



Ciudad Politécnica de la Innovación

WIICT 2016

Proceedings of the
Workshop on Innovation on Information
and Communication Technologies 2016

Editors:

Dr. Carlos Fernandez-Llatas
Dra. Maria Guillen

Committees

Organizing Committee

- Chair: Maria Guillem, Universitat Politècnica de València, Spain
- Alberto Bonastre, Universitat Politècnica de València, Spain
- Jose Manuel Catala, Universitat Politècnica de València, Spain
- Montserrat Robles, Universitat Politècnica de València, Spain
- Carlos Fernández-Llatas, Universitat Politècnica de València, Spain
- Jose Mariano Dahoui, Universitat Politècnica de València, Spain
- Isidora Navarro, Universitat Politècnica de València, Spain
- Manuel Traver Salcedo, Universitat Politècnica de València, Spain

Scientific Committee

- Chair: [Carlos Fernandez-Llatas](#), Universitat Politècnica de València, Spain
- Cenk Demiroglu, Ozyegin University, Turkey
- Johan Gustav Bellika, National Center of Telemedicine, Norway
- Yigzaw Kassaye Yitbarek, University of Tromso, Norway
- Raymundo Barrales, Universidad Autónoma Metropolitana de México
- [Frank Y. Li](#), University of Agder, Norway
- Antonio Martinez, Tecnologías para la Salud y el Bienestar S.A., Spain
- Jose Carlos Campelo, Universitat Politècnica de València, Spain
- Juan Vicente Capella, Universitat Politècnica de València, Spain
- Antonio Mocholí, Universitat Politècnica de València, Spain
- Sara Blanc, Universitat Politècnica de València, Spain
- Carlos Baraza, Universitat Politècnica de València, Spain
- Juan Carlos Ruiz, Universitat Politècnica de València, Spain
- Joaquin Gracia, Universitat Politècnica de València, Spain
- David De Andres, Universitat Politècnica de València, Spain
- Ricardo Mercado, Universitat Politècnica de València, Spain
- Miguel Rodrigo Bort, Universitat Politècnica de València, Spain
- Alejandro Liberos, Universitat Politècnica de València, Spain
- Jorge Pedrón, Universitat Politècnica de València, Spain
- Lenin Lemus, Universitat Politècnica de València, Spain
- Vicente Traver, Universitat Politècnica de València, Spain
- Pilar Sala, Universitat Politècnica de València, Spain
- Alvaro Martinez, Universitat Politècnica de València, Spain
- Jose Luis Bayo, Universitat Politècnica de València, Spain
- Juan Miguel García-Gomez, Universitat Politècnica de València, Spain
- Elies Fuster, Universitat Politècnica de València, Spain
- Carlos Saez, Universitat Politècnica de València, Spain
- Miguel Angel Rodriguez-Hernandez, Universitat Politècnica de València, Spain
- Canek Portillo Jimenez, Universitat Politècnica de València, Spain
- Ángel Perles, Universitat Politècnica de València, Spain
- Julian Camilo Romero-Chavarro, Universitat Politècnica de València, Spain
- Diego Boscá Tomás, Universitat Politècnica de València, Spain
- Luis Tello-Oquendo, Universitat Politècnica de València, Spain
- Israel Leyva-Mayorga, Universitat Politècnica de València, Spain
- Ángel Sanchis-Cano, Universitat Politècnica de València, Spain
- Luis José Saiz Adalid, Universitat Politècnica de València, Spain
- Gema Ibañez, Universitat Politècnica de València, Spain
- Pedro Yuste, Universitat Politècnica de València, Spain

Table of Contents

Article	Authors	Pages
A facial expression recognition tool to aid in depression assessment	Ernesto Cabrera Solis, Sabina Asensio-Cuesta, Rafael Prieto, Adrián Bresó	1-10
A Methodology to Guide IT Tools Development for Depression Assessment within a Specific Context	Sabina Asensio-Cuesta, Adrián Bresó, Rafael Prieto, Ernesto Cabrera Solis	11-23
Development of an acquisition electrophysiological signals system.	Ramón Albert, Andreu M. Climent, Miguel Rodrigo and María S. Guillem	24-31
Estimating the number of contending users for a single random access in LTE-A networks: The baseline for designing congestion control schemes at the evolved Node B *	Luis Tello-Oquendo, Israel Leyva-Mayorga, Vicent Pla, Jorge Martinez-Bauset and Vicente Casares-Giner	32-41
Drift clock analysis on distributed embedded systems for IoT applications	M. Navia, J.C. Campelo, A. Bonastre, R. Ors, J.V. Capella	42-49
Simulation of Telecom Traffic Sources in Hospitals	Yang LI, Angel Gomez-Sacristan, Miguel A. Rodriguez-Hernandez	50-59
A dynamic access class barring method to avoid congestion in M2M applications in LTE-A	Israel Leyva-Mayorga, Luis Tello-Oquendo, Vicent Pla, Jorge Martinez-Bauset, and Vicente Casares-Giner	60-69
Minimal Configuration of Body Surface Potential Mapping for Discrimination of Left and Right Atrial Drivers during Atrial Fibrillation	Miguel Rodrigo, Andreu M Climent, Alejandro Liberos, Francisco Fernández-Avilés, Felipe Atienza, María S Guillem and Omer Berenfeld	70-81
Towards Customizable Assisted Living with friendly environment using intelligent agents: A Survey	Sandra Lopez C., Rosario Baltazar F., Arnulfo Alanis, Lenin Lemus, Guillermo Méndez Z.	82-91
Process Choreography in Wireless Sensor and Actuator Networks: a proposal to achieve Network Virtualization	Sara Blanc, Jose Luis Bayo-Monton, Jose Carlos Campelo, Carlos Fernandez- Llatas	92-104
Assessment of Ad Hoc routing protocols for network deployments in disaster scenarios	Miquel Martinez, Yusheng Ji, Juan-Carlos Ruiz, and David de Andres	105-113

Optical Disdrometer Based on a Universal Frequency-to-Digital Converter	R. Barrales-Guadarrama, A. Mocholí-Salcedo, and F. Mocholí-Belenguer	114-123
Visual Interfaces for Clinical Archetype Editing	Christian Ponce, Jose Alberto Maldonado, Diego Bosca, Montserrat Robles	124-130
Multiscale technical systems (in particular, totality of cell telephones) reaffirms the existence of a fractal system of tripled periods	E. V. Kolombet, V. N. Lesnykh, O. Yu. Seraya, V. A. Kolombet, G. Verdú, V. Milian-Sanchez , A. Mocholí-Salcedo	131-140
Multi-core and Cluster-based Approaches to Enable Exhaustive Fault Injection on Highly Complex HDL Models	Ilya Tuzov, David de Andres, and Juan Carlos Ruiz	141- 150
A very fast Single Error Correction-Double Error Detection Code for short data words	J. Gracia-Morán, L.J. Saiz-Adalid, D. Gil-Tomás, J.C. Baraza, P.J. Gil-Vicente	151 - 159
High Power Monocycle Pulse Transmitter for Ultra Wideband Radar Sensor	Younes Ahajjam, José M. Catalá-Civera, Felipe Peñaranda and Abdellah Driouach	160 -167

A facial expression recognition tool to aid in depression assessment

Ernesto Cabrera Solis¹, Sabina Asensio-Cuesta², Rafael Prieto³, Adrián Bresó⁴

¹ Universidad Politécnica de Tulancingo. Calle Ingenierías # 100, Huapalcalco, 43629 Tulancingo, HGO, México

² Instituto Universitario de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas. Universitat Politècnica de València. Camino de Vera s/n. 46022 Valencia, España.

³ Departamento de Proyectos de Ingeniería. Universitat Politècnica de València. Camino de Vera s/n. 46022

⁴ Veratech for Health SL, Valencia, Spain

Abstract. The World Health Organization (WHO) estimates that Depression is a mental disorder that affects 350 million people worldwide. In this sense, it is grown the interest of developing tools to assist in depression assessment. This paper proposes a new automatic system to recognize emotions from facial gestures through images captured online during a depression assessment process. The facial expression recognition tool is based on the Microsoft Emotion API. The system generates indicators according with the facial emotions evaluation and relate then with a reference Depression Risk level. In the developed system the Depression Risk level is calculated applying the BDI-II questionnaire. The proposed tool is currently under validation process under the E-UNIHEALTH project.

1 Introduction

The human emotions can be expressed in many ways, such as facial expression. Usually people are able to determinate the emotion present in a human face. However, the problem arises when the objective is to determinate the emotion of a human within the voice, image or video through a computer system. One of the most studied cases in emotion recognition it is the usage of images of faces. For instance, Zhang [1] reports experiments for facial expression recognition using two types of features: geometric positions and Gabor wavelet. While Pantic et al. research [2] proposes an automatic system to recognize facial gestures in static images. Meanwhile, Kanade and Cohn [3] sustain that an automatic facial expression analysis could be applied in many areas such as emotion and paralinguistic communication, clinical psychology, psychiatry, neurology, pain assessment, lie detection, intelligent environments, and multimodal human computer interface.

The facial expression recognition may be considered as part of the pattern recognition that involves three principal stages: (1) image acquisition, (2) feature extraction

and (3) classification of the emotion. The Fig. 1 shows the main sequential stages in facial emotion recognition and the subtasks for each stage.

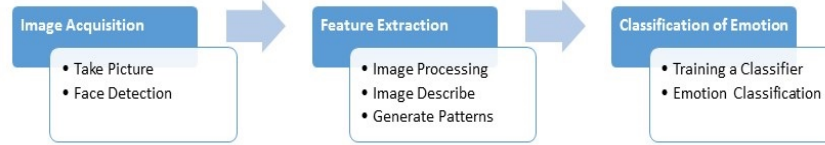


Fig. 1 Subtask for each of the three main stages involved in the facial expression recognition.

Regarding the facial expression recognition associated with depression illness, we found previous research studies. In several studies is reported a facial expressions recognition deficit associated with depressed people. Thus, Cognitive theories of depression suggest that perception is negatively biased in depressive disorder [4]. In studies [5] and [6] the authors studied the facial expression recognition in patients with depression; the authors of those works conclude that people with depression lose some ability to recognize facial expressions. In Stuhmann et al. and Christian G. Kohler et al. [7; 8] an exhaustive revision about the emotion recognition in depression is presented.

The proposal of this article is based on the analysis of emotions through images from a webcam capture to aid in the depression assessment. This proposal is described in the next section.

2 Materials and methods

This section is organized in two parts. First part is dedicated to describe the facial recognition system. The second part describes the procedure to obtain emotional indicators and how to relate them with a specific Depression Risk Level.

