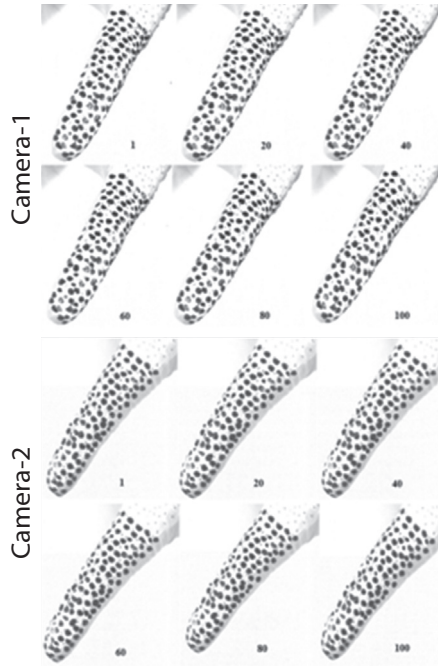


Fig. 3. Image frames for Finger Motion from Camera-1 and Camera-2

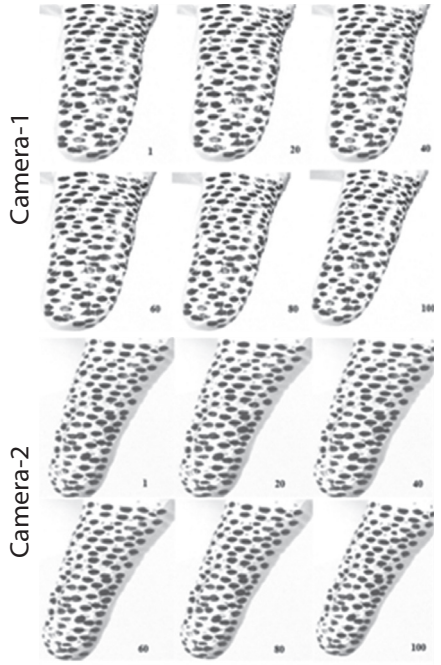


Fig. 4. Image frames for Finger Motion after Rotation from Camera-1 and Camera-2

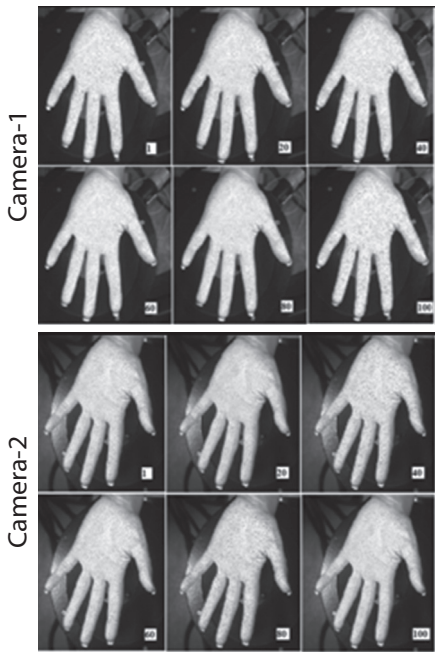


Fig. 5. Image frames for Hand Motion from Camera-1 and Camera-2

2.2. OBJECT SPECKLING AND TEMPLATE GENERATION

The algorithm follows the subset produced from a reference image in subsequent grayscale images. For the image correlation algorithm, objects are speckled in grayscale. Templates are developed from the first image of the reference camera. All images obtained from the left and right cameras, or camera-1 and camera-2, are correlated using the generated templates.

2.3. CORRELATION

The focus is adjusted by camera lenses, while the constant camera speed is essential for proper lighting. The Light should be sufficient for the templates to be visible and clear. The burst of 100 frames is captured and correlation is performed in all the frames as mentioned in Section 2.1. The CC is calculated as a measure of how efficiently the algorithm is implemented in images for template-based feature tracking.

3. EXPERIMENTAL

3.1. CALIBRATION

The 3D world coordinates (X_w, Y_w, Z_w) and the coordinates of the camera (X, Y, Z), need to be perfectly aligned. The grid also known as checker-board is used in calibration to calculate the intrinsic parameters e.g. focal length, principal point, distortion, and extrinsic parameters e.g. rotation and translation. A monochrome camera is preferred for perfect correlation. Two synchronized i-speed TR cameras (10000 FPS) and resolution (1280 x 1024 square pixels) with Nikon 50mm, 1:1.4D lenses are used in image correlation and for calibration and correlation. The lighting required in the process should be uniform and of adequate intensity, for the subset to correlate. Table (1), (2), and (3) shows the calibration data for finger motion, and Table (4), (5), and (6) shows the calibration for hand motion. The camera speed is selected based on the best object focus achieved in cameras [30]. The stereo camera calibration app in MATLAB is used for calibration. The checkerboard was placed in the field of view of cameras. Burst frame was set to unity and calibration grid was given small translation and rotation in random direction and images are acquired in both cameras using a common trigger. The angle between cameras was set to 34.5° at 1500fps and 42.6° at 2000fps. The angles were verified from extrinsic camera calibration parameters.

Table 1. Camera Calibration for 1500fps

Axis	Cam	Focal Length	Principal point
X	C1	2876.539 ± 15.705	666.061 ± 2.331
	C2	2151.659 ± 20.427	589.091 ± 2.017
Y	C1	3108.329 ± 16.365	623.902 ± 2.653
	C2	3095.646 ± 20.629	619.756 ± 2.169

Table 2. Distortion Parameters for 1500 fps

Axis	Cam	Radial	Tangential
X	C1	2.475 ± 0.139	0.001 ± 0.000
	C2	3.766 ± 0.206	0.004 ± 0.001
Y	C1	-433.388 ± 32.269	-0.007 ± 0.001
	C2	-903.849 ± 64.362	0.035 ± 0.001
Z	C1	23838.387 ± 2177.570	-
	C2	67071.699 ± 5985.477	-

Table 3. Position and Orientation of Camera-2 Relative to Camera-1 at 1500 fps

Axis	Rotation	Translation
X	-0.007 ± 0.001	-791.821 ± 3.883
Y	0.601 ± 0.002	5.363 ± 0.585
Z	-0.019 ± 0.001	345.907 ± 12.434

Table 4. Camera Calibration for 2000 fps

Axis	Cam	Focal Length	Principal point
X	C1	2883.706 ± 19.595	610.486 ± 7.263
	C2	2846.721 ± 32.167	586.3763 ± 17.936
Y	C1	2899.606 ± 20.078	565.609 ± 2.475
	C2	2788.549 ± 27.192	531.337 ± 6.315

Table 5. Distortion Parameters for 2000 fps

Axis	Cam	Radial	Tangential
X	C1	-1.748 ± 0.084	0.003 ± 0.001
	C2	0.556 ± 0.041	-0.007 ± 0.002
Y	C1	85.953 ± 9.265	-0.024 ± 0.002
	C2	-5.204 ± 1.036	-0.096 ± 0.004
Z	C1	-1648.520 ± 278.072	-
	C2	62.249 ± 14.187	-

3.2 OBJECT MOTION

Case-1: The motion of the finger in X, Y, and Z direction is captured, and required displacement in X_w, Y_w, and in Z_w is identified. Fig. 3 shows the image frames captured from both the cameras. It is found that the subsets formed in the process also capture the region that is not in the area of motion. Finger and hand motion were acquired in the 15th and 20th of a second respectively by adjusting camera speed. The short exposure time while capturing was chosen to acquire micro-scale motion in terms of 3D displacements.

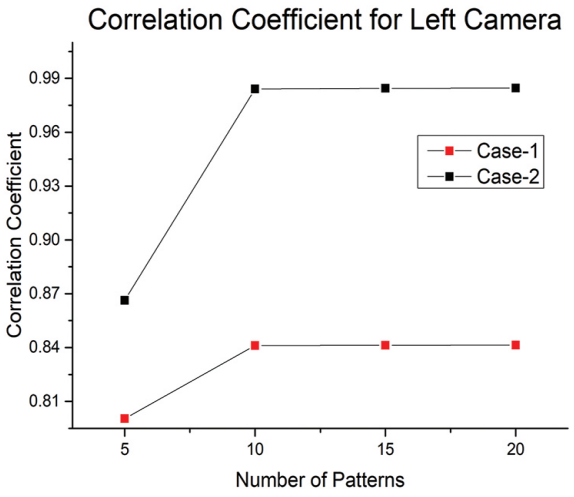


Fig. 6(a). Correlation Coefficient for Finger Motion from Camera-1

Table 7(a). Minimum Displacement for Finger Motion

Cases	Template	Minimum Displacement		
		X_w	Y_w	Z_w
Case-1	5	-127.810	-12.821	-1.092x10³
	10	-127.810	-12.821	-1.148x10³
	15	-142.345	-16.121	-1.189x10³
	20	-142.734	-16.607	-1.199x10³
Case-2	5	-101.921	-41.240	-1.209x10³
	10	-104.822	-59.539	-1.209x10³
	15	-138.121	-59.539	-1.216x10³
	20	-138.850	-59.651	-1.258x10³

Table 6. Position and Orientation of Camera-2 Relative to Camera-1 at 2000 fps

Axis	Rotation	Translation
X	-0.035 ± 0.002	-807.622 ± 8.697
Y	0.744 ± 0.006	-4.752 ± 1.142
Z	0.021 ± 0.001	273.609 ± 19.859

Hence for Case-2: rotation code is developed, by which excess area can be cropped easily. Fig. 4 shows the image frames after applying the rotation code. Case-3: Motion of a human hand is captured using the templates to further validate the proposed algorithm irrespective of object shape and movement. The hand is randomly translated in 3D space and motion is captured. The number of templates is increased to 50. Fig. 5 shows the image sets for hand motion. It was found that CC is almost similar for 15 and 20 templates as shown in Table 9.

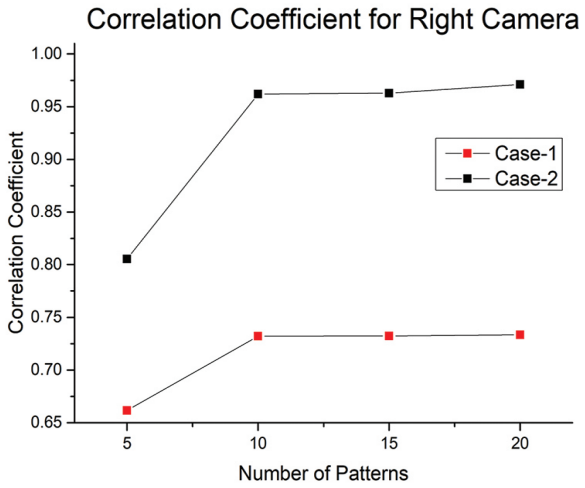


Fig. 6(b). Correlation Coefficient for Finger Motion from Camera-2

Table 7(b). Maximum Displacement for Finger Motion

Cases	Template	Maximum Displacement		
		X_w	Y_w	Z_w
Case-1	5	-38.129	258.973	-0.883x10³
	10	-38.129	258.973	-0.871x10³
	15	-42.333	268.923	-0.879x10³
	20	-42.793	269.134	-0.881x10³
Case-2	5	33.137	199.691	-0.845x10³
	10	33.137	201.359	-0.866x10³
	15	38.149	282.564	-0.889x10³
	20	38.154	282.987	-0.919x10³

The increment in CC with respect to the template size is attributed to the efficient matching of templates in the sequence images. To further propose the technique for large object motion estimation, a hand motion is captured and motion traversed in X, Y, and Z direction is calculated using 5, 10, 20, 30, 40, and 50 templates.

4. RESULTS AND DISCUSSION

4.1 MOTION MEASUREMENT

The correlation was performed for a fixed subset size of 10 and a threshold value >10. Tables 7 and 8 show displacement data for finger and hand motion up to 20 and 50 templates respectively. Six images are represented in Fig. 3, 4, and 5 out of 100 captured images for finger and hand motion considered.

4.2 EFFECT OF NUMBER OF TEMPLATES ON TEMPLATE TRACKING

Case-1 represents finger motion and the minimum and maximum displacement measurements are shown in Table 7(a) and 7(b). The finger image frames are rotated in Case-2 to eliminate excess area out of the region of interest. We observed a significant difference in CC. The increment in CC for left camera in case-2 is 8.2% for 5 templates, 17.0% for 10 templates, 17.2% for 15 templates and 20 templates. Similarly for right camera increment in CC is 21.6% for 5 templates, 31.3% for 10 templates, 31.5% for 15 templates and 32.4% for 20 templates. Fig. 6(a) and 6(b) show CC with the number of templates for Case-1 and Case-2 respectively.

Table 8(a). Minimum Displacement for Hand Motion

Case	Template	Minimum Displacement		
		X_w	Y_w	Z_w
Case-3	5	-266.769	99.169	-1.289x10³
	10	-209.746	207.174	-1.251x10³
	20	-301.008	144.432	-1.286x10³
	30	-335.446	-169.568	-1.277x10³
	40	-349.846	-239.659	-1.260x10³
	50	-349.979	-239.948	-1.255x10³

Table 8(b). Maximum Displacement for Hand Motion

Case	Template	Maximum Displacement		
		X_w	Y_w	Z_w
Case-3	5	-108.059	218.922	-1.228x10³
	10	-146.302	270.258	-1.256x10³
	20	-110.669	270.258	-1.226x10³
	30	-72.183	233.611	-1.222x10³
	40	-49.156	252.978	-1.266x10³
	50	-49.251	252.989	-1.269x10³

As shown in Tables 8 and 10, the displacement values and CC values for 40 and 50 templates are almost similar. We observed that the template number plays an important role in motion assessment, which is expressed in Table 10. The CC increases on increasing the number of templates and saturates at the value of 40 and further found similar for 40 and 50 templates. Fig 7(a) and 7(b) show the CC for hand motion.

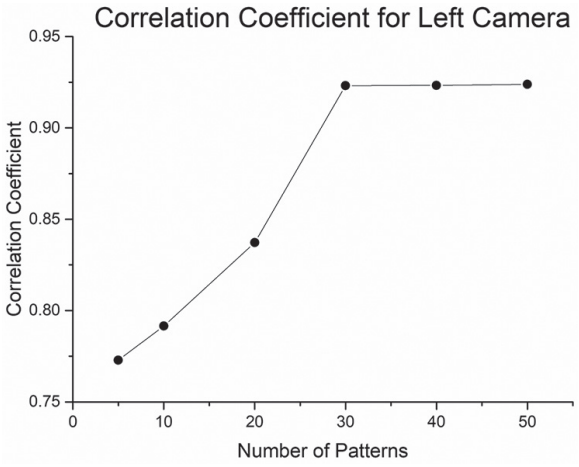


Fig. 7(a). Correlation Coefficient for Hand Motion from Camera-1

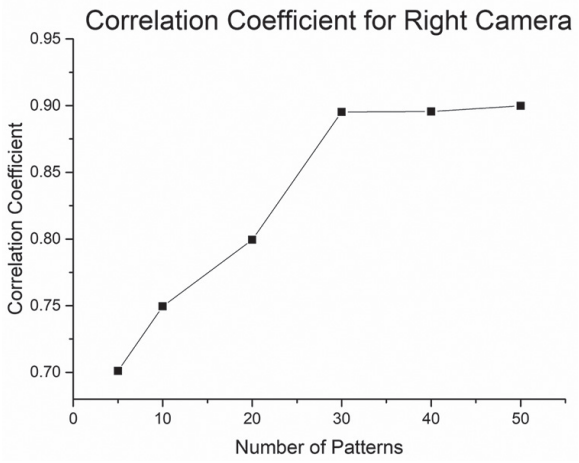


Fig. 7(b). Correlation Coefficient for Hand Motion from Camera-2

Template tracking is used to identify the location of templates with identical areas and different grayscale intensities. The area is in pixel units. The identical size templates were developed using a separate template-making algorithm. The correlation algorithm produces a match of a particular template only when it identifies the exact location of the template in the camera images.

Table 9. Correlation Coefficient for Finger Motion

Cases	Template	Correlation Coefficient	
		(Left Camera)	(Right Camera)
Case-1	5	0.801	0.662
	10	0.840	0.732
	15	0.841	0.733
	20	0.842	0.734
Case-2	5	0.867	0.805
	10	0.983	0.961
	15	0.986	0.964
	20	0.987	0.972

Table 10. Correlation Coefficient for Hand Motion

Cases	Template	Correlation Coefficient (Left Camera)	Correlation Coefficient (Right Camera)
Case-3	5	0.774	0.702
	10	0.793	0.751
	20	0.838	0.799
	30	0.921	0.895
	40	0.923	0.896
	50	0.924	0.898

4.3. VALIDATION OF THE ALGORITHM

The proposed technique was validated on a rigid aluminum block fixed using a fixture to the shaker. Half of the block was speckled using black and white paint whereas an accelerometer was mounted on the other half. A known frequency of 10 Hz was given through a function generator and a power amplifier. The motion of the block was captured using the same high-speed cameras and an image correlation algorithm was applied. The displacement response from both accelerometer and image correlation technique were compared. Fig. 8 shows the validation of the correlation algorithm.

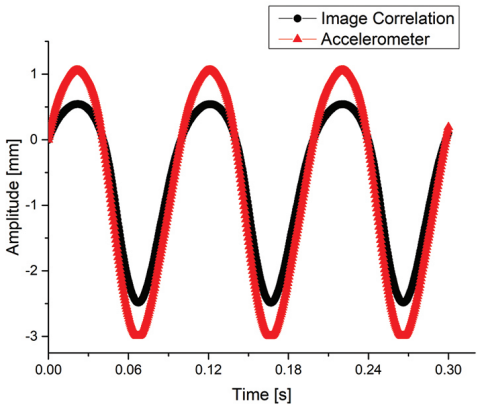


Fig. 8. Validation of Correlation Algorithm

5. CONCLUSION

The present study aimed to estimate object motion using the non-contact image correlation method. High-speed cameras are used in correlating the templates generated from the reference image to the images acquired from both cameras. Camera calibration is done before correlation to calculate the intrinsic and extrinsic parameters. The displacement was calculated using an image correlation technique in three cases: finger motion, rotating finger motion, and hand motion. Two high-speed cameras were synchronized in stereovision configuration to obtain CC. Case-1 and Case-2 represent finger motion with a constant shutter speed of 1500 fps. CC was found to increase up to 0.842 and 0.734 for cameras 1 and 2 respectively. For Case-2, an increase in CC goes up to 0.987 and 0.972 for cameras 1 and 2 respectively. The hand motion was studied at 2000 fps to further propose the algorithm for large objects motion measurement, CC was found to increase up to 0.924 and 0.898 for cameras 1 and 2 respectively, which is almost equivalent to that obtained in case-2. It was concluded that the proposed image correlation algorithm can be applied for motion estimation irrespective of speckle size for a constant template size. The CC was found to increase with the increase in template number.

6. ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Structures Lab at IIT Kanpur, where the experiments were carried out.

7. REFERENCES

[1] K. Singh, R. S. Ahamed, "Computationally efficient motion estimation algorithm for HEVC", Journal of Signal Processing Systems, 2017.

[2] Z. Wang, A. C. Bovik, A. R. Sheikh, E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", IEEE Transactions on Image Processing, Vol. 13, No. 4, 2004, pp. 1-14.

[3] H. Tekwani, K. Raj, "Compression using thresholding on signal/image by applying wavelet analysis", Proceedings of the IEEE International Conference on Computational and Characterization Techniques in Engineering and Sciences, Lucknow, India, 14-15 September 2018, pp. 17-20.

[4] E. Antonakos, J. Alabort-i-Medina, G. Tzimiropoulos, S. P. Zafeiriou, "Feature-based Lucas-Kanade and active appearance models", IEEE Transactions on Image Processing, 2015, pp. 1-16.

[5] Y. Kellar, A. Averbuch, "Global parametric image alignment via high-order approximation", Computer Vision and Image Understanding, Vol. 109, 2008, pp. 244-259.

[6] J. Shin, S. Kim, S. Kang, S. W. Lee, J. Paik, B. Abidi, M. Abidi, "Optical flow-based real-time object tracking using non-prior training active feature model", Real-Time Imaging, Vol. 11, 2005, pp. 204-218.

[7] G. D. Evangelidis, E. Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, No. 10, 2008, pp. 1-8.

[8] V. Mahalingam, k. Bhattacharya, N. Ranganathan, H. Chakravarthula, R. R. Murphy, K. S. Pratt, "A VLSI architecture and algorithm for Lucas-Kanade-based optical flow computation", IEEE Transactions on Very Large Scale Integration Systems, Vol. 18, No. 1, 2010, pp. 29-38.

[9] J. Guo, C. Zhu, "Dynamic displacement measurement of large-scale structures based on the Lucas-Kanade template tracking algorithm", Mechanical Systems and Signal Processing, Vol. 66-67, 2016, pp. 425-436.

[10] S. Baker, I. Matthews, "Lucas-Kanade 20 years on: a unifying framework", International Journal of Computer Vision, Vol. 53, No. 3, 2004, pp. 221-255.

[11] Y. Altunbasak, R. M. Mersereau, A. J. Patti, "A fast parametric motion estimation algorithm with illumination and lens distortion correction", IEEE Transactions on Image Processing, Vol. 12, No. 4, 2003, pp. 395-408.

[12] L. Jia, C. Tsui, O. C. Au, K. Jia, "A new rate-complexity-distortion model for fast motion estimation in HEVC", IEEE Transactions on Multimedia, 2018.

[13] Z. Pan, Y. Zhang, S. Kwong, "Efficient motion and disparity estimation optimization for low complexity multiview video coding", IEEE Transactions on Broadcasting, Vol. 61, No. 2, 2015, pp. 166-176.

[14] M. Drulea, S. Nedevschi, "Motion estimation using the correlation transform", IEEE Transactions on Image Processing, Vol. 22, No. 8, 2013, pp. 3260-3270.

[15] H. Yin, H. Jia, H. Qi, X. Ji, X. Xie, W. Gao, "A hardware-efficient multi-resolution block matching algorithm and its VLSI architecture for high definition MPEG-like video encoders", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 20, No. 9, 2010, pp. 1242-1254.

[16] S. Lee, "Fast motion estimation based on adaptive search range adjustment and matching error prediction", IEEE Transactions on Consumer Electronics, 2009.

[17] L. Shen, Z. Liu, T. Yan, Z. Zhang, P. An, "View-adaptive motion estimation and disparity estimation for low complexity multiview video coding", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 20, No. 6, 2010, pp. 925-930.

[18] Y. Huang, J. Ji, K. M. Lee, "Model-based digital image correlation for non-contact deformation measurement of strain field and mechanical property", IEEE Transactions on Industrial Informatics, 2018, pp. 1-9.

[19] W. Feng, Y. Jin, Y. Wei, W. Hou, C. Zhu, "Technique for two-dimensional displacement field determination using a reliability-guided spatial-gradient-based digital image correlation algorithm", Applied Optics, Vol. 57, No. 11, 2018, pp. 2780-2789.

[20] H. Tekwani, K. Raj, "Brightness intensity-based transient motion prediction", Smart Systems: Innovations in Computing. Smart Innovation, Systems and Technologies, Vol. 235, Springer, Jaipur, India, 22-23 January 2021, pp. 1-27.

[21] P. Premaratne, S. Ajaz, M. Premaratne, "Hand gesture tracking and recognition system using Lucas-Kanade algorithms for control of consumer electronics", Neurocomputing, Vol. 116, 2013, pp. 242-249.

[22] J. Weng, P. Cohen, M. Herniou, "Camera calibration with distortion models and accuracy evaluation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14 No. 10, 1992, pp. 965-980.

[23] S. Monin, E. Hahamovich, A. Rosenthal, "Single-pixel imaging of dynamic objects using multi-frame motion estimation", Scientific Reports, 2021.

[24] H. Tekwani, K. Raj, "Role of pattern characteristics in cross correlation based motion estimation", Indian Journal of Science and Technology, Vol. 14, No. 41, 2021, pp. 3114-3125.

[25] P. Reu, "Stereo-rig design: lens selection part-3", Experimental Techniques, Vol. 37, No. 1, 2013, pp. 1-3.

[26] Y. Gao, T. Cheng, Y. Su, X. Xu, Y. Zhang, Q. Zhang, "High-efficiency and high-accuracy digital image correlation for three-dimensional measurement", Optics Lasers Engineering, Vol. 65, 2015, pp. 73-80.

[27] M. A. Sutton, J. J. Orteu, H. W. Schreier, "Image correlation for shape, motion and deformation measurements", Springer, Berlin, 2009.

[28] W. LePage, „A practical guide to digital image correlation“, https://digitalimagecorrelation.org/, 2017.

[29] H. Tekwani, K. Raj, "Study of template compression in image correlation based motion estimation", International Journal of Electrical Engineering and Technology, Vol. 12, No. 6, 2021, pp. 308-319.

[30] E. M. C. Jones, M. A. Iadicola, "A good practices guide for digital image correlation", International Digital Image Correlation Society, 2018.

[31] H. Tekwani, K. Raj, "Non-contact motion estimation using image correlation", Proceedings of the IEEE International Conference on Signal Processing and Integrated Networks, Noida, India, 27-28 February 2020, pp. 122-125.

[32] P. Reu, "Stereo-rig design: camera selection part-2", Experimental Techniques, Vol. 36, No. 6, 2012, pp. 3-4.

Dynamic Load Balancing for Congestion Avoidance using Adaptive Neuro-Fuzzy Inference System in Mobile Communication Network

Original Scientific Paper

Prince David Collins
Federal University of Technology, Owerri,
Department of Electrical and Electronic Engineering
Owerri, Imo State, Nigeria
collinsglamour@gmail.com

Ezema, Longinus Sunday
Federal University of Technology, Owerri,
Department of Electrical and Electronic Engineering
Owerri, Imo State, Nigeria
ezemms@yahoo.com

Nkwachukwu Chukwuchekwa
Federal University of Technology, Owerri,
Department of Electrical and Electronic Engineering
Owerri, Imo State, Nigeria
nkwachukwu.chukwuchekwa@futo.edu.ng

Abstract *– Network congestion is one of the key challenges to mobile communication services. This is because of rapid and constant growth in the mobile subscriber base that has led to an increase in network traffic. The more the network traffic the more economic opportunity from the business perspective for the operators and the challenges it poses on the network. If not followed by network capacity expansion it will defiantly pose a serious network issue like network congestion that leads to call drop and poor Quality of Service (QoS). Many research works adopted handover mechanism as a means of reducing congestion in a network. However, the decision on when to initiate a handover, which cell to receive the Mobile Station (MS) and how to ensure that the QoS requirements are maintained are the paramount research questions that must be resolved. Therefore, a handover process using soft computing in the Adaptive Neuro-Fuzzy Inference System (ANFIS) to ensure balanced traffic load distribution in-network and reduce the probability of congestion is proposed. The study allows simultaneous evaluation of three major network parameters Received Signal Strength (RSS), Received Signal Quality (RxQual) and network traffic using ANFIS to improve system performance. The results show that when the Hysteresis value approaches 6 the handover processes are triggered. The hysteresis value of the concerned MS with neighbouring cells is considered to determine the most suitable cell to handover. This work will be able to achieve dynamic load balancing, congestion avoidance and avoided the 'ping pong' effect that is often an issue with handover with less computation. At the end customer satisfaction will be achieve Quality of experience (QoE).*

Keywords*: ANFIS, GSM, handover mechanism, mobile communication, network congestion*

1. INTRODUCTION

The GSM network, (second generation of mobile communication), over the years has become the bedrock of mobile communication. The GSM network has remained influential in the mobile communication world despite other technological upgrades like Wideband Code Division Multiple Access (WCDMA), Long Term Evolution (LTE) and others. The GSM has been the basics by which other technologies were developed. The success achieved by this system has led to greater commercialization of the mobile communication system and the gush in the number of subscribers scrambling for the limited network resources that must be shared among numerous users. If these hardware system resources and exchange of information are no longer sufficient to handle the available network users it leads to network congestion, call drop, handover failure, network failure etc [1, 2, 3].

Network congestion is one major factor among Received Signal Strength (RSS), coverage area, Received Signal Quality (RxQual), Signal to Noise Ratio (SNR), network failure, and insufficient network infrastructure that lead to call drop. In the rural areas, the call drop

mainly is as a result of lack of network coverage while in the urban area it is due to lack of enough network infrastructures to march the ever-increasing network traffic [4, 5]. However, this increase in network traffic has placed a serious burden on the limited network resources. The more the network traffic the more economic opportunity from the business perspective for the network operators and the challenges it poses on technology. This rapid and constant growth in subscriber base if not well followed by network capacity expansion will defiantly pose serious network issues. The main technical challenges due to the increase in network traffic are; the limited network resources, energy utilization and equipment cost [6].

There are many investigations on how to improve the efficiency and effectiveness of the limited network resources. This must be achieved in such a way that the QoS is not compromised. Network expansion is one method that can easily achieve this at the expense of cost. Network operators usually expand their capacity if there are significant increases in the subscriber base to march the expected capacity upgrade and it must be business-oriented from the economic perspective. This usually includes increasing the number of Base Stations (BS) or cell splitting techniques. Cell splitting is a technique of subdividing a congested cell (usually large – macro and micro-cells) into smaller cells (femtocells). Cell splitting is also a technique of combating network congestion. Each cell should have its base station at a reduced antenna height and transmitter power [7, 8, 9, 10]. Although, in cell splitting, the individual cell has a reduced channel as compared to the original large cell, its ability to increase capacity is by a multiplicity of cells separated by geographical distance to utilise the same channel at the same time with minimal or no interference from each other. The new cells have a smaller coverage area and less transmitter power. Therefore, the measure of capacity increase is dependent on the number of channels per unit area [11].

The issues with cell splitting are; (1) the frequency allocation will be repeated which will require the neighbouring cell to be recalibrated to avoid interference; (2) the handover rate will defiantly increase as it is more likely MS will cross to other cells' coverage area more often due to reduction in coverage radius; and (3) it will also lead to increase in energy consumption of the BS. The network topology and the country's law on BS installation can pose a limit to cell splitting [6, 9, 11].

Another technique of reducing the BS power consumption and network congestion for efficient network resource utilization is handover processes. The process of moving from one BS to another, while a call is ongoing, is called handover. Various handover types, techniques and processes are explained in much literature [12, 13, 14, 15].The MS during handover request that a new Traffic Channel (TCH) be assigned to it. The handover is complete if the MS is assigned a new TCH from another BS. If no TCH is available from the neighbouring BS, the handover is blocked which may lead to a call drop. The blocking of the ongoing call from handover is called handover blocking. This handover is used to maximize spectrum utilization of the system. This is initiated when the signal of the serving cell has dropped below the one from the neighbouring cells. Apart from this reason, a handover can also be initiated to balance the network load or load sharing [16, 17].

The decision on when to initiate a handover, which cell to receive the MS and ensure the QoS requirements are maintained are the paramount research questions that must be resolved. Many research works have considered different cell metrics for handover decision which includes SNR, Path-Loss (PL), the distance between the mobile station and the BS, Mobile Velocity (MV), RSS, traffic load and others. The traditional technique is a single criteria handover decision that compares the RSS from the current cell with that from the neighbouring cells. The single criteria do not put into consideration other cell parameters that could affect the handover. The issue with this technique is that the propagation related RSS fluctuation can result in a "ping pong" effect. It is not all neighbouring cells that can accept a call due to congestion or traffic load on the cell which may lead to handover blocking and eventual call drop [18]. However, it can be initiated as a load balancing to easy traffic on the congested cell to cell with less traffic load, hence easing congestion in the network in the process. Therefore, it is important to put into perspective other important cell metrics (RSS, RxQual and number free channel – traffic load) with a technique that can handle imprecision in some metrics for effective and efficient handover.

The main aim of this work is to develop a handover mechanism for dynamic load balancing using soft computing in ANFIS. The study allows simultaneous evaluation of several significant cell metrics to improve system performance in the handover procedure. This system will consider the traffic load (using several free channels) at the serving and neighbouring cell as well as another key handover parameter in RSS and RxQual. Moreover, in ANFIS, the need for human instinct in exploiting the tolerance of imprecision and in handling nonlinear problems cannot be overstressed. A proper handover algorithm should ensure balanced traffic load distribution in-network and reduce the probability of congestion. The "ping pong" effect will be eliminated due to multiple parameters evaluation in the handover decision. The system when implemented will also ensure reduction in energy consumption, balanced traffic load and reduced traffic congestion in the network without any intrusive network.

2. REVIEW OF RELATED WORKS

The following research works on congestion control through handover management in cellular networks have been reported in journal and conference publications.

An adaptive hysteresis based horizontal handoff algorithm for the GSM network was proposed in [19]. The authors developed a model to adaptively determine the hysteresis value (within the range of 20 to 0dB) using two variables; the distance between the MS and the serving BS and the radius of the BS coverage. The handover process is activated once the RSS from any neighbouring BS is more than the RSS from the serving BS plus the hysteresis value. This technique considered primarily the RSS and adjusted the handover hysteresis margin using the separation distance. The result analysis showed that it performed better than the traditional handover technique. However, it has been observed that the technique ignored the traffics on the neighbouring base stations. A handover to a BS with traffic congestion could lead to handover blocking or call drop.

In this paper [20], an algorithm for automatic off-line optimization of handover margins in GSM/General Packet Radio Services (GPRS) networks is proposed. The author adopted a simple load sharing rule on an adjacency basis to deal with spatial load variability in a mobile network. It also applied the discrimination of several periods using differentiated parameters setting to cope with variations of traffics during the day. The variation in handover margin is relative to the difference in blocking probability between the serving BS and the adjacent BSs without consideration to the instantaneous traffics. The analysis achieved a good result as it recorded a reduced call drop rate and were able to carry more traffic even though call quality deteriorated. There is a need to factor into the optimization algorithm one of the key performance indicators to improve the quality of calls.

In [21], an adaptive hysteresis margin and load balancing to manage congestion in a heterogeneous network was proposed. The authors also investigated major challenges of hysteresis margins and load balancing in a mobile network. The author used quality indicators in SNR to vary the handover margins (HM) for the handover technique. For the mobility load balancing, the cell individual offset (CIO) value and the traffic loads on the serving and new BS were used for the load balancing decision. The result analysis showed that the schemes improved the handover success rate and performed well. However, the techniques are computationally intensive.

The use of a soft computing technique that is based on the hybrid of Neural Network (NN) and Fuzzy Logic (FL) for load balancing in mobile communication was proposed in [22]. An artificial intelligence (AI) based on three calculated parameters; the virtual load of the serving BS, the number of unsatisfied users in the serving BS and the overall load state of the target BS, were used to determine the handover hysteresis margin adaptively. The simulation results appear to have a good performance, although some computations were unnecessary. The number of unsatisfied users is the network-wide parameter that should rather be substituted with SNR or still the BS throughput as a performance indicator.

In [23], a handoff decision in the cellular mobile system using NN, direct retry and load sharing technique was proposed. The parameters considered for the best technique are the RSS and traffic intensity of the BS. The NN technique outperformed the other two techniques. The issue with the algorithm is that it automatically blocks the handover of calls with RSS below -100dB. This might lead to a call drop as the RSS deteriorate further below the MS receiver sensitivity.

In [24], a NN approach to GSM traffic congestion prediction was proposed. The authors used real network traffic data to train NN to predict the probability of congestion in a BS. The result of the trained model when tested with new traffic data came close to a real network situation. A very good prediction technique applied but did not offer any solution to congestion management in the network.

In [25], the authors approached congestion management by developing a traffic class prioritization algorithm. The algorithm classified the real network traffic based on the order of importance. Priority Queuing (PQ), Weighted Fair Queuing (WFQ) and the hybrid of the two were developed. The performance evaluation was done based on the traffic served per class. The algorithms achieved a good classification and priority given to the traffic with higher priority. The prioritization algorithm ensures that instead of dropping less priority traffic that the less important traffic should be dropped. How that affect congestion is not stated. A hybrid system of the developed algorithm and handover mechanism for load balancing to achieve better congestion management performance is recommended.

An automatic handover control for distributed load balancing in mobile communication networks was proposed in [26]. The algorithm considered only the available channel in a BS in achieving its load balancing. The algorithm adopted a round-robin queuing method for calls waiting to be assigned a channel based on quantum and time slice variables. The performance evaluation was based on the call blocking rate. It performed very well when compared with a normal system without load balancing. The algorithm was applied regardless of any QoS parameters like RSS; SNR and others.

In [17], the work proposed a fuzzy logic multiple parameter handover algorithms based on RSS, traffic load, SNR and Path Loss (PL) of the serving BS for load balancing. The output of the fuzzy inference system is a handover decision based on the values of the four inputs parameters. This technique cannot be applied in determining the suitability of the target base stations because the requesting MS don't have the SNR and PL parameters used for the algorithm. MS can only measure these parameters from the serving BS. We recommend that the two parameters be removed as an input

to the fuzzy controller. The system will perform better if NN is applied to optimize the system.

Some research works proposed techniques for load balancing in 3GPP LTE. An algorithm was developed in [27] to find the appropriate handoff offset value between the congested BS and a potential target BS. In this technique, each eNodeB accumulates measured RSS report from the served MS and the same resource reports from neighbouring BS. If the traffic load of the selected BS exceeds a pre-set threshold, it commences transferring the MS to less loaded target BS. This handover procedure goes on pending when the traffic load of the sending BS becomes less than the overload threshold, or the traffic load of the target eNodeB surpasses a particular threshold. The algorithms are based on hard (traditional) computing which does not exploit the tolerance for exactness of the load balancing technique.