2.1 Procedure to obtain emotion from facial expression

The proposed system uses a webcam to take images of a person while performing an activity, such as answer a depression assessment questionnaire, a psychological interview, watch a video, etc. Then, images are sent to a server, which is responsible for processing them to determine the emotions presented in the face registered. After images server analysis, different emotions levels and a global emotional risk factor are obtained. The Fig. 2 shows the block diagram of this proposed system.

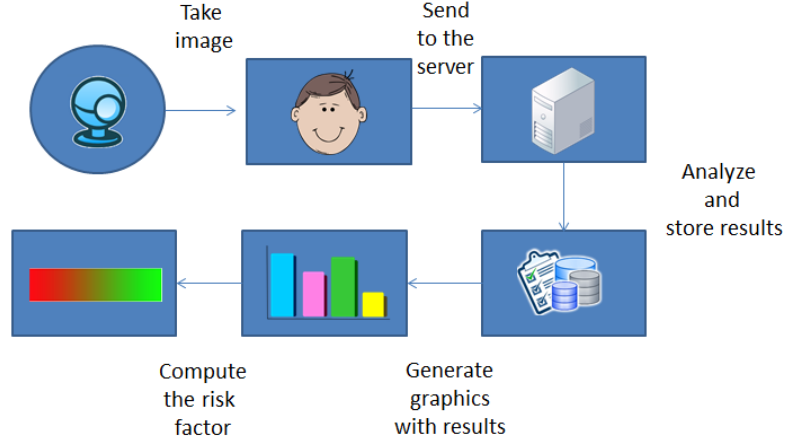


Fig. 2 Block diagram of the proposed system. An image taken from a webcam is the input, and an emotional risk factor is the output.

The first part of this procedure is based on the user images captures by a webcam. To manipulate the webcam is used the HTML5 API [6]. This API allows the user to manage the camera and microphone for recording video, taking pictures and recording audio. The API is supported by multiple platforms including mobile ones.

The captured images are sent to a server; the server saves the images, and sends an http request to Microsoft Emotion API [7]. This API takes a facial expression in an image as an input, and returns the confidence across a set of emotions for each face in the image.

Microsoft Emotion API [7] includes eight emotions, such as anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. Moreover, there are some Microsoft Emotion API conditions to take into account [8]:

- The supported input image formats includes JPEG, PNG, GIF (the first frame), BMP. Image file size should be no larger than 4MB.
- The detectable face size range is 36x36 to 4096x4096 pixels. Faces out of this range will not be detected.
- For each image, the maximum number of faces detected is 64 and the faces are ranked by face rectangle size in descending order. If no face is detected, an empty array will be returned.
- Some faces may not be detected due to technical challenges, e.g. very large face angles (head-pose), large occlusion. Frontal and near-frontal faces have the best results.
- The emotions contempt and disgust are experimental.

Regarding the Emotion API interpretation, the emotion detected should be understood as the emotion with the highest score. A unity-based normalization, feature scaling used to bring all values into the range [0, 1] is applied by the Emotion API.

The API response is in JavaScript Object Notation (JSON¹) format, which is syntax for storing and exchanging data and is an easier-to-use alternative to XML. The following Code 1 is an example of the API response in JSON format:

```
[
  {
    "faceRectangle": {
      "left": 68,
      "top": 97,
      "width": 64,
      "height": 97
    },
    "scores": {
      "anger": 0.00300731952,
      "contempt": 5.14648448E-08,
      "disgust": 9.180124E-06,
      "fear": 0.0001912825,
      "happiness": 0.9875571,
      "neutral": 0.0009861537,
      "sadness": 1.889955E-05,
      "surprise": 0.008229999
    }
  }
]
```

Code 1 JSON response example of 8 Emotional Values (EV) obtained from Microsoft Emotion API [8]

This set of values, collected during the user interaction with the system, may be plotted versus time, in order to help the user to understand the results.

2.2 Procedure to obtain emotional depression indicators

As has been mentioned in the previous section, the user is recorded during a stipulated period of time (D) to obtain a set of images (I) to be processed by Microsoft Emotion API. As a result, the system obtains emotional levels for each emotion in a given time (t). The Fig. 3 illustrates an example of this initial phase with three emotions (happy, sad, and surprise).

Secondly, user's depression risk level (DRL) need to be assess with a validated methods, such as BDI-II questioner. This depression level value is compared with the emotional indicators. This DRL value needs to be normalized within a range of]0, 1] and must be greater than 0.

Then, following emotional indicators are calculated:

¹ <http://www.json.org/>

- For each emotion (e):

- Average Emotion Value.

$$AVG_e = \frac{\sum_{i=1}^I EVe(i)}{I} \quad (1)$$

- Prevalent Average Emotion Value

$$\max_{i \in I} AVG_e(i) \quad (2)$$

- Peaks of emotion: maximum and minimum values.

$$\max_{i \in I} EVe(i) \quad (3) \quad \min_{i \in I} EVe(i) \quad (4)$$

- Emotion Variability value.

The total Emotion Variability of an emotion level (e) relatively to the duration of the observation is calculated by means of following equation.

$$EVariability(e) = \frac{\sum_{i=1}^I |EV(i)_{e+1} - EV(i)_e|}{D} \quad (5)$$

Similarly, a **Positive Emotion Variability** (+Variability) indicator is defined. In this case, emotion values variations are added only if EV (e+1) is higher than EV (e).

$$+EVariability(e) = \frac{\sum_{i=1}^I (EV(i)_{e+1} - EV(i)_e)}{D} \quad \text{if } EV(i)_{e+1} > EV(i) \quad (6)$$

Moreover, the **Negative Emotion Variability** (-Variability) indicator considers only transactions to lower emotions values.

$$-EVariability(e) = \frac{\sum_{i=1}^I (EV(i)_{e+1} - EV(i)_e)}{D} \quad \text{if } EV(i)_{e+1} < EV(i) \quad (7)$$

- Global values, for all emotions:

- Global Emotion Variability.

$$GlobalEVariability = \frac{\sum_{e=1}^E EVariability(e)}{E}$$

(8)

Similarly, Global Positive and Negative Variability are calculated.

Also, relative depression indicator values are obtained from the depression level risk DRL. Those relative indicators are calculated dividing each equation (1) to (8) by DRL value.

Finally, a global **Emotional Risk Factor** (ERF) is calculated applying equation (9):

$$ERF = \frac{\sum_{i=1}^I \sum_{e=1}^E Ce \cdot EV_e(i)}{I \cdot E} \quad (9)$$

Also, a **Depression Emotional Factor** (DEF) is defined in equation (10)

$$DEF = \frac{ERF}{DRL} \quad \text{where } DRL > 0 \quad (10)$$

Where (I) is the total number of captured images, (E) represents the number of emotions that can be recognized, for the Microsoft Emotion API is 8. Ce is a value depending on the emotion and should be setting by the clinic staff according with previous knowledge about depression and related emotions. An example of Ce parametrization is showed in Table 1, the more relation has an emotion with depression a higher value is assigned. $EV(i)e$ is the emotion value given by the Microsoft Emotion API for image i .

Table 1. Different values for the coefficient Ce according the emotion.

Emotion	Ce
anger	0.5
contempt	1
disgust	0.5
fear	0.5
happiness	0
neutral	0.25
sadness	1
surprise	0.25

The normalization of the Emotional Risk Factor (ERF) is required dividing it by the number of captured images (I) and emption (E). So that, ERF is within a range [0, 1].

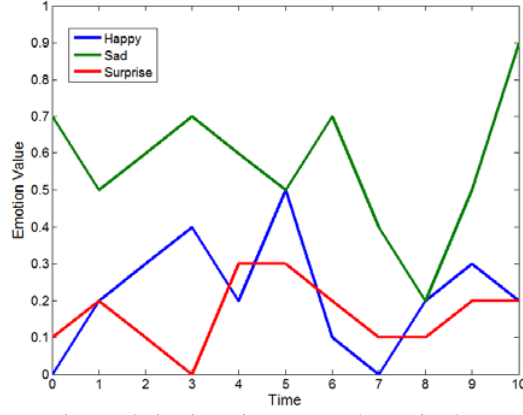


Fig. 3 Report of the emotion evolution in patient test. Each emotion has a value in a given time.

2.1.1 Beta tool and Results

A Beta tool has developed to be integrated in the E-UNIHEALTH Portal framework [9]. This first version of the tool includes an interface organized in three main areas as is showed in Fig. 4: (1) control buttons, (2) camera area, and (3) plot area. Actions: “Start Camera” to turn on the webcam, “Take Picture” begins the capture the images for analysis them, “Stop Camera” to turn off the webcam. These actions can be executed manually or automatically. In automatic case, we set a frequency to capture the emotions of the user during the development of the activity that you propose (as perform an evaluation questionnaire) Also the graphic report including the global emotional risk factor and the emotions levels registered in time scale.

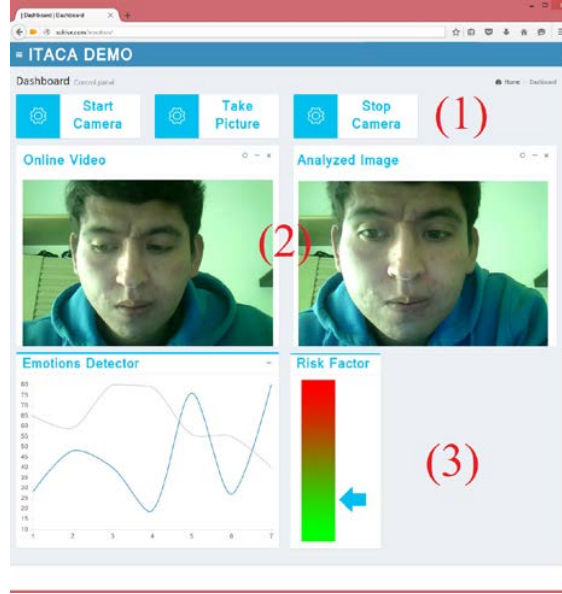





Fig. 4 Pilot System user interface, opened in the Firefox browser, divided in three principal parts: (1) three buttons on the top to manage the Pilot System. One to start the camera, one to take a picture and one to stop the camera. (2) A camera area divided into webcam online video (left) and analyzed picture (right). (3) And the graphical area, which shows the emotion values and the computed emotional risk factor.

The camera area is divided in two parts: one for the video and one for the picture. At the moment of pushing the “Start Camera” button the video captured with the webcam is painted on this place. The report graphical area shows the values of the recognized emotions, when the user pushed “Stop Camera” button. The inferred emotional risk factor is shown with the plot, as a summary of the interaction. This factor represents the user’s level of risk for depression, based on the detected emotions during the interaction.