This research seeks to provide a handover mechanism that will utilize multiple criteria to provide congestion management in most mobile communication. The ANFIS will be developed to handle load balancing in the network through a handover mechanism. ANFIS is chosen as the best technique to handle the imprecision in RSS and its ability to handle the on-linear problem. It is important to make a handover decision based on RSS, RxQual and traffic loads in BS as the three major metrics. Relying only on RSS might result in handing over to a network already congested thereby creating an imbalance in the network. Most of the reviewed works depend on one or two-parameter(s) to take handover decisions while a traditional method uses only the RSS.

3. SYSTEM AND PARAMETERS MODELLING

The proposed technique uses five-layer ANFIS with three inputs; –(1) the RSS; (2) the RxQual; and (3) the number of free channels. The problem of network congestion remains a critical issue due to the ever-growing size, speed and demand of mobile networks these days. Steps to congestion avoidance through uneven load distribution in cell cluster using handover process are:

- Locate the most congested Base Transceiver Station (BTS) within a cluster.
- Check the status of the BTS neighbours based on the following parameters; RSS, RxQual and the number of available channels
- Determine the Hysteresis value of the MS using the ANFIS techniques
- Initiate a handover process if the hysteresis value is above the pre-determine hysteresis value called handover hysteresis value (HHV).

Therefore, by determining the hysteresis value of mobile users of a congested BTS within a cluster, it is possible to move users with handover hysteresis value to BTSs with lesser traffic loads or better hysteresis value. Aside from locating a congested BTS, the handover mechanism can also be triggered once an MS reaches hysteresis value irrespective of the traffic load on the serving BTS. This ensures the QoS is not degraded as a result of the insufficient value of the parameters under consideration. This approach can be called dynamic load balancing which leads to more efficient radio resource utilisation [28]

For each channel, RxQual is measured, averaged without including the measurement during the previous period on that channel over the measurement period of length of the slow-associated control channel (SACCH) multi-frame. When the RxQual is accessed over the full-set and sub-set of the frame, eight levels of quality (0 – 7) are defined [29, 30]. This reflects the quality of voice connection where 0 is the best and 7 the worst. Each of these values matches up to an estimated number of bit errors in a burst. Therefore, it is measured based on Bit Error Rate (BER) with an acceptable value of "0 to 5" in a commercial mobile network [31]. RxQual is a part of the network measurement reports (NMR) and a measured metric that indicates the quality of the downlink model.

The RSS may be adopted as a measure in the radio frequency power control and handover process. The Root-Mean-Square (RMS) RSSI at the receiver input shall be measured by the MS and the BSS over a full range of -110dBm to -48dBm with a fixed accuracy of ± 4dB from -110dBm to -70dBm under normal conditions and ± 6dB under both normal and severe conditions over the full range. The RMS RSS at the receiver input shall be measured by MS above -48dBm to -38dBm with a fixed accuracy of ± 9dB under both normal and severe conditions [29],[30]. Hence, the range of RSS levels considered for this research is -120dBm to -30dBm.

It is possible to have excellent RSSI but no connection due to poor RxQual and vice versa. Both QoS parameters must be considered for a successful connection between the receiver and the transmitter. A variance of both quality parameters is significant for connection as a particular measurement characterize one moment which can vary radically over a short time. A stable connection requires consistency of both parameters.

Many factors influence the RSS and RxQual, including but not limited to;

- BTS load
- MS proximity to BTS
- Signal going through a mobile repeater
- Interference from competing signal
- Physical barriers (building, mountain, trees etc)
- Weather condition

Even if you have an excellent RSSI, you may not achieve maximum network speed due to the high vol-

ume of mobile users on the BTS (BTS load). The BTS may be congested leading to call drop or degradation in the network connection. Thus, the number of free channels is considered as a parameter for use in scheduling or handover mechanisms for mobility load balancing.

Table 1 is used to generate input/out data for ANFIS modelling using MATLAB. The ranges of these parameters are mostly standards which are classified into four linguistic variables such as 'insufficient', 'medium', 'Good' and 'Strong' to indicate the strengths of these input data for this work [32], [33]. It is a matter of necessity to generate an ideal input/output relationship data since we don't have real measured network input data with corresponding output data (Hysteresis) to use the model and train the ANFIS. Also generated is input/output data sufficiently different from the training data set for model validation and testing. The generated data completely represents the features of the data expected to get from a real cellular network if implemented. Based on the three input data with four linguistic variables, there can only be 64 possible permutations each with a distinct hysteresis value as shown in Table 2. The range of the hysteresis value is 0 to 10. The MATLAB simulated training data in table 3 is generated using Table 1 and 2.

Table 1. Data classification and FIS variables

RSSI(dBm)	RxQual	Free Channel	FIS Variables
RSSI<= -110	5<=RxQual<=7	0<=FC<=200	**Insufficient**
-109<=RSSI<=-91	3<=RxQual<=4	201<=FC<=450	**Medium**
-90<=RSSI<=-75	2<=RxQual<3	451<=FC<=700	**Good**
-74<=RSSI<=-30	0<=RxQual<=1	701<=FC<=1000	**Strong**

Table 2. Sample of 64 permutations for 3 inputs data with 4 linguistic variables

1.	If (RSS is Insuff) and (RxQUAL is Insuff) and (FChannel is Insuff) then (hysteresis is 10)
2.	If (RSS is Insuff) and (RxQUAL is Insuff) and (FChannel is Medium) then (hysteresis is 9.84375)
3.	If (RSS is Insuff) and (RxQUAL is Insuff) and (FChannel is Good) then (hysteresis is 9.375)
4.	If (RSS is Insuff) and (RxQUAL is Insuff) and (FChannel is Strong) then (hysteresis is 8.90625)
5.	If (RSS is Insuff) and (RxQUAL is Medium) and (FChannel is Insuff) then (hysteresis is 9.53125)
6.	If (RSS is Insuff) and (RxQUAL is Medium) and (FChannel is Medium) then (hysteresis is 8.28125)
7.	If (RSS is Insuff) and (RxQUAL is Medium) and (FChannel is Good) then (hysteresis is 7.65625)
8.	If (RSS is Insuff) and (RxQUAL is Medium) and (FChannel is Strong) then (hysteresis is 6.71875)
9.	If (RSS is Insuff) and (RxQUAL is Good) and (FChannel is Insuff) then (hysteresis is 9.0625)
10.	If (RSS is Insuff) and (RxQUAL is Good) and (FChannel is Medium) then (hysteresis is 7.5)
.....	
63	If (RSS is Strong) and (RxQUAL is Strong) and (FChannel is Good) then (hysteresis is 0.3125)
64	If (RSS is Strong) and (RxQUAL is Strong) and (FChannel is Strong) then (hysteresis is 0.15625)

Table 3. Sample of 1200 simulated training data

RSS (dBm)	RxQUAL	Free Chanel	Hysteresis
-65	5	815	4.6857
-57	4	505	2.5
-104	6	665	7.9688
-102	6	793	7.0313
-116	5	895	8.9063
-122	3	846	6.7188
-57	7	127	8.75
-99	4	276	4.2188
-92	2	996	2.3438
-37	1	190	4.375
-49	2	754	0.625
-64	2	114	5

4. ANFIS MODEL DEVELOPMENT

ANFIS is a hybrid intelligence system that comprises the artificial neural network (ANN) and Fuzzy Inference System (FIS). The ANFIS was first proposed in [34]. The ANN maps input space to an output space using a set of layered processing units called neurons that are interconnected by synaptic junction in parallel [35]. This is developed by passing raw data from an input layer through to its output layer continuously to produce ANFIS output which is compared with the target output. This computation using the optimization procedure expressed as the sum of the squared difference between the ANFIS output and the target output is used to adapt the synaptic connection (weight) so that ANFIS can learn the pattern of the input data (ie the learning process of ANFIS). The ANFIS in this process learns the pattern of information presented to the network and can predict the output of a new set of raw data presented to it after training [36, 37]. FISs are based on fuzzy logic, the fuzzy rule which can be viewed as computing with the word rather than the numbers and fuzzy reasoning or linguistic variable whose membership is a matter of degree. The use of word computing can be likened to human perception and exploit the tolerance for imprecise raw data using linguistic variables. The integration of the learning ability of ANN and knowledge representation for making a deduction from observation in FIS to form ANFIS of-

fers a superior technique that can dynamically model load balancing through handover mechanism [36, 38]. The ANFIS structure adopted for this work is a 5 layered feed-forward neural network and is implemented as a Takagi and Sugeno (TKS) fuzzy inference system which is more compact and computationally efficient than the Mamdani system [39]. To get the system output, a constant expression is added to the linear combination of the input variables for each rule while the final output is the weighted average. The system is a five-layered architecture with many nodes in each layer. For simplicity, the ANFIS structure is considered with two inputs x and y and output z for the TKS model. The common rule set of if-then rules is as follow;

If x is A1 and y is B1, then f1 = p1x + q1y + r1

If x is A2 and y is B2, then f2 = p2x + q2y + r2

The nodes in the first layer are used to generate membership grades of input variables. Adopted for node function implementation in this work is the Gaussian membership function that varies between 0 and 1.

$$o_{1,i}=\mu A_i(x), \text{ for } i=1,2$$
$$o_{1,i}=\mu B_{i-2}(y), \text{ for } i=3,4$$
$$\mu A(x) = \frac{1}{1+x\left|\frac{x-c_j}{a_i}\right|^{2b}} \quad (1)$$

Where x and y are inputs to node I, A and B are linguistic variables (insufficient, low, medium and strong), $o_{1,i}$ and μA_i (x) (orμB_{i-2} (y)) is ith node output and MF of x (or y) in the linguistic variable A_i (or B_{i-2}). The MF shape is determined by the parameter set (a_i,b_i,c_i). The parameters of this layer are called premises parameters. The layer two nodes determine the firing strength of the rule. This is the fixed node in nature whose output is the product of membership functions.

$$w_i = O_{2,1} = \mu A_i(x)\mu B_i(y), i=1,2 \quad (2)$$

Layer three comprises fixed nodes used to determine the ratio of the ith rule's firing strength to the total number of the firing strengths whose output is referred to as normalised firing strength.

$$O_{3,i} = \overline{w} = \frac{w_i}{w_1+w_2}, i = 1,2 \quad (3)$$

Where w_i and $\overline{w}$ are called firing strength and normalised firing strength respectively.

The mapping of the output membership functions (MF) is carried out at the fourth layer (also called the defuzzification layer) by its adaptive nodes

$$\overline{w}_i f_i = O_{4,i} = \overline{w}_i(pix + qiy + ri) \quad (4)$$

Where { pix + qiy + ri} is the parameter set of the layer four adaptive nodes and whose output resulting from the inference of the rules is called consequent parameters.

Lastly, the fifth layer consists of fixed nodes that determine the final/ANFIS output through weighted average.

$$O_{5,i} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (5)$$

4.1 HYBRID LEARNING ALGORITHM

The essence of learning in ANFIS is to tune all the adjustable parameters to ensure there are minimal errors between the ANFIS output and the target output. Here, the efficiency of the training is improved by using a combination of the Least-Square Algorithm (LSA) and Back Propagation Gradient Descent Algorithm (BPGDA). There are two steps in a hybrid optimization algorithm which include a forward pass (LSA) and a backward pass [40]. In the forward pass, the premises (antecedent) parameter is fixed while the LSA is used to optimize the consequent parameters.

In the backward pass (backpropagation training algorithm), the consequent parameters are fixed while the gradient descent is used to update the parameters of the premises. These two steps are repeated until optimum premises and consequent parameters are identified for the FIS system. The output of Fig 1 is expressed in equation 6.

$$\begin{aligned} f &= \overline{w}_1 f_1 + \overline{w}_2 f_2 \\ &= \overline{w}_1(p_1x + q_1y + r_1) + \overline{w}_2(p_2x + q_2y + r_2) \\ &= (\overline{w}_1x)p_1 + (\overline{w}_1y)q_1 + (\overline{w}_1)r_1 + (\overline{w}_2x)p_2 \\ &\quad (\overline{w}_2y)q_2 + (\overline{w}_2r_2) \end{aligned} \quad (6)$$

4.2. ANFIS CONFIGURATION

As discussed in the above section, the inputs to the ANFIS model are the RSSI, the RxQual and the number of free channels. The inputs data were first changed into linguistic variables each with Gaussian Membership (GM) function. The GM function was used for its smoothness, concise notation and nonzero at all times. These three crisp inputs parameters and four fuzzified variables gave rise to 64 knowledge-based or inference systems. The output of the inference engine is a fuzzy set derived from the application of fuzzy variables on the fuzzy rule. In practical case, the crisp output is what is required. The fuzzy output is, therefore, further defuzzified to produce crisp and quantifiable output which is the hysteresis value that determines the handover decision. The defuzzification technique adopted for the work is the centre-of-gravity (COG) which employs a weighted average.

The ANFIS model in Fig 1 is made up of 158 nodes, 64 linear parameters, 24 nonlinear parameters, 64 fuzzy rules and 88 total numbers of parameters. The total number of training data pairs is 1200, the same as the number of checking data pairs. To achieve good generalisation, it's normal to have the number of training data pairs far more than the total number of parameters. The ratio of the training data pairs to parameters numbers is 1200/88 = 13.6.

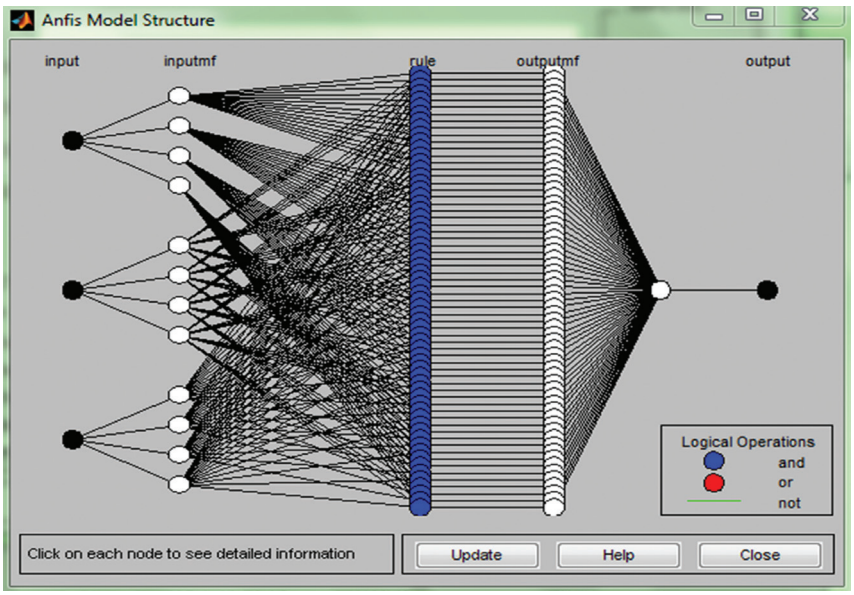


Fig. 1. ANFIS model structure for a dynamic handover decision.

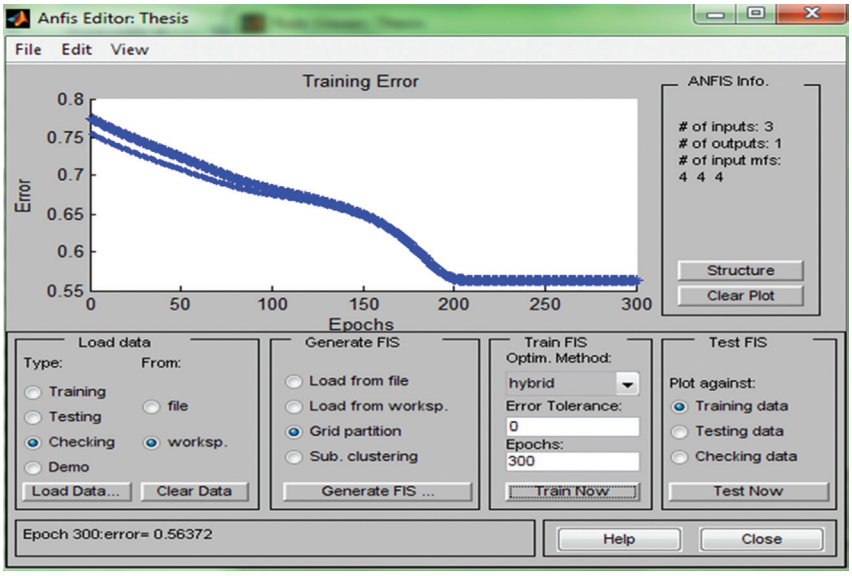


Fig. 2. ANFIS configuration for training FIS for handover decision

5. RESULT AND DISCUSSION

The developed ANFIS model employs the hysteresis value for QoS aware dynamic load balancing through the handover mechanism. The ANFIS was able to correctly predict the checking data output from the checking data input. The plot shows the ANFIS was correctly trained and can dynamically predict the hysteresis value from unknown data (data not used for the training).

Figs 3 to 5 are the 3D surface representation of ANFIS model prediction from a combination of two input variables from RSS, RxQual, and Free Channels. Each surface illustrates the impact of two input variables on the Hysteresis value. In cellular systems, RSS and RxQual are both fundamental quality performance indicators. The ANFIS system increases the hysteresis value as the RSS decreases and RxQual increases in value. It is very difficult to observe an overriding effect of one variable over the other when these are the predominant input metrics as shown in Fig. 3. A closer look at Fig 3 indicates that the RxQual is indeed the most critical factor in determining the hysteresis value.

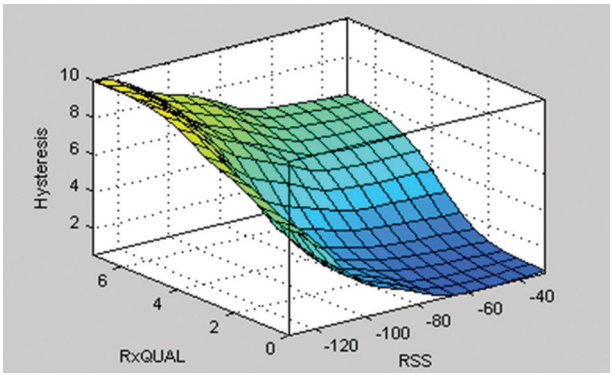


Fig. 3. ANFIS Model 3D Surface – Impact of RSS and RxQual on Hysteresis Value

As the number of free channels in a cellular cell increases, hysteresis value decreases, and vice versa. The number of free channels is not a quality parameter indicator but should be seriously considered in load balancing through the handover mechanism. The RSS has a dominant effect on hysteresis value over several free channels as illustrated in Fig. 4.

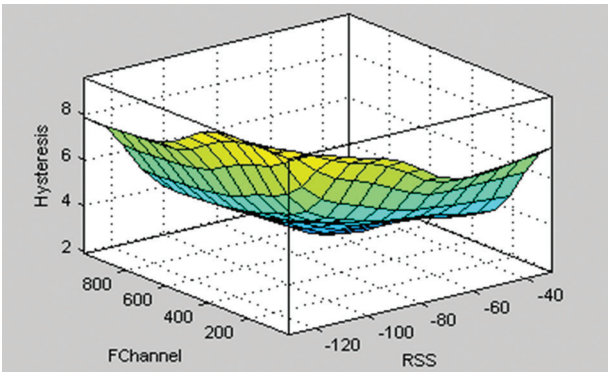


Fig.4. ANFIS Model 3D Surface – Impact of RSS and No. of Free Channel on Hysteresis Value.

Fig. 5. depicts the impact of received signal quality and the number of the free channel on hysteresis value. It can be observed that the RxQual has a more dominant effect over the number of free channels in determining the hysteresis value and subsequently the load balancing.

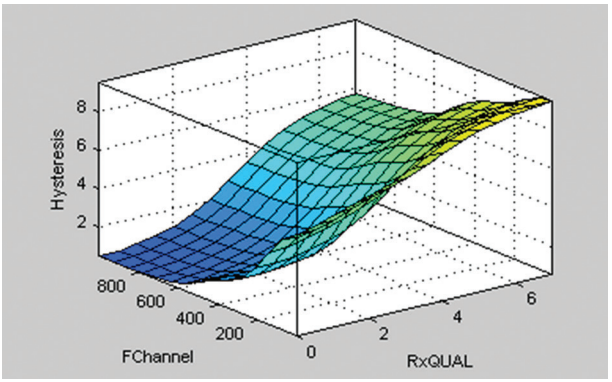


Fig.5. ANFIS Model 3D Surface – Impact of No. of Free Channel and RxQual on Hysteresis Value

Figures 6 to 8 illustrate individual parameter effects on load balancing and congestion avoidance. The blue colours indicate that the network is at the optimum condition concerning the three parameters under consideration. Handover is considered or triggered in the green colour while it is unexpected for MS to hang on to BS at the yellow colour. The yellow colors indicate critical network conditions and no communication can be established in this region. It also indicates that at least two of the parameters under consideration are in a poor state or insufficient for communication. The developed system will not allow the network to reach the yellow region through the handover mechanism.In Fig 6, it took up to -100dB for there to be an appreciable increase in hysteresis value. This is because above -100dB is considered sufficient for effective communication. At -110dB the RSS value becomes insufficient for communication and the load balancing is triggered at hysteresis value six (6) irrespective of other parameter values. The same can be said of received signal quality with a hysteresis value of 6 at RxQual 5 in Fig. 7. By cellular network standard, there should be no communication when RSS and RxQual are -110dB and 5 values respectively. Most mobile receivers have a sensitivity value of -115dB. RxQual value shows the higher impact on hysteresis value as there is an appreciable increase in hysteresis value halfway through the RxQual values. As expected the number of the free channel did not show any appreciable increase in hysteresis value until at 300 free channels from 1000.

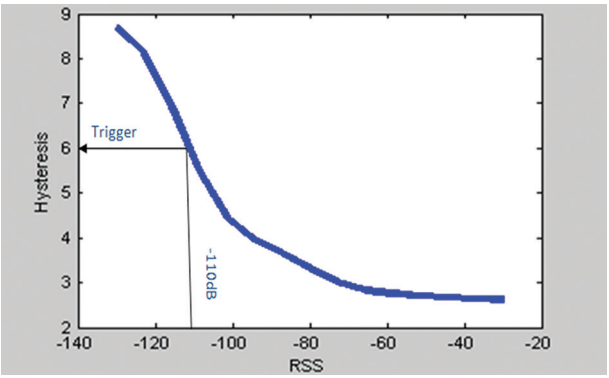


Fig. 6. The effect of RSS on Hysteresis Value

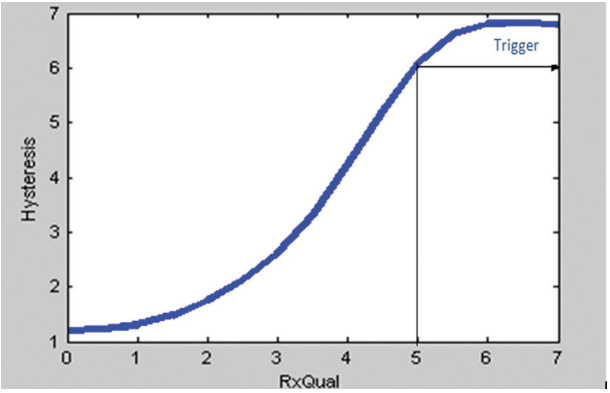


Fig. 7. The effect of RxQual on Hysteresis value

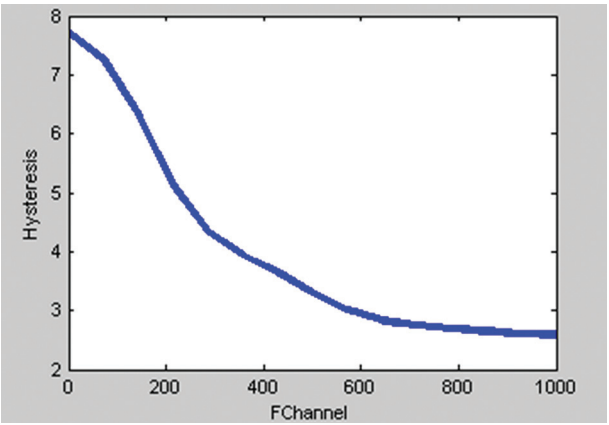


Fig. 8. The effect of numbers of free channels on Hysteresis value

Fig. 9 (a) shows the linear regression between the ANFIS output and the corresponding targets value

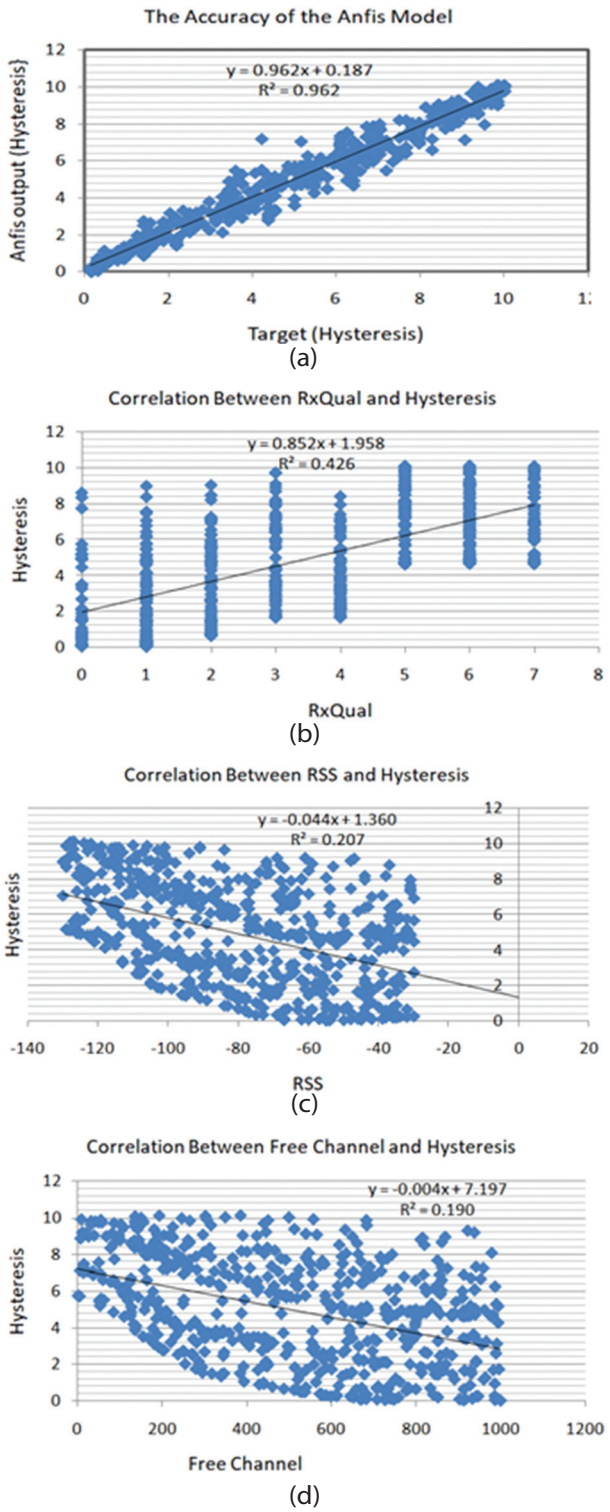


Fig. 9. Regression and Correlation between independents and dependent variables

In summary, these relationships illustrate that RSS and RxQual are the dominant parameters in determining the load balancing process and congestion avoidance, and should be controlled. The number of the free channel did not reflect well when compared with other load balance parameters in determining the hysteresis value. Based on illustration from From table 3 and figs 3 to 8, it can be observed that at Hysteresis value 6 one or two of the independent variables are almost insufficient to maintain good communication and to avoid call drop the handover mechanism should be triggered. In this handover process, the overall hysteresis state of the serving cell and its neighboring cells should be considered. A handover to the cell with a hysteresis value of 5 or less should always be considered. This is important to avoid excessive and ping-pong effects when the neighboring cell is highly loaded or has a hysteresis value of 6 or close. The increase in hysteresis value either sustains or triggers the load balancing process.

6. CONCLUSION

In this paper, ANFIS model prediction for dynamic load balancing and congestion avoidance through the handover mechanism is proposed. As can be seen in the results, the RxQual and RSS play a key role in determining the hysteresis value for load balancing complemented by the number of the free channel which ensures congestion avoidance. The results show that when the Hysteresis value approaches 6 the handover process should be triggered. The hysteresis value of the concern MS with neighboring cells is considered in determining the most suitable cell to handover to using the ANFIS technique, thereby achieving dynamic load balancing and congestion avoidance. Moreover, RSS and RxQual can also be used as the key performance indicators to decide scheduling and mobile assignment in a mobile communication network. Based on the simulation result, it is assumed that the performance of the system when implemented will be close if different from the result obtained in this work. This work was able to achieve dynamic load balancing, congestion avoidance and avoided the 'ping pong' effect that is often an issue with handover. This work also reduces the rate of call drop, call block and energy consumption of the base station associated with the fluctuation (ping pong effect) in conventional handover. Therefore, the subscribers' quality of experience (QoE) would have been enhanced. This work can be furthered by adopting a test-bed setup using a Wireless Network Simulation (WNS) test suite or Emulator.

7. REFERENCE

[1] T. Iliev, T. A Bikov, G. Mihaylov, E. Ivanova, I. Stoyanova, V. Keseev, "Algorithm for Congestion Control in LTE Mobile Network", Software Engineering and Algorithms in Intelligent Systems, Vol. 763, 2018.

[2] S. Kramer, S. Molnar, S. Solymoi, A. Mihaly, "On the Benefit of Multi-Domain Congestion Control in LTE Networks", Proceedings of the IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks, Limassol, Cyprus, 2019, pp. 1-6.

[3] S. Pa, S. Mandale, A. Adiwjaja, "A New Method for Congestion Avoidance in Wireless Mesh Networks", Proceedings of the 2nd International Conference on Data and Information Science, IOP Conference Series: Journal of Physics, Conference Series 1192, 10-14 October 2019, pp. 1-13.

[4] A. Singh, S. P Singh, U. N. Tripathi, M. Mishra, "Optimizing Call Drops in cellular Network using Artificial Intelligence Based Handover Scheme", International Journal of Advanced Research in Computer and Communication Engineering, Vol 6, No. 1, 2017.

[5] E. O. Oluwafemi, O. E. Nnonye, U. Okechukwu, A. L. Adewale, "Prediction of Call Drops in GSM Networks using Artificial Neural Network", Jurnal Teknologi dan Sistem Komputer, Vol 7, No. 1, 2019, pp. 38-46.

[6] C. O. Nnamani, C. L. Anioke, C. I. Ani, M. Anedda, M. Murroni, "Load-Shared Redundant Interface for LTE Access Network", https://www.researchgate.net/publication/318020116 (Online).

[7] Y. Zaki, T. Potsch, J. Chen, L. Subramanian, C. Gorg, "Adaptive Congestion Control for Unpredictable Cellular Networks", ACM SIGCOMM Computer Communication Review, Vol. 45, No. 4, 2015, pp. 509-522.

[8] N. S. Khan, M. Umar, K. Anand, S. K Shabia, "Early Prediction of Congestion in GSM based on Area location using Neural Network" International Journal of Advanced Research in Electronics and Communication Engineering, Vol. 5, No. 5, 2016.

[9] S. Alam, A. Mittal, M. G. Siddiqui, T. Qamar, "Capacity Improvement by Cell Splitting Technique in CDMA System over Telecommunication Network", International Refereed Journal of Engineering and Science, Vol. 2, No. 7, 2013, pp. 1-8.

[10] A. Ozovehe, O. U. Okereke, E. Anene, A. U Usman, "Busy Hour Traffic Congestion Analysis in Mobile Macrocells", Nigeria Journal of Technology, Vol. 36, No. 4, 2017, pp. 1265-1270.

[11] C. O. Ohaneme, G. N. Onoh, E. N. Ifeagwu, I. Eneh, "Improving Channel Capacity of A Cellular SystemUsing Cell Splitting", International Journal of Scientific and Engineering Research, Vol. 3, No. 5, 2012, pp. 1-8.

[12] R. Arshad, H. Elsawy, S. Sorour, T.Y Al-Naffouri, M. Alouini, "Handover Management in 5G and beyond", A Topology Aware Skipping Approach, IEEE Access, Vol. 4, 2016, pp. 9073-9081.

[13] R. A. Saeed, "Performance Comparison of Handover Mechanisms in Mobile Wireless Communication Networks for Broadband Wireless Access Systems", International Journal of Computer Communication and Informatics, 2020, pp. 25-37.

[14] U. Sehgal, R. Sharma, "An Efficient Communication MCHO Protocol used Cellular Networks Technology", International Journal of Computer Science and Mobile Computing, Vol. 6, No. 1, 2017, pp. 52-60.

[15] K. Yap, Y. Chong, W. Liu, "Enhanced Handover Mechanism using Mobility Prediction in Wireless Networks", PLOS ON, Vol. 15, No.1, 2020.

[16] S. Moshe, S. David, "New call blocking versus handoff blocking in cellular networks", Wireless Networks, Vol. 3, 1997, pp 15-27.

[17] Girma. S. T, Konditi. D. B. O and Ndunga, E. N, Fuzzy Logic Based Traffic Balancing in a GSM Network, Journal of Research in Engineering, Vol. 1, No. 2, 2014, pp. 63-74.

[18] O. A. Jeremiah, A. F. Olasebikan, O. Abidemi, "Handover in Mobile Wireless Communication network – A Review", International Journal of Advanced Engineering, Management and Science, Vol. 3, No. 9, 2017, pp. 934-940.

[19] S. Neeraja, G. S. B. Rao, A. L. Narayana. "Analysis of Adaptive Hysteresis Based Horizontal Handoff Algorithm for GSM", International Journal of Innovative research in Electrical, Electronics, Instrumentation and Control Engineering, Vol. 1, No. 9, 2013.

[20] M. Toril, S. Pedraza, R. Ferrer, V. Wille, "Optimization of handover margins in GSM/GPRS networks", Proceedings of the 57th IEEE Semi-annual Vehicular Technology Conference, Jeju, Korea, 22-25 April 2003.

[21] R. P. Ray, L. Tang, "Hysteresis Margin and Load Balancing for Handover in Heterogeneous Network", International Journal of Future Computer and Communication, Vol.4, No. 4, 2015

[22] A. A. Ateyero, M. K. Luka, "A Soft Computing Approach to Dynamic Load Balancing in 3GPP LTE", International Journal of Computer Application, Vol. 43, No. 19, 2012.

[23] P. Moungnoul, P. Pinprasert, T. Paungma, "Handoff decision in Cellular Mobile System using Neural Network", 2002, http://www.ecti-thailand.org/assets/papers/300_pub_24.pdf (Online).

[24] M. A. Raheem, O. U. Okereke, "A Neural Network Approach to GSM Traffic Congestion Prediction". American Journal of Engineering Research, Vol. 3, No. 11, 2014, pp. 131-138.

[25] K. O. Omotoye, M. A. Akogbe, O. M. Olaniyi, H. C. Inyiama, "Congestion Management on GSM Networks using Traffic Class Prioritization". International Journal of Engineering Research and Technology, Vol. 3, No. 12, 2014, pp. 174-178.

[26] D. F. A. Ali, F. A. Mustafa, M. O. M. Fdolelsid, "Automatic Handover Control for Distributed Load Balancing in Mobile Communication Network", 2016, Arxiv:1605.01228.

[27] L. Andreas, S. Szymon, J. Thomas, B Irina, "Load Balancing in Downlink LTE Self-Optimizing Networks", Proceedings of the IEEE 71st Vehicular Technology Conference, Taipei, Taiwan, 16 -19 May 2010.

[28] T. G Solomon, B. O. K. Dominic, N. N. Eward, "Fuzzy Logic Based Traffic Balancing in a GSM Network", Journal of Research in engineering, Vol. 1, No. 2, 2014, pp. 63-74.

[29] ETSI Technical Specification, "Digital Cellular Telecommunications System (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS), LTE, Service Accessibility (3GPP TS 22.011 version 13.4.0 Release 13), 2016, http://www.etsi.org/standards-search (Online).

[30] ETSI Technical Specification, "Digital Cellular Telecommunications System (Phase 2+); Radio Subsystem Link Control", 3GPP TS 45.008 version 9.4.0 Release 9, 2010, http://www.etsi.org/standards-search (Online).

[31] Radio Frequency Engineering: GSM Radio Parameter Details and Score Evaluation from Drive Test (DT) End, 2015, www.cwnabook.blogspot.com/2015/12/gsm-radio-parameter-details-and-score.html?m=1 (Online)

[32] S. Serkan, T. Nurettin, L. Bunyamin, Analysis of Mobile Coomunication Signals with Frequency Analysis Method, Gazi university Journal of Science, Vol. 25, No. 2, 2012.

[33] GSM/GPRS Modem, "GSM/GPRS Modem, Telemetry and Data Accessories "At + CSQ in dbm", 2011, http://www.gprsmodems.co.uk/images/csq1.pdf.

[34] J R Jang, "ANFIS: Adaptive Network-Based Fuzzy Inference System", IEEE Transactions on Systems, Man and Cybernetics, Vol. 23, No. 3, 1, 1993, pp. 665-685.