The Pilot System has been integrated in the E-UNIHEALTH Portal [10]. E-UNIHEALTH is a Web-Based Support System to prevent depression in the University Community through emotional virtual agents.

To test the Microsoft Emotion API were used a set of images taken of The Japanese Female Facial Expression (JAFFE) Database [11], and with some random images, some results achieved are show in the table 2.

Table 2. Some images proven with the Microsoft Emotion API and its respective values. The validation of the proposed system and the emotional risk factor is being performed combined with BDI-II questionnaire in a controlled group of people. Since BDI-II questionnaire is one of the most commonly instruments used to measure the severity of depression.

Image	Emotions Value
	"anger": 4.37488934e-12, "contempt": 2.34036987e-12, "disgust": 3.56840842e-11, "fear": 1.52212943e-12, "happiness": 1, "neutral": 3.3472658e-10, "sadness": 1.46328311e-10, "surprise": 1.870819e-9
	"anger": 0.000568888441, "contempt": 0.00767677976, "disgust": 0.001665115, "fear": 0.0121105909, "happiness": 0.0000225891345, "neutral": 0.137544319, "sadness": 0.8303469, "surprise": 0.0100648524
	"anger": 0.01703057, "contempt": 0.0000141991322, "disgust": 0.000676677329, "fear": 0.0007115363, "happiness": 0.000335023651, "neutral": 0.000461025425, "sadness": 0.00000516357932, "surprise": 0.9807658

4 Conclusions

The proposed automatic system to recognize emotions facial gestures could be a helpful tool to be integrated and complement the actual depression assessment validated methods. In addition to identifying the user emotions generated during the interaction, is able to infer a risk factor that can provide valuable information for the decision support systems focused on depression.

The proposed system is currently being validated under E-UNIHEALTH Project research focused on the Universitat Politècnica de València context. But the proposed system could be integrated with others web systems as an independent module. Moreover, the system applications are diverse such as marketing or human-machine interaction areas. The risk factor function could be redefined to suit the new desired purpose.

The results obtained in the Pilot System are quite efficient in the sense of emotion recognition and calculating the emotional risk factor, working almost in real time. Further research will confirm the system utility in the depression assessment field.

References

1. Zhang, Z. Feature-based facial expression recognition: Sensitive analysis and experiments with a multilayer perceptron. *International Journal of Pattern Recognition and Artificial Intelligence*, 13(6) (1990) 893-911
2. Pantic, M., Rothkrantz, L. J. M.: Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on systems, man, and cybernetics-part b: cybernetics*, 34(3) (2004)
3. Tian, Y., Kanade, T., Cohn, J.F.: Facial expression recognition. In: *Handbook of face recognition*. Springer, (2011), 487–519
4. Münkler, P., Rothkirch M., Dalatim Y., Schmackm K., Sterzerm P.: Biased Recognition of Facial Affect in Patients with Major Depressive Disorder Reflects Clinical State. *PLoS ONE* 10(6): e0129863. doi:10.1371/journal.pone.0129863 (2015).
5. Mikhailova, E. S., Vladimirova, T. V., Iznak, A. F., Tsu-sulkovskaya, E. J., Sushko, N.V.: Abnormal recognition of facial expression of emotions in depressed patients with major depression disorder and schizotypal personality disorder. *Biological Psychiatry*. 40(8) (1996) 697–705
6. Rubinow, D. R., Post, R.M.: Impaired recognition of affect in facial expression in depressed patients. *Biological Psychiatry*. 31(9) (1992) 947–953
7. Stuhmann et al.: Facial emotion processing in major depression: a systematic review of neuroimaging findings. *Biology of Mood & Anxiety Disorders* (2011) 1:10.
8. Christian G. Kohler et al.: Facial emotion perception in depression and bipolar disorder: A quantitative review. *Psychiatry Research* 188 (2011) 303–309
9. The HTML 5 JavaScript API Index. Retrieved from: <http://html5index.org/index.html>
10. Microsoft Cognitive Services. Retrieved from: <https://www.microsoft.com/cognitive-services/en-us/emotion-api>
11. Microsoft Emotion API. Retrieved from: <https://dev.projectoxford.ai/docs/services/5639d931ca73072154c1ce89/operations/563b31ea778daf121cc3a5fa>
12. Beck, A. T., Steer, R. A., y Brown, G. K. Manual. BDI-II. Inventario de Depresión de Beck-II (Adaptación española: Sanz, J., y Vázquez, C.). Madrid: Pearson Educación. (2011).
13. Asensio-Cuesta, S., Bresó, A., Miralles, C., García, J. M., Robles, M.: An eHealth Portal to prevent depression in the University Community through emotional virtual agents. *WIICT2015* (2015) 9-16
14. Valstar, Michel, et al. AVEC 2013: the continuous audio/visual emotion and depression recognition challenge. En *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*. ACM, 2013. p. 3-10.
15. Bettadapura, Vinay. Face expression recognition and analysis: the state of the art. arXiv preprint arXiv:1203.6722, 2012.

A Methodology to Guide IT Tools Development for Depression Assessment within a Specific Context

Sabina Asensio-Cuesta¹, Adrián Bresó², Rafael Prieto³, Ernesto Cabrera Solís⁴

¹Instituto Universitario de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas. Universitat Politècnica de València. Camino de Vera s/n. 46022 Valencia, España.

²Veratech for Health SL, Valencia, Spain

³Departamento de Proyectos de Ingeniería. Universitat Politècnica de València. Camino de Vera s/n. 46022

⁴Universidad Politécnica de Tulancingo. Calle Ingenierías # 100, Huapalcalco, 43629 Tulancingo, HGO, México

Abstract. The approach of most popular assessment methods for depression, often questionnaires, is general; they are standard methods not adapted to a specified user population or context. In the other hand, Informatics Technologies are more and more applied to implement those standard methods. However, while the most popular questionnaires have been validated during decades, the new technology applied to implement them not always is properly tested. Thus, IT needs to be validated as an adequate mean for implementing depression assessment methods in a context and for its user's profiles. Therefore, we detect this lacking and we present a new Methodology to guide IT tools developments for depression assessment within a specific context. The Methodology is applied in a case of study that involves Teaching and Research Staff (PDI) in the Universitat Politècnica de València (UPV) context.

1 Introduction

The prevalence of depression in Spain is usually quite high. In a study of more than 5,000 people, the ESEMeD project found that major depressive episode is the most common mental disorder in Spain, with a 12-month prevalence of 3.9% and a lifetime prevalence of 10.5% [1]. The diagnosis of major depressive episode proposed by the DSM-IV, explain of depression to a severe degree of episodic disorder. However, we know that there may be more moderate and even lower levels of depression, which, without becoming an acute clinical picture, may also have effects on mood and work performance. Thus, for example, from the psychology of organizations and work, particularly with the study of mobbing, is often explore some behaviors associated with depression and apathy or lack of encouragement and crying or wanting to mourn, among others [2], without being associated necessarily a diagnosis of major depressive episode. This perspective points to a preventive purpose of working therapeutically depressive symptoms before they increase in intensity and cause greater discomfort.

The solutions developed and evaluated in a specific context may not be valid for other environment conditions. If the solution is too general, it can lead to errors, biases, or exclusions. In general, the more defined and bounded has been the approach, more efficient and valid the results.

In mental health, specifically in Major Depression Disorder (MMD) or depression, the use of medical questionnaires for assessing depression is the most common technique to estimate the user's status and monitoring the evolution of this pathology. Most of these questionnaires [3 - 6] are applied in any environment, regardless of the particularity of the targeted population. These questionnaires are based in the symptoms identified in the most accepted diagnostic guides, published by the American Psychiatric Association (APA¹) and by the World Health Organization (WHO²). The last published standard classification of mental disorders from APA, is the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5 [7]). Regarding WHO, the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) is the most updated. Both criteria for diagnosing depressive episodes are based on symptoms (see Table 1). ICD-10 and DSM-5 overlap considerably but have some differences of emphasis. In ICD-10 identified ten symptoms, and the patient must have two from the core, plus at least two of the remaining seven symptoms. However, DSM-5 defined nine symptoms, and the patient must have at least one symptom from the core and five or more from the non-core symptoms. Both diagnostic systems require symptoms to have been present for at least two weeks to make a diagnosis.

Table 1. Comparison for depression diagnostic symptoms of the current two major classification systems (ICD-10 and DSM-5). * Core symptoms.

Num.	ICD-10	DSM-5
1	Depressed mood*	Depressed mood by self-report or observation made by others *
2	Loss of interest *	Loss of interest or pleasure *
3	Reduction in energy *	Fatigue/loss of energy
4	Recurrent thoughts of death or suicide	Recurrent thoughts of death, suicidal thoughts or actual suicide attempts
5	Diminished ability to think/concentrate or indecisiveness	Diminished ability to think/concentrate or indecisiveness
6	Change in psychomotor activity with agitation or retardation	Psychomotor agitation or retardation
7	Sleep disturbance	Insomnia/hypersomnia
8	Change in appetite with weight change	Significant appetite and/or weight loss
9	Loss of confidence or self-esteem	Worthlessness/excessive or inappropriate guilt
10	Unreasonable feelings of self-reproach or inappropriate guilt	-

¹ <https://www.psychiatry.org/>

² <http://www.who.int/en/>

Most of the questionnaires for assessing depression correlate highly in clinical diagnosis of depression [8 - 10]. However, some studies have reported a lack adaptability of these questionnaires to the targeted population [11 - 12]. Questionnaires do not take into account the particularities of the context and individuals, which are evaluated. This is a problem because each individual may have different alterations over the same symptom. Culture, gender, context, background, race, age or ethnicity may be report differences in the diagnostic symptoms [13 - 14]. For example, women with depression are more likely to have feelings of sadness, worthlessness, and excessive guilt. However, NIMH³ asserts that men are more likely to be very tired, irritable, lose interest in once-pleasurable activities, and have difficulty sleeping.

The reported findings in literature, require a rigorous analysis for the personalization of these self-reported tools to the population-targeted domain, in order to become more reliable and accurate. We propose a Methodology in order to develop and validate a new screening tool for the assessment of the depression, adapted for the targeted people. In addition, our Methodology identifies and highlights the most relevant risk factors related with depression, for the specific studied domain.

In the rest of the paper, we review published findings related with the prevalence of depression for different studied contexts, in order to strengthen the argument that environmental and individual factors, affect the clinical status of the user.. In Section 3 the relevance of the context is analyzed from the depression point of view. In Section 4, the proposed Methodology is presented and applied in a case of study. Finally, conclusion and future work are given in Section 5.

2 Depression assessment background

MMD is the leading mental disorders that generate disability and shorten life. Is expected that will be the second in the disease global burden ranking in 2020 [15], and the first in 2030 [16]. MMD is a serious public health concern in worldwide, which affects more than 350 million people [16]. In general, MMD is characterized by sadness, loss of interest or pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, feelings of tiredness, and poor concentration. However, there are differences in the depression rates attending to social, demographic, and economic criteria such as gender, age, marital status, residence, ethnicity, educational level, race, ethnicity, income, or profession. The values of these criteria may alter the risk factors of depression. So is very important take into account these criteria in order to obtain adjusted and realistic assessments. In general, low Socioeconomic Status (SES) is generally associated with high psychiatric morbidity, more disability, and poorer access to health care [17 - 20]. Specifically, Subjective Social status (SSS) -an individual's perception of her socioeconomic standing- is a robust predictor of physical health in many societies [21 - 22]. We can found many studies about the association between these criteria and depression. Low educational level, white women, young adults, and those not currently married are criteria linked to higher rates of suffer depressive syn-

³ <http://www.nimh.nih.gov/>

drome [23]. Employment status and the work motivation are fundamentals to mental health recovery [24 - 25]. Instead, Mihai et al. [26] published that no exist statistical differences regarding depression rate in employed and unemployed persons in a concrete scenario. Bautista et al. [27] correlated the prevalence of depressive symptoms in the Mexican population (aged 12 to 65 years) with socio-demographic and personal factors. Significant differences were obtained. For example, depression prevalence was higher in widowed men, divorced women, people with lower education, residents in unsafe places, and for women in central region of Mexico. All of this information could be taking account in order to provide personalized prevention programs, such as adapted for region-specific in women. Similarity, gender differences in 23 European countries were analyzed in [28]. In which, marriage (or cohabiting with a partner) and good socioeconomic position were identified as the strongest association with depression, decreasing the prevalence in both genders. Additionally, gender differences in depression were detected higher in Eastern and Southern European countries. In MMD, the screening tools and rating scales can be used to help measure a user's mental health symptoms, and/or measure progress after interventions (these tools are not use for diagnostic proposals). Attending the relevance of the commented criteria around the users (social, demographic, and economic), we highlight than in general the most used self-assessment depression questionnaires not include any considerations. For example, Hamilton questionnaire is applied for a young, adults and older users, but it no personalize neither the content of the questionnaire and the assessment process. Instead, the Kutcher Adolescent Depression Scale (KADS) is focused only in adolescent people; and Geriatric Depression Scale (GDS) on geriatric population. Authors as Serra et al. [29] have been focus on an adaptive algorithm for the assessment of depression, following Formal Psychological Assessment (FPA) methodology. FPA consist in organized a set of items from self-report questionnaires with the identified symptoms from diagnostic criteria such as DSM-5. FPA provides to collect structured and qualitative information, no redundant and adaptive. In summary, the selections of the questions are based on the previous user answers. If more questions are needed, this algorithm should provide its.

3 Depression and Context

Depression is a multifactorial illness associated with wide range of environmental, social, family, and individual factors. In the other hand, a Context can be defined as an environment or situation in which a fact is considered. The context often explains part of the fact and helps to allows understanding it. A context may be anything like a location, activity, time, relation among them, etc. [30]. So that, depression associated factors can be defined as depression context. Next, the most relevant depression factors are described.

Culture and ethnicity are important aspects of health and illness. A new branch of medicine, known as ethnomedicine, focuses on the role of culture, perception, and context in shaping someone's physical and mental health.

Cultural identity often influences the degree to which a particular individual shows somatic (physical) symptoms of depression. In other words, some cultures are more comfortable reporting depressive symptoms that are physical in nature rather than mental. For example, many depressed Chinese people complain of bodily discomfort, feelings of inner pressure, and symptoms of pain, dizziness, and fatigue. Similarly, depressed Japanese individuals often complain of abdominal, headache, and neck pain symptoms. Even within Western countries where depressive disorders are more "acceptable", researchers have theorized that some chronic conditions (chronic pain, fibromyalgia, chronic fatigue syndrome) may be more somatic (physical) forms of a mood disorder than actual physical problems [31].

Family context has been shown through empirical studies to be a central determinant in the mental health of youth throughout the past decade. Marshall and Henderson [32] studied the association between attachment, parental depression, divorce, family climate, and siblings, and the risk of adolescent depression.

Psychosocial stressors involve individual internal and external aspects that interact to produce depression. An association between the prevalence of depression and the social class has been described. In the other hand, Gender Roles causes major depression in women. The presence of Violence is directly related to depression, like overload, stress, and little social support. Social support is the most important factor of vulnerability in depression. The very low level of support in the working class women when they have a child, it is enough to suggest an explanation for the increased risk of depression in this period of life [33].

Finally, stress is defined and classified from different perspectives. Stress is a cause of tension and demand or an adaptive mechanism. Stress and the presentation of the first episode of depression are directly related, however when recurrences are present significant stress loses strength among the factors that trigger it [34].

4 Methodoly

Based on the relevance of the context from depression perspective a Methodology to guide IT tools development is presented. The Methodology starts with the context and user profiles definition. After that, a selection of depression assessment validated methods need to be performed and its context validation designed. Then, IT technologies to be included into the tool for depression assessment aim must be selected. Additionally, an IT technology validation procedure must be defined for testing in the context. Finally, IT tool development requirements and specifications based on previous steps must be carefully defined. Figure 1 shows the Methodology proposed schema.

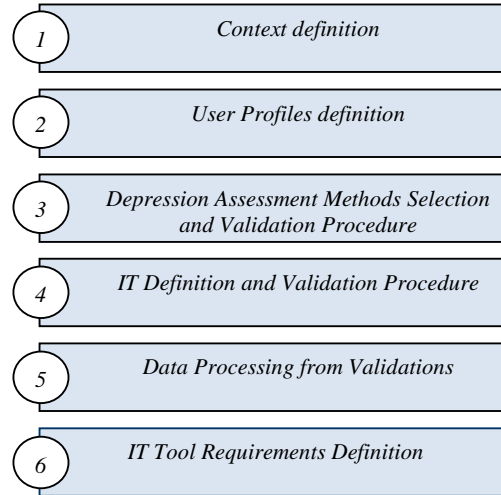


Fig. 1. Methodology scheme to guide IT tools development for depression assessment in a context

4.1 Case of study

A case of study has been defined based on the proposed Methodology. The study is centered in the University Community. The University context selected for the pilot study has been the Universitat Politècnica de València (UPV).

The case of study is focused in the Non-Tenure positions, including the very recent figure of Interim Assistant Professor (Contratado Doctor interino). This selection is based on previous study conducted in the Universitat Jaume I context [33]. This study indicates that **Teachers Assistants (AY)** are exposed to greater psychosocial risks by the presence of high levels of emotional overload and quantitative overload (considered these Labor demands or stressors), higher levels of burnout (emotional exhaustion and cynicism) as well as higher levels of anxiety and depression at work, followed by **Part time University Lecturer (ASO)** working full time [12].

Following sections describe the methodology steps 1 (Context definition) to 4 (IT Definition and Validation Procedure) in the case of study. Steps 5 (Data Processing from Validations) and 6 (IT Tool Requirements Definition) are planned as future work.

4.1.1 Context Definition

The context definition starts collecting actual teaching categories in the Universitat Politècnica de València. Those categories are established in the Spanish Organic Law 4/2007 (LOU 2007). The positions are organized in three groups: Tenured positions, Non-tenure positions, and others. Figure 2 and Table 2 show Tenured and Non Tenured positions distribution in UPV, in global percentage and absolute counting by school respectively: Full Professor (CU), Associate Professor (TU), Full Professor University School (CEU), Associate Professor University School (TEU), Assistant Professor (COD), Assistant Professor (AYD), Collaborator Lecturer (COL), Teaching Assistant (AY), Part time University Lecturer (ASO), Emeritus (EM), and Visiting (VIS).

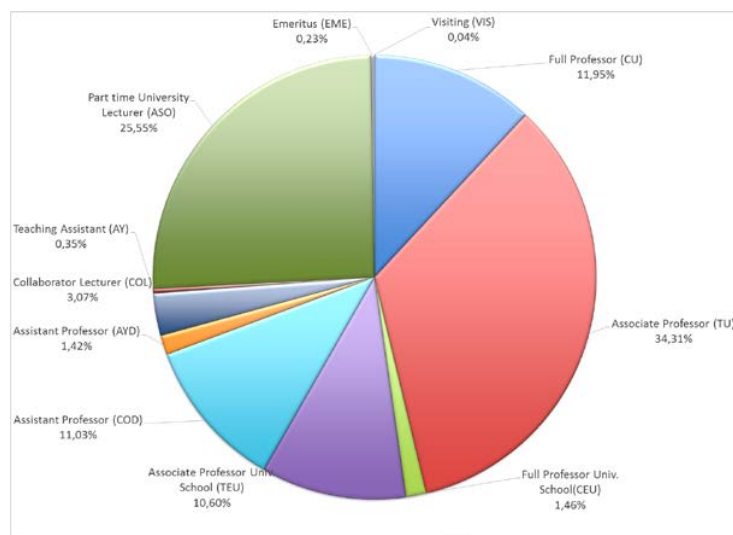


Fig.2. Positions distribution in the Universitat Politècnica de Valencia (Universitat Politècnica de Valencia, 2016)

Table 2. University Schools and Faculties teaching cathegories distribution in UPV (2016).

	CU	TU	CEU	TEU	COD	AYD	COL	AY	ASO	EM	VIS	
Higher Polytechnic School of Alcoy	6	52	4	25	22	2	4	1	81	0	0	197
School of Architecture	18	49	0	32	16	7	15	0	124	0	0	261
School of Building Engineering	0	17	4	30	5	2	10	0	46	0	0	114
School of Design Engineering	27	76	10	34	41	11	10	3	73	0	0	285
School of Informatics	27	141	4	32	23	0	5	0	11	0	0	243
School of Agricultural Engineering and Environment	58	115	11	18	43	4	6	1	30	5	0	291
School of Civil Engineering	29	50	0	25	10	2	3	2	88	0	1	210
School of Telecommunications Engineering	33	68	0	12	3	0	1	0	9	0	0	126
School of Engineering in Geodesy, Cartography and Surveying	3	23	1	13	3	1	3	0	0	0	0	47
School of Industrial Engineering	70	142	3	23	38	6	5	1	103	1	0	392
Higher Polytechnic School of Gandia	5	41	1	13	28	1	12	0	42	0	0	143
Faculty of Business Administration and Management	11	29	0	6	18	1	4	1	45	0	0	115
Faculty of Fine Arts	24	90	0	13	37	0	2	0	13	0	0	179
Total	311	893	38	276	287 ¹	37	80	9	665	6	1	2603

¹Since Royal Decree-Law 14/2012, of 20 April, on urgent measures to rationalize public spending in education two different Assistant Professor (COD) positions exists: Iterim: 17; Non Iterim 270.