[35] L. S. Ezema, C. I. Ani, "Artificial Neural Network Approach to Mobile Location Estimation in GSM Network", International Journal of Electronics and Telecommunications, Vol. 63, No. 1, 2017, pp. 39-44.

[36] A. A Atayero, M. K Luka, "Adaptive Neuro-Fuzzy Inference System for Dynamic Load Balancing in 3GPP LTE", International Journal of Advanced Research in Artificial Intelligence, Vol. 1, No. 1, 2012.

[37] F. Castellanos, N. James, "Average Hourly Wind Speed Forecasting with ANFIS", Proceedings of the 11th Americas Conference on Wind Engineering, San Juan, Puerto Rico, 22-26 June, 2009.

[38] S. Huang, "Cross-Layer Congestion Control with Deep Neural Network in Cellular Networks, KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science, Stockholm, Sweden, 2019.

[39] D. S Reddy, C Chandrasekhar, "Support Vector Regressive Dragonfly Optimized Shift Invariant Deep Neural Learning Based Handover for Seamless Data Delivery in Heterogeneous Network", International Journal of Computer Network and Applications, Vol 7, No. 4, 2020.

[40] V. Vaidhehi, "The Role of Dataset in Training ANFIS System for Course Advisor", Department of Computer Science, Christ University, Bangalore, 2016

Reinforcement Learning based Gateway Selection in VANETs

Original Scientific Paper

Hasanain Alabbas
Budapest University of Technology and Economics,
Faculty of Electrical Engineering and Informatics, Department of Networked Systems and Services
H-1117, Budapest, Hungary
Al-Qasim Green University
Computer Center
Babel, Iraq
hasanain@hit.bme.hu

Árpád Huszák
Budapest University of Technology and Economics,
Faculty of Electrical Engineering and Informatics, Department of Networked Systems and Services
H-1117, Budapest, Hungary
huszak@hit.bme.hu

Abstract *– In vehicular ad hoc networks (VANETs), providing the Internet has become an urgent necessity, where mobile gateways are used to ensure network connection to all customer vehicles in the network. However, the highly dynamic topology and bandwidth limitations of the network represent a significant issue in the gateway selection process. Two objectives are defined to overcome these challenges. The first objective aims to maximize the number of vehicles connected to the Internet by finding a suitable gateway for them depending on the connection lifetime. The second objective seeks to minimize the number of connected vehicles to the same gateway to overcome the limitation of gateways' bandwidth and distribute the load in the network. For this purpose, A gateway discovery system assisted by the vehicular cloud is implemented to find a fair trade-off between the two conflicting objectives. Proximal Policy Optimization, a well-known reinforcement learning strategy, is used to define and train the agent. The trained agent was evaluated and compared with other multi-objective optimization methods under different conditions. The obtained results show that the proposed algorithm has better performance in terms of the number of connected vehicles, load distribution over the mobile gateways, link connectivity duration, and execution time.*

Keywords*: Gateway selection, Reinforcement learning, Proximal policy optimization, VANET*

1. INTRODUCTION

1. INTRODUCTION

Vehicular Ad hoc network (VANET) is one of the interesting fields in the Intelligent Transportation System (ITS) that exploits the moving vehicles as mobile nodes in the network. Because of its wide applications of increasing safety for drivers, reducing car accidents, and providing Internet to users, it has attracted researchers' interest [1]. VANET infrastructure is composed of two communication entities, On-Board Unit (OBU), which is integrated inside the vehicles, and Road Side Unit (RSU), which is mounted on the roadsides or near traffic intersections [2]. These communication entities allow two types of communications. Vehicle-to-Vehicle (V2V), which enables the vehicular nodes to contact each other directly, and Vehicle-to-Infrastructure (V2I), in which vehicles can communicate with RSUs [3]. Providing vehicles with a permanent internet connection has become an urgent necessity to feed drivers with relevant road information and offer a comfortable trip for passengers [4][5]. Providing the Internet for vehicles requires finding a suitable gateway. Unfortunately, the implementation of this goal is facing many challenges, most notably the highly dynamic topology of the network and the bandwidth limitations [3][6][7]. Most gateway discovery techniques are based on Inquiry and Solicitation messages sent and received between the vehicular nodes to find a suitable gateway [8][9]. These techniques have many issues (broadcast storm problem, overhead) when the nodes increase [10]. The progress in cloud computing and making it compatible with ITS provides a valuable opportunity to benefit from cloud computing resources utilized by VANET

services [11]. Many research efforts have been made in this area, which produced a new paradigm called Vehicular Cloud (VC) [12][13]. VC offers many features such as collecting vehicle information, optimizing traffic control, and detecting congestion [14]. Due to the massive services and features provided by VC, some studies have invested it to perform more complex computations and find efficient solutions to improve gateway selection and address overload problems. However, the disadvantage of these solutions is that they do not find a gateway with the highest link connectivity duration (LCD), because they don't take the nature of roads and their vulnerability to traffic congestion into account. The efficiency of these solutions decreases in urban area so that the execution time increases catastrophically with the number of vehicles. VC is exploited to build a novel model by using reinforcement learning. The proposed model is designed to optimize gateway discovery by maximizing LCD and minimizing bandwidth overload. For this purpose, Multi-Objective Reinforcement Learning (MORL) is used. In this paper, the Proximal Policy Optimization (PPO) model is adopted to train the agent due to its better performance compared to other standard reinforcement learning algorithms. To the best of our knowledge, the previous studies of gateway selection use current speed, direction, and distance as crucial factors in the selection, ignoring other factors that have a significant influence on selection like road density and intersections. The main contributions of this paper are as follows:

1. The algorithm implicitly takes into consideration factors related to road density and the impact of intersections in addition to the traditional factors (speed, distance, and direction).
2. The decision of electing the gateways is based on the real and actual time of the link connectivity duration between the vehicles, so the algorithm gives better results in terms of stability and scalability.

The rest of this paper is arranged as follows. Section 2 reviews some related literature on VANET and gateway selection solutions. Section 3 presents the proposed system model components in detail and discusses the reinforcement learning method used in the model. Section 4 evaluates the presented technique with other existing multi-objective optimization solutions. The conclusion and future work are presented in section 5.

2. RELATED WORKS

Providing permanent access requires finding a suitable gateway, which has a direct connection to the Internet. In literature, the gateway can be a stationary station (RSUs, cellular base stations), which is considered as a part of vehicular network infrastructure [15]. In [16], the authors suggested a Fuzzy QoS-balancing Gateway Selection algorithm to connect the vehicles to the LTE infrastructure. The proposed algorithm employs the distributed LTE Advanced eNodeBs as stationary gateways to meet the vehicles' needs. The connections between the vehicles and LTE advanced eNodeBs are either directly or by choosing a relay gateway. Fuzzy logic is applied to select the best gateway based on signal strength, load, link connectivity duration, and QoS traffic classes. However, the drawback of these kinds of solutions is the handover of connections that are generated as a result of the vehicles' high speed compared to the fixed road infrastructure. Moreover, the gateway selection process uses a reactive approach, where the vehicles broadcast the Solicitation message to seek a suitable gateway, which engenders a high amount of overhead.

The study [17] proposed a routing protocol for mobile gateway discovery to ensure Internet access for vehicles in the area where it is not available. Many parameters have been adopted, namely robust parameters, like received signal strength (RSS), trust connection, the number of hops, and route lifetime, to establish a robust route protocol for mobile gateway discovery. Two stages are defined to set the routing protocol. The first stage is the gateway selection process which starts when the moving vehicles toward the UMTS base station enter its coverage area. These vehicles declare themselves as mobile gateways if the received signal strength of UMTS is greater than the RSS threshold. Relays selection represents the second stage in which the mobile gateways select relays based on robust parameters. The simulation results exhibit good performance in terms of packet delivery and decreasing the overhead when applied in a highway scenario. However, the proactive and reactive strategies used can decrease the throughput when the vehicle's number increases. Idrissi et al. [18] used the vehicular cloud architecture to develop the gateway discovery system. They adapted a multi-criteria decision approach known as Preference Ranking Organization METHod for Enrichment of Evaluations (PROMETHEE) to find the best gateway. Several criteria are considered, representing the difference between the customer vehicles that tend to get access to the Internet and the mobile gateways that have a direct Internet connection. These criteria have been used to increase the number of connected customer vehicles and decrease the traffic routed by the mobile gateway. However, the gateway selection mechanism in this study lacks the use of optimization techniques. Sara Retal and Abdellah Idrissi proposed a method to improve the mobile gateways selection [19]. Multi-Objective Optimization is considered to overcome the weakness of the previous study by maximizing the number of connected vehicles and minimizing the overload of the gateways. Different models are used to find the best solution, which represents a trade-off solution of different conflicting objectives. To the best of our knowledge, this study is considered one of the pioneering studies in the scope of mobile gateway selection techniques; therefore, we will adopt one of its used model, namely Integer Optimization Problem (IOP), as

a benchmark. This algorithm can perform well when vehicles keep a relatively constant speed and direction, especially on highways. Still, they do not perform well in urban areas because of the roads' nature and intersections, which significantly affect vehicles' variation of speed and direction. Moreover, the execution time of the gateway selection process is relatively high, and it increases significantly when the vehicles increase. A novel model is presented to discover the mobile gateways by using reinforcement learning.

3. GATEWAY SELECTION ARCHITECTURE

The main objective of this study is to find a suitable Mobile Gateway (MG) upon request from Client Vehicles (CVs). MGs are vehicles with direct Internet access, whereas CVs represent all vehicles that have no direct connection. In the analyzed urban scenario, we assume that the public transport buses are equipped with Internet access and can serve as MGs. Their convenient speed, which usually does not exceed 40 km/h, and their regular geographic distribution in urban areas make them have a highly predictable day-to-day pattern [20]. As shown in Fig. 1, the proposed system consists of CVs, MGs, VANETs infrastructure (4G/5G base station, RSU), and Vehicular Cloud (VC). The CVs and MGs can communicate with each other via V2V connection, while VANETs infrastructure represents the link between the VC from one side and CVs and MGs on the other side. VC consists of two servers in which, The Registrar server monitors the VANETs environment, collects the vehicles' information, and registers it in a dataset. In contrast, the Agent server is in charge of gateway discovery for vehicles trying to access the Internet.

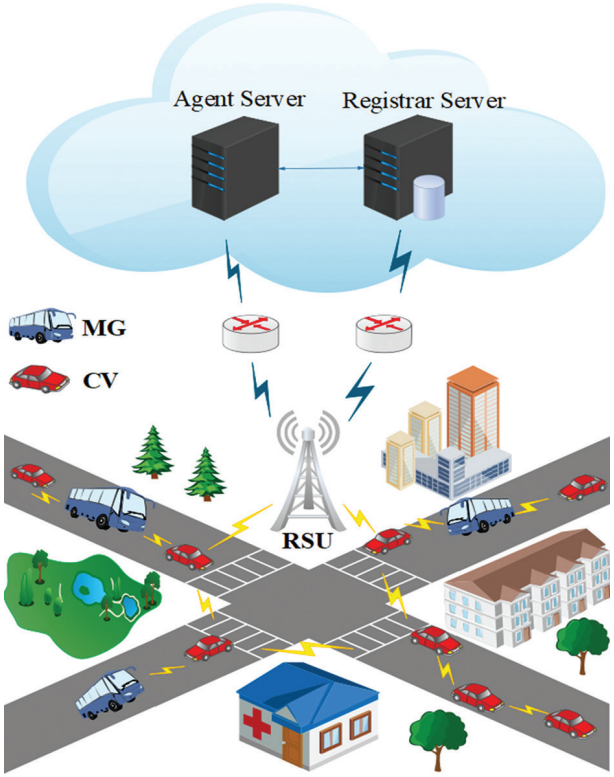


Fig. 1. System architecture.

The Registrar Server collects the necessary information on CVs and GWs related to speed, geographical location, direction, and the link connectivity duration (LCD) between CVs with all MGs. This information is stored in a database containing records generated for each CV. Each registry contains the difference between one CV and all the MGs in terms of the geographical location (longitude and latitude), speed, direction, and the traffic amount routed by the MG and LCD. This information is collected and stored in the database periodically so that as soon as the process of filling in the data of the current record is completed, the process of adding a new record begins. The gateway discovery system is built in the Agent server by using reinforcement learning. The main goal was to achieve two contradicting objectives by finding the best trade-off between them. These objectives are:

- Objective 1: Increasing the number of CVs connected to the MGs with the highest LCD.
- Objective 2: Minimizing the traffic volume routed by MGs by decreasing the number of CVs connected to the same MG.

In the training phase, the RL agent starts to adapt and learn from the environment of VANET based on the data collected by the registrar server. The RL agent will be able to find the best MG for each CV when the training phase ends. The role of the Agent server is to select the best GW for each CV requesting Internet access.

3.1. REINFORCEMENT LEARNING

Reinforcement learning (RL) is a branch of machine learning that imitates human behavior in acquiring skills by planning for the future and deciding based on it in a specified environment. The main objects in an RL problem are the agent and the environment. The concepts (state (*S*), action (*A*), reward (*R*)) represent the interaction of the agent with its environment. The agent monitors the environment state (s_t) and takes action (a_t) at the time (t), which causes a state transition to a new state (s_{t+1}). The correctness of the decision taken is determined by the reward (rt) given to the agent. The reward function $R(r|s, a, s')$ represents the immediate reward probability for state transition [21], [22] as shown in Equation (1):

$$R(r \mid s, a, s') = \Pr\left(r_t = r \mid s_t = s, a_t = a, s_{t+1} = s'\right), \quad (1)$$

the policy $\pi(a, s)$ defines the behavior of the agent depending on its observations. The mapping between the action (a) and the state (s) is modeled by the policy $\pi(a, s)$, which represents the action (a) probability as follows:

$$\pi(a \mid s) = \Pr\left(a_t = a \mid s_t = s\right), \quad (2)$$

the agent explores the optimal policy $\pi^*(a, s)$ by maximizing accumulated discounted reward for each $s \in S$ and $a \in A$ shown in Equation (3):

$$\pi^*(a \mid s) = \underset{\pi(a|s)}{\arg max} \sum_{t=t_0}^{t_{end}} \gamma^{t-t_0} r_t, \quad (3)$$

where $\gamma \in (0,1)$ is the discount factor and t is the time horizon. Policy optimization algorithms can be classified into two categories which are value-based algorithms and policy-based algorithms. Compared with the value-based algorithm, policy-based algorithms have better convergence and are more convenient for large action spaces. Proximal Policy Optimization (PPO) [23] algorithm is an actor-critic method that combines the value-based and the policy-based algorithm. Two neural networks are applied. The first one, named *actor*, takes the state *(s)* as entries and outputs the policy $\pi(a, s)$, while the second one, named *critic*, optimizes *V(s)* that measures the goodness of the action (a). PPO uses the advantage *A(s, a)* to reduce the estimation variance, which is expressed in the following:

$$A(s_t, a_t) = Q(s_t, a_t) - V(s_t)\ , \quad (4)$$

$$Q(s_t, a_t) = r_t + \sum_{i=1}^{T-1} \gamma^i r_{t+i} + \gamma^{t+T} V(s_{t+T})\ , \quad (5)$$

where *Q(s,a)* represents the cumulative discount reward when action at is taken for the state s_t, while *V(s)* represents the baseline, this technique allows updating the policy network in a direction that chooses better actions. PPO uses the trust-region (TRPO) method to ensure that the new updated policy never goes far away from the current policy, making it more stable and reliable. The primary objective function of PPO is denoted as $L^{CLIP(\theta)}$:

$$L^{\mathrm{CLIP}(\theta)} = \mathbb{E}_\tau\big[min\big(R_t A_t, \mathrm{clip}\ (R_t, 1-\epsilon, 1+\epsilon)\big)A_t\big], \quad (6)$$

where A_t is an abbreviation of $A(s_t, a_t)$, ϵ denotes a small positive constant, and the policy ratio (R_t) measures the similarity between the updated policy and old policy as shown in Equation (7):

$$R_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\mathrm{old}}}(a_t|s_t)}\ , \quad (7)$$

while clipping function, clip (R_t, 1-ϵ, 1+ϵ) ensures the R_t moving inside the interval [1- ϵ, 1+ ϵ]. For these reasons, the PPO algorithm is adopted in the proposed gateway selection system, namely PPO-GS. Three components should be defined carefully to enable the agent to sense the environment and make the right decision: state, action, and reward.

3.2. DEFINITION OF OBSERVATION STATE

The vehicles in VANET are classified into two categories MGs and CVs. The state (st) will be created for each CVi that needs Internet access and looks for a connection to a suitable MGj. It represents the relationship between the CV and all the MGs in terms of geographical location, speed, and available bandwidth. The state is expressed by the entries as follow:

$$X = \begin{bmatrix} Lo_{i1} & Lat_{i1} & V_{i1} & \theta_{i1} & T_1 \\ Lo_{i2} & Lat_{i2} & V_{i2} & \theta_{i2} & T_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Lo_{ij} & Lat_{ij} & V_{ij} & \theta_{ij} & T_j \end{bmatrix} \quad (8)$$

- $Lo_{ij} = Lo_i - Lo_j$, where Lo_i and Lo_j denote the longitude of CV_i and MG_j, respectively.
- $Lat_{ij} = Lat_i - Lat_j$, where Lat_i and Lat_j denote the latitude of CV_i and MG_j, respectively.
- $V_{ij} = V_i - V_j$, where V_i and V_j denote the velocity of CV_i and MG_j, respectively.
- $\theta_{ij} = \theta_i - \theta_j$, where θ_i and θ_j denote the direction of CV_i and MG_j, respectively.
- T_j denotes the traffic volume routed by MG_j.

The difference in longitude and latitude is applied between CVs and MGs rather than distance. By using the difference of the coordinates, the algorithm can decide whether the MG is at the front of the CV or not. The number of entries in the state increases with the increase in the number of MGs since the state represents the relationship of each CV to all MGs. Since the relationship of the CV to each MG is represented by five parameters *S*=(Lo,Lat,V,θ,T), the total number of entries to represent the state is $|S|\cdot|MG|$,where $|S|$ is the number of parameters used to describe a state, while $|MG|$ is the number of MGs. The number of GWs distributed in the environment is 20, so the agent state is composed of 100 entries.

3.3. DEFINITION OF AGENT ACTION AND REWARDS

The decision taken in the gateway selection system is defined as the agent actions. In the proposed system, the set of actions represents the number of MGs. Action $a = \{a_1\ a_2\ a_3\ ...\ a_n\}$, where a_1 represents the selection of MG_1 while a_n represents the selection of MG_n. The reward is assigned based on two metrics: the first metric is the connection lifetime between the CV and MG, whereas the second is the traffic amount routed by each MG. The first parameter aims to increase the number of vehicles connected to the Internet, while the second one aims to reduce the CV_s linked to the same MG. Setting the reward function with two contradicting objectives needs to apply Multi-Objective Reinforcement Learning (MORL). MORL aims to find a trade-off solution for multiple conflicting objectives. Several approaches can be used to achieve this goal [24]. Weighted Sum Approach (WSA) is applied to solve the multi-objective problem in the reward function. The weight (*w*) of each metric is set to define the objectives preferences as shown in Equation (9):

$$R = w_1 \cdot lcd_{ij} + w_2 \cdot T_j, \quad (9)$$

where lcd_{ij} denotes the link connectivity duration value between CV_i and MG_j, while T_j represents the traffic volume routed by MG_j. w_1 and w_2 take values between 0 and 1 depending on the importance of the objectives so that $w_1 + w_2 = 1$.

The reward value is positive when the action is valid. Else, the reward is negative. The positive reward ranges in value between 0 and 20, while the negative reward is (-4).

The negative reward is applied in two cases:

1. The agent selects an MG which is out of the CV coverage area.
2. The agent selects an MG that does not have enough amount of traffic.

Setting the reward in this manner motivates the agent to find an MG with the best link connectivity duration and the least amount of traffic handled by it. In contrast, the negative reward ensures that the agent avoids choosing an MG out of the communication range of CV, or an MG cannot dedicate a channel to a CV. As is known, there are no rules that specify the reward value. Any value can be used, provided the agent is learning correctly. The reward value was adopted as mentioned above after training the agent several times with different rewards values because the assumed value showed a faster response from the agent to learning.

3.4. AGENT STATE PARAMETERS

In the proposed system, training the agent relies entirely on the dataset generated by the registrar server. The dataset represents an enormous number of snapshots taken from the VANET environment. Each entry in the dataset consists of two main parts: 1) the first part represents the state s, which involves the amount of difference for each CV with all MGs in terms of longitude, latitude, speed, and direction, as well as the available amount of bandwidth for each MG. 2) the second part which contains the LCD values. PPO is employed to maximize the MGs selection return. The reward r is a multi-objective reward in which the agent tends to find an MG for a CV with the maximum LCD and minimum number of CVs connected to it. The episodic environment is considered during agent training so that each episode consists of 100 steps (finding MGs for 100 CVs). The VANET environment is variable and highly dynamic as it depends on moving nodes in which it is difficult to determine the future return. So in an environment as dynamic as the presented work, it is appropriate for the RL agent to maximize the current reward rather than the cumulative discount reward. This is done by adopting $\gamma = 0$. To prove our hypothesis, we applied different gamma values during the training process. We found that the lower the gamma value, the faster the agent learns. Fig. 2 shows the highest reward collected when $\gamma = 0$.

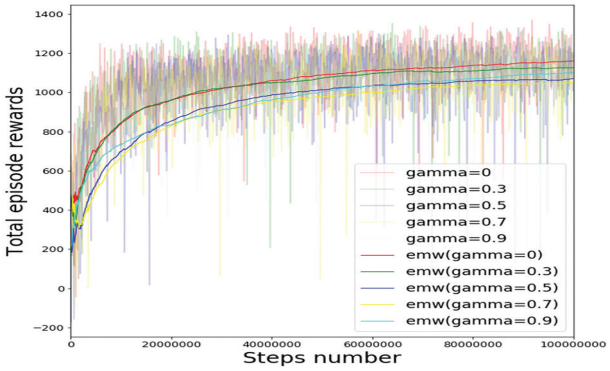


Fig. 2. Training the agent with different gamma values

4. RESULTS

This section exhibits the performance evaluation of the proposed algorithm. For comparison purposes, the well-known gateway selection method (A multi-objective optimization system for mobile gateways selection in vehicular Ad-Hoc networks) [19] is simulated, namely the MOO algorithm. The MOO technique has similar properties as this work. Therefore, it is adopted as a benchmark. It has a centralized algorithm to make a decision, while the rest of the recent studies are decentralized algorithms in which the gateway selection depends on sending messages between vehicles. This work is implemented by using the Python programming language. The stable-Baseline3 library is used to implement and train the RL agent [25], whereas Gurobi Optimizer is executed to solve the multi-objective optimization problems used in the MOO solution. The vehicles' mobility and their behavior are simulated by using SUMO. The simulation is performed under an urban area map of 1500 m x 1500 m using Open Street Map (OSM). OSM provides free maps from all around the world, which makes the simulation more realistic. The urban area environment is adopted as shown in Fig. 3 because it can be taken as a true scale of how successful an algorithm is. It is more challenging than the highway environment because of the road's nature (congestion in a rush, speed limit, intersections, etc.).

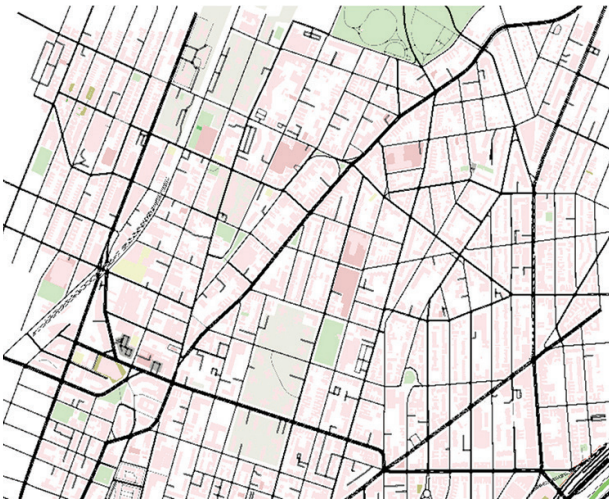


Fig. 3. Map from OSM

The number of MGs deployed in the simulation network is 20, while the number of CVs is 60-100. They are deployed randomly in the simulation to evaluate all the proposed models and compare them with the MOO method. By fixing the number of MGs, whereas the CVs number is variable. The entire simulation parameters are listed in Table 1. The proposed models are compared to the MOO models (MOO1, MOO2, and MOO3) to evaluate the performance. Several metrics have been used for the performance evaluation, including the number of connected vehicles (CVs), the distribution of CVs among MGs, LCD between CVs and MGs, and the execution time.

Table 1. Simulation Parameters

Parameters	Setting
Transmission Range	500 m
X-coordinate	0-1500 m
Y-coordinate	0-1500 m
Vehicles speed	0-20 m/s
MGs Number	20
CVs Number	60-100

In the simulation, different weights are applied, as mentioned in section 3.3, to determine the objectives preferences by utilizing the weighted sum approach. According to these weights, three methods are implemented in the presented study: PPO-GS1, PPO-GS2, and PPO-GS3. In PPO-GS1, the reward function maximizes the first objective only. In PPO-GS2, the reward function takes into consideration the two objectives but with more preference for the first one. Meanwhile, the objectives in PPO-GS3 take the same priority. On the other hand, three approaches which are called MOO1, MOO2, and MOO3, are defined in the MOO algorithm as detailed in Table 2.

Table 2. Objectives weights

		W1	W2
PPO-GS1	MOO1	1	0
PPO-GS2	MOO2	0.7	0.3
PPO-GS3	MOO3	0.5	0.5

PPO-GS1 has the highest number of connected CVs because it makes the decision based on objective1 and does not consider the MGs bandwidth limitation. The relying on objective1 in selecting MGs causes inequality and a wide variation in the distribution of CVs over the MGs, as shown in Fig. 4.

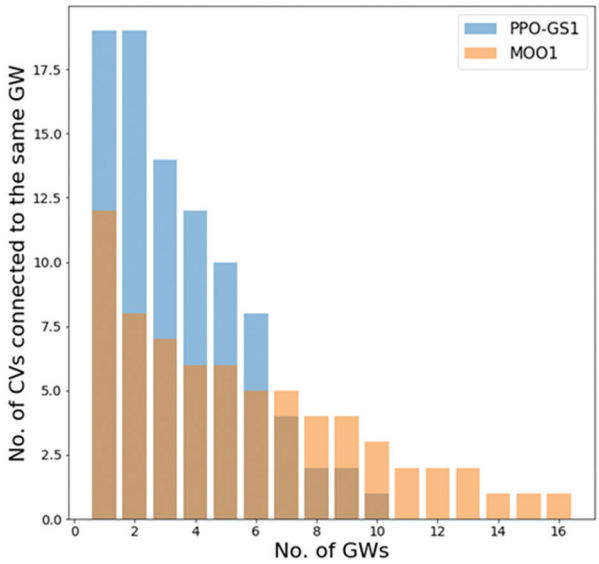


Fig. 4. CVs distribution with w_1=1 and w_2 =0

Fig. 5 and Fig. 6 show that the number of CVs connected to the same MGs decreases as the weight of the second objective increases, which takes into consideration the limitation of MGs bandwidth; therefore, PPO-GS3 and MOO3 exhibit more equitable distribution in comparison with other solutions.

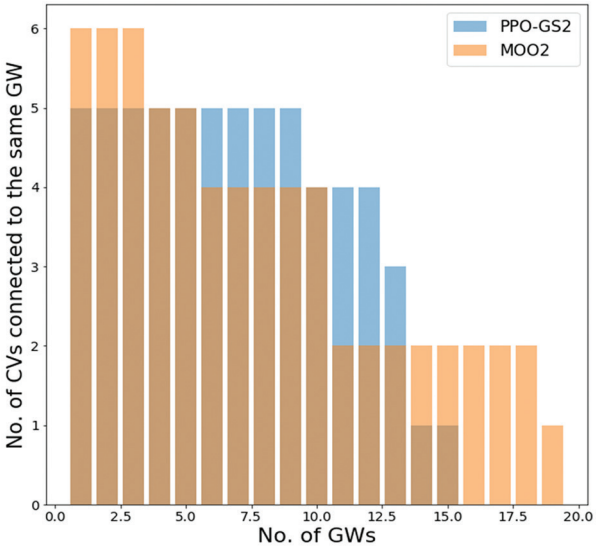


Fig. 5. CVs distribution with w_1=0.7 and w_2 =0.3

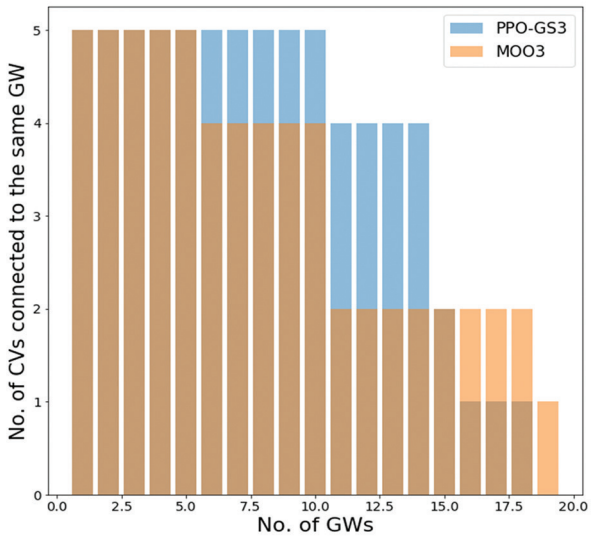


Fig. 6. CVs distribution with w_1=0.5 and w_2 =0.5

Fig. 7 shows that the new proposed technique, in general, has better performance in increasing the number of connected CVs in comparison with MOO solutions. The reason is due to the fact that the MOO algorithm uses the constraints to ensure that the MGs and associated CVs have similar speeds and directions. Consequently, CVs with a large difference in speed and direction cannot find a suitable MG. All the approaches were tested applying the same conditions. Each scenario was executed and evaluated multiple times so that each point in the plot shows the mean of 15 executions. The number of CVs in each scenario is varied from 50 to 100, whereas the number of MGs is fixed to

20. Most of the charts are submitted with a 95% confidence interval. Regarding LCD, it is essential to find MGs to the CVs with the highest LCD because this ensures stable Internet connections for CVs.

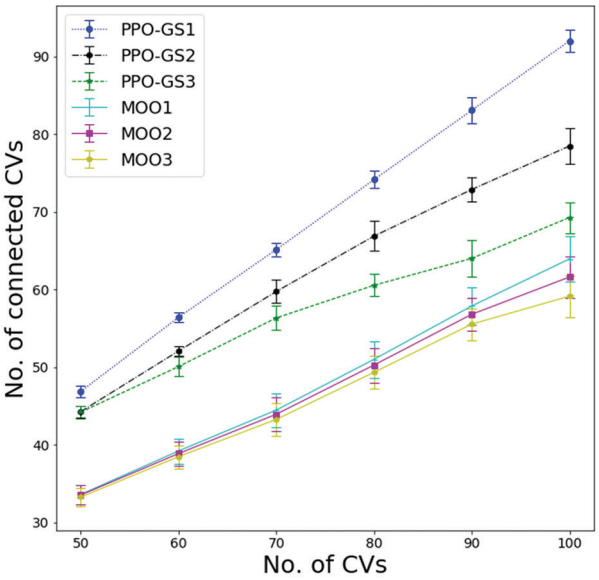


Fig. 7. Number of connected CVs

As Fig. 8 shows, the percentage of MGs selection with the best LCD in the new proposed algorithm is not affected by the increase of the number of CVs compared to the MOO algorithm, which decreases when the number of CVs increases. In Fig. 9, the execution time of the proposed algorithm is low and almost unaffected by increasing the CVs number. In contrast, the MOO solution's execution time is high and affected drastically by increasing the number of CVs. The high increase in execution time means the MOO algorithm is impractical to be applied in the gateway selection system, which needs to be executed in real-time, especially in an urban area where the vehicles' density is high.

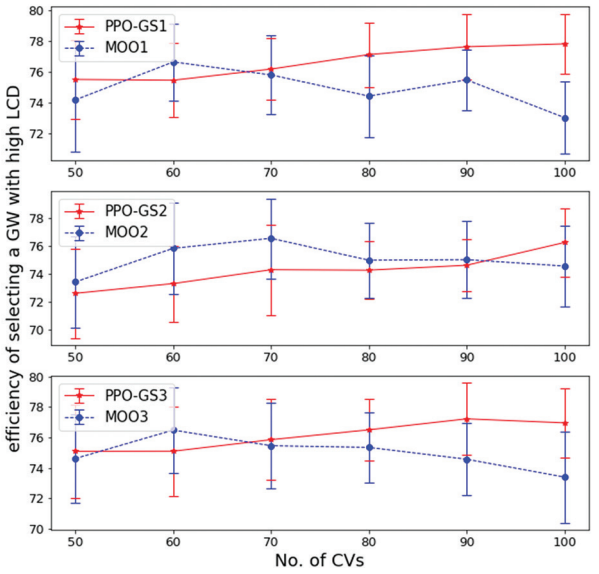


Fig. 8. The percentage of the gateway selection with the highest LCD.

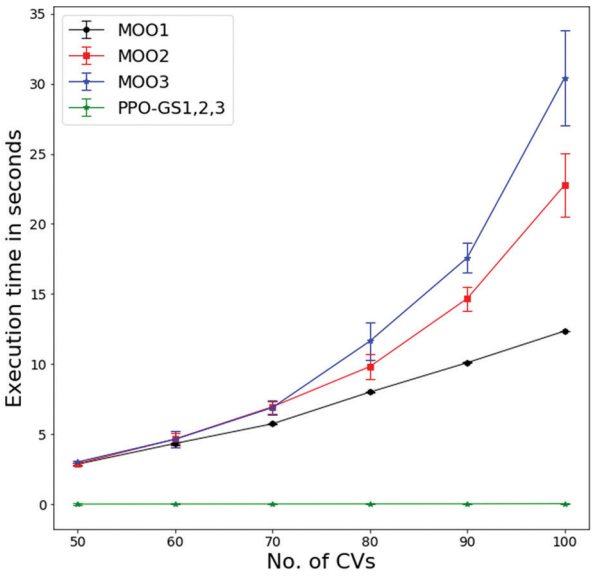


Fig. 9. Execution time.

5. CONCLUSION AND FUTURE WORK

In this article, a new gateway selection algorithm is presented with the aim of finding the best mobile gateway for vehicles in need of Internet access. For this purpose, an integrated system is proposed using two cloud servers. The first one collects all necessary information about CVs and MGs, while the second one, where the gateway discovery system resides, uses the data collected by the first server to train the agent. RL uses two objectives to optimize the gateway discovery system. These two objectives are the link connectivity duration between the vehicles in need of the Internet and the gateways and bandwidth limitation of the gateways. The weighted sum approach is employed to find a trade-off between the two contradicting objectives. The proposed algorithm uses the proximal policy optimization strategy to implement and train the agent. Different agents are created based on the objectives' preferences. Compared with the existing mobile gateway selection algorithms, the simulation results show that the proposed approach is effective in terms of increasing the number of connected vehicles, distributing the traffic among gateways, and reducing the execution time. In future work, we plan to add more parameters to the agent state to make it more expressive, like road density and the number of neighbors for each gateway.

6. REFERENCES

[1] M. Lee, T. Atkison, "VANET applications : Past, present, and future", Vehicular Communications, Vol. 28, 2021, p. 100310.

[2] H. Alabbas, Á. Huszák, "A New Clustering Algorithm for Live Road Surveillance on Highways based on DBSCAN and Fuzzy Logic", International Journal of Advanced Computer Science and Applications, Vol. 11, No. 8, 2020, pp. 580–587.

[3] F. Cunha et al., "Data communication in VANETs: Protocols, applications, and challenges" Ad Hoc Networks, Vol. 44, pp. 90–103, 2016.

[4] Y. Fangchun, W. Shangguang, L. I. Jinglin, L. I. U. Zhihan, S. U. N. Qibo, "An Overview of Internet of Vehicles", China Communications, Vol. 11, no. 10, pp. 1-15, 2014.

[5] P. Anusha, S. K. Shabanabegum, R. Pavaiyarkarasi, E. Seethalakshmi, and K. Vadivukkarasi, "Smart internet of vehicle maintenance system", Materials Today: Proceedings, 2020.

[6] M. Alawi, E. Sundararajan, "Gateway Selection Techniques in Heterogeneous Vehicular Network: Review and Challenges ", Proceedings of the IEEE 6th International Conference on Electrical Engineering and Informatics, Langkawi, Malaysia, 25-27 Nov. 2017, pp. 1-6.

[7] S. Sharma, B. Kaushik, "A survey on internet of vehicles : Applications, security issues & solutions", Vehicular Communications, Vol. 27, 2021, p. 100289

[8] Y. Lin, J. Shen, H. Weng, "Gateway Discovery in VANET Cloud", Proceeding of the IEEE International Conference on High Performance Computing and Communications, Banff, AB, Canada, 2-4 September 2011, pp. 951-954.