4.1.2 User Profiles Definition

In the context under study the user profile has been organized in six categories. The objective of the profiles data is to be able to correlate depression assessment results with context individual characteristics in following processing data step.

The profiles defined are: Academic profile, Diseases profile, Depression Risk factors profile, Family profile, Employment profile, Medication profile

4.1.3 Depression Assessment Methods Selection and Validation Procedure

There seems to be fairly widespread agreement when defining indicators that characterize depression. In a review for the validity depression content, through different instruments most used self-applied, Sanz-Izquierdo and Garcia-Vera [35] corroborate a high consistency between those described by the manuals used clinical criteria (DSM IV and ICD-10) and the main questionnaires, including the BDI, BDI-II (Beck Depression inventory), CES-D (Center for Epidemiologic Studies Depression Scale), SDS (Zung Self-Rating Depression Scale) and inventories as the MMPI-II (Minnesota Multiphasic Personality Inventory-II) or the MCMI-III (Millon Clinical Multiaxial Inventory-III). They found a general tendency to respect these criteria when measuring depression, so it is common to find that all these instruments take into account: (1) sad or depressed mood, (2) loss of interest or pleasure (3) significant profit-loss, or weight, (4) insomnia or hypersomnia, (5) psychomotor agitation or retardation, (6)

fatigue or loss of energy, (7) feelings of worthlessness or excessive guilt, (8) ability to think or concentrate, and (9) recurrent thoughts of death. They conclude that all these inventories that have the highest content validity is the BDI-II. Then we will present the most representative criteria that will be used in the present study to assess the presence and degree or intensity that symptoms of depression:

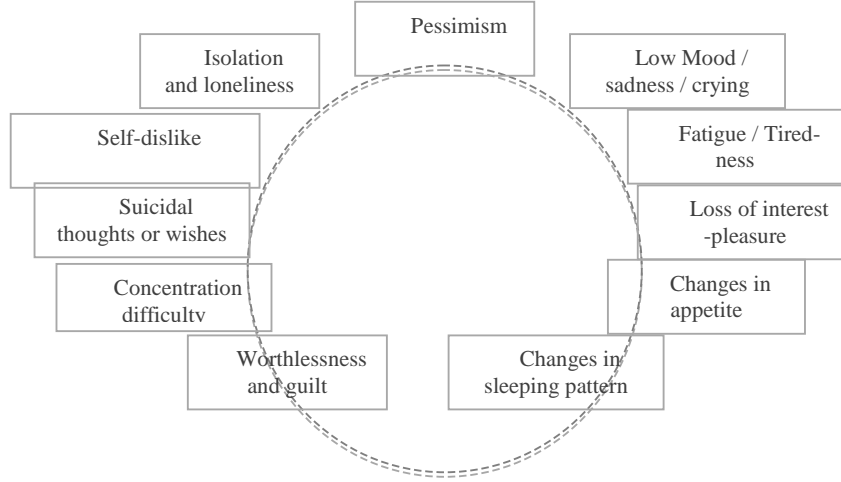


Fig. 3. Content validity criteria for the diagnosis of Depression

Non-Tenured positions volunteer are being recruited in UPV. They are taken into account demographics such as age, gender, years of work, medical history, and position held within the university, type of contract and category (Professor, Assistant doctor, etc.) within their user's profiles.

The BDI-II has been the most widely used in empirical research of depression [35]. Use the translated into Spanish and the standards proposed by Sanz et al version [36], with high internal consistency, α : 0.87. The test consists of 21 items and the measurement range is from 0 to 63. Higher scores represent greater severity of depression. Each item is answered on a 4-point scale from 0 to 3, except items 16 (changes in sleep patterns) and 18 (changes in appetite) containing 7 categories. It can be answered pencil or orally. It indicates the presence or absence of depressive symptoms and their severity.

They are provided two conditions for evaluation of depression and differ in how the instrument is given to evidence if there are differences: **On-line Condition:** The evaluation will be conducted with the response of the BDI in on-line format of the questionnaire to be answered by computer. We have omitted researchers. Facial gestures will be recorded and will be analyzed by the emotion recognition program; **Face-to-face Condition:** The evaluation will be orally by an experimenter who will ask the questions and record answers of the BDI. In this situation the facial gestures that will be analyzed by the emotion recognition program is also recorded.

Some authors have found that applying by computer BDIs, statistics properties are maintained, no clutch suggests further studies especially for the diagnosis of suicidal ideation [37].

4.3.4 IT Definition and Validation Procedure

An online self-assessment depression tool has been integrated in the E-UNIHEALTH portal [38] dedicated to prevent depression in the University Community through emotional virtual agents. The tool is organized in next modules: Data Configuration Module, Assessment Questionnaires Module, and Emotional Module.

The Assessment Questionnaires Module includes a selection of the most representative questionnaires applied in depression assessment. This selection is the result of a deep research work about depression assessment. For the case of study the BDI-II, online version has been included.

Finally, an Emotional Module includes a tool base on Microsoft Emotion API [39], [40] has been integrated in the questionnaires' online tools. The tool system uses a webcam to take images of a person while performing the depression assessment questionnaire. After image processing different emotions levels and a global emotional risk factor are calculated.

5 Conclusions and Future Work

The proposed Methodology summarized in (1) the analysis of the targeted context and in the particular individuals, in order to identify the most relevant features, requirements or limitations; (2) the definition of IT tool; (3) evaluation for novel tool is performed in order to evaluate control group (using traditional questionnaire) versus intervention group (using proposed new method); (4) evaluation results are analyzed and correlate with the initial identified factors. These findings we should inform themselves whether there are risk factors of context that may influence depression assessments of individuals. That is, if there are risk factors inherent to this context, it should be taken into account to make a better evaluation. Moreover, initial identified factors could be redefining in order to improve the performance.

The case of study presented is currently in process. Next steps will finally permit to validate the Methodology. Moreover a new depression assessment tool will be available to prevent depression in the UPV context for teaching staff group.

6 References

1. Haro, JM., Palacin, C., Vilagut, G., Martínez, M., Bernal, M., Luque, I., Codony, M., Dolz, M., Alonso, J. (2006). Prevalence of mental disorders and associated factors: results from the ESEMeD-Spain study. *Medicina Clínica*, 126, (12) 445-451.

2. Piñuel, I., & Cantero, A. O. (2002). La incidencia del mobbing o acoso psicológico en el trabajo en España: Resultados del barómetro CISNEROS II sobre violencia en el entorno laboral. *Lan harremanak: Revista de relaciones laborales*, 7, 35-62.
3. Hamilton, M. (1960). A rating scale for depression. *Journal of neurology, neurosurgery, and psychiatry*, 23(1), 56.
4. Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The Phq 9. *Journal of general internal medicine*, 16(9), 606-613.
5. Bech P, Rasmussen N-A, Olsen LR, Noerholm V, Abildgaard W. The sensitivity and specificity of the Major Depression Inventory, using the Present State Examination as the index of diagnostic validity. *J Affect Disord* 2001; 66: 159-164 PMID 11578668
6. Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients. *Journal of personality assessment*, 67(3), 588-597.
7. First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2014). Structured clinical interview for DSM-5 disorders. New York: New York Psychiatric Institute, Biometrics Research Department.
8. Chaudron, L. H., Szilagyi, P. G., Tang, W., Anson, E., Talbot, N. L., Wadkins, H. I., Tu, X., & Wisner, K. L. (2010). Accuracy of depression screening tools for identifying postpartum depression among urban mothers. *Pediatrics*, 125(3), e609-e617.
9. Kramer, T. L., Owen, R. R., Wilson, C., & Thrush, C. R. (2003). Relationship between self-report and clinician-rated impairment in depressed outpatients. *Community mental health journal*, 39(4), 299-307.
10. Trivedi, M. H., Rush, A. J., Ibrahim, H. M., Carmody, T. J., Biggs, M. M., Suppes, T., Crismon, M.L., Shores-Wilson, K., Toprac, M.G., Dennehy, E.B., Witte, B., & Kashner, T.M. (2004). The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. *Psychological medicine*, 34(01), 73-82.
11. Ward, E. C., & Brown, R. L. (2015). A culturally adapted depression intervention for African American adults experiencing depression: Oh Happy Day. *American Journal of Orthopsychiatry*, 85(1), 11.
12. Cifre, E., & Llorens, S. (2002). Burnout en profesores de la UJI: un estudio diferencial. Exposición presentada en el Fòrum de Recerca, 7.
13. Boundless. "Influences of Culture and Gender in Psychotherapy." Boundless Psychology. Boundless, 20 May. 2016. Retrieved 25 May. 2016 from <https://www.boundless.com/psychology/textbooks/boundless-psychology-textbook/treating-psychological-disorders-19/introduction-to-the-treatment-of-psychological-disorders-99/influences-of-culture-and-gender-in-psychotherapy-380-12915/>
14. Hersen, M., & Thomas, J. C. (2007). *Handbook of clinical interviewing with adults*. Sage Publications.
15. Murray, C. J., & Lopez, A. D. (1996). *Global burden of disease* (Vol. 1, pp. 118-201). Cambridge, MA: Harvard University Press.
16. Mathers, C., Fat, D. M., & Boerma, J. T. (2008). *The global burden of disease: 2004 update*. World Health Organization.
17. Lorant, V., Delière, D., Eaton, W., Robert, A., Philpott, P., & Ansseau, M. (2003). Socio-economic inequalities in depression: a meta-analysis. *American journal of epidemiology*, 157(2), 98-112.
18. Lorant, V., Croux, C., Weich, S., Deliege, D., Mackenbach, J., & Ansseau, M. (2007). Depression and socio-economic risk factors: 7-year longitudinal population study. *The British Journal of Psychiatry*, 190(4), 293-298.