[9] M. H. Badole, T. Raju, "Protocol design for an efficient Gateway Discovery & Dispatching for Vehicular Ad Hoc Network", International Journal of Application or Innovation in Engineering & Management, Vol. 3, No. 2, 2014, pp. 117–120.

[10] R. Ghebleh, "A comparative classification of information dissemination approaches in vehicular ad hoc networks from distinctive viewpoints: A survey", Computer Networks, Vol. 131, 2018, pp. 15-37,

[11] R. Hussain, J. Son, H. Eun, S. Kim, H. Oh, "Rethinking Vehicular Communications: Merging VANET with Cloud Computing", Proceeding of the 4th IEEE International Conference on Cloud Computing Technology and Science, Taipei, Taiwan, 3-6 December 2012, pp. 606-609.

[12] M. Gerla, E. Lee, G. Pau, U. Lee, "Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds", Proceeding of the IEEE World Forum on Internet of Things, Seoul, Korea (South), 6-8 March 2014, pp. 241-246.

[13] M. Chaqfeh, N. Mohamed, I. Jawhar, Jie Wu, "Vehicular Cloud data collection for Intelligent Transportation Systems", Proceeding of the IEEE 3rd Smart Cloud Networks & Systems, Dubai, United Arab Emirates, 19-21 December 2016, pp. 1-6.

[14] S. Bitam, A. Mellouk, S. Zeadally, "VANET-cloud: a generic cloud computing model for vehicular Ad Hoc networks", IEEE Wireless Communications, Vol. 22, No. 1, 2015, pp. 96-102.

[15] D. Abada, A. Massaq, A. Boulouz, M. Salah, "An Adaptive Vehicular Relay and Gateway Selection Scheme for Connecting VANETs to Internet via 4G LTE Cellular Network", Proceeding of the IEEE International Conference of Computer Science and Renewable Energies, Agadir, Morocco, 22-24 July 2019, pp. 1-8.

[16] G. Zhioua, H. Labiod, N. Tabbane, S. Tabbane, "FQGwS: A gateway selection algorithm in a hybrid clustered VANET LTE-advanced network: Complexity and performances", Proceeding of the IEEE International Conference on Computing, Networking and Communications, Honolulu, HI, USA, 3-6 February 2014, pp. 413-417.

[17] B. Sharef, R. Alsaqour, M. Alawi, M. Abdelhaq, "Robust and trust dynamic mobile gateway selection in heterogeneous VANET-UMTS network", Vehicular Communications, Vol. 12, 2018, pp. 75–87.

[18] A. Idrissi, S. Retal, H. Rehioui, A. Laghrissi, "Gateway selection in Vehicular Ad-hoc Network", Proceeding of the IEEE 5th International Conference on Information & Communication Technology and Accessibility, Marrakech, Morocco, 21-23 December 2015, pp. 1-5.

[19] S. Retal, A. Idrissi, "A multi-objective optimization system for mobile gateways selection in vehicular Ad-Hoc networks", Computers and Electrical Engineering, Vol. 73, 2019, pp. 289–303.

[20] G. Setiwan, S. Iskander, S. S. Kanhere, Q. J. Chen, "Feasibility Study of Using Mobile Gateways for Providing Internet Connectivity in Public Transportation Vehicles", Proceedings of the International Conference on Wireless Communications and Mobile Computing, British Columbia, Canada, July 3-6, 2006, pp. 1097–1102.

[21] R. S. Sutton, A. G. Barto, "Reinforcement learning: An introduction", 2nd Edition, MIT Press, 2018.

[22] H. Dong, H. Dong, Z. Ding, S. Zhang, Chang, "Deep Reinforcement Learning", Springer, 2020.

[23] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, "Proximal policy optimization algorithms", arXiv Prepr. arXiv1707.06347, 2017.

[24] C. Liu, X. Xu, D. Hu, "Multiobjective reinforcement learning: A comprehensive overview", IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 45, No. 3, 2014, pp. 385–398.

[25] A. Raffin, A. Hill, M. Ernestus, Gleave, A. Gleave, A, Kanervisto, N. Dormann, Stable Baselines3, https://github.com/DLR-RM/stable-baselines3 (accessed: 2021).

Modified Dijkstra Shortest Path Algorithm for SD Networks

Original Scientific Paper

Haitham M. Abdelghany
Electronics and Communication Engineering Department, Faculty of Engineering, Mansoura University, El-Mansoura, Egypt
habdelghany@ outlook.com

Fayez W. Zaki
Electronics and Communication Engineering Department, Faculty of Engineering, Mansoura University, El-Mansoura, Egypt
fwzaki2017@gmail.com

Mohammed M. Ashour
Electronics and Communication Engineering Department, Faculty of Engineering, Mansoura University, El-Mansoura, Egypt
mohmoh2@yahoo.com

Abstract *– This paper uses a modified Dijkstra shortest path method for considering cumulative delays rather than bandwidth in software-defined networks. To implement the proposed method, an open-source Ryu controller is used, and a Mininet tool is used to emulate the topology. The proposed method is compared with the traditional Dijkstra's algorithm to demonstrate its performance. This comparison shows that the modified Dijkstra's algorithm provides higher performance of the different cumulative delays. Several experiments were conducted to evaluate the performance of the proposed method using three parameters (bandwidth, transfer rate and jitter). In addition, the cumulative distribution function is calculated using the parameters to show its distribution through the experiment period.*

Keywords*: Dijkstra shortest path, Software-Defined Networking, Ryu, Mininet, Jitter, cumulative distribution function.*

1. INTRODUCTION

Software-defined networking (SDN), considered the next generation of networking [1-3], aims to separate the control plane and the data plane. The control plane is moved to the central unit, called the controller, while the other switches act as forwarders [4]. Separating the control plane makes it easy to develop novel techniques that give such networks the flexibility to respond to network changes.

SDN technology allows network programmers to deploy the controller in a flexible way through programming languages such as Python and Java. For example, programmers can apply load-balancing techniques and intrusion prevention through programming [5].

In this paper, the widest Dijkstra shortest path method [6] is modified to forward load based on link delay. Python is used to implement the modified method and evaluate its performance by comparing it with the traditional Dijkstra's method. The emulation is carried out using the Mininet tool. The modified method outperforms the original one.

As reported in [7], there are some problems in deploying Dijkstra's method [8] and the altered Floyed–Warshall shortest path method in OpenFlow. The altered Dijkstra's method in [7] is not the same as the method proposed here. The proposed method is therefore compared with the traditional Dijkstra's method.

The remaining sections of this paper are organized as follows. Section 2 introduces SDN and the Mininet emulator. Section 3 discusses the related work. Section 4 presents the modified widest Dijkstra's method and its deployment. Section 5 reports the emulation results. Section 6 describes the analytical model of SDN. Finally, Section 7 presents the conclusion of this work.

2. SOFTWARE-DEFINED NETWORKING AND MININET

In SDN, forwarding decisions are taken by the central unit, consisting of controllers [9], while the other network devices are forwarders. Fig. 1 shows the main concept of SDN. Communication from the controllers to the network device is called the southbound interface. The

most used protocol is the OpenFlow protocol [10,11] that may have a single or many flow tables and group tables. As shown in Fig. 2, the OpenFlow protocol allows the controllers to add or remove entries in the table.

When data reach the switch, the OpenFlow switch searches for a record in the flow table. If there is no record, the data are returned to the controller according to the routing policy [12].

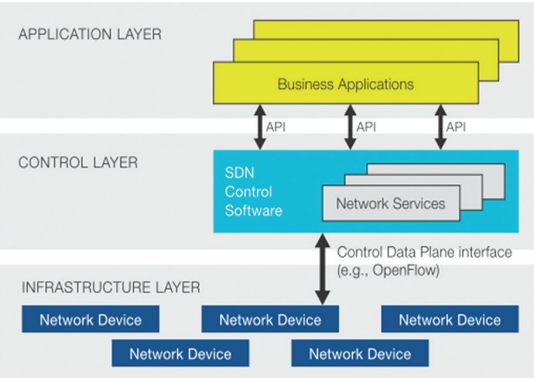


Fig. 1. The main concept of SDN [13]

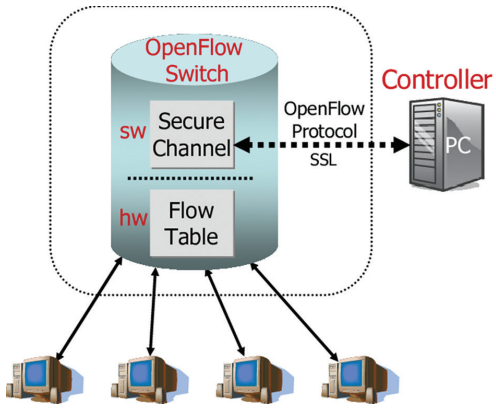


Fig. 2. The OpenFlow switch and the controller [14]

Mininet [6,15] is the network emulator that runs in Linux and is popular in SDN research. It is also widely used by researchers to emulate Open vSwitch and virtual hosts. As the complex topology depends on the specification of the server used, Mininet allows the researcher to create the topology using a Python script.

3. RELATED WORK

The Dutch computer researcher E. W. Dijkstra introduced the traditional Dijkstra algorithm in 1959. It has been used in many fields, such as mobile communication, computer networking, geographic information science and transportation. Dijkstra is a mathematical algorithm used to calculate the shortest path between two nodes in a system. The traditional algorithm was widely used in networking systems. Open Shortest Path First [16] mainly relies on Dijkstra's algorithm to calculate the best route from the source to a destination. The pseudocode for the traditional Dijkstra's algorithm is shown below.

Input: P, k **Output: dis[V], pre[V]**
1: for each v in P(V)
2: dis[v] ← ∞
3: pre[v] ← Ø
4: put distance at u node into Q set
5: while (Q!=Ø)
6: u← Min distance(Q)
7: for each v of u
8: if dis[v] > dis[u]+ew[u,v] then
9: dis[v]←dis[u]+ew[u,v]
10: pre[v]←dis[u]

In [15], J. R. Jiang and a group of researchers proposed an algorithm that extends the traditional Dijkstra's algorithm by adding predefined values of the node weights to prioritize the traffic [17]. Their extended Dijkstra's algorithm outperforms the traditional Dijkstra's algorithm. The weighted Dijkstra's algorithm was implemented using the Abilene topology [18]. This is because the extended Dijkstra's algorithm calculates the distance based on the weights of the nodes in the network and takes node weight into account, whereas the traditional Dijkstra's algorithms do not consider the edge weight or load balancing [19].

Laberio [20] and Lobus [21] proposed load-balancing algorithms for SDN that use the path and link employment to optimize the network throughput. The other algorithm is the service-based load-balancing algorithm [22]. In service-based algorithms, the flow is related to fixed services. Using the service-based load-balancing algorithm is specified to network devices as switches and routers with particular services to increase throughput of the network. Both methods consider the route as the important way of achieving the optimal throughput.

4. IMPLEMENTATION OF THE MODIFIED DIJKSTRA'S ALGORITHM

4.1 MODIFIED EXTENDED DIJKSTRA'S ALGORITHM

Given a single source S and weighted graph G = (V,E), the pseudocode is as shown below.

Input: G=(V, E), ed, nd, h **Output: delay[|V|], p[|V|]**
1: delay [h]←0; delay [t]←∞, for each t≠h, t€V
2: insert t with key delay [t] into the priority queue Q, for each t€V
3: while (Q≠Ø)
4: t←Extract-Min(Q)
5: for each v adjacent to t
6: if delay [v] > delay [t]+ed[t,v]+nd[t] then
7: delay [v]←delay [t]+ed[t,v]+nd[t]
8: p[v]←delay [t]

The difference between the modified extended Dijkstra and the extended Dijkstra is that in the former the route is selected based on distance instead of the delay. The importance of the proposed method is evident when the delays are different on equal bandwidth links. Thus, the selection is made based on delay. Table 1 shows a brief comparison of the traditional Dijkstra and modified Dijkstra 3.

Table 1. Comparison of traditional Dijkstra and modified Dijkstra 3

Algorithm	Traditional Dijkstra	Modified Dijkstra 3
Criteria for Path Selection	Distance	Delay
Initiation of Destination Node	+ve Infinity	-ve Infinity
Load Balancing	No	Yes
Weighted Nodes	No	Uses weights for nodes

5. ANALYTICAL MODEL

As shown in Fig. 3, in the OpenFlow SDN model, the controller is connected to a number of switches. It is assumed that packet arrival follows a Poisson distribution with an arrival rate λ_i. Packets that do not have matching entries are probably sent to the controller. Thus, the arrival rate is λ_i.*p, λ_i.*(1-p), and the processing time is exponential $1/\mu_i$ for the switches. The average service time for the controller is equal to $1/\mu_c$ where μ_i is the processing rate of the switches and μ_c is the processing time of the controller.

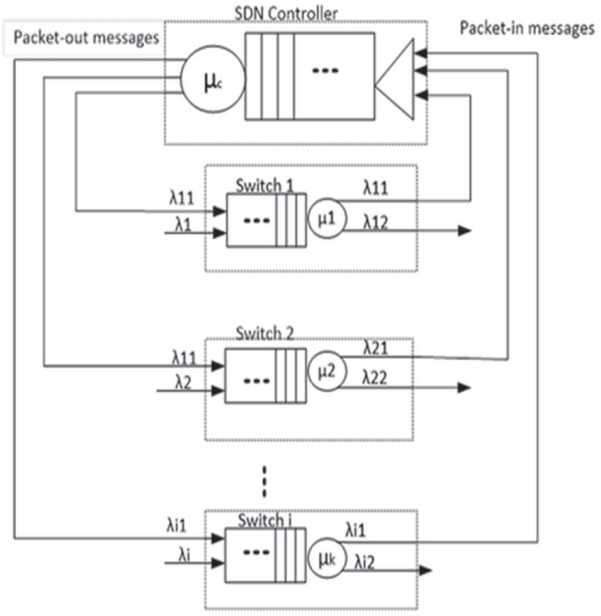


Fig. 3. OpenFlow SDN analytical model [23]

5.1. SWITCH PERFORMANCE

The flow tables are not always the same and can be changed based on link delay, and the processing time is assumed to follow an exponential distribution.

The OpenFlow switches and the controller can be modelled with a M/H_2/1 queue [14]. This means that the arrival of packets λ_i. and the serving rate are hyper-exponential with two-phases.

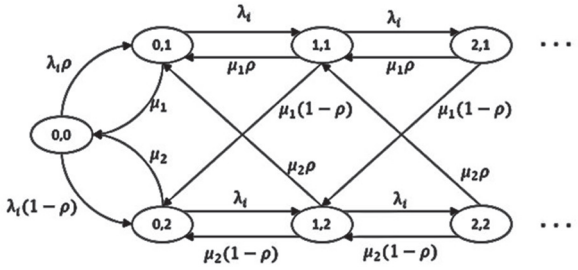


Fig. 4. State diagram of an M/H_2/1 queue [24]

In the state diagram in Fig. 4, p is the probability that data are processed at rate μ_1, and 1−p is the probability of receiving a service rate of μ_2.The stationary probability π(i) is a vector and can be represented as

$$\pi^i = (\pi_0{}^i, \pi_1{}^i, \pi_2{}^i, \ldots\ldots, \pi_k{}^i). \quad (1)$$

$$\rho = {}^{\lambda}/_{\mu} < 1. \quad (2)$$

$$\pi_0 = 1\text{-}\rho. \quad (3)$$

$$\pi_k = (1\text{-}p)\,\rho^k. \quad (4)$$

$\pi_k(1)$ is the *k* packet probability in the i_{th} switch.

The mean value of packets in the queueing system is

$$N_i = \sum_{k=o}^{\infty} k\pi_k^i \quad (5)$$

$$N_i = \sum_{k=o}^{\infty} k(1-\rho)\rho^k \quad (6)$$

For *k* = 0, the product is equal to 0, and so the summation can begin from *k* = 1:

$$N_i = (1-\rho)\sum_{k=1}^{\infty} k\rho^k = (1-\rho)\rho\sum_{k=1}^{\infty} k\rho^{k-1} \quad (7)$$

$$\text{With } k\rho^{k-1} = \frac{d\rho^k}{d\rho},$$

$$N_i = (1-\rho)\rho\sum_{k=1}^{\infty}\frac{d\rho^k}{d\rho}\rho^k \quad (8)$$

$$N_i = (1-\rho)\rho\frac{d\rho}{d\rho}\sum_{k=1}^{\infty}\rho^k \quad (9)$$

So,

$$\sum_{k=1}^{\infty}\rho^k = \sum_{k=0}^{\infty}\rho^k - 1 = \frac{1}{1-\rho} - 1 = \frac{\rho}{1-\rho} \quad (10)$$

$$N_i = \frac{\rho}{1-\rho} \text{ while } \rho < 1 \text{ and } \rho = {}^{\lambda}/_{\mu}. \quad (11)$$

From Little's formula, the mean processing delay in the i_{th} switch may be expressed as

$$W_{si} = \frac{1}{\lambda}\mu_i = \frac{1}{\mu-\lambda} \quad (12)$$

The average processing delay of packets by all switches can also be given as

$$W_s = \sum_{i=1}^{n}\frac{\lambda_i}{\sum_{i=1}^{n}\lambda_i} W_{si} \quad (13)$$

6. EMULATION

Mininet was used to implement the experiments of the virtual environment. The experiments were hosted on a custom-built server with Intel Core i5 CPU at 2.5 GHz, 8 GB RAM with a solid-state drive for storage and switching. The Ubuntu Server 20.04 LTS distribution was installed on the machine. Then the Open-V-Switch software was installed. One controller, three switches and three hosts were used, as shown in Fig. 5. The emulation parameters are shown in Table 2. Iperf was used to test the bandwidth, transfer rate and jitter. In the experiments, host 3 acted as a server, while hosts 1 and 2 acted as clients.

Table 2. Simulation parameters

Delay on Edges	1 ms
Delay Between Switches	1-3 ms
Number of Hosts	3
Number of Switches	3
Controller	Ryu 4.34
Testing Tool	Iperf

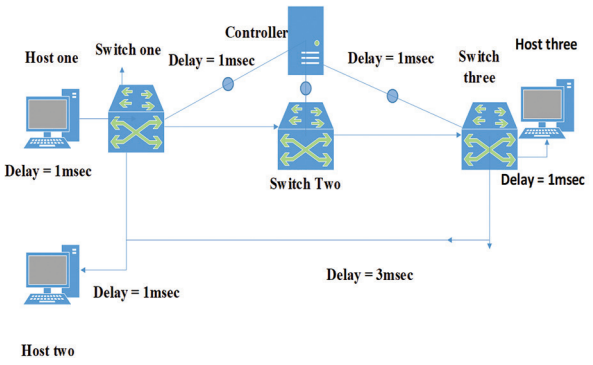


Fig. 5. The topology used in the experiment

6.1 EMULATION RESULTS

Fig. 5 shows that, when host 1 sent data to host 3, the link delay was 4 ms plus the processing time of every switch. Host 1 has two routes. The first route is the direct link from switch 1 to switch 3, which has a link delay of 4 ms plus the processing time of every switch. The other route, which is through switch 2, has link delay of 4 ms plus the processing time of every switch. The proposed algorithm load-balances the traffic between the two routes, whereas the traditional Dijkstra's algorithm does not. Iperf was used to measure the bandwidth, transfer rate and jitter. The results in Figs. 6, 7 and 8 show that the modified extended Dijkstra outperforms the extended Dijkstra when two hosts send data at the same time to the same host (host 3). In the preceding experiments, the cumulative distribution function (CDF) of the parameters was measured to describe the distribution of the parameters across the whole time.

As shown in Fig. 6, the CDF for the bandwidth from host 1 to host 3 was between 46 Mbps and 56 Mbps when the proposed method was used. In the traditional Dijkstra, the CDF for the bandwidth from host 1 to host 3 varied in a wide range from 20 Mbps to 43 Mbps.

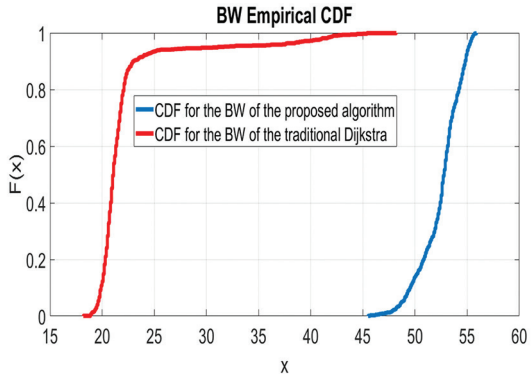


Fig. 6. The CDF for the bandwidth in Mbps for host 1 when the proposed method is applied vs traditional Dijkstra

As shown in Fig. 7, the CDF for the transfer rate lies between 5.5 Mbytes and 6.7 Mbytes using the proposed method. The CDF for the transfer rate varies widely from 2.3 Mbytes to 5 Mbytes using the traditional Dijkstra's algorithm.

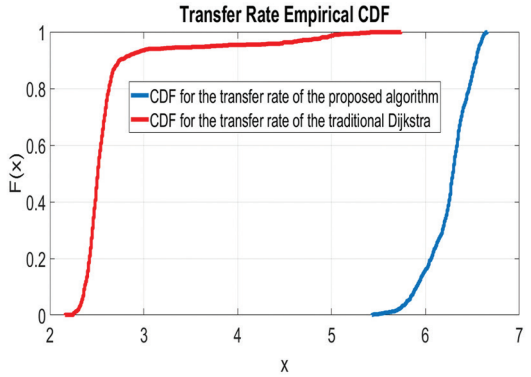


Fig. 7. The CDF for the transfer rate in MBytes for host 1 when the proposed method is applied vs traditional Dijkstra

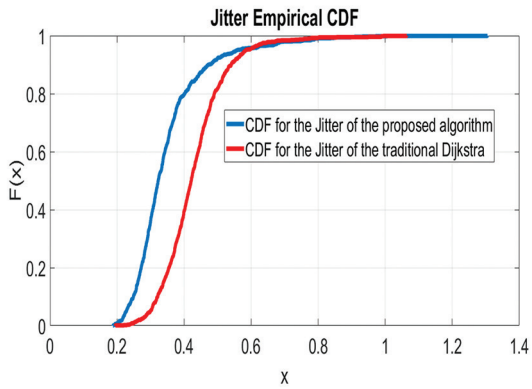


Fig. 8. The CDF for the jitter in ms for host 1 when the proposed method is applied vs traditional Dijkstra

Finally, as shown in Fig. 8, the CDF for the jitter is between 0.2 ms and 0.6 ms using the proposed method. The CDF for the jitter varies from 0.3 ms to 0.8 ms when traditional Dijkstra is used. Furthermore, Table 3 summarizes the average values of bandwidth, transfer rate and jitter for host 1 when traditional Dijkstra is used versus the modified Dijkstra.

Table 3. Average parameter values

Parameter	Traditional Dijkstra	Modified Dijkstra 3
Bandwidth (Mbps)	22.1213	52.4679
Transfer Rate (MB)	2.6367	6.2553
Jitter (ms)	0.4303	0.3501

7. CONCLUSION

This study proposes and implements a modified extended Dijkstra's algorithm. The main test showed that the proposed method outperforms the traditional Dijkstra's algorithm. In the previous experiments, the CDF of the selected parameters was used to evaluate the performance of the proposed method against the traditional Dijkstra's algorithm. It is important to mention that every experiment was repeated 10 times and held for about one hundred seconds. The proposed method uses a Ryu controller implemented in Python, and the Mininet tool was used to emulate the network topology. In the future, a more complex design will be proposed to further test the performance of the proposed method.

8. REFERENCES

[1] S. Rowshanrad, S. Namvarasl, V. Abdi, M. Hajizadeh, M. Keshtgary, "A survey on SDN, the future of networking", Journal of Advanced Computer Science & Technology, Vol. 3, No. 2, 2014, pp.232-248.

[2] A. Hakiri, A. Gokhale, P. Berthou, D. C. Schmidt, T. Gayraud, "Soft- ware-defined networking: Challenges and research opportunities for future internet", Computer Networks, Vol. 75, No. 24, 2014, pp.453-471.

[3] J. Pan, S. Paul, R. Jain, "A survey of the research on future internet architectures", IEEE Communications Magazine, Vol. 49, No. 7, 2011, pp. 26-36.

[4] A. Lara, A. Kolasani, B. Ramamurthy, "Network innovation using open- flow: A survey", IEEE Communications Surveys & Tutorials, Vol. 16, No. 1, 2014, pp. 493-512.

[5] S. Ahmad, A. H. Mir, "Scalability, consistency, reliability and security in SDN controllers: A survey of diverse SDN controllers", Journal of Network and Systems Management, Vol. 29, No. 1, 2021, pp. 1-59.

[6] E. Dijkstra, "A note on two problemsecin connexion with graphs", Numerische Mathematik, Vol. 1, No.1, 1959, pp. 269-271.

[7] A. Furculita, M. Ulinic, A. Rus, V. Dobrota, "Implementation issues for Modified Dijkstra's and Floyd-Warshall algorithmsecin OpenFlow," Proceedings of the RoEduNet International Conference 12th Edition: Networking in Education and Research, 2013, pp. 141-146.

[8] A. Rus, V. Dobrota, A. Vedinas, G. Boanea, M. Barabas, "Modified Dijkstra's algorithm with cross-layer QoS", ACTA TECHNICA NAPOCENSIS, Electronics and Telecommunications, Vol. 51, No. 3, 2010, pp. 75-80.

[9] Open Network Foundation (ONF) Website (SDN whitepaper), https://www.opennetworking.org/sdn-resources/sdn-definition (accessed: 2021)

[10] A. Greenberg et al. "A clean slate 4D approach to network control and management", Computer Communication Review, Vol. 35, No. 5, 2005, pp. 41-54.

[11] M. Casado, M. J. Freedman, J. Pettit, J. Luo, N. McKeown, S. Shenker, "Taking control of the enterprise", Computer Communication Review, Vol. 37, No. 4, 2007, pp. 1-12.

[12] Open Networking Foundation, "OpenFlow Switch Specification version 1.4.0", October 14, 2013.

[13] Floodlight OpenFlow Controller—Project Floodlight, Big switch network, http://www.projectfloodlight.org/floodlight (accessed: 2019)

[14] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, J. Turner, S. Shenker. "Openflow Enabling innovation in campus networks", Computer Communication Review, Vol. 38, No. 2, 2008, pp 69-74.

[15] J. Jiang, H. Huang, J. Liao, S. Chen, "Extending Dijkstra's Shortest Path Algorithm for Software Defined Networking", Proceedings of the 16th Asia-Pacific Network Operations and Management Symposium, Hsinchu, Taiwan, 2014, pp. 1-4.

[16] J. Moy, "OSPF: Anatomy of an Internet Routing Protocol", Addison-Wesley, 2000.

[17] P. Tantisarkhornkhet, W.Werapun, B. Paillassa "SDN experimental on the PSU network", Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems, Phuket, Thailand, 2016, pp. 1-6.

[18] Abilene Network, https://en.wikipedia.org/wiki/Abilene_Network (accessed: 2021)

[19] W. Yahya, A. Basuki, J. R. Jiang, "The extended Dijkstra's-based load balancing for openflow network", International Journal of Electrical and Computer Engineering, Vol. 5, No. 2, 2015, pp. 289-296.

[20] H. Long, Y. Shen, M. Guo, F. Tang. "LABERIO: Dynamic load-balanced routing in OpenFlow-enabled networks", Proceedings of the IEEE 27th International Conference on Advanced Information Networking and Applications, Barcelona, Spain, 2013, pp. 290-297.

[21] N. Handigol, S. Seetharaman, M. Flajslik, N. McKeown, R. Johari, "Plug-n-Serve: Load-balancing web traffic using Open Flow", Demo at ACM SIGCOMM, August 2009.

[22] M. Koerner, O. Kao, "Multiple service load-balancing with OpenFlow", Proceedings of the 13th IEEE International Conference on High Performance Switching and Routing, Belgrade, Serbia, 2012, pp. 210-214.

[23] S. Muhizi, G. Shamshin, A. Muthanna, R. Kirichek, A. Vladyko, A. Koucheryavy, "Analysis and Performance Evaluation of SDN Queue Model", International Federation for Information Processing, 2017, pp. 26-37.

[24] Z. Shang, K. Wolter, "Delay evaluation of openflow network based on queueing model", http://arxiv.org/abs/1608.06491 (accessed: 2021)

Secure Complaint Management System against Women Harassment at Workplace Using Blockchain Technology

Original Scientific Paper

Md. Mijanur Rahman
Southeast University,
Assistant Professor, Department of Computer Science and Engineering
Banani, Dhaka, Bangladesh
mijanur.rahman@seu.edu.bd

Md. Moshiul Azam
Southeast University,
Student, Department of Computer Science and Engineering
Banani, Dhaka, Bangladesh
azammoshiul8@gmail.com

Faria Sanjida Chowdhury
Southeast University,
Student, Department of Computer Science and Engineering
Banani, Dhaka, Bangladesh
fariachy102@gmail.com

Abstract *– Since the Industrial era, women are playing a significant role in the workforce to move the world forward. Their increasing contribution in various fields has earned a fortune for the global economy. Despite that, women constantly face more obstacles than men in the workplace. When half of the population are mistreated because of gender inequality, the economy of any nation is supposed to collapse. One of the biggest barriers for women in their careers is workplace harassment. Workplace harassment may include physical, verbal or nonverbal harassment that not only have an adverse effect on a woman's career, mental health and physical health but also organizational reputation. A common way to make a complaint in most organizations is to fill up a complaint form, email or go directly to the competent authority and complain. But victims often hesitate to complain because their identity might get revealed or their documentary evidence might be tampered. As a result, most of the harassers get through very easily. To resolve this problem, this paper presents a blockchain-based anonymous, transparent and secure platform where women can easily complain against their harassers. To keep the platform secure and reliable, a two-level hierarchical model is introduced, where level-1 is the Human Resources (HR) and level-2 is the Higher Authority. In level-1, victims can anonymously complain to HR and in Level-2, victims can complain with their identity revealed to higher authority. This way, the proposed platform ensures women of a healthy work environment and provides all necessary support to stand up against injustice in the workplace.*

Keywords*: Blockchain, Anonymity, Hyperledger Fabric, Workplace Harassment, Women Harassment*

1. INTRODUCTION

Workplace harassment has been a serious issue for millions of working women across the world. It is defined as an offensive behaviour towards an employee by another to hurt them physically or mentally on purpose [1]. Workplace harassment includes many different forms of harassment such as insulting [2], bullying [3], teasing, mobbing [4], threatening, work abuse, physical abuse, sexual advance [5], etc. In a developed country like the USA, one out of three women claims to have been sexually harassed in the workplace [6]. There are also other cases of women falling victim to physical, verbal and non-verbal harassment. Such incidents cause negative effects on the victim's physical, emotional and occupational well-being [7].

To help working women in the workplace, there are some existing countermeasures against workplace harassment, such as, anti-harassment and anti-dis-

crimination policy, monitoring system, report tracking system, and grievance procedures. Normally, a victim complains to Human Resource Personnel directly via a complaint form, email or hotline. But many victims hesitate to complain this way or have a fear of their identity being exposed. In some organizations, web-based applications and mobile applications are used to report such harassment. Some IT industries, for example, Speakfully [8] and #NotMe [9] offer services to employers and employees to deal with harassment in the workplace. Female employees can use these platforms for reporting and documentation at their convenience. But there is a possibility of data being altered or tampered. So, the security and reliability of these platforms are questionable.

Using Blockchain-based complaint systems in the workplace to overcome such limitations can be a remarkable solution for all of us. In 2008, Bitcoin, a peer-to-peer cryptocurrency based on blockchain technology, was first introduced by Satoshi [10]. Blockchain is now widely acknowledged in various fields because of its more secure, transparent, and tamper-proof ledger [11]. It also has prominent features like anonymity and autonomy. This distributed public ledger uses the Merkle tree and Hash function for its encryption and depends on consensus mechanisms such as Proof of Work (PoW), Proof of Stake (PoS), Proof of Concept (PoC), etc., as required. As the complaint management system for workplace harassment needs to be anonymous, reliable, secure, transparent and tamper-proof, the features of blockchain technology meet all its requirements.

There are two challenges in building a secure, reliable and tamper-proof complaint management system.

- The complaint may contain the victim's identity or other sensitive information that the victim wants to keep hidden. But, the authenticity and reliability of anonymous complaints cannot be guaranteed.
- The complaint needs to be transparent and tamper-proof as the authenticity and reliability of the complaint are the highest priorities. This is why traditional centralized systems are not considered to be secure and trusted.

Therefore, we proposed to build a decentralized complaint management system based on blockchain technology. Considering that blockchain has the trust of people for having decentralized, transparent, tamper-proof and trustless architecture, and it supports anonymity to assist a victim in filing their complaint without revealing their identity are two main reasons to utilize its properties in our system.

Our main research question is:

(1) How can blockchain help in supporting workplace harassment complaint procedures using its decentralized, secure, tamper-proof and trustless properties?

The objectives of our proposed system are as follows:

(1) Developing a blockchain-based complaint management system for employees and employers to deal with harassment in the workplace.

(2) Implementing two different levels to help victims complain either by keeping their identity hidden or revealing it.

2. RELATED WORKS

In 2020, Bárbara Aburachid Rocha proposed an Ethereum blockchain-based system for workplace harassment complaints and evidence tracking [12]. It documents the whole procedure while creating and tracking evidence of all the actions until the complaint procedure is resolved. The proposed architecture has two proof-of-concept. One is a fixed system that follows the guidelines of the Code of Practice Detailing Procedures for addressing Bullying in the Workplace from Ireland and the other is a flexible system that can be used in different procedures.

The Ministry of Women and Child Development, Government of India has an online workplace harassment complaint management system named SHe-Box for women where they can file complaints regarding sexual harassment at the workplace [13]. Any female employee regardless of any sector can use SHe-box. Victim has to provide her name, designation, contact number, email, identification number, accused's name, description, organization details, etc., to create a user id and file a complaint. She can also see the status of her complaint. However, SHe-box is a traditional centralized system, it is not transparent, tamper-proof and trustworthy.

Speakfully is a third party enabled app that helps employees report any incidents related to harassment and discrimination with documentation and support [8]. It offers different pricing solutions for individuals and organizations. Anonymous reporting with documents, case management, messaging with employees, pulse surveys, feedback, etc., are some of the features it provides. Here, victims have to enter their experiences in a document that can include text, image, audio, video, etc. They have to share personal data like name, email, contact number, address, company name, designation, IP address, etc., and report to their HR team. Such solution platforms are not very reliable as they work as a middleman.

3. SYSTEM BACKGROUND

In this section, the system background of our proposed system is explained.

A. Blockchain Technology

In 1991, two Bellcore researchers, Stuart Haber and W. Scott Stornetta established the first concept of blockchain technology [14]. Blockchain is a type of distribut-

ed ledger technology where every block stores data or transactions and system validators validate each block with a consensus mechanism. The first block of a chain is called "Genesis Block" [15]. To retain data, a block is recognized with hash and previous hash. Each block's previous hash is linked with the earlier block's hash. This is how blocks create chains (Fig. 1).

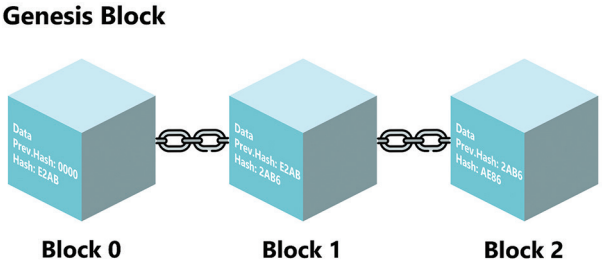


Fig. 1. Blockchain Technology

Blockchain is decentralized, immutable, secure, transparent and anonymous [11], [16].

Blockchain can be either permissioned blockchain or permissionless blockchain. Permissioned Blockchain is a closed ecosystem where only selected members can participate in the system with permission whereas permissionless blockchain is open for all. Anyone can participate in the system and validate any transaction [17].

B. Docker

Docker is a container similar to a virtual machine that allows developers to share containers and applications inside them among their peers [18]. It is an open-source project and works as a platform to build a server-client relationship. Docker images create containers [19] that can be considered as running instances of the base image. Alike an OS image, docker image executes code. It has multiple layers; each one is formed on top of the previous layer with commands while creating it. So, to share a code, developers have to share the image.

C. Hyperledger Fabric

Hyperledger Fabric is a popular project, introduced by IBM, with the intent of implementing a group of modular blockchain based applications within a single framework [20]. It is a partial permission based private blockchain, where every member has to get permission from the network, but the degree of permission varies for different users of different applications. Chaincode, aka, smart contract is used here in Docker containers [21]. It allows implementing applications in any programming language like Google Go, Node.js, Java etc.

D. CouchDB

CouchDB is a document-oriented database [22] that is supported by Hyperledger Fabric as a state database [23]. It is in JSON format that provides rich queries in opposition to chaincode making the queries more effective. Fabric stores the transaction data of its blocks in CouchDB as key and value pairs for user queries.

4. OUR PROPOSED FRAMEWORK

Our proposed system includes five types of entities-

- Victim
- Level-1
- Level-2
- Investigator
- Harasser/ Accused

Victim can complain on our proposed system in two ways.

1) Complaining anonymously to level-1: Victim wants the HR to warn their harasser.

2) Complaining with identity to level-2: In any serious case, the victim wants the higher authority to take immediate action against the offender.

A. Complaining Anonymously to Level-1

The following use case diagram (Fig. 2) shows how victim, level-1 and accused interact on our platform.

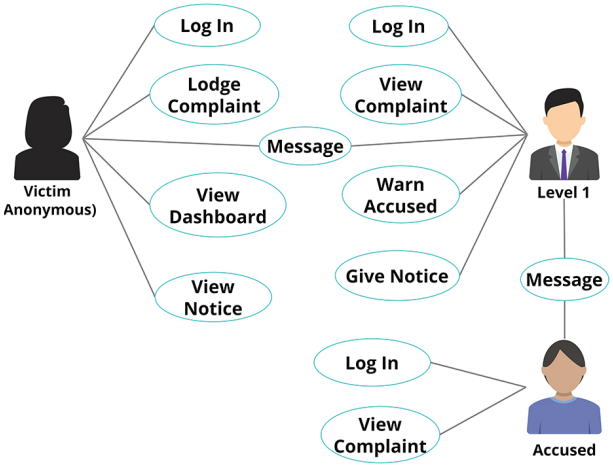


Fig. 2. Use Case Diagram for Level-1

Our proposed system for level-1 is divided into two systems.

a. Victim and Accused Interaction System

After logging in, the victim can anonymously lodge a complaint against her offender. She can choose what kind of complaint she wishes to file, usually minor cases of harassment. According to the complaint, Level-1 will give a warning to the accused. Both the level-1 and harasser will not know the identity of the victim.

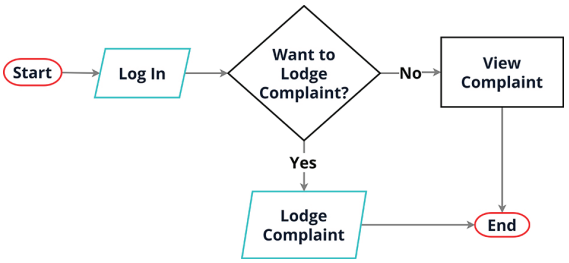


Fig. 3. Flow Chart of Victim and Accused Interaction System for Level-1

The following section describes Victim and Accused Interaction System (Fig. 3)

Lodge Complaint: If a victim wants to complaint, she has to provide complaint details, name of the accused and choose the complaint category.

View Complaint: Accused can view complaints filed against them.

b. Level-1 Interaction System

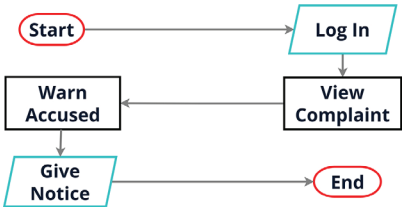


Fig. 4. Flow Chart of Level-1 Interaction System

The following section describes Level-1 Interaction System (Fig. 4)–

View Complaint: Level-1 can view complaints filed by victims and take action accordingly.

Warn Accused: If a complaint is filed against an employee, level-1 warns him.

Give Notice: Level-1 gives a notice about the action taken on the complaint.

B. Complaining with identity to Level-2

The following use case diagram (Fig. 5) shows how victim, level-2, accused and investigator interact on our platform.

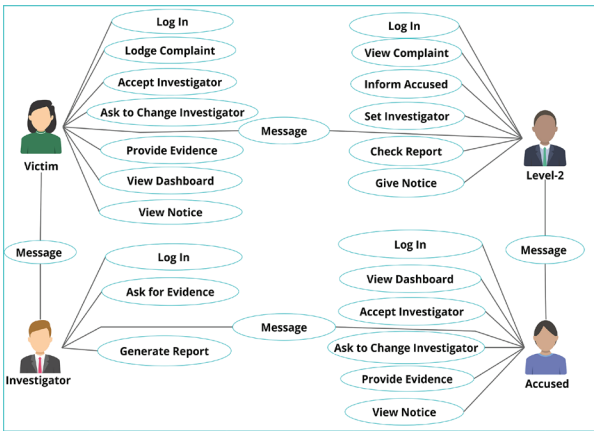


Fig. 5. Use Case Diagram for Level-2

Our proposed system for level-2 is categorized into three systems.

a. Victim and Accused Interaction System:

After logging in, the victim can lodge a complaint against her perpetrator disclosing her own identity. She can choose what kind of complaint she wishes to file. As for the accused, he will get a notification if he is alleged to have harassed a woman.

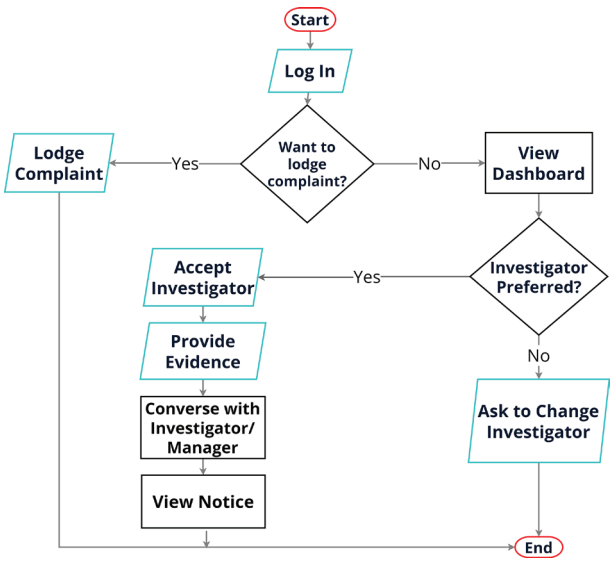


Fig. 6. Flow Chart of Victim and Accused Interaction System for Level-2

The following section describes Victim and Accused Interaction System (Fig. 6)-

Lodge Complaint: To lodge a complaint, a victim has to provide complaint details, name of the accused and choose the complaint category.

View Dashboard: Victim and accused can view their alleged complaint and get notified about the course of action.

Accept Investigator: Victim or accused will accept the investigator appointed by level-2 if they agree with the investigator to work on the matter.

Ask to Change Investigator: If any of them does not prefer the investigator, they can request level-2 to change the appointed investigator.

Provide Evidence: Victim or accused can provide their evidence to the investigator to conduct a thorough investigation.

Message: Victim and accused can message level-2 or investigator if needed.

View Notice: Victim, accused and other employees of the company will know what action has been taken on the complaint.

b. Level-2 Interaction System:

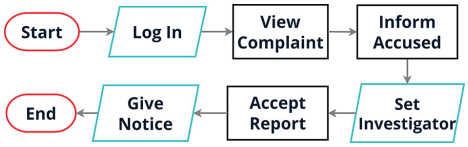


Fig. 7. Flow Chart of Level-2 Interaction System

The following section describes Level-2 Interaction System (Fig. 7).

View Complaint: Level-2 can view complaints filed by victims and take actions accordingly.

Inform Accused: If a complaint is filed against an employee, level-2 informs him.

Set Investigator: Level-2 appoints an investigator to scrutinize the evidence.

Message: Level-2 can message both the victim and the accused if needed.

Accept Report: Level-2 accepts the report generated by the investigator.

Give Notice: Level-2 gives a notice about the action taken on the complaint.

c. Investigator Interaction System

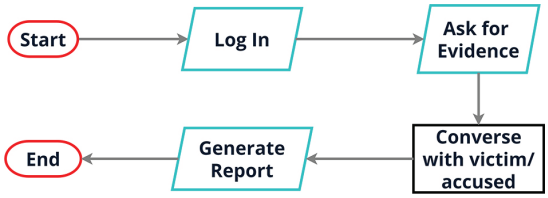


Fig. 8. Flow Chart of Investigator Interaction System

The following section describes Investigator Interaction System (Fig. 8).

Ask for Evidence: Investigator asks for evidence from the victim or the accused that may help him carry out the investigation.

Message: Investigator can message the victim and the accused if needed.

Generate Report: After completing the investigation, the investigator generates the investigation report to level-2.

There might be a situation in level-1 where the accused is not at fault and they want to take action against this false accusation. In this case, the accused can also complain to level-2.

While dealing with such a case, level-2 may need the previous complaint (Fig. 9)

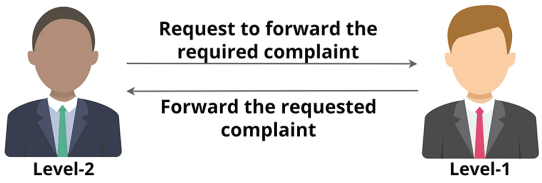


Fig. 9. Flow Chart of the Interaction between Level-1 and Level-2

After the investigation is done, the authority either takes disciplinary action or legal action to support their victim at all costs. This way, an organization can confirm a safe workplace along with its positive reputation.

5. IMPLEMENTATION

Our proposed platform is a web application and uses Hyperledger Fabric blockchain to keep all records.

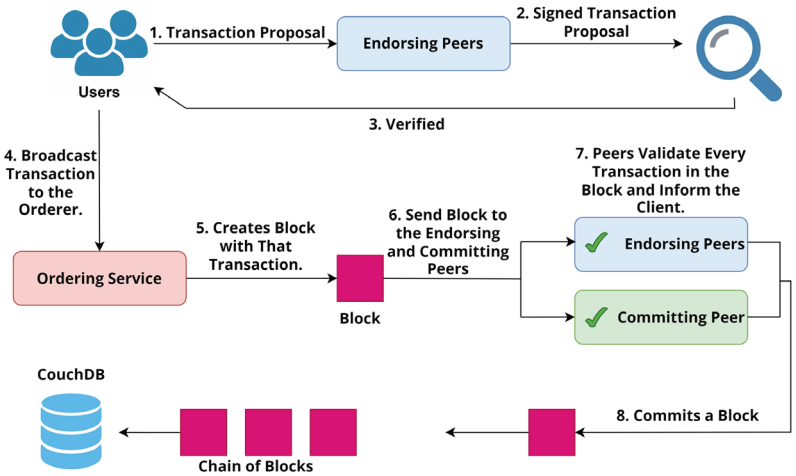


Fig. 10. System Architecture

The following section describes the system architecture (Fig. 10)

Before interacting with the Hyperledger Fabric Network, the system verifies user (Victim/Level-1/Level-2/Accused/Investigator) identity using Membership Service Provider. After the user identification is confirmed, the network controls the user's access. When a user initiates a transaction, the transaction is broadcasted to the Endorsing Peers in form of a proposal. After receiving the proposal, Endorsing Peers execute the chaincode and return its consequences to the user. Then the user verifies the proposal response with the help of a consensus mechanism and broadcasts the transaction to the Ordering Service. Orderer creates a new block with that transaction and dispatches that block to the Endorsing and Committing Peers. Peers validate all the transactions in the block and notify the user too. Finally, committing nodes keep a copy of the block in their ledger.

In this proposed application, Next.js is used to design the front-end while the Fabric network acts as the back-end. To implement the Hyperledger Fabric Network, the chaincode is written in Golang which runs in a secured Docker Container. The Node SDK is used to set up communication between the front-end and back-end. It helps to query all the functions and properties from the Fabric Network and communicate with the CouchDB.

Some code snippets of our proposed platform are given below.

LodgeComplaint(): This function (Fig. 11) takes the complainant's ID, department, harasser's name, harasser's department, Type of complaint and complaint details from the complainant to lodge a complaint. Complainants can choose at which level they want to complain.

```
// This function helps to Lodge new Complaint
func (pc *ComplaintSmartContract) LodgeComplaint(ctx contractapi.TransactionContextInterface,
    complainantNum string, department string, harasserName string, harasserDepartment string,
    typeComplaint string, complaintDetails string, complaintTo string) error {
    var count int = 1
    complaintIterator, err := ctx.GetStub().GetStateByRange("", "")
    if err != nil {
        return err}
    defer complaintIterator.Close()
    for complaintIterator.HasNext() {
        complaintResponse, err := complaintIterator.Next()
        if err != nil {
            return err}
        if complaintResponse!=nil{
            count=count+1}
    }
    id := strconv.Itoa(count)
    comp := Complaint{
        ID:                 id,
        Complainantnum:     complainantNum,
        Department:         department,
        HarasserName:       harasserName,
        HarasserDepartment: harasserDepartment,
        TypeComplaint:      typeComplaint,
        ComplaintDetails:   complaintDetails,
        ComplaintTo:        complaintTo,
    }
    complaintBytes, err := json.Marshal(comp)
    if err != nil {
        return err}
    return ctx.GetStub().PutState(id, complaintBytes)
}
```

Fig. 11. Function to Lodge a Complaint

ViewAllComplaintsLevel1(): Level-1 uses this function (Fig. 12) to view all the complaints registered for them by the complainants. But here the complainant's ID is kept hidden, so here level-1 cannot see it and find out the victim's identity.

```
// This function returns all the existing complaints registered for Level 1
func (pc *ComplaintSmartContract) ViewAllComplaintsLevel1(ctx contractapi.
    TransactionContextInterface) ([]*ComplaintL1, error) {
    complaintIterator, err := ctx.GetStub().GetStateByRange("", "")
    if err != nil {
        return nil, err}
    defer complaintIterator.Close()
    var complaints []*ComplaintL1
    for complaintIterator.HasNext() {
        complaintResponse, err := complaintIterator.Next()
        if err != nil {
            return nil, err}
        var Cl *ComplaintL1
        err = json.Unmarshal(complaintResponse.Value, &Cl)
        if err != nil {
            return nil, err}
        if Cl.ComplaintTo == "Level1" {
            complaints = append(complaints, Cl)}
    }
    return complaints, nil
}
```

Fig. 12. Function to View all Complaints by Level-1

ViewAllComplaintsLevel2(): This function (Fig. 13) is used to view all the complaints registered for Level-2. Being the top authority, level-2 can see every information including the complainant's ID.

```
// This function returns all the existing complaints registered for Level 2
func (pc *ComplaintSmartContract) ViewAllComplaintsLevel2(ctx contractapi.
    TransactionContextInterface) ([]*Complaint, error) {
    complaintIterator, err := ctx.GetStub().GetStateByRange("", "")
    if err != nil {
        return nil, err}
    defer complaintIterator.Close()
    var complaints []*Complaint
    for complaintIterator.HasNext() {
        complaintResponse, err := complaintIterator.Next()
        if err != nil {
            return nil, err}
        var Complaint *Complaint
        err = json.Unmarshal(complaintResponse.Value, &Complaint)
        if err != nil {
            return nil, err}
        if Complaint.ComplaintTo == "Level2" {
            complaints = append(complaints, Complaint)}
    }
    return complaints, nil
}
```

Fig. 13. Function to View all Complaints by Level-2

ViewComplaintByID(): Level-1 and level-2 use ViewComplaintByID (Fig.14) to view a complaint by its unique ID.

```
// This function helps to view the Complaint by Id
func (pc *ComplaintSmartContract) ViewComplaintById(ctx contractapi.
    TransactionContextInterface, id string) (*Complaint, error) {
    complaintJSON, err := ctx.GetStub().GetState(id)
    if err != nil {
        return nil, fmt.Errorf("Failed to read the data from world state",err)}
    if complaintJSON == nil {
        return nil, fmt.Errorf("the Complaint %s does not exist", id)}
    var Complaint *Complaint
    err = json.Unmarshal(complaintJSON, &Complaint)
    if err != nil {
        return nil, err}
    return Complaint, nil
}
```

Fig. 14. Function to View a Complaint by ID

ForwardComplaint: This function (Fig.15) is used to forward a complaint. When level-2 needs a previous complaint to solve an issue and makes a request for it to level-1, level-1 uses this function to forward the complaint to level-2. Though level-1 does not know the identity of the complainant, the copy of the complaint discloses the identity only to level-2 after forwarding.

```
// This functions helps to forward the complaint to Level2
func (pc *ComplaintSmartContract) ForwardComplaint(ctx contractapi.
    TransactionContextInterface, id string) error {
    complaint, err := pc.ViewComplaintById(ctx, id)
    if err != nil {
        return err}
    complaint.ComplaintTo = "Level2"
    complaintJSON, err := json.Marshal(complaint)
    if err != nil {
        return err}
    return ctx.GetStub().PutState(id, complaintJSON)
}
```

Fig. 15. Function to Forward a Complaint to Level2

6. RESULT

The proposed system has four modules– User, Level-1, Level-2, and Investigator. Level-2 is the highest level administrator while level-1 is the second level administrator. Level-1 and level-2 both control and manage system data. But level-1 has some restrictions here. Users are those who are employees of the company and are pre-registered in the system. They either file their complaints or get accused of their faults in the system.

On the other hand, an investigator is assigned to the system by level-2 to investigate a complaint.

When a user logs in, they can see the dashboard (Fig. 16). Here, they can file a complaint, check the statuses of their other active complaints or view any notice, such as warning for an accusation (only the accused can see it) and legal action or disciplinary action against other convicted employees.

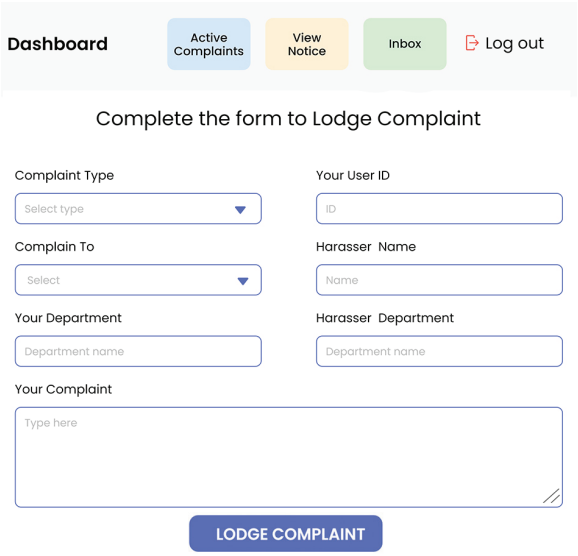


Fig. 16. User Interface Dashboard

The following image (Fig. 17) is the dashboard for level-1 and level-2. They can view the complaint list, check reports generated by the investigator, view notices and use inbox to communicate with the users. They can also search a complaint by its unique id. If need be, they can forward a complaint to themselves or ask to be forwarded.

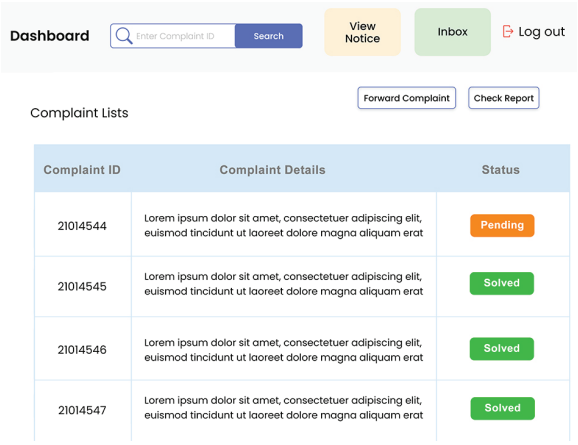


Fig. 17. Level-1 and Level-2 Interface Dashboard

Level-1 and level-2 can access the details of any complaint from the complaint list. At Level-1, the identity of the victim is kept hidden (Fig. 18) while level-2, as the higher authority, can see the identity (Fig. 19). Both levels perform specific functions based on their administrative capacities.

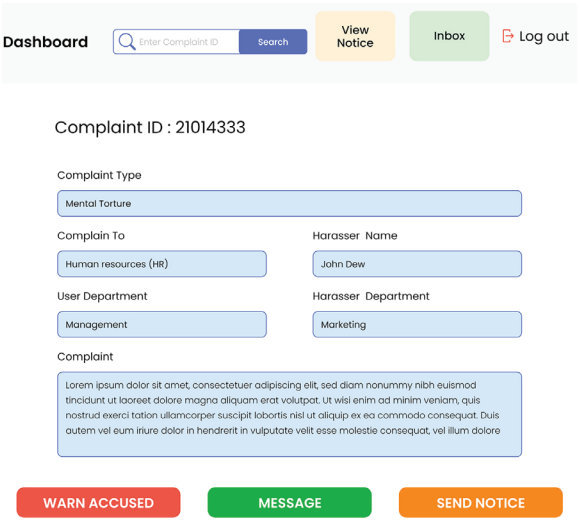


Fig. 18. Complaint Details Interface for Level-1

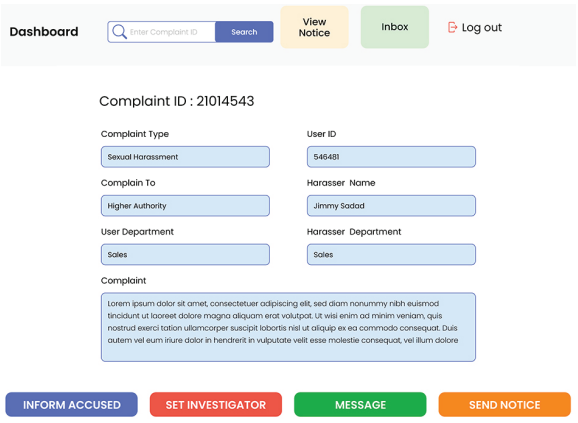


Fig. 19. Complaint Details Interface for Level-2

After being appointed by level-2, the investigator can view the details of the complaint on their dashboard (Fig. 20). They can ask for evidence from the victim and accused and may even message them if necessary. At the end of the investigation, they have to make a report to Level-2.

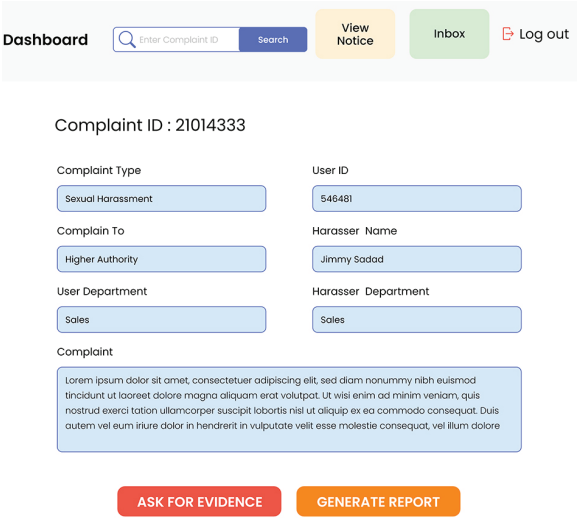


Fig. 20. Interface Dashboard for Investigator

7. CONCLUSION

This paper proposes a blockchain-based women harassment complaint system in the workplace for the documentation and management of complaints. The proposed system helps the victims to file their complaints in a trustless and tamper-proof environment while availing the prominent attributes like decentralization, anonymity, immutability, transparency, reliability and security of blockchain. It is composed of two hierarchical levels where level-1 ensures the victims' anonymity by letting them file complaints without revealing their identity and level-2 assists them in legitimately filing complaints with identity to the higher authority and documenting evidence. It not only makes the complaint management system more secure, productive and simple but also protects the victim from the next threat of the accused by preserving the victim's privacy. Furthermore, our system makes it easier for the HR and the higher authority to handle the complaints more efficiently. To cap it all, our proposed platform will ensure a safe working environment for women which will have a significant impact on the nation. Since very few papers have focused on this issue to solve with blockchain, our paper will contribute greatly to support the working women.

8. REFERENCES:

[1] K. Aquino, K. Lamertz, "A relational model of workplace victimization: Social roles and patterns of victimization in dyadic relationships", The Journal of applied psychology, Vol. 89, No. 6, 2004, pp. 1023–1034.

[2] S. Einarsen, "Harassment and bullying at work: A review of the Scandinavian approach", Aggression and Violent Behavior, Vol. 5, No. 4, 2000, pp. 379–401.

[3] J. E. Bartlett, M. E. Bartlett, "Workplace bullying: An integrative literature review", Advances in Developing Human Resources, Vol. 13, No. 1, 2011, pp. 69–84.

[4] S. B. Matthiesen, S. Einarsen, "Bullying in the workplace: definition, prevalence, antecedents and consequences", International Journal of Organization Theory & Behavior, Vol. 13, No. 2, 2010, pp. 202–248.

[5] A. Amin, M. S. Darrag, "Sexual Harassment in the Egyptian Workplace: A Literature Review and Research Agenda", Review of Management, Vol. 1, No. 4, pp. 25–38, 2011.

[6] B. L. Carvalho, M. E. Griffith, "More than One in Three Women Report Sexual Harassment In the Workplace", www.maristpoll.marist.edu (accessed: 2021)

[7] V. E. Sojo, R. E. Wood, A. E. Genat, "Harmful Workplace Experiences and Women's Occupational Well-Being: A Meta-Analysis", Psychology of Women Quarterly, Vol. 40, No. 1, 2016, pp. 10–40.

[8] "Speakfully | Workplace Reporting and Resource Platform." https://www.speakfully.com/ (accessed: 2021).

[9] "#NotMe." https://not-me.com/en/ (accessed: 2021).

[10] C. S. Wright, "Bitcoin: A Peer-to-Peer Electronic Cash System", SSRN Electronics Journal, 2019.

[11] K. Sultan, U. Ruhi, R. Lakhani, "Conceptualizing blockchains: Characteristics & applications", Proceedings of the 11th IADIS International Conference Information Systems, 2018, pp. 49–57.

[12] B. A. Rocha, "WORKPLACE HARASSMENT COMPLAINT EVIDENCE TRACKING SYSTEM USING BLOCKCHAIN", 2020,: https://harvest.usask.ca/handle/10388/13244 (accessed: 2021)

[13] "MINISTRY OF WOMEN & CHILD DEVELOPMENT." http://shebox.nic.in/ (accessed: 2021).

[14] S. Haber, W. S. Stornetta, "How to time-stamp a digital document", Journal of Cryptology volume, Vol. 3, No. 2, 1991, pp. 99–111.

[15] M. Crosby, Nachiappan, P. Pattanayak, S. Verma, V. Kalyanaraman, "BlockChain Technology: Beyond Bitcoin", 2016.

[16] H. F. Atlam, A. Alenezi, M. O. Alassafi, G. B. Wills, "Blockchain with Internet of Things: Benefits, challenges, and future directions", International Journal of Intelligent Systems and Applications, Vol. 10, No. 6, 2018, pp. 40–48.

[17] "Permissioned and Permissionless Blockchains: A Comprehensive Guide." https://www.blockchain-council.org/blockchain/permissioned-and-permissionless-blockchains-a-comprehensive-guide/ (accessed: 2021).

[18] C. Anderson, "Docker", IEEE Software, Vol. 32, No. 3, 2015, pp. 102–105.

[19] D. Bernstein, "Containers and cloud: From LXC to docker to kubernetes", IEEE Cloud Computing, Vol. 1, No. 3, 2014, pp. 81–84.

[20] "What is hyperledger fabric? | IBM." https://www.ibm.com/topics/hyperledger (accessed: 2021).

[21] M. Sethumadhavan, "On Blockchain Applications: Hyperledger Fabric And Ethereum", International Journal of Pure and Applied Mathematics, Vol. 118, No. 18, 2018, pp. 2965–2970.

[22] J. C. Anderson, J. Lehnardt, N. Slater, "CouchDB : the definitive guide", O'Reilly Media, Inc, 2010.

[23] E. Androulaki et al. "Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains", Proceedings of the 13th EuroSys Conf. EuroSys, Porto, Portugal, 23-26 April 2018, pp. 1-15.

A Low-Cost IoT Based Buildings Management System (BMS) Using Arduino Mega 2560 And Raspberry Pi 4 For Smart Monitoring and Automation

Original Scientific Paper

Muhammad Uzair
Electrical Engineering Department, Faculty of Engineering, Islamic University of Medina
Abo Bakr Al Siddiq, Medina, Saudi Arabia
uzair91@hotmail.com, muzair@iu.edu.sa

Karam Manaf Al-Janadi
Electrical Engineering Department, Faculty of Engineering, Islamic University of Medina
Abo Bakr Al Siddiq, Medina, Saudi Arabia
Aljanadi.karam.o@gmail.com

Salah Yacoub Al-Kafrawi
Electrical Engineering Department, Faculty of Engineering, Islamic University of Medina
Abo Bakr Al Siddiq, Medina, Saudi Arabia
Kafsalah@msn.com

Ibrahim Abdulrahman Al-Bulushi
Electrical Engineering Department, Faculty of Engineering, Islamic University of Medina
Abo Bakr Al Siddiq, Medina, Saudi Arabia
albulushiibrahim@gmail.com

Abstract *–This work presents an internet of things (IoT) based building management system (BMS) for monitoring, control, and energy management in buildings to provide an efficient way of energy utilization. Existing systems mainly provide monitoring of different parameters with limited controlling/automation functions. Existing solutions also do not provide automatic decision-making, advanced safety management, and resource tracking. However, the proposed system provides a comprehensive way of monitoring, controlling, and automatic decision making regarding different environmental and electrical parameters in buildings, i.e., temperature, humidity, dust, volt, etc., by using a low-cost wireless sensor network (WSN). The architecture of the proposed system consists of five layers and uses analog sensors which are connected to Arduino Mega 2560 microcontrollers for data collecting, NodeMCUs ESP8266 for wireless communication, Raspberry Pi4 microcomputers for decision making, and nod-RED dashboard which runs locally on a Raspberry Pi 4to provide a friendly end-user interface. The system also uses the Message Queuing Telemetry Transport (MQTT) communication protocol through Wi-Fi and completely relies on the local devices in the architecture and does not need cloud computing services. The proposed system provides two different kinds of automation, i.e., safety automation for the safety of different devices with advanced features, and energy automation. The proposed system is also able to provide humidity control inside a room and to track and count the available resources in any facility. The proposed system is low cost, scalable, and can be used in any building. Simulation results show that the proposed system is highly efficient.*

Keywords*: Automation, Building Management Systems, Energy Management, Internet of Things, Monitoring*

1. INTRODUCTION

Energy saving is one of the most critical challenges of this century. The increase in energy consumption, especially in residential buildings, has made this issue more critical. This is because residential/commercial buildings have now become an essential part of our community to live peacefully. However, many other factors, i.e., maintenance, power line health, etc., are also needed to be taken care of along with energy consumption for a smooth life in these buildings. Globally, the residential sector consumes around 41.4% of total energy consumption, which is considered a high percentage of energy usage. Therefore, many countries are trying to reduce this percentage by making an accurate analysis of different kinds of loads [1,2].

There is also a growing concern for the mismatch between the energy efficiency expected for residential buildings and the real output assessment, usually referred to as the performance gap, i.e., actual usage and expected usage. Typically, the reasons for the performance gap are distributed among multiple factors such as incorrect assumptions and insufficient forecast of the energy models at the design stage, the differ-

ence between the design and construction stage, not operating the system according to the instructions, and installation & maintenance challenges/issues, etc. [3]. Minimizing the performance gap is also a critical challenge for public-private social housing, where energy efficiency initiatives are the primary driver. It has also been proved that the high energy efficiency of residential/commercial buildings is considered as a leverage for the governments that need to be taken care of [4].

Similarly, with the increase of the population and the evolution of the world, many smart concepts have also been developed by the scientific community, i.e., smart houses, smart vehicles, smart health, etc. These smart concepts enable people to live comfortably by providing a wide range of accessibility. Furthermore, the scientific community is also developing a new concept of smart buildings. In smart buildings, an intelligent computer-based system, i.e., building management system (BMS), consisting of hardware & software will be installed to monitor and control different appliances (Heating, ventilation and air-conditioning, power systems, etc.) and services (security access control, elevator, and safety system, etc.) to increase the energy efficiency of the buildings. BMS coordinates all the appliances and services within a building to work/operate as a single complete integrated system.The main purpose of this concept is to reduce the energy consumption of the residential/commercial buildings up to a reasonable level by limiting the operation of the appliances according to a specific need, i.e., temperature, etc. This concept is a framework that incorporates remote control technologies to allow the end-users to monitor and manage energy consumption operations for the appliances [5]. Overall, BMS provides efficient usage of energy, healthy impact on the environment, peace of mind, improved security &better appliance life. Although BMSs existed since the 80s, approximately only 15% of buildings use this technology due to its overpriced costs and limited functionality [6]. Up to 30%, saving can be achieved in energy bills by using BMS, but only 9% of the world's energy is controlled by the BMS [7]. A more unified approach is required by manufacturers to build BMSs to provide more functions and enhance compatibility.

Moreover, the use of information and communication technology, i.e., IoT, is used to improve the efficiency of energy consumption in the building sector. IoT refers to the millions of physical devices that are connected to the internet to operate/control them wirelessly. The IoT technology provides a wide coverage area, low energy consumption, low cost, and outstanding connectivity. IoT concept capitalizes a vast majority of analog sensors as the main component for its operation [8]. Merging the residential buildings with the IoT concept offers an opportunity for the better use of the building's energy. To optimize the energy usage in buildings, many solutions have been developed previously such as applying a monitoring system that uses sensor devices to analyze the data, i.e., humidity, heat, vibration, current, voltage and pressure, etc., taken from these buildings. Monitoring and controlling these parameters makes it much easier to deal with the energy performance [9,10].

The existing literature generally discusses building automation systems (BAS) and building management systems (BMS) for the management of different building parameters. Most of the solutions in the existing BMS do not provide the automation feature, except a few of the papers discuss the automation with limited goals or as future work. The existing solutions leave the decision-making to the end-user as discussed in [11, 12]. Similarly, building automation systems (BAS) as discussed in [13] provide only limited energy and safety automation, i.e., turning on/off the lights if the system senses a human walk-in corridor, automatic water spray if smoke sensors detect fire, etc. BAS also does not provide facility management and enhancement of interior comfort. Moreover, the majority of the commercially available BMS systems are designed to implement very specific functions chosen by the manufacturer. If more functions are required, then the provider will charge extra or the order will be refused. These systems also suffer from inefficient automation, increased maintenance & security costs, etc. [14]. However, this work presents a comprehensive BMS system that covers all the deficiencies in the existing BAS and BMS by introducing an extremely compatible/scalable system due to fast data collection and automatic controlling.

The proposed system is not only capable of comprehensive monitoring and controlling, but also provides a new feature of taking decisions automatically, i.e., automation according to the desired goal and threshold set by the user. The proposed system also provides resource tracking and system safety with advanced features. The proposed system utilizes IoT technology for the monitoring, controlling, and smart automation of the building ambient parameter measurements, i.e., current, voltage, vibration, dust level, humidity, temperature, water pressure, etc., to provide efficient energy utilization in these residential buildings. Automation has an impactful role in energy management because machines, i.e., computers, etc., are more consistent in performing as compared to humans and also reduce exposure to control problems [15].

The work presents two kinds of automation features, i.e., safety automation and energy management automation. The parameters are periodically measured in the safety automation based on some predetermined value and the system can take up to four unique decisions for the safety of any device. The proposed system checks/monitors the devise's reading variation from 0 to 30%, and takes appropriate actions accordingly at a different level of variation, i.e., less than 10%, 20%, and 30%, etc., to provide safety of these devices. For energy management automation, the user can set a daily wattage limit for each room/unit by using the dashboard. If the wattage readings exceed the specified limit, the system tunes down the wattage of the room to the de-

sired level. Simulation results show that 90 Wh energy can be saved if a 60 W bulb is used for 10 hours a day. Similarly, simulation results show that the proposed system saves up to around 2600 Wh energy for a room in one day. These energy savings can be further enhanced/adjusted depending on the end-user's comfort level. Similarly, for resource tracking, a sensing circuit is placed in the storage room that can track and count the items to detect how many objects are available at any time. The proposed system is capable of tracking up to 540 units in any building, and it is easily extendable. The proposed system also keeps humidity inside a room within limits, i.e., between 35-55%, by activating and deactivating the humidifier through an automatic control mechanism.

The architecture of the proposed system consists of five different layers which are the perception layer, transport layer, middleware layer, application layer, and business layer. We chose to build our system based on five layers to elaborate the system more comprehensively and efficiently. The paper also discusses in detail how the data is collected from analog sensors, then processed and uploaded to the MQTT broker. The processing unit in the system, i.e., the middleware layer, is programmed to take actions according to the predefined thresholds regarding safety and energy by enhancing the IoT data analytic level in the system. Mainly, the proposed BMS goal is to reduce the energy consumption of buildings and provide an automation system that can predict any emergency before its occurrence. We used a Wi-Fi network for high-speed communication which is very essential for safety automation as compared to the existing solutions which use Zigbee or Lora communication which has a low speed. Based on the described features, the proposed system provides enhanced safety, more energy saving by using automation, and efficient resource tracking as shown by simulations.

This article furthermore includes multiple sections which are arranged as follows. Section 2 presents the related work. Section 3 presents the architecture of the system and discusses different layers. Section 4 discusses the data collection and decision-making procedure, including the idea of how the different devices communicate with each other. Section 5 shows the simulation of the proposed system and section 6 is the conclusion.

2. LITERATURE REVIEW

The existing literature discusses very little work related to the IoT in BMS. One of the papers [16] introduced an industrial IoT system that allows operators to control industrial applications using the Raspberry Pi 3 B. The paper claims that the IoT in buildings significantly reduces operating expenditures. The proposed approach only focuses on industry-related applications and monitors just three parameters which reduces the feasibility of the system. In [17], an efficient and low-cost IoT-based building monitoring system is implemented by using a Raspberry Pi microcomputer which collects the environmental parameter such as temperature and humidity. An advantage of this system is that it eliminates the need for high-power computers to analyze the collected data. The system uses the Zigbee protocol for communication which considers a battery-friendly protocol, but this protocol has short-range coverage as compared to Wi-Fi or Lora.