19. Goodman, E., Slap, G. B., & Huang, B. (2003). The public health impact of socioeconomic status on adolescent depression and obesity. *American journal of public health*, 93(11), 1844-1850.
20. Singh-Manoux, A., Adler, N. E., & Marmot, M. G. (2003). Subjective social status: its determinants and its association with measures of ill-health in the Whitehall II study. *Social science & medicine*, 56(6), 1321-1333.
21. Nobles, J., Weintraub, M. R., & Adler, N. E. (2013). Subjective socioeconomic status and health: relationships reconsidered. *Social Science & Medicine*, 82, 58-66.
22. Demakakos, P., Nazroo, J., Breeze, E., & Marmot, M. (2008). Socioeconomic status and health: the role of subjective social status. *Social science & medicine*, 67(2), 330-340.
23. Craig, T. J., & Van Natta, P. A. (1979). Influence of demographic characteristics on two measures of depressive symptoms: the relation of prevalence and persistence of symptoms with sex, age, education, and marital status. *Archives of General Psychiatry*, 36(2), 149-154.
23. Chen, F. P., Samet, S., Gorroochurn, P., & O'Hara, K. M. (2015). Characterizing Factors of Employment Status in Persons With Major Depressive Disorder. *Evaluation & the health professions*, 0163278715619738.
23. Chen, F. P., Samet, S., Gorroochurn, P., & O'Hara, K. M. (2015). Characterizing Factors of Employment Status in Persons With Major Depressive Disorder. *Evaluation & the health professions*, 0163278715619738.
25. Farré, L., Fasani, F., & Mueller, H. F. (2015). Feeling Useless: The Effect of Unemployment on Mental Health in the Great Recession.
26. Mihai, A., Ricean, A., & Voidazan, S. (2014). No significant difference in depression rate in employed and unemployed in a pair-matched study design. *Frontiers in public health*, 2.
27. Fleiz Bautista, C., Villatoro Velázquez, J., Icaza, M., Medina, M. E., Moreno López, M., Gutiérrez López, M. D. L., & Oliva Robles, N. (2012). Sociodemographic and personal factors related to depressive symptomatology in the Mexican population aged 12 to 65. *Revista Brasileira de Psiquiatria*, 34(4), 395-404.
28. Van de Velde, S., Bracke, P., & Levecque, K. (2010). Gender differences in depression in 23 European countries. Cross-national variation in the gender gap in depression. *Social Science & Medicine*, 71(2), 305-313.
29. Serra, F., Spoto, A., Ghisi, M., & Vidotto, G. (2015). Formal Psychological Assessment in evaluating depression: a new methodology to build exhaustive and irredundant adaptive questionnaires. *PloS one*, 10(4), e0122131.
30. Subbaraj, R., & Venkatraman, N. (2015). A Systematic Literature Review on Ontology Based Context Management System. In *Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2* (pp. 609-619). Springer International Publishing.
31. Nemade, R., Reiss, N.S., Dombeck, Mark. *Sociology Of Depression – Effects Of Culture*. 2007.
32. Marshall, C., Henderson, J.: *The Influence of Family Context on Adolescent Depression: A Literature Review*. *Canadian Journal of Family and Youth*, 6(1), (2014), 163-187 ISSN 1718-9748 University of Alberta
33. Montesó- Curto, P., Ferré-Grau, C., Lleixà- Fortuño, C., Albaric Riobóo, N., Espuny-Vidal, C. *Factores Sociológicos*. *Anuario Hojas de Warml* nº 16, 2011
34. Vargas-Navarro, P., Latorre-López, D., Parra-Cárdenas, S.. *Psychosocial stressors and major recurrent depression*. *Revista Salud Bosque*. Volumen 1. Número 2. Págs. 39-53
35. Sanz, J., Izquierdo, A., García-Vera, M.P. (2013). Una revisión desde la perspectiva de la validez de contenido de los cuestionarios, escalas e inventarios autoaplicados más utilizados en España para evaluar la depresión clínica en adultos. *Psicopatología Clínica, Legal y Forense*, 13 (2) 139-175.

36. Sanz, J., Perdigón, L. A., y Vázquez, C. (2003). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 2. Propiedades psicométricas en población general. *Clínica y Salud*, 14, 249-280.
37. Holländare, F.; Andersson, G.; Engström, I. (2010) A comparison of psychometric properties between internet and paper versions of two depression instruments (BDI-II and MADRS-S) administered to clinic patients. *Journal Of Medical Internet Research*. 12(5):e49. Doi:10.2196/jmir.1392
38. Asensio-Cuesta, S., Bresó, A., Miralles, C., García, J. M., Robles, M.: An eHealth Portal to prevent depression in the University Community through emotional virtual agents. *WIICT2015* (2015) 9-16
39. Microsoft Cognitive Services. Retrieved from: <https://www.microsoft.com/cognitive-services/en-us/emotion-api>
40. Microsoft Emotion API. Retrieved from: <https://dev.projectoxford.ai/docs/services/5639d931ca73072154c1ce89/operations/563b31ea778daf121cc3a5fa>

Development of an acquisition electrophysiological signals system.

Ramón Albert¹, Andreu M. Climent², Miguel Rodrigo¹ and María S. Guillem¹

¹ Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politècnica de València, Valencia, Camino de Vera s/n, 46022 Valencia, España.

Phone: +34-963-877-968; e-mail: ramalmar@epsa.upv.es

² Cardiology Department, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain.

Abstract. Many experiments in cardiology for the purpose of studying heart diseases require a complex system to obtain a multitude of cardiac electrophysiological measurements simultaneously with a high sampling frequency and low noise level. However, commercially available biopotential measurement systems are expensive, often not customizable and commercial software normally does not provide access to information collected in real time. In this work a 256 channel electrophysiological data acquisition system has been developed using standard and commercially available components. This system is capable of recording many types of electrophysiological signals including ECG and EGM. In addition, our system supports the connection of individual electrodes or intracavitary catheter for electrical mapping of the atria. A software driver to manage system using a custom C++ software also has been developed to control the acquisition system with an advanced signal analysis software. We tested the complete system using reference signals and performed real recordings giving satisfactory results. This system is upgradeable, customizable and less expensive than similar commercial products, and can encourage extending the use of novel diagnostic techniques in clinical practice.

1 Introduction

Cardiovascular diseases are becoming in the major cause of death in the developed countries [1], and many efforts are focused to research into these diseases. For this reason, clinical electrical mapping techniques have appeared over the last years. One of these techniques is the BSPM (Body surface potential mapping) used to study the heart's electrical activity non-invasively [2] and providing more information than a standard ECG. This technique uses from 32 to 256 electrodes distributed around the torso surface in order to capture all the electrical information of the heart available on the body surface, which improves the detection of cardiac diseases. This technique has been used for localizing the dominant atrial regions in patients with atrial fibrillation [3-5] and for identification of patients at risk for ventricular tachycardia [6-7]. Alternatively, cardiac mapping can make use of intracavitary catheters by positioning a

catheter array with multiple electrodes into the cardiac chamber. This technique reveals information of the electrical activity of the heart and can guide the cardiac procedures [8- 11].

These techniques require a complex electronic system to obtain a multitude of measures simultaneously with a high sampling frequency and low noise level. Commercially available biopotential measurement systems are expensive, often not customizable and commercial software normally does not provide access to information collected in real time. To overcome these limitations, a custom system to acquire and processing biopotentials signals have been developed. This system has to accomplish the following conditions: must be able to acquire up to 256 channels of electrophysiological signals, must be compatible with individual electrodes or intracardiac catheters, must acquire electrophysiological signals with a minimum sample rate of 1 KHz and has to be compatible with advanced signal analysis software.

2 Methods

Figure 1 shows a general diagram of the designed system, where the role of each function module can be observed. The system consists of a main module for the acquisition and an interface module to connect electrodes. The custom-made software was programmed in C++ to interact with MATLAB in order to record and process the signals data obtained.

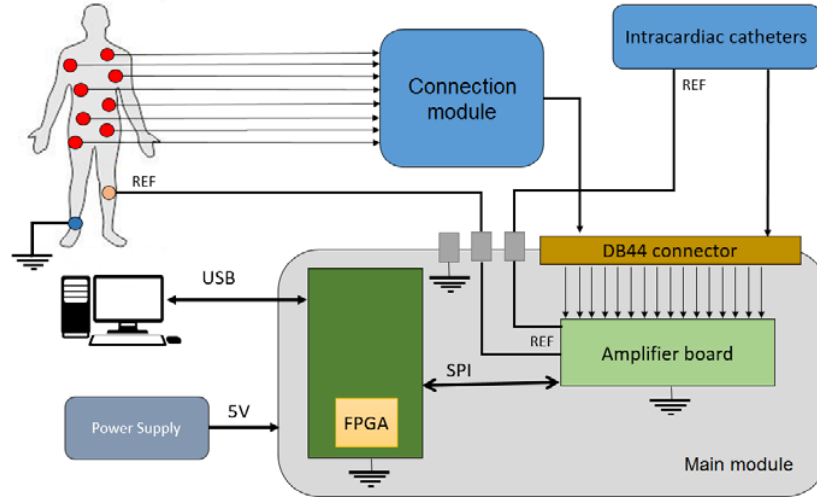


Fig. 1. System diagram

2.1 Main module

This module contains the main system board and signal stage amplifiers. The connection with the electrodes is performed on the front of the box through DB44 ports. At the rear there are located the power Jack, USB connection, switch and operation LED. For data acquisition, a RHD2164 amplifier board (Intan technologies) has been implemented. This prefabricated module includes the RHD2164 Chip capable of monitor and record signals from 64 low-noise amplifier channels at sampling rates ranging from 1kHz to 30kHz per channel, this board includes modules for signal digitation. To provide a way to manage the data acquired from this amplifier stages, a RHD2000 USB Interface Board has been installed. This board has a FPGA (Field-programmable gate array) capable of collecting digital data sent by the amplification stages and send them to the PC through the USB port. This board is able to manage up to four amplifier stages of 64 channels each.



Fig. 2. Acquisition system.

2.2 Connection module for BSPM recordings

In order to connect the multitude of electrodes required for a BSPM recording, a connection module has been developed. A module to connect up to 64 banana connectors for individual input channels has been constructed plus two ground electrodes. These two reference connections are available in the main module and in the connection module.

An example of a possible electrode configuration is shown in (Fig. 3). These special electrodes are attached to the patient's skin by an adhesive that ensures good electrode-skin contact.

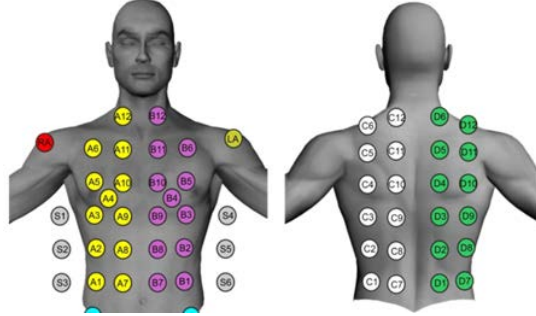


Fig. 3. ECG electrodes position for a BSPM recordings.

2.3 Connection to Intracavitary catheters

Our system supports the connection of 64- pole basket catheter (Constellation, Boston Scientific, Natick, MA, USA) to the main box through DB44 compatible connection. This feature allows the use of this system to obtain the intracavitary electrical activity simultaneously to BSPM recordings in real time. The reference signal of the intracavitary catheter can be set to either an electrode from the same catheter or a secondary reference catheter. The reference connections are plugged directly on the main module.

2.4 Software

An acquisition software was programmed in C++ to record and process the amplified signals (Fig. 3). This software creates a link between the acquisition card and MATLAB. Once registration is concluded, signals are stored in the MATLAB workspace as a matrix. The data acquisition process can be automated for the desired application, enabling data acquisition with real-time processing.

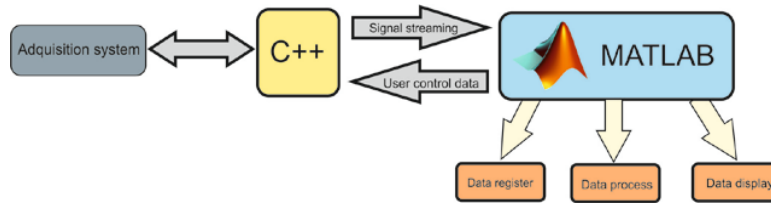


Fig. 3. Acquisition system software diagram

3 Results

In order to know the deviation in the reading, a synthetic ECG signal was compared with its recording by the developed device (Fig. 4). This reference signal was synthesized using an Arduino Mega and a DAC (Digital to analog converter) connected to an amplifier input through 1/1000 voltage divider.

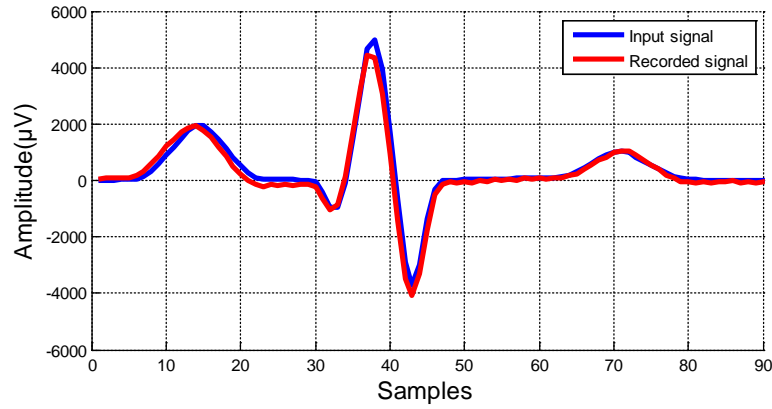


Fig. 4 Comparison between the synthetic input signal and the signal recorded by the acquisition system.

The registered signal was compared with the original synthesized signal. The average difference between the two signals was $1.81 \pm 1.95\%$.

Furthermore, to check the proper functioning of the system in real conditions, a BSPM recording was performed. A total of 59 electrodes were connected to the skin: 54 electrodes from the BSPM distribution, 3 leads for the Wilson Central Terminal in shoulders and left leg, an amplifier reference in the right leg and the system ground in the right leg. Figure 5 shows the ECG recording at the A10 channel, corresponding to the standard V1 lead together with the surface potential distribution obtained on the full lead set at three representative instants marked in the V1 trace.

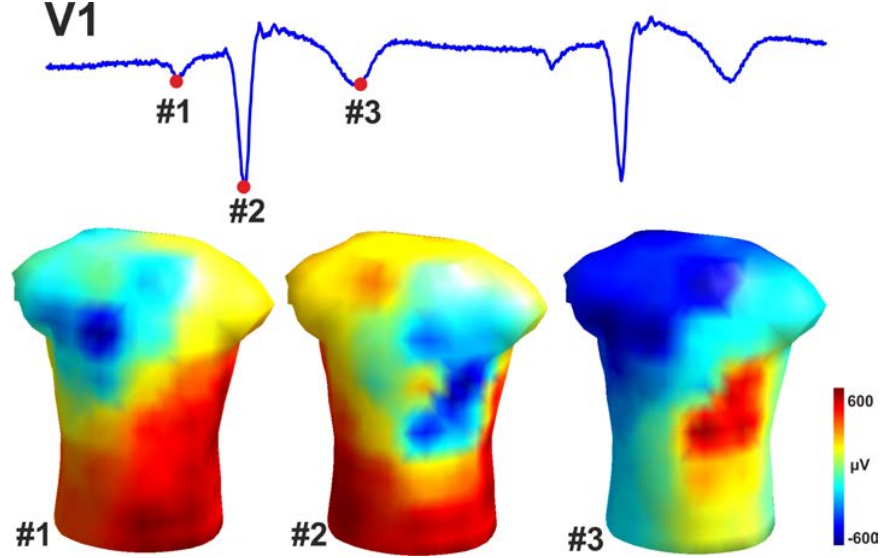


Fig. 6. ECG at A10 lead (top) and potential maps at 3 time instants (bottom): (#1) atrial depolarization, (#2) ventricular depolarization and (#3) ventricular repolarization.

4 Discussion

In this paper, an acquisition system for electrophysiological signals has been developed and tested. This system allows acquiring up to 256 electrophysiological signals with a varying sample rate from 1kHz to 30kHz, and is capable of recording many types of electrophysiological signals including ECG and EGM. The system is compatible with individual electrodes or intracardiac catheters, allowing the option of BSPM recording simultaneously to invasive measurements. Finally custom software was programmed to control the device using MATLAB which allows controlling the recording parameters.

Some advantages have been achieved in the presented system respect to commercial systems. Whereas commercial systems commonly restrict real-time access to the recorded signals, our software provide free access to information collected in real time, enabling data acquisition with real-time processing through custom scripts programmed in MATLAB. Moreover, the reduced economic charge of our system compared to commercial systems can encourage extending the use of novel techniques such as the BSPM in clinical practice.

The outcomes of our device should be tested in a clinical scenario with real patients in order to ensure its capability to be used in clinical practice. To that purpose, previously presented signal analysis techniques for both non-invasive [2-5] and invasive [7] signals should be integrated and tested in this system in authorized clinical trials. Last-

ly, our system is in the process to pass the test of European conformity in order to commercialize the device.

5 Conclusion

At the present work an acquisition electrophysiological signals system has been designed and tested. The different modules constituting the system are upgradeable and customizable according to the needs of each application. Furthermore, the dedicated software allows researchers to easily create custom scripts to interact with the system to read and process data in real time. Moreover, this system reduces the economic charge that similar products involve, so it may suppose the extension of techniques such as the BSPM in clinical practice.

Acknowledgments

This work has been funded by the Spanish Ministerio de Economía y Competitividad: Programa Promoción de Empleo Joven e Implantación de la Garantía Juvenil 2015 PEJ-2014-A-81398

References

1. Baena Díez. J.M., Val García. J.L., Tomàs. Pelegrina. J., Martínez. J.L., Martín. Peña. R., González. I., Quintana. E.M., Sajkiewicz. M., Borona.t A., Alvarez B., Forcadell. P., Rovira. M., Oller. M.: Cardiovascular disease epidemiology and risk factors in primary care. *Rev Esp Cardiol* (2005) Apr; 58(4):367-73.
2. Rodrigo. M., Pedrón-torrecilla. J., Hernández. I., Liberos. A., Climent. AM., Guillem. M.S.: Data analysis in cardiac arrhythmias. *Methods Mol Biol* (2015) 1246:217-35. doi: 10.1007/978-1-4939-1985-7_14.
3. Pedrón-Torrecilla. J., Rodrigo. M., Climent. A.M., Liberos. A., Pérez-David. E., Bermejo. J., Arenal. Á., Millet. J., Fernández-Avilés .F., Berenfeld. O., Atienza. F., Guillem. M.S.: Non-invasive Estimation of Epicardial Dominant High-Frequency Regions During Atrial Fibrillation. *J Cardiovasc Electrophysiol* (2016) Apr; 27(4):435-42. doi: 10.1111/jce.12931.
4. Guillem. M.S., Quesada. A., Donis. V., Climent. A.M., Mihi. N., Millet. J., Castells. F.: Surface Wavefront Propagation Maps: Non-invasive characterization of atrial flutter circuit. *International Journal of Bioelectromagnetism*. (2009) Vol. 11, No. 1, pp.22-26.
5. Guillem. M.S., Climent. A.M., Millet. J., Arenal. Á., Fernández-Avilés. F., Jalife. J., Atienza F., Berenfeld. O.: Noninvasive localization of maximal frequency sites of atrial fibrillation by body surface potential mapping. *Circ Arrhythm Electrophysiol* (2013) Apr; 6(2):294-301. doi: 10.1161/CIRCEP.112.000167.
6. Hubley-Kozey. C., Mitchell. L., Gardner. M., Warren. J., Penny. C., Smith E., Horacek. B.: Spatial features in body-surface potential maps can identify patients with a history of sustained ventricular tachycardia. *Circulation*. vol. 92 (1995) pp. 1825–1838.

7. Rodrigo. M., Climent. A.M., Liberos. A., Calvo. D., Fernández-Avilés. F., Berenfeld. O., Atienza. F., Guillem. M.S.: Identification of Dominant Excitation Patterns and Sources of Atrial Fibrillation by Causality Analysis. *Ann Biomed Eng.* (2016) DOI: 10.1007/s10439-015-1534-x.
8. Bailon. R., Olmos. B., Horacek. M., Laguna. P.: Identification of patient at risk for ventricular tachycardia by means of body surface potential maps. *Conf. Computers in Cardiology.* Vol. 30 (2003) pp. 217–220.
9. Atienza. F., Climent. A.M., Guillem. M.S., Berenfeld. O.: Frontiers in Non-invasive Cardiac Mapping: Rotors in Atrial Fibrillation-Body Surface Frequency-Phase Mapping. *Card Electro-physiol Clin* (2015) Mar 1;7(1):59-69.
10. Nademanee. K., McKenzie. J., Kosar. E., Schwab. M., Sunsaneewitayakul. B., Vasavakul T., Khunnawat. C., Ngarmukos. T.: A new approach for catheter ablation of atrial fibrillation: mapping of the electrophysiologic substrate. *J Am Coll Cardiol* (2004) Jun 2; 43(11):2044-53.
11. Atienza. F., Almendral. J., Jalife. J., Zlochiver. S., Ploutz-Snyder. R., Torrecilla. E.G., Arenal. A., Kalifa. J., Fernández-Avilés. F., Berenfeld. O.: Real-time dominant frequency mapping and ablation of dominant frequency sites in atrial fibrillation with left-to-right frequency gradients predicts long-term maintenance of sinus rhythm. *Heart Rhythm* (2009) Jan; 6(1):33-40. doi: 10.1016/j.hrthm.2008.10.024. Epub 2008 Oct 22
12. Sanders. P., Berenfeld. O., Hocini. M., Jaïs. P., Vaidyanathan. R., Hsu. L.F., Garrigue. S., Takahashi. Y., Rotter. M., Sacher. F., Scavée. C., Ploutz-Snyder. R., Jalife. J., Haïssaguerre. M.: Spectral analysis identifies sites of high-frequency activity maintaining atrial fibrillation in humans. *Circulation* (2005) Aug 9; 112 (6):789-97.