Paper [18] demonstrated the rapid growth of IoT technology in BMS and reviews the related works of IoT and big data analytics in smart buildings. The paper implies that integrating IoT in building management will lead to the improvement of residents' comfort by simplifying the building management. However, the huge amount of collected data is a drawback in this BMS system. Paper [19] utilized a web-based tool that stores, collects, and represents the energy data of buildings from heterogeneous and dynamic sources to enhance the interactivity of building energy management systems. The purpose of such a system is to contribute to energy saving and behavioral change, i.e., finding monitoring points and monitoring objects for intelligent power supply, etc. However, the presented solution provides only monitoring options, which limits the system potentials such as controlling.

A real-time IoT platform for building health monitoring was proposed in [20]. The system encompasses Piezoelectric sensors (PZT), an Arduino, and a NodeMCU ESP8622 Wi-Fi chip. The PZT is used to generate and receive Lamb waves to determine and analyze the health of the concrete structure. The presented IoT could detect the cracks present in the building and other potential problem areas. In [21], an intelligent power monitoring of building equipment based on IoT technology is presented. The solution focuses on implementing a monitoring system for indoor electrical equipment. Paper [22] introduced a system for data acquisition and control for homes by using the principle of IoT. The system implements a Zigbee approach and a single chip controller to handle the data connection process and implement various types of sensors to obtain the readings. However, the proposed system does not provide a clear circuitry or simulations.

Similarly, paper [23] proposed a design of smart panels that could be used for building management purposes. These smart panels control loads using the principle of IoT claiming that it increases the protection of the operator by allowing them to observe and control the loads from a distance. The system uses NodeMCU to monitor the main electrical parameters such as current and voltage. However, the system is only aimed to control loads with limited functionality, i.e., connect and disconnect. Paper [24] illustrated a monitoring solution for zero energy building management by using power line communication (PLC) and android application. This building management system enables the monitoring and control of the power usage, through an android application to enhance the sustainability of

the building. However, the PLC technology can only be installed in existing power lines. A smart building automation system (BAS)was also proposed in [25]. BAS controls the lighting including artificial lighting i.e., on/off & dimming control, etc. The system also controls air conditioners with some safety features i.e., fire alarm and gas alarm, etc. A limitation of the system is that it presents fewer sensory and automation options for a building automation system.

Paper [26] presented a monitoring and diagnostic system to improve the building operations. This system provides a prototype information monitoring and diagnostic system (IMDS). The IMDS consists of multiple high-quality sensors, data collecting software & hardware, and a web-based remote system for information accessibility. The web-based remote enables the user to identify and control problems. However, the system only shows an improvement in the building's climate. Paper [27] discussed a remote monitoring system for building automation, i.e., BAS by using VxWorks. The BAS consists of a PC104 bus, which adopts VxWorks operating system and AMD processor. The system helps to improve the automation of several operations. Paper [28] proposed a monitoring and controlling system for house lights and appliances using the principle of IoT. The proposed system used mega Arduino and sensors for data collection alongside MIT app inventor to design the interface of the system. The Bluetooth was used in the proposed system for communication which limited the range and ultimately provided inaccurate current and voltage readings.

Paper [29] discussed an IoT system based on cloud computing that would help in efficient building energy management for countries that encounter electrical power shortages such as Nigeria. The proposed IoT system supports utilizing the little available energy by automation and it collects data for future planning. The paper indicated that one shortcoming of the system is the method of sensing which is not cost-effective for their native environment. Paper [30] introduces a BAS system by using two approaches, i.e., programmable logic controller and image processing. The proposed system uses motion and light sensors to indicate the presence of individuals. For AC controlling, the system uses image processing to indicate the number of individuals in an area to control the temperature and operation state (on/off). However, the proposed system is not able to monitor the operational states of the loads, i.e., no monitoring of the health of the electric machines.

The comprehensive literature review indicates that the current presented building management systems (BMS) are mainly aiming to monitor specific ambient parameters such as temperature, humidity, and/or provide limited controlling and automation, i.e., leaves the decision making to the user. Also, the current building automation systems (BAS) have limited goals of controlling and automating. Only the on/off capability of switches, adjustment of the air conditioner according to the number of people in the room kind of functions can be performed which ultimately limits the overall system efficiency. These systems do not provide a complete understanding of the operating loads' states and conditions, i.e., machine health, etc., and do not provide facility management options at all. However, the proposed system in this work does not only provide all the features of BMS and BAS systems with better monitoring and controlling of different parameters but also provides the decision automatically, i.e., automation, to achieve the desired goal and threshold set by the user. The proposed system also provides automatic safety with advanced features for different equipment/machines along with resource tracking.

3. ARCHITECTURE

IoT architecture consists of many elements such as sensors, actuators, protocols, cloud services, layers. gateways, etc., which are connected to collect, store, and process data to perform desired functions sent via a user application. The distinguished layers in the IoT architecture use different protocols and gateways to keep track of the system's consistency.

At the beginning of the new era of IoT technology implementations, the proposed IoT architectures generally consisted of three layers, i.e., perception, network, and application layer [31]. However, these architectures are not able to provide a complete visualization of the IoT main components such as data processing and security [32]. Therefore, researchers have modified and added two more layers known as transport and middleware layers to acquire a comprehensive conceptualization of the IoT system [33]. Table 1 shows the different IoT layered architectures [34,35]. In this paper, 5-layer architecture is used and presented from the bottom (perception layer) to the top (business layer) to cover all the necessary details in the implementation process. Fig. 1. shows the complete architecture of the proposed solution.

Table 1. IoT architecture.

3-Layer architecture	4-Layer architecture	5-Layer architecture
Application layer	Application layer	Business layer
Network layer	Processing layer	Application layer
Perception layer	Network layer	Middleware layer
--	Perception layer	Transport layer
--	--	Perception layer

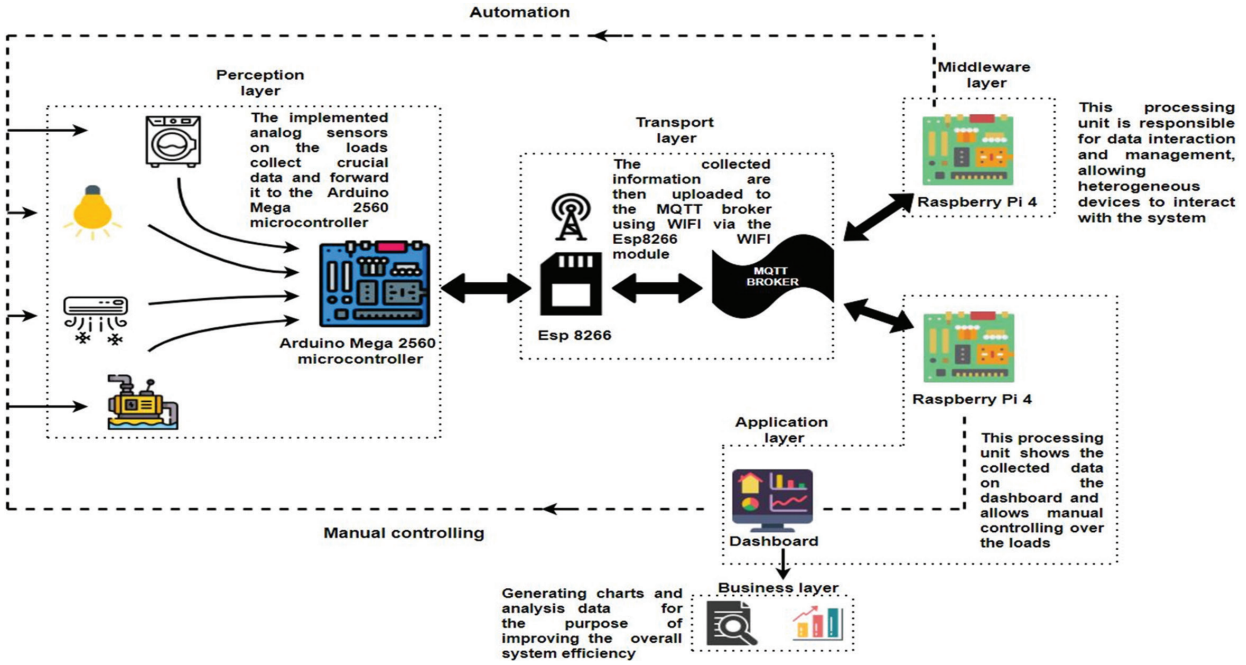


Fig. 1. The complete architecture of the proposed solution.

3.1 PERCEPTION LAYER

The perception layer, i.e., the recognition layer, is the lowest layer of the conventional IoT architecture. The main objective of this layer is to collect useful data from the environment, i.e., physical parameters, other smart objects, etc. The perception layer collects information using interlinked data collection technologies such as heterogeneous devices, environmental sensors, and WSNs. The collected data is converted to digital data and transferred to different layers, i.e., application and middleware, etc., in the network for further processing [36].

This form of data collection raised a lot of security-related concerns, i.e., eavesdropping, node capture, fake node & malicious, Replay & timing attack, etc., which leads to the development of various security approaches to protect the perception layer from unauthorized access. The process of gaining authentication and authorization is known as access control which is achieved by two main methods Role-Based Access Control (RABC) and Attribute-Based Access Control (ABAC). In the RBAC approach, the roles of the users are already defined, limiting the number of resources authorized to the user and allowing him to access specific information related to his role. On the other hand, ABAC grants the user's resources or data access based on attributes, i.e., user attributes, environmental attributes, resource attributes, etc. ABAC approach provides a much-reduced risk by making the access more flexible and scalable supporting both fine-grained access and dynamic extension for a large scale number of users [37].The proposed system also uses the ABAC approach due to its sufficient scalability and flexibility in providing fine-grained access control alongside the ability to operate in a complex system with a large number of users making it more suitable and adaptable to the access control of IoT. To obtain an accurate analysis of the load behavior in our design various types of sensors were implemented to collect important information during the load operation time as shown in Table 2.

Table 2. Implemented sensors

Sensors	Monitoring
TMP36 Temperature Sensor	Ambient Measurement
GP2Y1010AU0F Dust Sensor	Ambient Measurement
AHT10 High Precision Digital Temperature and Humidity Sensor Measurement Module I2C Communication	HVAC Unit Room Humidity Monitoring
SW-420 Vibration Sensor	HVAC units-Water Pump Motors
SEN0257 Gravity Water Pressure Sensor	Water Pump Motors
BMP180 Water Pressure Sensor	Water Pump Motors
ZMPT101B Single Phase AC Voltage Sensor	Electrical products-HVAC units
Allegro ACS712 AC/DC Current Sensor	Electrical products-HVAC units

For environment analysis, temperature, humidity, vibration, and pressure sensors are implemented along with voltage and current sensors for electrical analysis. Each of these sensors is distributed on loads based on the parameter that defines its behavior. For example, a water pump motor behavior is defined by water pressure, motor temperature, and vibration setting. A reference point is set for all these parameters and the operation of the motor is analyzed based on these parameters which provide us a clear view of the motor operation condition. These readings would also make it easier for the operator to define any occurring liabilities. The readings from the sensor are collected, organized, and converted to digital data by using an Arduino Mega microcontroller 2560 which is physically connected to analog sensors. The Arduino mega microcontroller can handle up to 16 analog signals simultaneously providing great sensing expandability to cover a complete residential unit or a room. Fig. 2. shows the connectivity between the sensors and Arduino mega 2560.

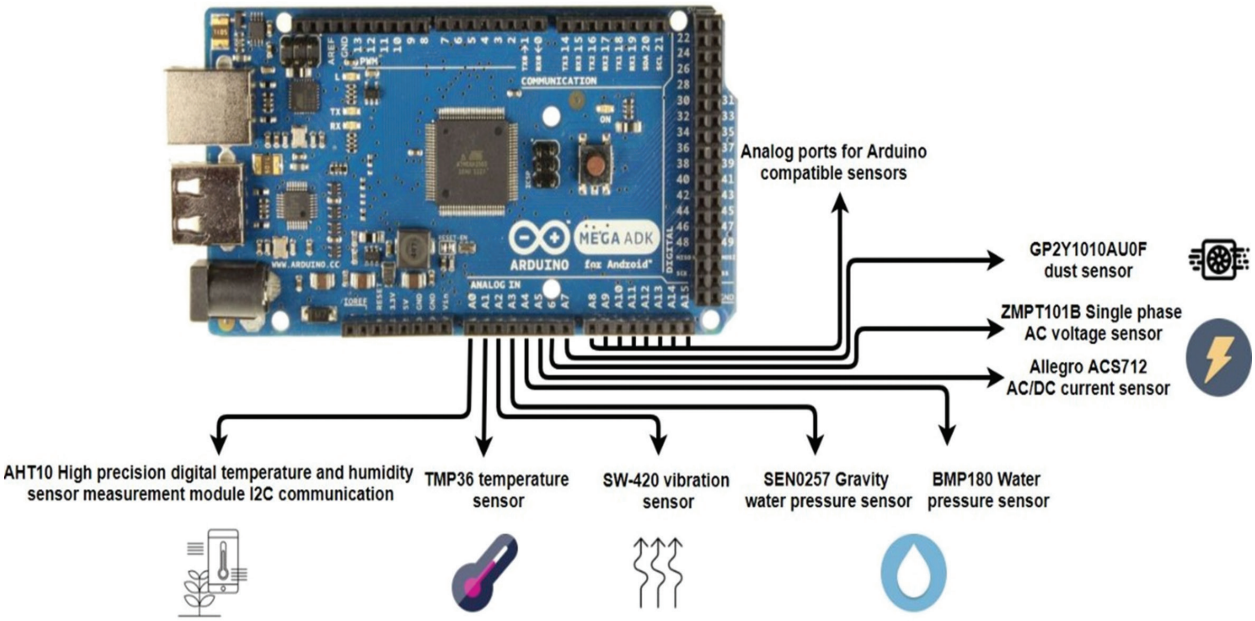


Fig. 2. Connectivity between analog sensors and Arduino Mega Microcontroller

3.2 TRANSPORT LAYER

The transport layer plays a major role in the WSNs that is established by the IoT system [38]. After collecting the ambient parameter measurements by the sensors in the perception layer, the transport layer is responsible for transmitting the collected data from the Arduino mega platform to the processing unit i.e., the middleware, etc., and vice versa. Several networks have been used in this layer such as Bluetooth, Zigbee, Lora, Wi-Fi, and cellular. Lora network has been used due to its capability to transmit data over long distances which could reach several kilometers with low energy consumption as proposed in [39,40]. However, since it uses low energy, the bandwidth of Lora channels is small compared with other networks such as Wi-Fi and Zigbee and the data rate is very low i.e., 0.3-50 kbps [41].

The low data rate can only be accepted if the IoT system is used for monitoring. However, the proposed BMS system in this work also provides controlling in addition to monitoring to provide simultaneous bidirectional communication for better management which cannot be provided by the Lora network. Therefore, the proposed transport layer (Ergo) uses Wi-Fi as a communication network. The advantage of the Wi-Fi network is that all smart buildings nowadays have Wi-Fi coverage and there is no need for additional gateways (routers) in suchbuildings while using Wi-Fi as needed by Lora or Zigbee networks. Additionally, the current Wi-Fi standards provide a bandwidth of 2.4-5 GHz, which is sufficient for faster communications without any delay, and further enhances user experience [42]. Fig. 3. demonstrates the range of networks correlated with their bandwidth [43-45].

To grant Wi-Fi connectivity to the Arduino Mega 2560, NodeMCU ESP8622 is attached to the Arduino microcontroller. The ESP8622 is a low-cost firmware kit that is used in Arduino IoT projects and communicates with sensors via serial communication and is responsible for transferring the data from the Arduino Mega to the gateways i.e., routers. The gateway is the main bridge that links between the middleware and sensors, microcontrollers, actuators, and relay modules. The communication protocol, which is followed by the Wi-Fi network is the MQTT network protocol, which is an IoT machine-to-machine connectivity protocol and is used for bidirectional publish-subscribe data transport as illustrated in Fig. 4. MQTT is also capable of providing connections at remote locations having constrained resources and at low bandwidth environments.

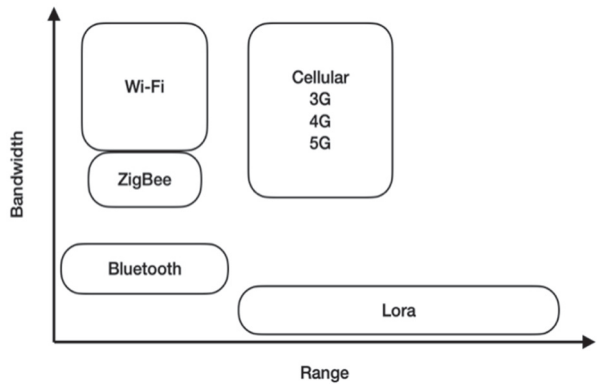


Fig. 3. Wireless networks' bandwidth correlated to their range.

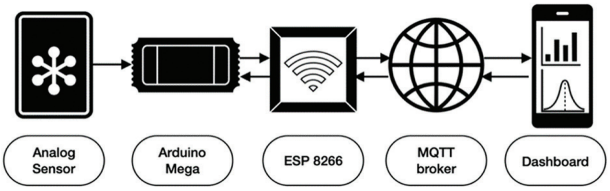


Fig. 4. MQTT communication between sensors and dashboard.

3.3 MIDDLEWARE LAYER

Heterogeneous device interaction within the IoT network is a complicated process since each device provides a unique data type, operational behavior, and supports different communication protocols to perform a different operation for the required interaction. These issues also arise when these heterogeneous devices interact with the data collection devices which collect information from the surrounding environment. All of these issues are addressed in the middleware layer of the IoT, which is also known as the processing layer. This layer stores, analyzes, and processes a large amount of data coming from the transport layer, and also manages and provides several services to lower layers. The middleware layer also fulfills the IoT domain for various applications such as managing a large amount of information, conditional awareness, adoption, scalability, cloud computing, databases, bid data processing modules privacy, security, automation, device locating and management, etc. Therefore, the middleware layer is a software platform represented by some sort of computing process where all the heterogeneous sensor domains of the applications are joined to create a processing unit that is responsible for the devices' information interactions and management in a unified way. [46].

In the proposed system, we selected a low-cost Raspberry Pi 4microcomputer to behave as the middleware layer due to its ability to perform different functions, i.e., structure complex data models, supports various operating systems, supports multiple coding languages, and can handle huge processing power in a compact board, etc. The Raspberry Pi 4 analyzes, processes and manages the information which is uploaded to the MQTT broker. The gathered and processed data is also saved in a comma-separated values (CSV) file, i.e., a file that saves data in the form of spreadsheets and tables. It is always easy to import and export data using programs that similarly save data. Therefore, storing data as a CSV file assist in communication between the devices in the network as a unified data language. This middleware approach allowed our system to collaborate in communication and provide information exchange along with massive scaling flexibilities of devices, which is due to the huge computational power capability of the Raspberry Pi 4 to operate a large number of devices at the same time [47].

In the proposed system, two controllers are used (Arduino Mega and Raspberry Pi), as each one of them has unique functionality and tasks to perform. The Arduino is responsible for collecting the environmental signals through sensors and converts these signals into digital form for further processing. However, the Arduino hardware is limited and cannot be used for automation or dashboard operations. Therefore, Raspberry Pi is used, which works like a computer or processing unit for the overall system. Raspberry Pi processes the collected data, makes it visualize on the dashboard, and performs automatic decisions. Therefore, Arduino's task is to do datafication, and Raspberry's task is to do automation.

A primary concern in the proposed system is handling a large amount of data coming from the sensors attached to the loads as Arduino mega is collecting data continuously and a huge amount of data will be accumulated as time will pass. To overcome this challenge, the system is programmed to extract only useful information by managing the raw data into more managed groups (CSV) files and removing needless loads of data as discussed in paper [48]. Similarly, in cloud-based application network delay is generally higher due to the sharing of the same communication link in the cloud by multiple loads. This issue is resolved by implementing a specific mega Arduino microcontroller for each building unit which will increase the bandwidth available for each load and will reduce the round-trip time and network congestion.

3.4 APPLICATION LAYER

The application layer provides monitoring and controlling. This layer is the end-user dashboard layer, i.e., interface, which enables the users to interact and communicate with the IoT devices in the network, as it is capable of providing application-specific services to end-user such as smart homes, smart health, and smart cities, etc. It displays sensor readings, which are uploaded to the MQTT broker in an organized and simple way, i.e., easily understandable to the end-user. Furthermore, this layer empowers the user to remotely control the devices in the system, i.e., moving the actu-

ator parts for motion purposes, switching the electrical relay modules for energy management purposes, etc. The design of the application layer heavily relies on the understanding of the end-users and their main goal of using the application.

In this work, the application layer displays the environmental measurements of the building such as temperature, humidity, power, etc. The dashboard is programmed to notify the end-user in case of any abnormal measurements and the IoT system immediately acts to protect the electrical devices accordingly. Moreover, in the dashboard, the user can control the electrical devices which are connected to the system such as air conditioning units to achieve facility and energy management. Additionally, the dashboard also shows the total number of units/resources in the warehouse for tracking and counting purposes, i.e., resource location tracking/counting, etc. We built the application layer using the Node-RED platform. This platform is an open-source flow-based programming tool for connecting hardware devices and is capable of providing a set of nodes to quickly create a live data dashboard. The dashboard will run locally on a Raspberry Pi 4 microcomputer as a web server which can be accessed via the Wi-Fi network of the building through computers or mobile phones. Fig. 5. shows the proposed IoT dashboard which displays the available monitoring and control options in the system.

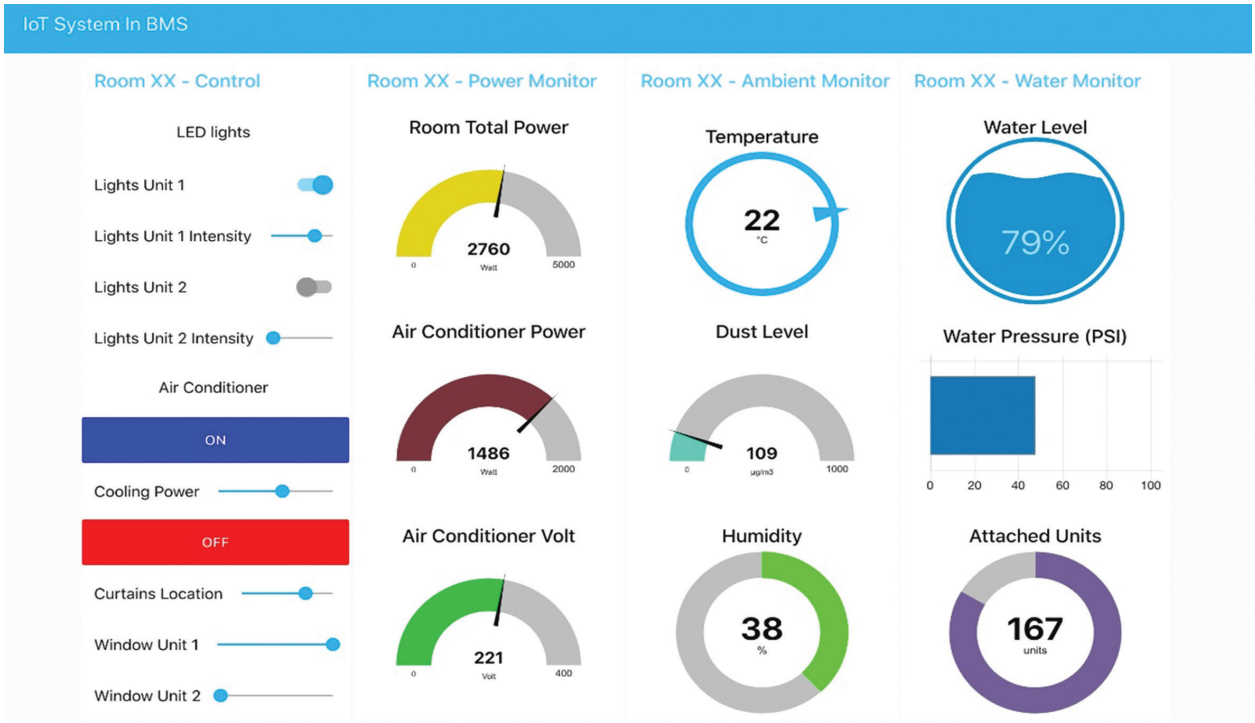


Fig. 5. Proposed IoT Nod-RED dashboard built on Raspberry Pi 4.

3.5 BUSINESS LAYER

The top-level layer in the IoT architecture is the business layer which acts as a manager of the whole system. This layer manages and controls applications, business, and profit models of IoT, along with the user's privacy. Hence, the business layer coordinates business logic, top-level requirements, and various other management operations to achieve a successful and durable architecture that provides continuing values to the system and operator. The business layer takes the responsibility for the management of the overall IoT system by creating graphs, flowcharts, business models, and much more based on the information gathered from the analyzed applications as this layer can create, store, and change the acquired information. This whole process ultimately increases the system efficiency by introducing future actions and strategies that would contribute to achieving success for the implemented IoT system.

4. DATA COLLECTION AND DECISION MAKING

4.1 ANALOG MEASUREMENTS

The flow chart for data collection is shown in Fig. 6.The data collection process starts with the perception layer in which the analog sensors collect the ambient measurements and send them to the Arduino mega 2560, which converts the analog signals into digital signals. The number of required microcontrollers for each unit, i.e., area or room, depends on the application. It is sufficient to provide each room with one Arduino mega to collect temperature, humidity, dust, water pressure, water level, voltage, current, and power of the room in residential applications since each Arduino mega microcontroller is capable of processing up to 16 analog signals.

A NodeMCU ESP8266 which is physically connected to the Arduino mega receives the digital signals via se-

rial communication. Both Arduino mega and ESP8266 are programmed in C++ language. After receiving the data, the ESP8266 uploads the readings of each room to its specified topic in the MQTT broker and each sensor reading to its assigned subtopic in the form of topic/subtopic as demonstrated in Fig. 7., which shows the readings of temperature, dust and humidity of a room xx with date and time. Since MQTT is a publish-subscribe communication protocol that is implemented in the transport layer of the proposed IoT network, the ESP8266 is publishing the readings to the predefined broker and any device which is subscribed to the broker can receive the published information. Subscribing to the MQTT broker can be achieved via different devices and platforms such as personal computers, mobile phones, the Nod-RED dashboard, and the middleware of the system. Both middleware and application layers are subscribed to the topics and receive data directly from the broker. The display elements on the dashboard of the system are subscribed to the subtopics of the broker to display the readings directly in the user application and the middleware computer (programmed in Python), the data is imported from the broker for further analysis and decision making such as automation as shown in Fig. 6.

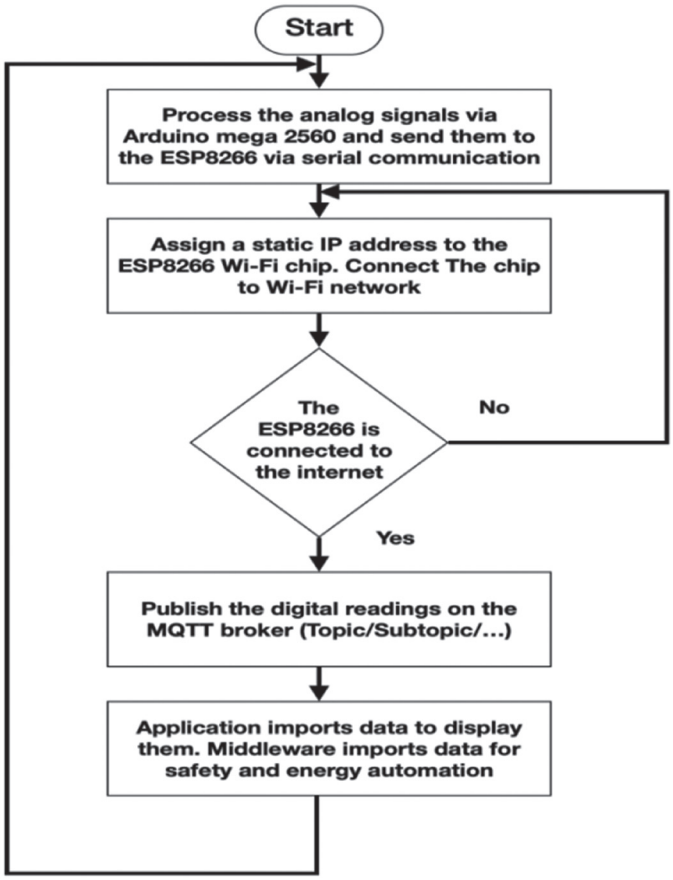


Fig. 6. Data collecting flowchart

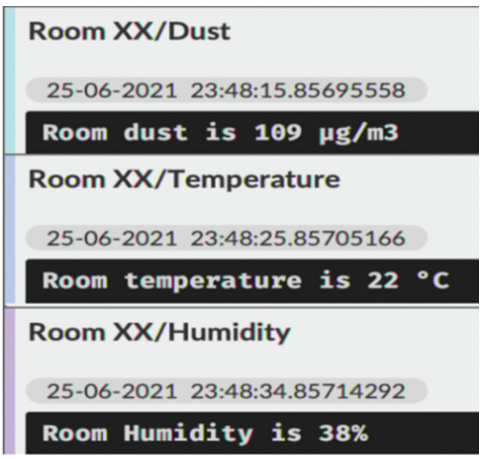


Fig. 7. Sensor reading uploaded to the broker

4.2 SAFETY AUTOMATION

The new key aspect of the system is also its capability to perform actions without human interaction, i.e., automation. The goal of automation is to enhance the efficiency of energy utilization and keep the loads in healthy conditions in case of any malfunction. The proposed system is capable of providing two kinds of automation, i.e., safety-automation, and energy-conserving automation. In safety automation, two kinds of operations can be done as shown in Fig. 8b.Either the load can be shut down in case of major malfunctions or the operation intensity can be reduced temporarily, if the problem can be fixed by the operator in a short time. The system sends a notification of the problem in both cases informing the operator.

To obtain the safety automation, the decision-making process is implemented in the middleware layer of the IoT system using the Raspberry Pi 4 microcomputer by adjusting a predetermined range of the parameters utilized by the loads that would determine the load condition. The decision-making process also varies between loads depending on the number of parameters used in making that decision and the impact of the malfunction itself. For example, a slight increase in the air conditioner unit temperature wouldn't be sufficient to start the decision-making process, but on the other hand, a slight change in the humidity and vibration of the air conditioner units would stimulate the IoT network to perform the required action. However, dramatic change in any parameters is enough to activate the automation process and the corresponding action will depend that how much change has occurred in the parameter reading. In our proposed system, a measured parameter is compared with the predefined limits, and if there is a 10% increase or decrease in the reading, i.e., variation, no action is taken, and only a notification is sent to the operator/management. If there is a 20% increase or decrease in the reading, i.e., variation, operation intensity is reduced and this reduction continues until the reading tolerance is below 20%. The operator/management is also informed during this process. If the reading increases or decreases more than 30%, i.e., variation, immediate action is taken to shut down the load for maintenance purposes due to some major malfunction. The decision-making process combines multiple parameter readings to implement the suitable action based on a predetermined function applied in the Raspberry Pi 4 which is implemented in each unit (e.g., room) to remotely control the electrical switches i.e., relay module connected to the loads, etc.

4.3 ENERGY AUTOMATION

Energy conserving automation is shown in Fig. 8a. The system integrates energy-saving methods via automation by measuring power consumption, i.e., a way to improve energy management in buildings. In every room, a Raspberry Pi 4 device will be programmed by using the python language in the middleware layer to collect and analyze the power behavior of the room. The analysis of the power is conducted to generate a mathematical estimation, i.e., estimating average power consumption, to take actions automatically. Estimating the average energy consumption in each unit helps to attain a goal limit via the dashboard that the middleware layer must achieve to minimize energy consumption. The business layer then presents the energy consumption of a certain unit, i.e., room, in the dashboard using statistical charts/figures. In our work, the proposed system calculates the current power consumption in a room, as shown in Fig. 8a. Then the proposed system estimates the power consumption of the room for the next one hour by generating a mathematical model.

The estimation is established periodically with a period defined according to the building nature. In our work, we consider that a period of one hour is sufficient for the residential buildings. Thus, for each hour the system calculates the average power of the room for 2 minutes and estimates the power for the next 58 minutes. Each room has its own Raspberry Pi 4 processing unit as a middleware layer which generates a mathematical estimation using linear regression and acts accordingly. Then the system measures the consumed energy of the room, and If energy consumption is more than the limits, the system automatically starts to tune down the lightning and cooling power of the air conditioner of that room to curtail the energy consumption. For example, if the average daily consumption of a room is 23,000 Wh, the management can set a goal of 20,000 Wh as a way to reduce electricity consumption. If energy consumption increases beyond this limit, the system automatically starts to tune down to curtail 3,000 Wh daily.

This approach is implemented without compromising the interior comfort or the intended goal of the loads. The maximum level of tuning down can be defined by the users to maintain the convenience of the system through the dashboard where the user specifies the minimum threshold level for cooling power or lightning, i.e., the automation system cannot go below. However, if energy consumption is close to the limits, the analytic tools also notify the user/management to take necessary actions, i.e., turning off any appliance, etc., to conserve the energy.

4.4 RESOURCE TRACKING

The proposed system also provides a resource tracking feature as shown in Fig. 9. With analog inputs, the digital ports of the Arduino Mega are utilized in collecting data for resource tracking. The digital pins are connected to a high input voltage, i.e., 3.3 volts, and a mechanical switch is placed in between the input voltage and the digital pin. The mechanical switch is located beneath the resources such as paint cans. If the resource is placed at its designated location, then the mechanical switch will connect to the circuit and 3.3 volts will be applied to the pin which is associated with the resource, and it will indicate that the item is attached or vice versa. This approach will help in counting the number of available resources in certain locations such as warehouses, buildings, etc. Implementing resource tracking via this method in buildings will help in minimizing errors and provide constant feedback about the status of the resources regarding their location and quantities. This approach is scalable because each Arduino mega has 54 digital inputs, and 10 Arduino mega are sufficient to track up to 540 units. Future work can be enhanced by using computer vision in which counting is achieved via visual computing through cameras and this approach can cover larger areas with massive scalability. Another advantage of the system is its flexibility, and it can be used for counting different resources, i.e., paint cans, soft drinks, water bottles, devices, etc.

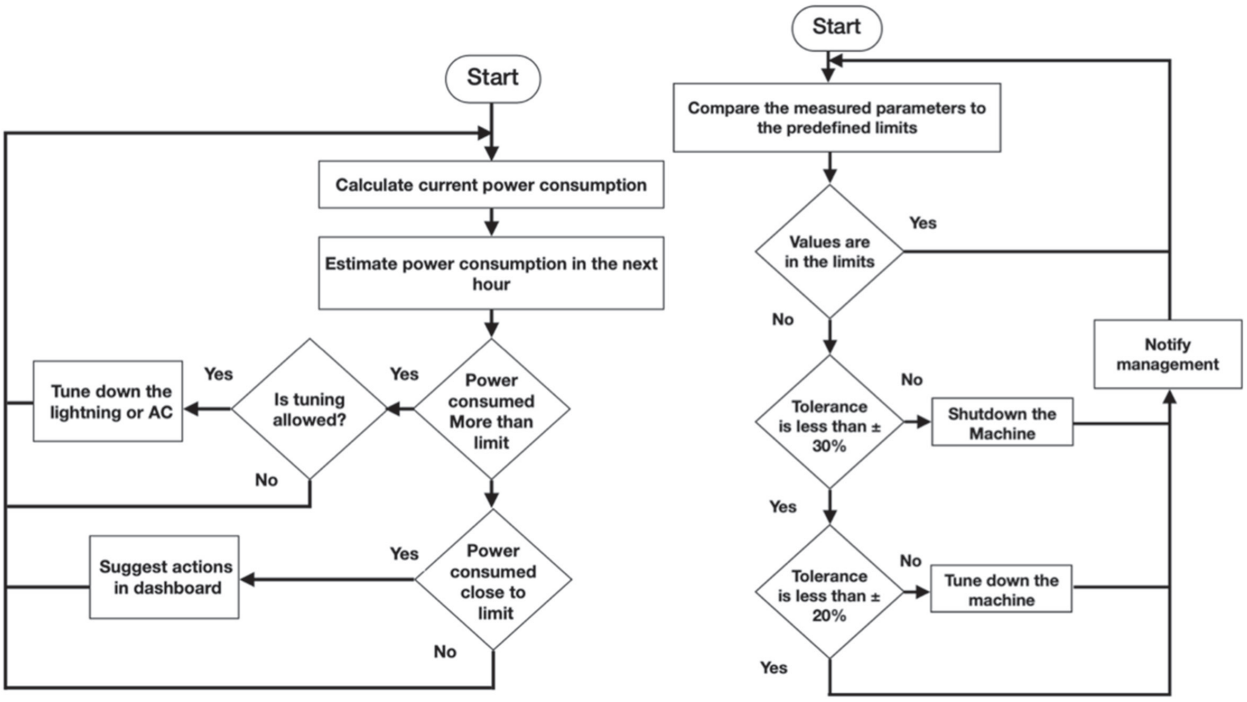


Fig. 8. Automation flowcharts. (a) Energy conversion automation flowchart (b) Safety automation flowchart

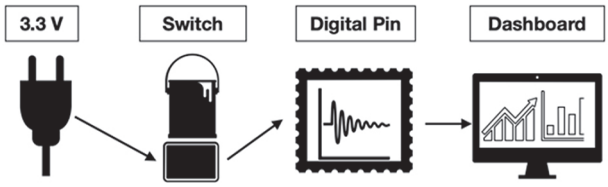


Fig. 9. Resources tracking mechanism

5. SIMULATION

Proteus 8 professional software is used to create the circuitry for simulation. This software provides a wide range of electronic devices, i.e., Arduino mega 2560, NodeMCUs, environmental sensors, etc. The user can also download multiple electronic device libraries to encompass all kinds of electronic tools. The proposed BMS monitoring system is composed of Arduino mega 2560, NodeMCU, vibration sensors, current sensors, water pressure sensors, humidity sensors, temperature sensors, dust sensors, and voltage sensors as mentioned in Table 3. These components are interconnected together, and the readings are displayed in the dashboard as well as in the simulation module.Fig. 10. presents the simulation module of the proposed BMS system. The NodeMCU communicates/transfers the readings collected by analog sensors to the dashboard through the transmitter and the receiver pin terminals i.e., TX and RX. The vibration sensor (SW-420) senses if there is a high or low vibration in the place and communicates with the dashboard through the NodeMCU. It is connected to a logic state switch which sends a signal of zeros and ones. If the signal is zero, no vibration will be detected and if the signal is logic 1,the vibration will be detected. Humidity and temperature sensors (DH22) deliver the humidity and temperature readings to the NodeMCU. The current sensor (ACS712) sends the reading of the loads' current to the NodeMCU and is displayed in the dashboard. The current readings are taken from different wires in the buildings to monitor the energy state of the building. The water pressure sensor (BMP180) senses the water pressure and sends the readings to NodeMCU. The pressure readings give us an idea about the water level in the tanks. The voltage sensor (ZMPT101B) senses the voltage in the wires and reports to NodeMCU. In the simulation module, the sensed voltage was air conditioner voltage. Dust sensor (GP2Y1010U0F) is an optical dust sensor that measures the dust level by using lenses and it can sense the dust level in the whole room, i.e., existing BMS systems do not provide this feature. The sensing module will be placed in each room in the building and all rooms in the building will report findings to the dashboard. With this method, we can predict any unusual phenomena and take actions accordingly before the occurrence of any emergency as well as monitor the energy behavior of the whole building.

For energy management, Fig. 10., also shows two lightning units, i.e., 60 W bulbs, which have been illuminated with and without automation. The bulb used without automation consumes 60W, and the bulb used with automation was illuminated at 85 percent working capacity, i.e., intensity, with a set value of operating at 51 watts, i.e., saving 9 W. This intensity does not compromise the comfort of the consumer either it affects the lighting of the room as it is providing a reasonable brightness. Moreover, this intensified value can be adjusted by the user to any desired level, keeping in view his comfort level. Therefore, the system tunes down the wattage of one light bulb by 9W, i.e., a set value in the dashboard. Hence, if a light bulb in a room is used for 10 hours a day, the system saves around 90Wh per day with a single 60 W bulb.

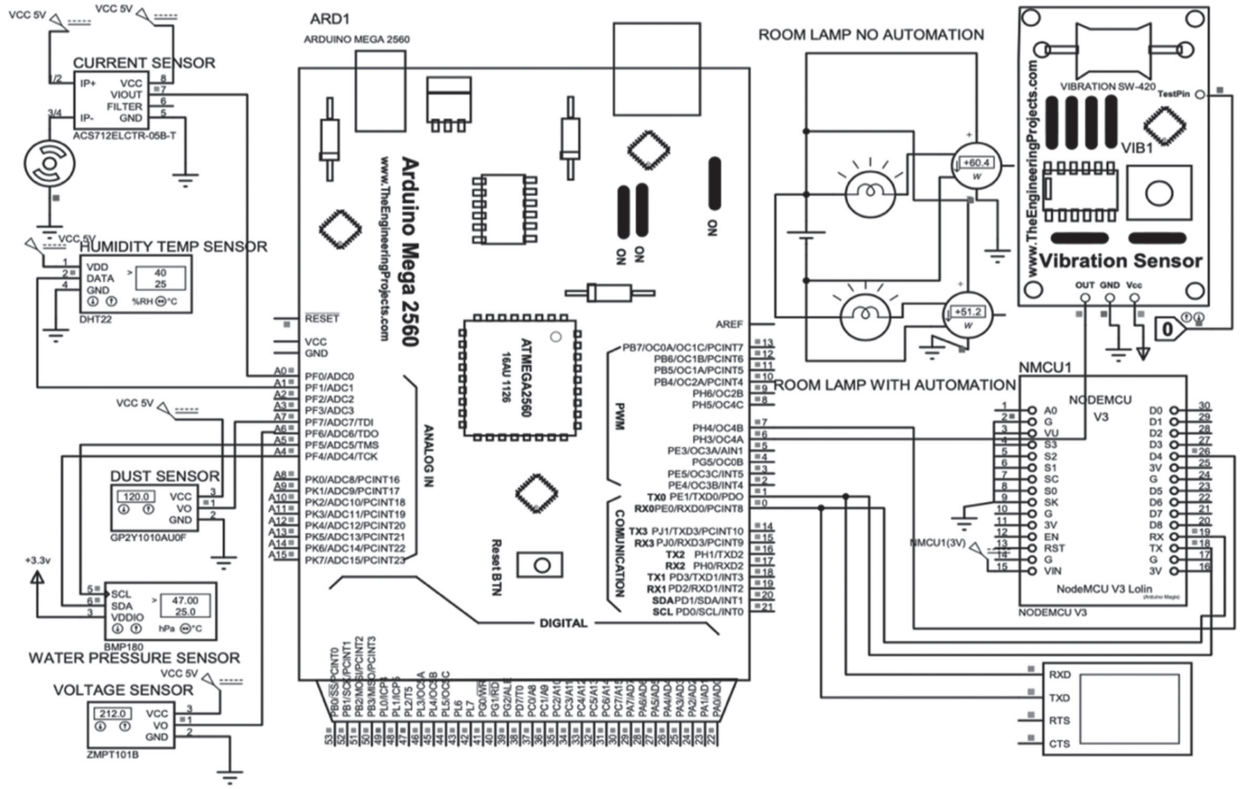


Fig. 10. Simulation module of the BMS monitoring system

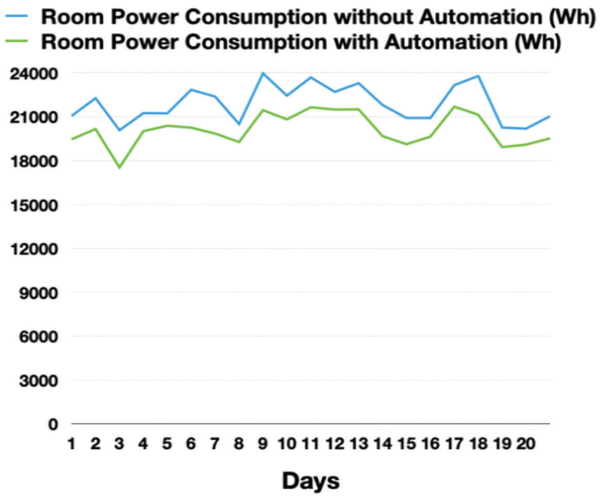
Based on the same measurement and to show the impact of energy automation in our system, we have also generated random power consumption values by using Python software for twenty days as shown in Table 3.The values are generated for two scenarios, i.e., with and without automation, according to a certain power consumption range for a room in a building. The power consumption values of a room without automation are randomized in the range of 20,000-24,000 Wh per day, which is generally a power consumption of a room in residential buildings taking into account lights, air conditioning, etc. On the other hand, the power consumption with automation is set in the range of 17,000-21,000 Wh per day, i.e., a reduction of 3,000 Wh per day. This value is chosen because we can save up to 3060 W in one day in a single room, if there are four 60 W bulbs and a 1200 W air conditioner which are used for 15 hours/day at a working capacity of 85 percent, without compromising the comfort of the room. However, other multiple factors can affect the expected reduced value such as the nature of the building, daily usage, seasons, number of appliances, etc. For example, the actual power consumption of day 9 is 23,964 Wh without automation. Now, if the user sets the limit to 20,000 Wh in the dashboard, the system will try to tune down the power as much as possible to reduce up to 3,964 Wh, but it is only able to curtail 2,516 Wh. This is due to the reason that the user also specified different working capacities, i.e., intensity, of different appliances according to the comfort level of the room and the system did not comprise on the comfort of the room. Table 3 represents the simulating results for the proposed BMS system indicating the amount of power consumed with and without automation along with energy-saving. The table clearly shows that the power consumption with automation is always less than the power consumption without automation and there is always energy savings every day ranging from 1091Wh (Min) to 2651Wh (Max). This energy-saving can be further enhanced/adjusted depending on the end-user requirement and room comfort level and it is completely automated by the system. Fig. 11. shows the graphical representation of the Table 3 values, i.e., power comparison with and without automation, and energy saved for every day by using automation.

Table 3. Simulation of the energy consumption with and without automation

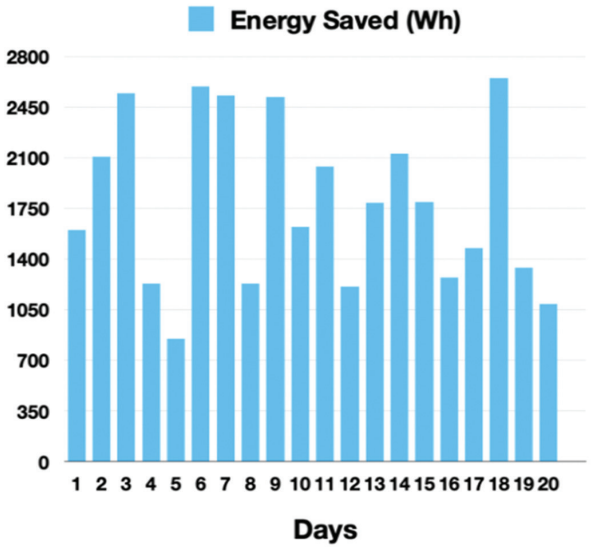
Days	Room Power Consumption Without Automation (Wh)	Room Power Consumption with Automation (Wh)	Energy Saved (Wh)
1	21058	19458	1600
2	22264	20156	2108
3	20077	17532	2545
4	21247	20016	1231
5	21228	20377	851
6	22843	20252	2591

7	22374	19846	2528
8	20498	19270	1228
9	23964	21448	2516
10	22438	20818	1620
11	23684	21646	2038
12	22700	21489	1211
13	23291	21501	1790
14	21796	19669	2127
15	20912	19119	1793
16	20909	19640	1269
17	23163	21689	1474
18	23782	21131	2651
19	20258	18919	1339
20	20184	19093	1091

Furthermore, the proposed system has a humidity control mechanism as shown in Fig. 12. that can monitor the relative humidity level in a room and maintain it in a predefined range, which is usually around 30-60%, as recommended by the United States Environmental Protection Agency [49]. Too much humidity, i.e., more than 60%, can cause mold, trap dirt, air pollution, etc. Similarly, too little humidity, i.e., less than 30%, can cause health issues like throat problems, eye irritation, etc. The proposed system measures and maintains the humidity level within the defined range, i.e., 30-60% [49]. In this way, the proposed system is capable of preventing structural damage and health problems by maintaining a balanced air quality inside buildings.



(a)



(b)

Fig. 11. (a) Power comparison with and without automation. (b) Energy saved by automation.

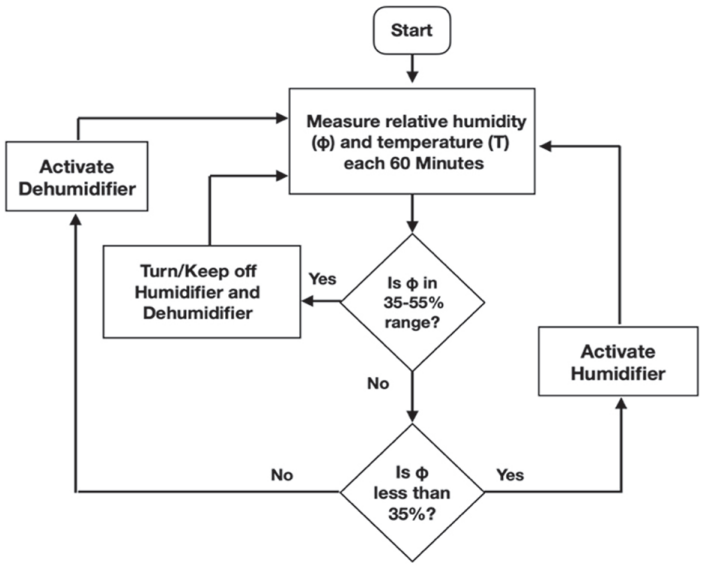


Fig. 12. Humidity control flowchart

AHT10 sensor is used to measure the temperature and relative humidity of the city of Medina, KSA. Table 4 shows the measured values for the last 24 hours, while Fig. 13. shows the graphical relationship between these two measured values. The proposed system maintains the RH value between 35-55% thresholds. Therefore, whenever the measurements go beyond these two limits as shown by red dotted lines, the system takes an automated action to keep the values within the defined range, i.e., activate humidifier if the relative humidity is less than 35%, and activate dehumidifier if the relative humidity is more than 55 %. The blue line represents the regression plot, which is based on the last 24 observations. It is used to predict the behavior of RH based on temperature, and it keeps on changing according to the new input, i.e., when measurement number 25 is recorded, the measurement number 1 is omitted to show the graph for only the last 24 observations.

Table 4. Sensors measurements (last 24 hours)

Temperature (°C)	Relative Humidity (%)	Temperature (°C)	Relative Humidity (%)
25	56	18	55
25	53	17	56
23	49	17	58
23	45	19	59
22	39	21	59
21	36	23	63
21	35	24	63
21	36	25	65
20	41	27	69
19	45	27	69
19	48	27	64
18	51	27	59

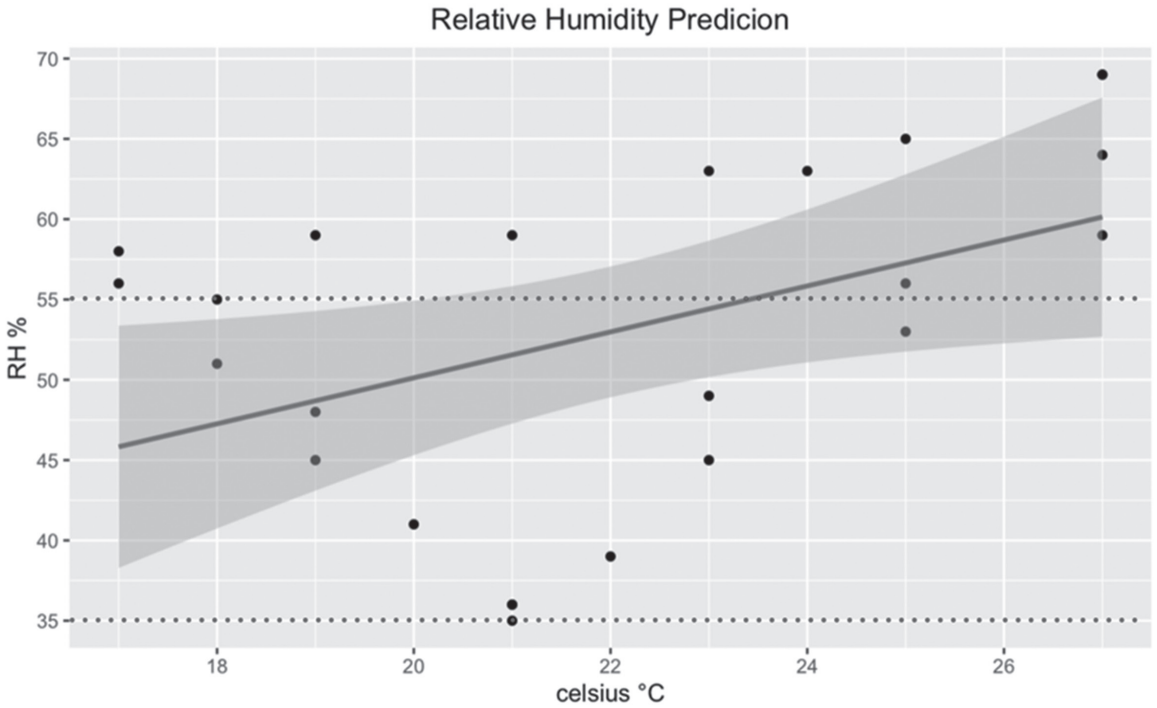


Fig. 13. Relationship between temperature & relative humidity

6. CONCLUSION

In residential/commercial buildings or large-scale facilities, many loads are installed in difficult locations other than the accessible loads and are very difficult to access for maintenance/change. Therefore, implementing BMS in buildings is very essential for controlling and monitoring all kinds of loads which can improve the overall efficiency of the facility and can also reduce operational cost, i.e., maintenance, etc. Existing IoT-based management systems generally provide only monitoring of different parameters, i.e., temperature, voltage, current, etc. The control/automation operations in these systems are generally very limited, i.e., manual

performancecontrol either by increasing or decreasing the amount of output for each load or switching these loads on/off, etc. Similarly, these systems do not provide a complete understanding of the operating loads' states and conditions, i.e., machine health, facility management options, etc.Our primary contribution in this work is that the proposed system does not only provide a better and more comprehensive way of monitoring and controlling different kinds of parameters in residential buildings, but also provides the ability to take decisions automatically. The proposed system also provides an advanced level of safety to different loads automatically.

The architecture of the proposed system uses five layers instead of three or four layers as used by other systems to comprehensively obtain the overall process of the system. After a comprehensive review of the existing literature, the paper presents how different parameters, i.e., current, voltage, temperature, humidity, power, vibration, pressure, etc., are monitored by using the proposed system with full detail. Then, the paper describes two different kinds of automation, i.e., safety and energy-saving automation, as compared to the most existing systems which provide only a limited automation solution. With these two automation actions, the proposed system is not only able to monitor and safely control the loads but also operates the existing loads in a highly efficient manner, i.e., energy-efficient, etc.

Simulations were carried out to validate the energy management aspect of the proposed system. Simulation results show that at least 90Wh energy can be saved per day without compromising any kind of comfort of the end-user when a 60 W light bulb is used with proposed automation for 10 hours a day. This saving can be further adjusted/enhancedby the end-user depending on the user's comfort level. Moreover, simulations were also carried out in two different ways, i.e., with and without automation, etc., to further validate the energy management capability of the proposed system in a room for twenty days. Simulation results show that the proposed system used with automation is capable of providing a minimum energy efficiency of 1091 Wh to a maximum of 2651 Wh per day while keeping the comfort of the room at standard level. These results clearly show that the proposed system is not only capable of monitoring and controlling, but also provides efficient energy saving with an advanced level of automationas compared to the existing solutions, which is another contribution of this work.

Another contribution of this work is that the proposed system is also able to track and count up to 540 units (easily expandable) in any facility, i.e., buildings, warehouses, factories, etc., which further enhances the efficiency of the proposed system. The proposed system is also capable of keeping humidity levels inside a room between 35-55% thresholds by automatically activating and deactivating the humidifier. The proposed system is also capable of providing the dust level in a room as compared to existing BMSs which do not have this feature. The proposed system also provides constant feedback about the loads' state and notifies the operator in case of any unexpected behavior. In this way, the proposed system is capable of operating the loads in the best possible state which does not only increases the life of the loads but also enhances the overall efficiency of the system along with energy-saving. The proposed system can be installed in any large-scale facility, which will reduce the operational cost, i.e., electricity, etc., and also the time and effort of human monitoring.

Funding Statement: The authors received no specific funding for this study.

7. REFERENCES

[1] N. Amani, A. A. R. Soroush, "Effective energy consumption parameters in residential buildings using building information modeling", Global Journal of Environmental Science & Management, Vol. 6, No. 4, 2020, pp. 467-480.

[2] E. Sirombo, M. Filippi, A. Catalano, A. Sica, "Building monitoring system in a large social housing intervention in Northern Italy", Energy Procedia, Vol. 140, No. 6, 2017, pp. 386-397.

[3] P. de Wilde, "The gap between predicted and measured energy performance of buildings: A framework for investigation", Automation in Construction, Vol. 41, No. 3, 2014, pp. 40-49.

[4] V. Marinakis, H. Doukas, C. Karakosta, J. Psarras, "An Integrated System for Buildings' Energy-Efficient Automation: Application in the Tertiary Sector", Applied Energy, Vol. 101, 2013, pp. 6-14.

[5] F. Terroso-Saenz, A. González-Vidal, A. P. Ramallo-González, A. F. Skarmeta, "An open IoT platform for the management and analysis of energy data", Future Generation Computer Systems, Vol. 92, No. 2, 2019, pp. 1066-1079.

[6] L. Vermorel, "Why are so many buildings filled with BMS ruins?", www.blog.wattsense.com (accessed: 2021)

[7] Wattsense Building Management Systems, "What is a BMS?", https://wattsense.com/bms/ (accessed: 2021)

[8] V. Sadhu, X. Zhao, D. Pompili, "Energy-Efficient Analog Sensing for Large-Scale and High-Density Persistent Wireless Monitoring", IEEE Internet of

Things Journal, Internet of Things Journal, Vol. 7, No. 8,2020, pp. 544-559.

[9] R. Casado-Vara, A. Martin-del Rey, S. Affes, J. Prieto, J. M, Corchado, "IoT network slicing on virtual layers of homogeneous data for improved algorithm operation in smart buildings", Future Generation Computer Systems, Vol. 102, 2020, pp. 965-977.

[10] C. H. Yang, K. C. Lee, S. E. Li, "A mixed activity-based costing and resource constraint optimal decision model for IoT-oriented intelligent building management system portfolios", Sustainable Cities and Society, Vol. 60, No. 3, 2020, pp. 655-671.

[11] J. Dongo, L. Foltete, C. Mahmoudi, F. Mourlin, "Distributed Edge Solution for IoT based Building Management System with NDN", Proceedings of the 11th Global Information Infrastructure and Networking Symposium, Paris, France, 23-26 June 2019, pp. 1-5.

[12] J. Yu, M. Bae, H. Bang, and S. Kim, "Cloud-based building management systems using short-term cooling load forecasting", Proceedings of the 9th IEEE Globecom Workshops, Atlanta, GA, USA, 9-13 December 2013, pp. 896-900.

[13] A. Ruiz, P. Martín, E. Salmerón, J. Guadix, J, "Building automation system with adaptive comfort in mixed mode buildings", Sustainable Cities and Society, Vol. 43, No. 2, 2018, pp. 77-85.

[14] M. Hannan, "A Review of Internet of Energy Based Building Energy Management Systems: Issues and Recommendations", IEEE Access, Vol. 6, No. 2, 2018, pp. 38997-39014.

[15] M. Villalon, J. Villalobos, E. Rivas, "IoT-Based System to Monitor and Control Household Lighting and Appliance Power Consumption and Water Demand", Proceedings of the 7th International Conference on Renewable Energy Research and Applications, Paris, France, 14-17 October 2018, pp. 785-790.

[16] A. S. Kulkarni, V. P. Kumar, T. S. Srujan, R. Rizwaan, S. Puneeth, "Implementation of industrial Iot for monitoring and controlling of temperature, pressure, and detection of volatile gas using Raspberry Pi 3B", Journal on Electronics Engineering, Vol. 8, No. 1, 2017, pp. 28-34.

[17] V. S. Perumal, K. Baskaran, S. K. Rai, "Implementation of effective and low-cost building monitoring system (Bms) using Raspberry Pi", Energy Procedia, Vol. 143, No. 3, 2017, pp. 179-185.

[18] A. Daissaoui, A. Boulmakoul, L. Karim, A. Lbath, "IoT and big data analytics for smart buildings: a survey", Procedia Computer Science, Vol. 170, No. 3, 2020, pp. 161-168.

[19] V. Marinakis, H. Doukas, "An advanced IoT-based system for intelligent energy management in buildings", Sensors, Vol. 18, No. 2, 2018, pp. 610-622.

[20] D. Misra, G. Das, D. Das, "An IoT based building health monitoring system supported by cloud", Journal of Reliable Intelligent Environments, Vol. 6, No. 4, 2020, pp. 141-152.

[21] L. Yu, B. Nazir, Y. Wang, "Intelligent power monitoring of building equipment based on internet of things technology", Computer Communications, Vol. 157, 2020, pp. 76-84.

[22] Z. Hu, "A data acquisition and control system in smart home based on the internet of things", International Journal of Simulation: Systems, Science & Technology, Vol. 17, No. 7, 2016, pp. 17.1-17.5.

[23] B. Artono, N. A. Hidayatullah, B. Winarno, "Smart panel system using internet of things", EAI Endorsed Transactions on Internet of Things, Vol. 5, No. 17, 2019, pp. 1-5.

[24] J. H. Huh, N. D. Park, "Decrepit building monitoring solution for zero energy building management using PLC and android application", Sustainability, Vol. 12, No. 5, 2020, pp. 561-574

[25] T. Sharmin, M. Gül, X. Li, V. Ganev, I. Nikolaidis, M. Al-Hussein, "Monitoring building energy consumption, thermal performance, and indoor air quality in a cold climate region", Sustainable Cities and Society, Vol. 13, 2014, pp. 57-68.

[26] M. A. Piette, S. K. Kinney, P. Haves, "Analysis of an information monitoring and diagnostic system to improve building operations", Energy and Buildings, Vol. 33, No. 8,2001, pp. 783-791.

[27] L. Wenjiang, D. Nanping, F. Tongshun, "Design of the embedded remote monitor system for building automation system based on the VxWorks",

Proceedings of the 2nd Asia-Pacific Conference on Computational Intelligence and Industrial Applications, Wuhan, China, 28-29 November 2009, pp. 436-438.

[28] M. H. Bola, E. N. Onwuka, S. Zubair, "An Efficient Energy Management in Buildings Using IoT - A Survey", Proceedings of the 15th International Conference on Electronics, Computer and Computation, Abuja, Nigeria, 10-12 December 2019, pp. 1-6.

[29] P. Brown, T. Loy, H. Pham, P. Sivabalan, "Automation and management control in dynamic environments: managing organisational flexibility and energy efficiency in service sectors", The British Accounting Review, Vol. 52, No. 2, 2019, pp.101-118.

[30] P. Nangtin, J. Nangtin, S. Vanichprapa, "Building automation system for energy saving using the simple PLC and VDO analytic", Proceedings of the1st International Workshop on Advanced Image Technology, Chiang Mai, Thailand, 7-9 January 2018, pp. 1-4.

[31] P. Sethi, S. R. Sarangi, "Internet of things: architectures, protocols, and applications", Journal of Electrical and Computer Engineering, Vol. 17, 2017, pp. 1-26.

[32] B. C. Kavitha, R. Vallikannu, "The internet of things model architectures for customized applications: A review", International Journal of Simulation: Systems, Science & Technology, Vol. 19, No.6, 2018, pp. 1-6.

[33] M. Burhan, R. A. Rehman, B. Khan, B. S. Kim "IoT Elements, Layered Architectures and Security Issues: A Comprehensive Survey", Sensors, Vol.18, No. 9, 2018, pp. 2796.

[34] W. Kassab, K. Darabkh "A-Z survey of internet of things: architectures, protocols, applications, recent advances, future directions and recommendations", Journal of Network and Computer Applications, Vol. 163, 2020, pp. 125-137.

[35] G. Codeluppi, A. Cilfone, L. Davoli, G. Ferrari, "LoraFarM: A LoRaWAN-based smart farming modular IoT architecture", Sensors, Vol. 20, No. 7, 2020, pp. 545-567.

[36] Y. M. Tukur, D. Thakker, I. U. Awan, "Ethereum blockchain-based solution to insider threats on perception layer of IoT systems", Proceedings of the 2nd IEEE Global Conference on Internet of Things, Dubai, UAE, 4-7 December 2019, pp. 1-6.

[37] N. Ye, Y. Zhu, R. C. Wang, R. Malekian, Q. M. Lin, "An efficient authentication and access control scheme for perception layer of internet of things", Applied mathematics & information sciences, Vol. 8, No. 4, 2014, pp. 1-8.

[38] K. Guleria, A. K. Verma, "Comprehensive review for energy efficient hierarchical routing protocols on wireless sensor networks", Wireless Networks, Vol. 25, No. 3, 2019, pp. 1159-1183.

[39] A. Mahmood, et al., "Scalability analysis of a LoRa network under imperfect orthogonality", IEEE Transactions on Industrial Informatics, Vol. 15, No. 3, 2019, pp. 1425-1436.

[40] A. Javed, H. Larijani, A. Wixted, "Improving energy consumption of a commercial building with IoT and machine learning", IT Professional, Vol. 20, No. 5, 2018, pp. 30-38.

[41] F. Adelantado, "Understanding the limits of LoRaWAN", IEEE Communications Magazine, Vol. 55, No. 9, 2017, pp. 34-40.

[42] M. Yang, J. Zhou, "A compact pattern diversity MIMO antenna with enhanced bandwidth and high-isolation characteristics for WLAN/5G/WiFi applications", Microwave and Optical Technology Letters, Vol. 62, No. 6, 2017, pp. 2353-2364.

[43] T. Janssen, N. BniLam, M. Aernouts, R. Berkvens, "LoRa 2.4 GHz communication link and range", Sensors, Vol. 20, No. 16, 2020, pp. 134-143.

[44] B. Xia, N. Qi, "Low-power 2.4GHz ZigBee transceiver with inductor-less radio-frequency front-end for Internet of things applications", IET Circuits, Devices & Systems, Vol.12, No. 2, 2020, pp. 209-214.

[45] M. Dahiya, "A short range wireless network: Bluetooth", International Journal of Advanced Research in Computer Science, Vol. 8, No. 3, 2017, pp. 37-41.

[46] S. Bandyopadhyay, M. Sengupta, S. Maiti, S. Dutta, "Role of middleware for internet of things: A study",

International Journal of Computer Science and Engineering Survey, Vol. 2, No. 3, 2011, pp.94-105.

[47] M. A. Chaqfeh, N. Mohamed, "Challenges in middleware solutions for the internet of things", Proceedings of the 8th International Conference on Collaboration Technologies and Systems, Denver, CO, USA, 21-25 May, pp. 21-26.

[48] C. Feng, M. Adnan, A. Ahmad, A. Ullah, H. U. Khan "Towards Energy-Efficient Framework for IoT Big Data Healthcare Solutions", Scientific Programming, Vol. 2020, pp. 1-9.

[49] United States Environmental Protection Agency, "Why and Where Mold Grow", www.epa.gov/mold/mold-course-chapter-2, (accessed: 2021)

Optimization Algorithms based compensation of mismatches in Time interleaved Analog to Digital Converters - A Review

Review Paper

Venkata naga chakravarthi Manepalli
Department of Electronics and Communication Engineering,
DR.Y.S.R. ANU College of Engineering and Technology,
Guntur, India-522510
chakrimvn@gmail.com

Chandramohan Bhuma
Department of Electronics and Communication Engineering, Bapatla Engineering College,
Bapatla, India-522101
chandrabhuma@gmail.com

Abstract *– Time interleaved analog to digital converters (TIADCs) play a significant role in signal processing wherever higher sampling rates are required. However, TIADCs suffer from various mismatches like sampling time, dc offset, gain, bandwidth etc. This results in generation of erroneous signal. Numerous methods were proposed for estimation of these mismatches and for correction of the erroneous signal. Optimization algorithms like genetic algorithm (GA), differential evolution (DE) algorithm were also used for estimation of mismatches. An overview of these algorithms and their performance comparison, with respect to various signal quality metrics like signal to noise ratio (SNR), signal to noise, distortion ratio (SNDR) etc, is given in this article.*

Keywords*: Time interleaved ADCs, mismatches, estimation, correction*

1. INTRODUCTION

Time interleaved analog to digital converters employ N parallel low rate ADCs which raise the overall sampling rate. In N-channel TIADCs a low rate channel of ADC operates with F_S/N sampling frequency. The block diagram of N-Channel TIADCs is shown in Fig.1. Each channel of TIADCs has a sampling time of NTs, and after reconstruction the overall system sampling time obtained is T_S. In the figure, $x(t)$ is the input applied to all ADCs and $y(nT)$ is the output after combining the outputs of low rate ADCs. However, the generated output is erroneous due to the mismatches existing in TIADCs. The most commonly occurring mismatches are gain, dc offset, sampling time and bandwidth mismatches. The mismatches among the channels' components result in generation of mismatch error in each channel. The mismatch error is dissimilar for different channels. The difference between the slope of ideal transfer characteristic and transfer characteristic with gain errors is known as gain mismatch. Similarly dc offset, bandwidth and timing mismatches are the differences in dc offsets, bandwidths and sampling times respectively.

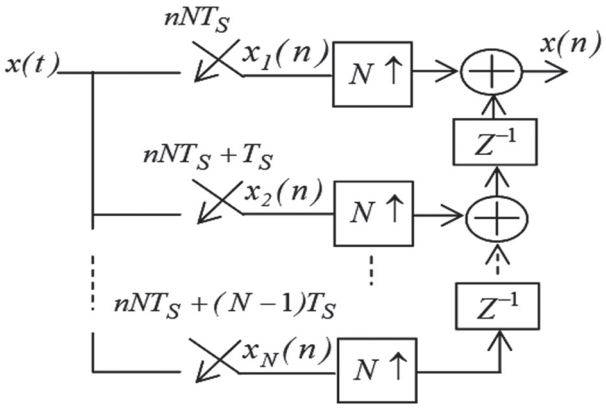


Fig. 1. N-Channel TIADCs

TIADCs with mismatches in sampling time, gain and dc offset is shown in Fig.2. Here T_i, Δg_i and Δx_i represent sampling time, gain and dc offset mismatches of i^{th} channel respectively. Various methods were suggested for reducing the effect of mismatches and to get compensated output. The output thus compensated has better SNDR (signal to noise and distortion ratio) and SFDR (spurious free dynamic range).

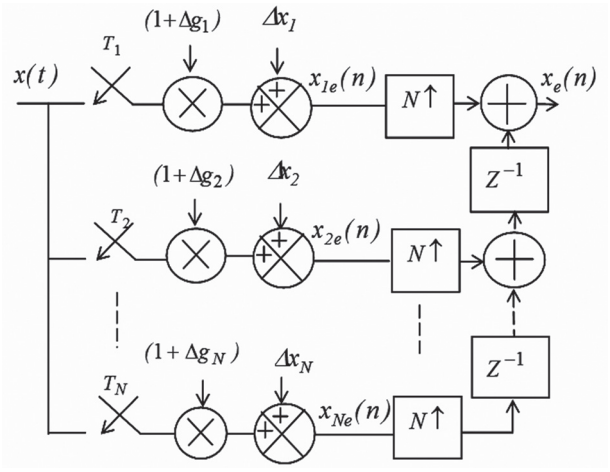


Fig. 2. TIADCs with mismatches

2. OPTIMIZATION ALGORITHMS USED FOR COMPENSATION OF MISMATCHES

Although there are numerous techniques existing for compensation of mismatches, the focus here is on optimization algorithms used for compensation. Optimization algorithms can be utilized to solve problems in fields like economics, sciences, engineering etc. Due to the growth of competition in all disciplines, the function of optimization has become still more prominent as the aim is to reduce the price of a product or wants to assign the resources wisely. The algorithms that are applied nowadays include genetic algorithm, differential evolution algorithm, ant colony optimization, particle swarm optimization, teacher-learner based optimization algorithm etc. Among of these algorithms, genetic and differential evolution algorithms were used for calibration of mismatches in TIADCs.

Differential evolution algorithm was applied for estimation of sampling time mismatch, dc offset and gain mismatches in 7 and 8-channel TIADCs in an OFDM system with 4-QPSK modulation. The method is evaluated by applying monotonic sinusoidal and image signals to the OFDM system [1]. Genetic algorithm based mismatch calibration was used for TIADCs with frequency dependent mismatches [2]. The sampling time mismatches were estimated by using modified differential evolution algorithm. An FPGA board was used to realize the algorithm and correction was applied in MATLAB [3].

A pilot signal was sent to train the neural network and then the sampled signals were directly calibrated with the trained network [4]. A differential evolution (DE) algorithm based estimation of sampling time mismatches was implemented [5]. Genetic Algorithm was used to estimate sampling time mismatches and fractional delay filters were used to correction [6]. A mean square algorithm was used to estimate the timing-error. The curve fitting algorithm was used to acquire the optimal sampling points and the best length of iterations [7]. Apart from these methods which used optimization algorithms, many other techniques were proposed which are summarized briefly below.

The first sub-ADC was chosen as a timing reference and first order Taylor approximation was used to estimate the errors in the samples due to the sampling time mismatches. Correction was applied by subtracting the error from erroneous samples [8]. Error estimation and calibration method using blind calibration technique was proposed to suppress the spurious components and restore the dynamic performance of sub-ADCs [9]. A digital background calibration method which makes use of Hadamard transform for calibrating gain, timing mismatches and dc offset mismatch cancellation by averaging was proposed [10]. Fast Fourier Transform Algorithm (FFT) based blind calibration algorithm was proposed for calibration of gain, offset and timing mismatches in a TIADC system with two channels [11]. Bandwidth mismatch optimization was performed using simulated annealing algorithm (SAA) [12]. The calibration of TIADCs and the challenges confronted by background calibration methods were enclosed in [13]. The channel mismatches were recalculated by using the calibration signal, when channel parameters change dynamically in TIADCs [14].

An autocorrelation function of each ADC's output was used to determine the timing skew mismatch between adjacent channels in TIADCs [15]. The analog transfer function of each channel was modeled including the sampling time and correction filters were designed for cancellation of distortions [16]. A technique which uses statistics of the input signal to continuously estimate and eliminate the conversion errors resulting from offset, gain, and timing mismatch was proposed [17].

A fully digital background algorithm was used to estimate and correct the timing mismatch errors between four interleaved channels [18]. The timing skew mismatches estimation methods for operating in the background, i.e., without interrupting the normal conversion, was reported [19]. A systematic overview of various calibration methods for timing skew in time interleaved analog-to-digital converters (TIADCs) was provided [20]. Architectures of 4-channel mixing-filtering-processing (MFP) digitizers were presented and evaluated [21].

Even though distinct techniques were followed in different optimization algorithms, almost all of them have a common configuration, given in Fig.3. The different operations carried out in the algorithms are described below.

Population initialization is the process of assigning a prescribed number (population) of random solutions to whatever parameter which is to be optimized. The initialized values are represented as a vector X_{ik}. Here the index *i* represents the i^{th} parameter and the *'k'* represents the k^{th} value in a population of *'p'*. Genetic operators like crossover, mutation etc are used to generate new solutions during reproduction process. The resulting vector after these operations is represented as X'_{ik}. The quality of the candidate solutions in the current population is determined by calculating the fitness values.

The fitness values are calculated using fitness function, which is a function of error signal. In the present case error signal *e(n)* is the difference between the output signal $x_p(n)$ from TIADCs without mismatches and the corrected signal $x_c(n)$ from TIADCs with mismatches. Based on the values of fitness function, the selection operator selects the candidate solutions which are to be retained in the next generation. The reproduction, fitness evaluation and selection operations run continuously in a loop, known as iterations, until the fitness of the solution is satisfactory and this solution gives the estimated values of the mismatch parameters.

Major optimization algorithms engage similar operations explained above and they only differ in operations carried out during reproduction process. The optimization algorithms which were suggested for compensation of mismatches in TIADCs and results obtained are enclosed in the following sections.

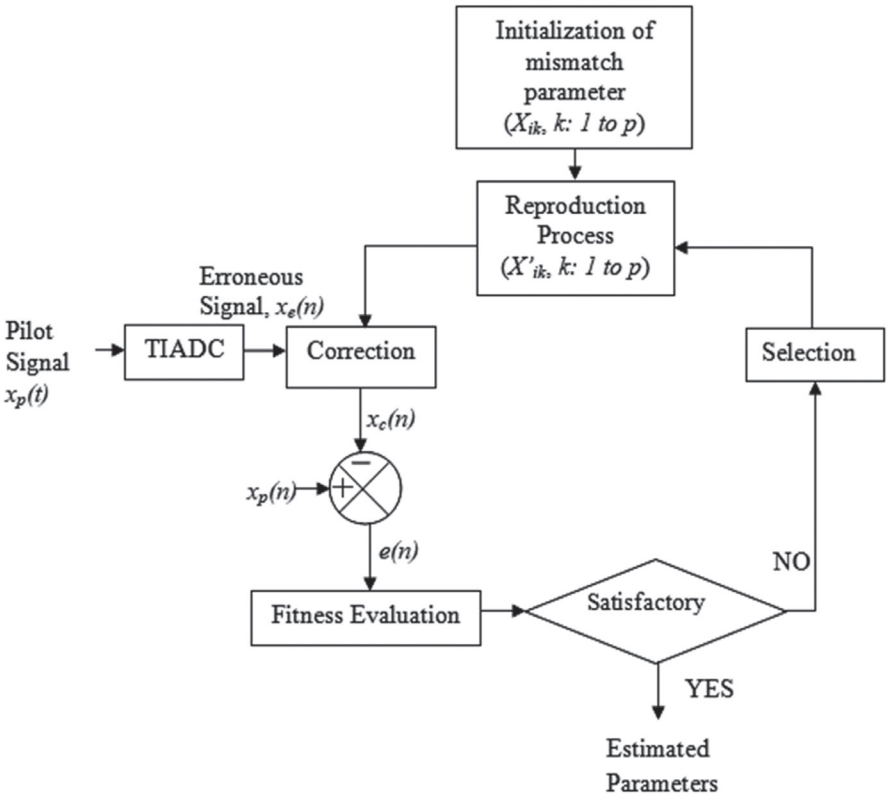


Fig. 3. Generalized block diagram for estimation of mismatches using optimization algorithms

2.1. DIFFERENTIAL EVOLUTION ALGORITHM

2.1.1. Algorithm

The differential evolution (DE) algorithm successfully estimates the mismatches with least number of iterations. As mentioned in the previous section that the algorithms differ in their operations carried out during reproduction process, DE algorithm executes two operations called as mutation and recombination, during reproduction process. During mutation operation, the difference of two random vectors with a weight of F is added to the third vector X_{ik}. The weight F of the difference is called as mutation factor. The resulting vector Y_{ik} is called as donor vector and is given by (1).

$$Y_{ik} = X_{ik} + F[X_{ik}(rand1) - X_{ik}(rand2)] \quad (1)$$

During recombination the trial vector X'_{ik} is obtained by recombining the elements of the target vector X_{ik}, and donor vector Y_{ik} as given in (2). Elements of the donor vector Y_{ik} are moved to trial vectorX'_{ik} with probability called as cross-over ratio (CR).

$$X'_{ik} = \begin{cases} Y_{ik} & if \quad rand(p) \leq CR \\ X_{ik} & if \quad rand(p) > CR \end{cases} \quad (2)$$

The selection operation is analogous to any other algorithm as explained in the previous section.

2.1.2. Estimation and correction of mismatches

The sampling time mismatches were estimated for 4-channel TIADCs using the proposed DE algorithm and was implemented in MATLAB [5]. The erroneous signal in each channel was corrected by delaying the signal with a delay equal to estimated sampling time mismatch. For delaying the signal, a fractional delay filter of 6 coefficients was used.

The suggested algorithm was assessed with various input signals like speech, sinusoidal and AM. The evaluation was carried out by calculating signal quality measuring parameters like SNR and SNDR. Table 1 shows the functional parameter values used the current work.

Table 1. DE algorithm parameters

S. No.	Parameter	Value
1.	Mutation Factor (F)	0.8
2.	Cross Over Ratio (CR)	0.4
3.	Population (p)	20
4.	Channels (N)	4

The initial population p of the parameter X_{ik} was taken as 20. A 6-tap fractional delay filter was used for correction of errors due to sampling time mismatches. The correction was implemented with TIADC input signals such as sinusoidal, speech, and amplitude modulated (AM) signals.

During the estimation of sampling time mismatches using DE, the convergence of the sampling time mismatches ($\Delta t_i'$) for a 4-channel TIADCs is shown in Fig. 4, where the actual mismatches introduced were (0.00, 0.02, 0.01, 0.015). The number of iterations (it_{max}) was the termination criterion which in this case was equal to 10.

The termination criterion can be either a predefined number of iterations or fitness function value. If the termination criterion is minimization of fitness function value, then the variation of number of iterations required for various fitness function values is shown Fig. 5.

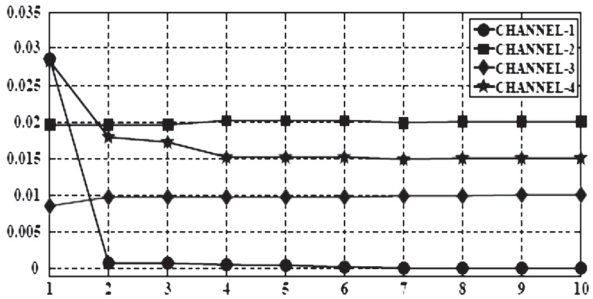


Fig. 4. Convergence of sampling time offsets during estimation by DE

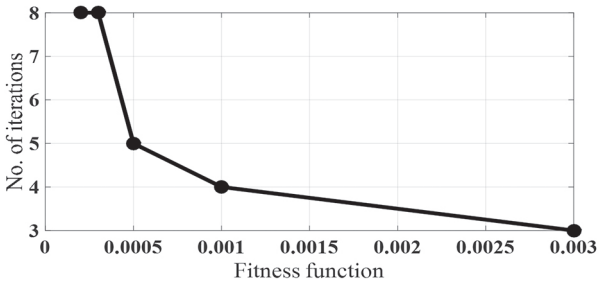


Fig. 5. Change in no. of iterations w.r.t fitness function

2.1.3. Estimation and correction in an OFDM receiver

The sampling time mismatch, gain mismatch and dc offsets were estimated for 7-channel TIADCs using the proposed DE algorithm. The algorithm was implemented in MATLAB. The DE algorithm specific parameters are listed in the following Table 2.

Table 2. DE algorithm parameters used for OFDM

Parameter	Range	Value Used
Mutation Factor, F	[0-2]	0.7
Cross Over Ratio, CR	[0-1]	0.5
Initial population, p	-	20

The termination criterion in this case is the number of iterations. The proposed algorithm used 50 iterations for estimation of mismatches. The number of iterations increases if the values of the mismatches to be estimated are very small.

The estimation was accurate and the estimated mismatches were very close to the introduced mismatches in the TIADCs. The convergence of the estimated dc offsets ($\Delta I_i'$), gain mismatches ($\Delta g_i'$) and sampling time mismatches ($\Delta t_i'$) towards their actual values (introduced) in 7- channel TIADCs are shown in Figures 6(a),(b)&(c) respectively. The mismatches in TIADCs were corrected in OFDM channel having awgn noise in the range 4-20dB. The SNDR for 7-channel TIADCs was improved by 42dB and 54dB for 6-bit and 8-bit precisions respectively for a sinusoidal input signal with a normalized frequency of 0.45.

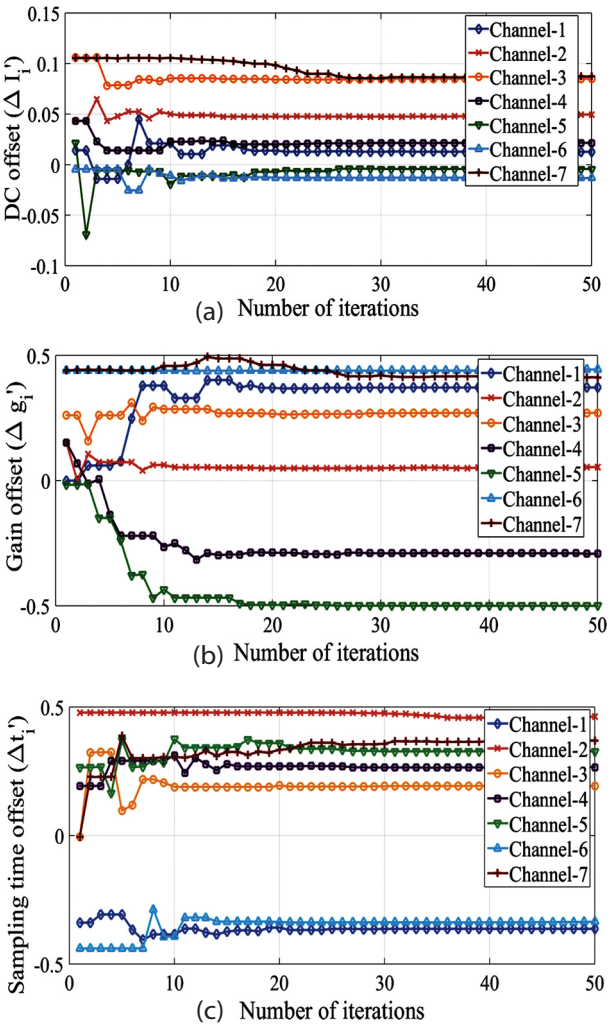


Fig. 6 (a),(b),(c). Convergence of dc, gain and sampling time offsets

2.1.4. DE algorithm implemented on FPGA

The DE algorithm implemented here was modified slightly and the differences are given below.

- Order in which the selection and reproduction operations were carried out.
- During reproduction only mutation was performed. There was no recombination operation.

A Field programmable gate array (FPGA) board was used for implementing the algorithm and correction was applied in MATLAB. The power utilization of FPGA kit and IO ports used for TIADCs with 4 and 2 channels are tabulated in Table 3.

Accurate estimation of sampling time mismatches was achieved using the algorithm. The algorithm employs a sinusoidal signal as a pilot signal. The correction was carried out by applying sinusoidal as well as speech input signals to TIADCs. SNR, SNDR, SFDR and PSNR were determined for evaluating the performance of the algorithm. A significant improvement was noticed in the parameters mentioned above. VHDL programming was used for implementing the algorithm for TIADCs with 2 and 4 channels. The algorithm was checked on NEXSYS-4 Double Data Rate (DDR) FPGA board. Fig. 7 shows the FPGA board used.

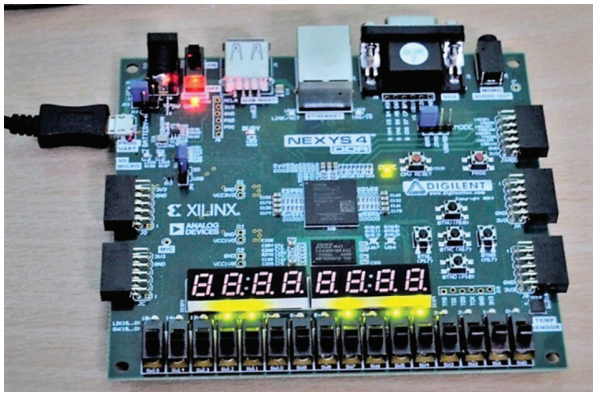


Fig. 7. NEXSYS-4 DDR FPGA board

In this technique, sinusoidal input signal was sampled and the samples were represented in 10 bit digital form. The coefficients of a 6-tap fractional delay filter, used for correction of sampling time mismatches, were generated in MATLAB. The synthesis results and summary of various parameters used during implementation are given in Table 3.

As the estimated sampling time mismatches $(\Delta t_i')$ were in 10-bit binary form, for TIADCs with 2 and 4 channels the number of output ports required was 20 and 40 respectively.

The signal at the input of each channel in TIADCs was delayed with the estimated sampling time mismatch of $\Delta t_i'$. The convergence of the sampling time mismatches $(\Delta ti')$ during their estimation for 4-channel TIADCs is shown in Fig. 8. The actual mismatches introduced were (0.01, 0.04,-0.04, 0.12).

Table 3. Implementation parameters and synthesis results

Parameter	Channels	
	N = 4	N = 2
Test Signal (Sinusoidal) Normalized Frequency (f)	0.03125	0.03125
ADC precision (No of bits)	10	10
Fractional Delay Filter Coefficients	6	6
I/O Ports used on FPGA Board	40	20
I/O Utilization	27%	13%
Total On-chip power	57mW	57mW

As the estimated sampling time mismatches $(\Delta t_i')$ were in 10-bit binary form, for TIADCs with 2 and 4 channels the number of output ports required was 20 and 40 respectively.

The signal at the input of each channel in TIADCs was delayed with the estimated sampling time mismatch of $\Delta t_i'$. The convergence of the sampling time mismatches $(\Delta t_i')$ during their estimation for 4-channel TIADCs is shown in Fig. 8. The actual mismatches introduced were (0.01, 0.04,-0.04, 0.12).

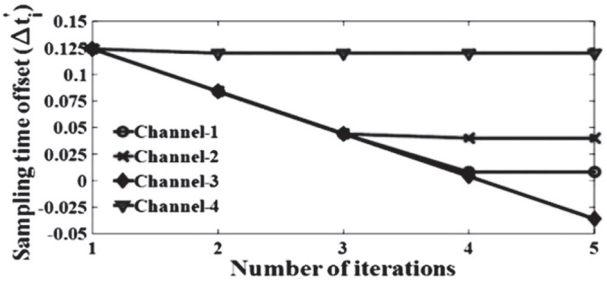


Fig. 8. Convergence of sampling time offsets

2.2. GENETIC ALGORITHM

2.2.1. Algorithm

The Genetic algorithm (GA) was also employed for estimation of sampling time mismatch and perfect reconstruction of the signal from its samples was accomplished. In GA the first step is initialization, where the mismatch parameters are initialized with a random set of values. The initialized values are known as chromosomes (Chr). The next step involves reproduction process, in which two operations called as crossover and mutation are performed. These two operations are described underneath.

Cross over: Any two initialized chromosomes (Chr) in binary form are selected randomly. After that, by choosing a cross over point (COP) the bits either before or after the cross over point are interchanged. The COP is chosen randomly and some times more than one COP are chosen. The cross over operation is illustrated in Table 4. The chromosomes (Chr) resulting from cross over are called as child chromosomes.

Table 4. Cross over operation

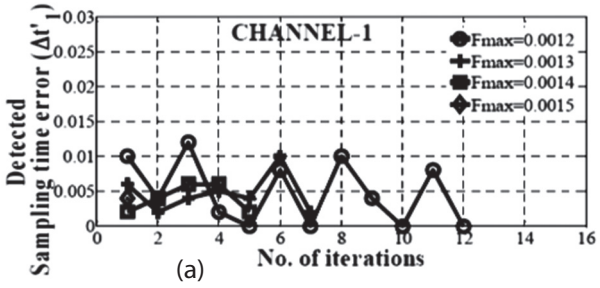
Chromosome	Binary							
25	0	0	0	1	1	0	0	1
54	0	0	1	1	0	1	1	0
						COP		

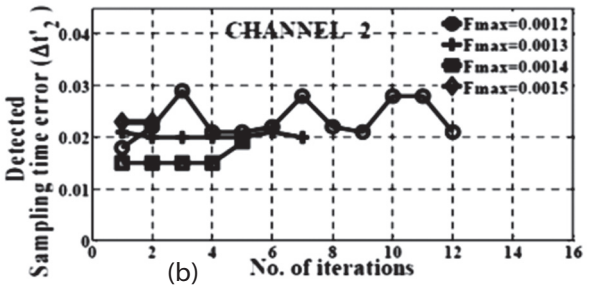
Mutation: Once the child chromosomes are obtained, one bit is chosen randomly in each child chromosome and it's mutated i.e., replace '0' with '1' and '1' with '0'.

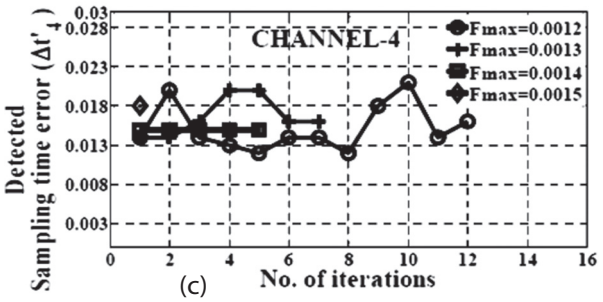
After performing these two operations the chromosome values are once again converted back to decimal values.

2.2.2. Estimation and correction using GA

GA based algorithms were used for calibration of sampling time mismatches in TIADCs with four channels [2, 6]. The pilot signal used was a single tone sinusoidal signal. The sampling time mismatches (Δt_i) in TIADCs were estimated with an accuracy of 3 decimal points. The correction was carried out with a fractional delay filter of 6 coefficients. An improvement in SNR was observed after correction. The convergence plots of estimated sampling time mismatches (Δt_i) for TIADCs four channels are shown in Fig. 9.(a),(b),(c),(d). Here the introduced mismatches are (0.000, 0.020, 0.010, 0.015) and the fitness evaluation parameter (F_{max}) are varying from 0.012 to 0.014.







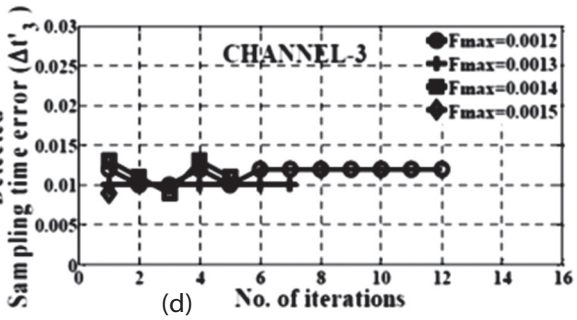


Fig.9. (a), (b), (c), (d) Convergence of sampling time offsets for 4-channel TIADCs

The number of iterations required for convergence with respect to fitness function values is shown in Fig.10.

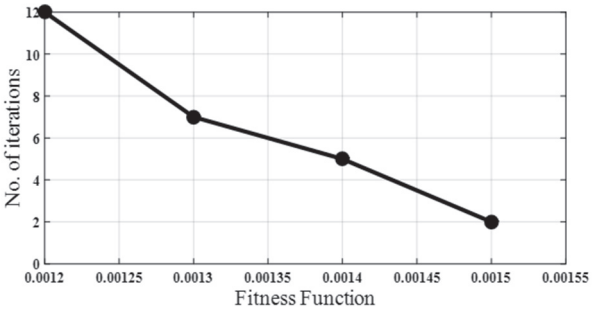


Fig.10. Variation in no. of iterations w.r.t fitness function

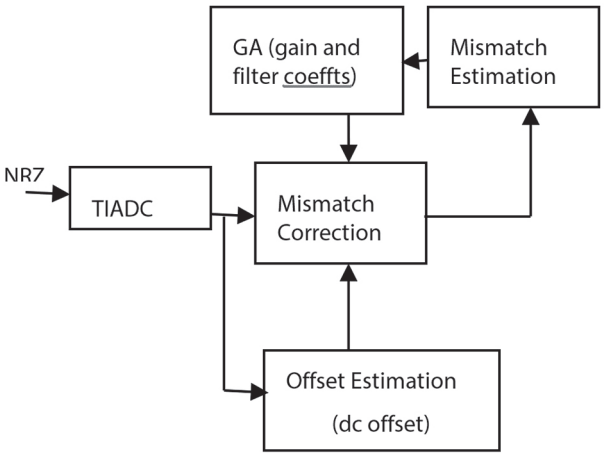


Fig.11. Estimation and correction

A foreground calibration mechanism employing GA together with a error estimation and correction method was proposed [2]. Joint estimation of static and frequency-dependent parameters was carried out by the employed detection technique comprising of GA. The proposed technique designed a notch filter to estimate the mismatches from the TIADC output. The presence of mismatches results in mismatch spurs in TIADC output. A notch filter was used to estimate the mismatches by eliminating the input frequency. Further, the power of the signal, with mismatch only, was pulled out and the total amount of mismatch was estimated. A mul-

tiple notch filter was designed with different cut-off frequencies. A foreground multitone signal, which was generated by a non-return-to-zero (NRZ) source, was applied as TIADC input. The above method corrects the offset and gain mismatches by direct multiplications and additions. Constant coefficient derivative filters were used to calibrate the frequency-dependent mismatches. The gain and filter polynomial coefficients were determined using GA and the channels' offset mismatches (dc offsets) were estimated by taking the mean of the output from each channel. The block diagram of this method is shown in Fig. 11.

3. PERFORMANCE COMPARISON OF THE ALGORITHMS

The optimization algorithms used for estimation and subsequent correction of mismatches in TIADCs are compared with respect to quality of the reconstructed signal. Signal quality metrics like signal to noise ratio (SNR), signal to noise and the distortion ratio (SNDR) and also number of iterations were used for comparison. Sinusoidal signal (Single tone) was applied as input signal to 4-Channel TIADCs and the reconstructed signal's quality was evaluated. Table 5 show cases the performance comparison of different algorithms.

Table 5. Performance comparison algorithms

	Algorithm				
		DE		GA	
Reference .No	[3]	[1]	[5]	[6]	[2]
No. of Channels (N)	4	7	4	4	4
Mismatch type	Timing	Dc , gain and timing	Timing	Timing	Timing, dc, gain and bandwidth
Filter Length	6	-	6	6	12
Normalized Frequency (f) range	0-0.45	0-0.45	0-0.45	0-0.4	0-0.5
Max.SNR (dB) (For single tone input)	67	-	63	71	65
Max. SNDR (dB) (For single tone input)	70	54	100	-	-
No. of iterations for convergence	5	30 (For 3 parameters)	4	12	100 (For 4 parameters)

4. CONCLUSION AND FUTURE SCOPE

Optimization algorithms applied in signal processing provides a scope for interdisciplinary research and application of those algorithms in solving diverse engineering problems. Optimization algorithms like DE and GA were used effectively for estimation of mismatches in TIADCs. Upon observation of the convergence plots, DE algorithm required less number of iterations and hence it is fast converging for the current problem sphere. However, as the algorithms were tested by simulation, the efficacy of these algorithms can be judged with more reliability by implementing them in hardware. The correction applied using the estimated mismatches provided a noteworthy improvement in signal quality in terms of SNR and SNDR. Moreover, since a wide range of optimization algorithms are available, the effectiveness of other algorithms, apart from the algorithms covered in the current review, can be tested.

5. REFERENCES

[1] M. V. N. Chakravarthi, B. Chandramohan, "Calibration of offsets in Time Interleaved ADCs in an OFDM receiver using differential evolution algorithm", Optoelectronics, Instrumentation and Data Processing, Vol. 57, No. 3, 2021, pp. 287-296.

[2] Y. A. Tavares, M. Lee, "A Foreground Calibration for M-Channel Time-Interleaved Analog-to-Digital Converters Based on Genetic Algorithm", IEEE Transactions on Circuits and Systems I: Regular Papers, Vol. 68, No. 4, 2021, pp. 1444-1457.

[3] M. V. N. Chakravarthi, B. Chandramohan, "Estimation and correction of sampling time offsets in TI-ADCs using optimization algorithm", International Journal of Innovative Technology and Exploring Engineering, Vol. 9, No. 4, 2020, pp. 1217-1222.

[4] Q. Yongtao, J. Zhou, Y. Liu, Y. Huangfu, "A Novel Calibration Method of Gain and Time-skew Mismatches for Time-interleaved ADCs Based on Neural Network", Proceedings of the IEEE MTT-S International Wireless Symposium, Guangzhou, China, 19-22 May 2019, pp. 1-3.

[5] M. V. N. Chakravarthi, B. Chandramohan. "Estimation of sampling time offsets in an N-channel time-interleaved ADC network using differential evolution algorithm and correction using fractional delay filters", Machine Intelligence and Signal Analysis, pp. 267-278. Springer, Singapore, 2019.

[6] M. V. N. Chakravarthi, B. Chandramohan, "Detection and Correction of Sampling-Time-Errors in an N-Channel Time-Interleaved ADC using Genetic Algorithm", Proceedings of the 14th IEEE India Council International Conference, Roorkee, India, 15-17 December 2017, pp. 1-6.

[7] W. Jun-Shan, S.-J. Liu, P.-P. Qi, "An optimization algorithm for timing-error estimation in time-interleaved ADCs", Proceedings of the IEEE 11th International Conference on Solid-State and Integrated Circuit Technology, Xi'an, China, 29 October - 1 November 2012, pp. 1-3.

[8] C. Shuai, L. Wang, H. Zhang, Rosanah Murugesu, Dustin Dunwell, Anthony Chan Carusone, "All-digital calibration of timing mismatch error in time-interleaved analog-to-digital converters", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 25, No. 9, 2017, pp. 2552-2560.

[9] G. Jian, P. Ye, H. Zeng, Jinpeng Song, Wentao Wei, Kuojun Yang, "An adaptive calibration technique of timing skew mismatch in time-interleaved analog-to-digital converters", Review of Scientific Instruments, Vol. 90, No. 2, 2019, p. 025102.

[10] T. Van-Thanh, V.-P. Hoang, V.-P. Pham, C.-K. Pham, "An Improved All-Digital Background Calibration Technique for Channel Mismatches in High Speed Time-Interleaved Analog-to-Digital Converters", Electronics, Vol. 9, No. 1, 2020, p. 73.

[11] B. Xu, H. Hu, W. Li, F. Liu, "Blind calibration method for two-channel time-interleaved analog-to-digital converters based on FFT", Journal of Electronic Testing, Vol. 34, No. 6, 2018, pp. 643-650.

[12] L. Jian, J. Li, S. Wu, N. Ning, Y. Liu, "A Bandwidth Mismatch Optimization Technique in Time-Interleaved Analog-to-Digital Converters", Journal of Circuits, Systems and Computers, Vol. 28, No. 6, 2019, p. 1950090.

[13] C. D. Kenneth, J. P. Keane, S. Lewis. "Calibration and dynamic matching in data converters: Part 2: Time-interleaved analog-to-digital converters and background-calibration challenges", IEEE Solid-State Circuits Magazine, Vol. 10, No. 3, 2018, pp. 61-70.

[14] L. Sujuan, N. Lv, H. Ma, A. Zhu, "Adaptive semiblind background calibration of timing mismatches in a two-channel time-interleaved analog-to-digital converter", Analog Integrated Circuits and Signal Processing, Vol. 90, No. 1, 2017, pp. 1-7.

[15] L. Xin, J. Wu, C. Vogel, "A Background Correlation-Based Timing Skew Estimation Method for Time-Interleaved ADCs", IEEE Access, Vol. 9, 2021, pp. 45730-45739.

[16] M. Pietro, F. Rosato, A. Trifiletti, "New models for the calibration of four-channel time-interleaved ADCs using filter banks", IEEE Transactions on Circuits and Systems II: Express Briefs, Vol. 65, No. 2, 2017, pp. 141-145.

[17] M. Hamidreza, M. Yargholi, M. Yavari, "Digital blind background calibration of imperfections in time-interleaved ADCs", IEEE Transactions on Circuits and Systems I: Regular Papers, Vol. 64, No. 6, 2017, pp. 1504-1514.

[18] A. Abbaszadeh, E. N. Aghdam, Alfredo Rosado-Muñoz, "Digital background calibration algorithm and its FPGA implementation for timing mismatch correction of time-interleaved ADC", Analog Integrated Circuits and Signal Processing, Vol. 99, No. 2, 2019, pp. 299-310.

[19] A. Sarkis, M. Ehsanian, "Generalized Method for Extraction of Offset, Gain, and Timing Skew Errors in Time-Interleaved ADCs", IEEE Transactions on Circuits and Systems II: Express Briefs, Vol. 67, No. 7, 2019, pp. 1214-1218.

[20] L. Xin, C. Huang, D. Ding, J. Wu, "A review on calibration methods of timing-skew in time-interleaved ADCs", Journal of Circuits, Systems and Computers, Vol. 29, No. 2, 2020, p. 2030002.

[21] L. Sujuan, H. Ma, N. Lyu, H. Wang, "Adaptive blind timing mismatch calibration with low power consumption in M-channel time-interleaved ADC", Circuits, Systems, and Signal Processing, Vol. 37, No. 11, 2018, pp. 4861-4879.

INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING SYSTEMS

Published by Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia.

About this Journal

The International Journal of Electrical and Computer Engineering Systems publishes original research in the form of full papers, case studies, reviews and surveys. It covers theory and application of electrical and computer engineering, synergy of computer systems and computational methods with electrical and electronic systems, as well as interdisciplinary research.

Topics of interest include, but are not limited to:

- Power systems
- Renewable electricity production
- Power electronics
- Electrical drives
- Industrial electronics
- Communication systems
- Advanced modulation techniques
- RFID devices and systems
- Signal and data processing
- Image processing
- Multimedia systems
- Microelectronics
- Instrumentation and measurement
- Control systems
- Robotics
- Modeling and simulation
- Modern computer architectures
- Computer networks
- Embedded systems
- High-performance computing
- Parallel and distributed computer systems
- Human-computer systems
- Intelligent systems
- Multi-agent and holonic systems
- Real-time systems
- Software engineering
- Internet and web applications and systems
- Applications of computer systems in engineering and related disciplines
- Mathematical models of engineering systems
- Engineering management
- Engineering education

Paper Submission

Authors are invited to submit original, unpublished research papers that are not being considered by another journal or any other publisher. Manuscripts must be submitted in doc, docx, rtf or pdf format, and limited to 30 one-column double-spaced pages. All figures and tables must be cited and placed in the body of the paper. Provide contact information of all authors and designate the corresponding author who should submit the manuscript to https://ijeces.ferit.hr. The corresponding author is responsible for ensuring that the article's publication has been approved by all coauthors and by the institutions of the authors if required. All enquiries concerning the publication of accepted papers should be sent to ijeces@ferit.hr.

The following information should be included in the submission:

- paper title;
- full name of each author;
- full institutional mailing addresses;
- e-mail addresses of each author;
- abstract (should be self-contained and not exceed 150 words). Introduction should have no subheadings;
- manuscript should contain one to five alphabetically ordered keywords;
- all abbreviations used in the manuscript should be explained by first appearance;
- all acknowledgments should be included at the end of the paper:
- authors are responsible for ensuring that the information in each reference is complete and accurate. All references must be numbered consecutively and citations of references in text should be identified using numbers in square brackets. All references should be cited within the text;
- each figure should be integrated in the text and cited in a consecutive order. Upon acceptance of the paper, each figure should be of high quality in one of the following formats: EPS, WMF, BMP and TIFF;
- corrected proofs must be returned to the publisher within 7 days of receipt.

Peer Review

All manuscripts are subject to peer review and must meet academic standards. Submissions will be first considered by an editor-in-chief and if not rejected right away, then they will be reviewed by anonymous reviewers. The submitting author will be asked to provide the names of 5 proposed reviewers including their e-mail addresses. The proposed reviewers should be in the research field of the manuscript. They should not be affiliated to the same institution of the manuscript author(s) and should not have had any collaboration with any of the authors during the last 3 years.

Author Benefits

The corresponding author will be provided with a .pdf file of the article or alternatively one hardcopy of the journal free of charge.

Units of Measurement

Units of measurement should be presented simply and concisely using System International (SI) units.

Bibliographic Information

Commenced in 2010.
ISSN: 1847-6996
e-ISSN: 1847-7003

Published: semiannually

Copyright

Subscription Information

The annual subscription rate is 50€ for individuals, 25€ for students and 150€ for libraries.

Postal Address

Faculty of Electrical Engineering,
Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia
Kneza Trpimira 2b
31000 Osijek, Croatia

IJECES Copyright Transfer Form

(Please, read this carefully)

This form is intended for all accepted material submitted to the IJECES journal and must accompany any such material before publication.

TITLE OF ARTICLE (hereinafter referred to as "the Work"):

COMPLETE LIST OF AUTHORS:

The undersigned hereby assigns to the IJECES all rights under copyright that may exist in and to the above Work, and any revised or expanded works submitted to the IJECES by the undersigned based on the Work. The undersigned hereby warrants that the Work is original and that he/she is the author of the complete Work and all incorporated parts of the Work. Otherwise he/she warrants that necessary permissions have been obtained for those parts of works originating from other authors or publishers.

Authors retain all proprietary rights in any process or procedure described in the Work. Authors may reproduce or authorize others to reproduce the Work or derivative works for the author's personal use or for company use, provided that the source and the IJECES copyright notice are indicated, the copies are not used in any way that implies IJECES endorsement of a product or service of any author, and the copies themselves are not offered for sale. In the case of a Work performed under a special government contract or grant, the IJECES recognizes that the government has royalty-free permission to reproduce all or portions of the Work, and to authorize others to do so, for official government purposes only, if the contract/grant so requires. For all uses not covered previously, authors must ask for permission from the IJECES to reproduce or authorize the reproduction of the Work or material extracted from the Work. Although authors are permitted to re-use all or portions of the Work in other works, this excludes granting third-party requests for reprinting, republishing, or other types of re-use. The IJECES must handle all such third-party requests. The IJECES distributes its publication by various means and media. It also abstracts and may translate its publications, and articles contained therein, for inclusion in various collections, databases and other publications. The IJECES publisher requires that the consent of the first-named author be sought as a condition to granting reprint or republication rights to others or for permitting use of a Work for promotion or marketing purposes. If you are employed and prepared the Work on a subject within the scope of your employment, the copyright in the Work belongs to your employer as a work-for-hire. In that case, the IJECES publisher assumes that when you sign this Form, you are authorized to do so by your employer and that your employer has consented to the transfer of copyright, to the representation and warranty of publication rights, and to all other terms and conditions of this Form. If such authorization and consent has not been given to you, an authorized representative of your employer should sign this Form as the Author.

Authors of IJECES journal articles and other material must ensure that their Work meets originality, authorship, author responsibilities and author misconduct requirements. It is the responsibility of the authors, not the IJECES publisher, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it.

- The undersigned represents that he/she has the authority to make and execute this assignment.
- For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.
- The undersigned agrees to indemnify and hold harmless the IJECES publisher from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.

_____ _____

Author/Authorized Agent **Date**

CONTACT
International Journal of Electrical and Computer Engineering Systems (IJECES)
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Josip Juraj Strossmayer University of Osijek
Kneza Trpimira 2b
31000 Osijek, Croatia
Phone: +38531224600,
Fax: +38531224605,
e-mail: ijeces@ferit.hr