Estimating the number of contending users for a single random access in LTE-A networks: The baseline for designing congestion control schemes at the evolved Node B [★]

Luis Tello-Oquendo, Israel Leyva-Mayorga, Vicent Pla, Jorge Martinez-Bauset
and Vicente Casares-Giner

ITACA. Universitat Politècnica de València. Camino de Vera s/n.
46022 Valencia, Spain

Abstract. Machine-type communications (MTC) have the potential to generate a myriad of connection requests in a short period. Using cellular networks to provide MTC connectivity presents numerous advantages such as coverage, roaming support, well-developed charging, QoS, security solutions, among others. Nevertheless, critical problems like congestion and overload of radio access and core networks need to be addressed for efficient cellular MTC. In LTE-A, the physical random access channel (PRACH) is used by MTC devices (UEs) to access the network. For doing so, a UE randomly chooses a preamble from a pool of preambles and transmits it during the PRACH. The evolved Node B (eNB) acknowledges the successful reception of a preamble if only one UE transmits that preamble. To increase the success rate of a massive number of access attempts is necessary to design congestion control schemes. For that, the key piece of information is the total number of UEs competing in the PRACH. To estimate this number at the eNB, we find the joint probability distribution function (PDF) of the number of successful and collided preambles within a random access slot. Then, we design a maximum likelihood estimator using this PDF. To further improve our estimation, we propose an iterative approach. Numerical results showcase the accuracy and usefulness of the proposed method even if the number of access attempts is significant.

Keywords: LTE-A networks, machine-type communications (MTC), maximum likelihood estimation, random access, random access channel (RACH) procedure.

1 Introduction

Nowadays, the use of cellular network technologies such as Long Term Evolution Advanced (LTE-A) for providing machine-type communications (MTC) has

[★] This research has been supported in part by the Ministry of Economy and Competitiveness of Spain under Grants TIN2013-47272-C2-1-R and TEC2015-71932-REDT. The research of L. Tello-Oquendo was supported in part by Programa de Ayudas de Investigación y Desarrollo (PAID) of the Universitat Politècnica de València. The research of I. Leyva-Mayorga was partially funded by grant 383936 CONACYT-Gobierno del Estado de México 2014.

attracted significant attention from both the research community and the industry. The widely deployed infrastructure is a major driving force that motivates MTC application developers to adopt cellular networks for their numerous remote monitoring and controlling applications [7, 16, 19]. However, critical problems like congestion and overload of radio access and core networks need to be addressed for efficient cellular MTC [15, 25].

In LTE-A, a random access procedure is initiated when a MTC device (named UE herein) desires to access the cellular network. The Physical Random Access Channel (PRACH) is used to signal a connection request. For that, the evolved Node B (eNB) has multiple preambles in a preamble pool available for initial access to the network; these preambles are generated by Zadoff-Chu (ZC) sequences due to their good correlation properties [4, 14]. The random access procedure consists of a four-message handshake. In *Msg1*, a UE transmits a randomly chosen preamble from the preamble pool during one of the available random access slots. A preamble will be detected at the eNB if it is not chosen by more than one UE in the same random access slot. Then, the eNB sends a random access response (RAR) message, *Msg2*, which includes one uplink grant for each detected preamble. *Msg2* is used to assign time-frequency resources to the devices for the transmission of *Msg3*. Next, the UEs that received an uplink grant send their connection request message, *Msg3*, during the resources specified by the eNB. Finally, the eNB responds to each *Msg3* transmission with a contention resolution message, *Msg4*. The interested reader is referred to [1, 3, 5, 6, 11, 12, 17, 20] for further details.

The contention-based random access procedure detailed before is similar to the slotted ALOHA protocol [10]; after the transmission of *Msg1*, a UE can be connected if there is no collision. In this sense, the random access procedure can be seen as a multichannel ALOHA [9], where congestion control is studied by estimating the number of arrivals or the number of UEs that send preambles to the eNB [13].

The estimation of the network load is a challenging task when the network includes MTC, due to the high (and unpredictable) number of UEs expected to access the cellular network simultaneously. In recent research, approaches using non-trivial combinatorics have been developed to derive the probability distribution of the number of successful and collided users in one-shot random access [21, 22]. However, the computational complexity of these procedures is considerably high and may not be suitable in the presence of thousands of UEs.

In this paper, we propose a simple scheme to estimate the network load. First, we design a recursive approach to find the joint probability distribution function (PDF) of successful, collided, and not used preambles in a random access slot as the number of UEs increases in the system. Then, based on this information available at the eNB, we find the number of contending UEs in a random access slot employing the maximum likelihood estimation to reduce the complexity during its on-line operation. Finally, we use an iterative approach to improve further the estimation.

The remainder of the paper is organized as follows. In Section 2, the system model and the problem formulation are presented. In Section 3, the method for estimating the number of contending users at eNB and the iterative approach for

refining its accuracy are detailed. Numerical results are presented in Section 4. Finally, the conclusions and future work are presented in Section 5.

2 System Model

This paper considers a fixed number of UEs performing random access to an LTE-A eNB. It is assumed that all the UEs fall within the coverage of just one eNB. According to the LTE-A standard [3, 5, 6], when a UE intends to establish a connection to an eNB, the UE initiates a random access procedure by sending a preamble to the eNB, *Msg1*, via a time-frequency radio resource called physical random access channel (PRACH). There are up to 64 orthogonal preambles available to the UEs per cell. Preambles are split into two sets:

- Contention-free: it is used for critical situations such as handover, downlink data arrival or positioning, where there is a coordinated assignment of preambles so collision is avoided. The eNB can only assign these preambles for specific slots to specific UEs. They can only use them, if assigned by the eNB, and for the specific slots assigned.
- Contention-based: it is the standard mode for network access (there are more preambles in this set). Preambles are selected in a random fashion, so there is risk of collision, i.e., multiple UEs in the cell might pick the same preamble signature and the eNB would assign the same physical resources to both UEs; therefore contention resolution is needed.

Herein, we focus on the *Msg1* of the contention-based random access procedure, and assume that UEs that have transmitted successfully the *Msg1* will successfully complete the random access procedure [8, 23]. The time is divided into independent fixed-length random access slots (RAS). Regarding the channel resources, the eNB has a set of r preambles $\{1 \leq r \leq 54\}$ available for contending UEs in each RAS, the remaining 10 out of 64 preambles are used in the contention-free mode by the eNB. Further, we assume that the cell is large enough, which allows the eNB to differentiate the preambles chosen by more than one UE, i.e. collisions in the preamble space can be detected [8, 25].

As explained before, if a device wants to connect to the eNB, it sends a preamble, chosen in a random fashion among all r preambles, in the first coming RAS. If the eNB receives this preamble without collision, the connection from the UE to the eNB can be established. However, if multiple UEs have transmitted the same preamble, a collision occurs, and the eNB ignores the collided preamble. A device learns the success or failure of its random-access attempt immediately at the end of the RAS [13, 22].

2.1 Capacity of the PRACH

In [18], it is found that the capacity of the PRACH, L , defined as the maximum expected number of preambles transmitted by a single UE in a RAS, approximately corresponds to the maximum number of stationary UE arrivals per RAS that the PRACH can handle efficiently, \hat{L} . In other words, the performance of the PRACH drops whenever the number of UEs (N)

that begin its RA procedure at the each and every RAS is $N \geq \hat{L}$. If r is the number of available preambles and N_i is the number of preamble transmissions at the i th RAS, the expected number of preambles transmitted by a single UE is $N_i(1 - 1/r)^{N_i-1}$ and its maximum, L , is achieved when $N_i = [\log(r/[r-1])]^{-1} \approx r$, given as follows

$$L = \left[\log \left(\frac{r}{r-1} \right) \right]^{-1} \left(1 - \frac{1}{r} \right)^{[\log(\frac{r}{r-1})]^{-1}-1}, \quad (1)$$

which results, for instance, in $L = 20.05$ preambles transmitted by a single UE at a given RAS when $r = 54$ (see Fig. 1(a)). Furthermore, we observed that (1) can be approximated as follows

$$L \approx r \left(1 - \frac{1}{r} \right)^{r-1} \approx \frac{r}{e}. \quad (2)$$

Hence, assuming a typical PRACH configuration (*PRACH configuration index* 6, in conformance to the LTE-A specification [2, 3]), the PRACH can handle a maximum of $\hat{L} \approx 20.05$ stationary UE arrivals per RAS and, given that RA slots occur every $T_{\text{RAS}} = 5$ ms, a maximum of $\hat{L}/T_{\text{RAS}} = 4010$ stationary UE arrivals per second.

3 Estimating the Number of Contending UEs at eNB

Some congestion control mechanisms, such as access class barring and extended access barring in LTE-A [3], require complete awareness of the input load to optimize their parameters or adapt them to the network status. The input load depends on the arrival process of the access requests and can only be estimated from the information available at the eNB during a time slot, i.e., the number of success/collided/not-used preambles in an RAS [12, 24].

Let us focus on a single RAS with r preambles, and, to simplify the notation, let the random variable N be the number of UEs which randomly select their transmission preamble among the r preambles available in the RAS. Denote by S the random variable representing the number of preambles successfully transmitted, C the random variable representing the number of collided preambles and U the random variable representing the number of not-used preambles. The conditional joint probability that models the fact that exactly $S = s$ and $C = c$ out of $N = n$ UEs transmit their preambles successfully and with collision, respectively, becomes

$$P_n(s, c) = P(S = s, C = c, U = r - s - c | N = n) = \frac{\binom{r}{s} \binom{c}{n-s-2c}}{\binom{n}{r}}, \quad (3)$$

therefore, the conditional probability that exactly s out of the n UEs successfully transmit their preamble in a RAS comprising r preambles is given by

$$P_n(s) = P(S = s | N = n) = \sum_c P_n(s, c). \quad (4)$$

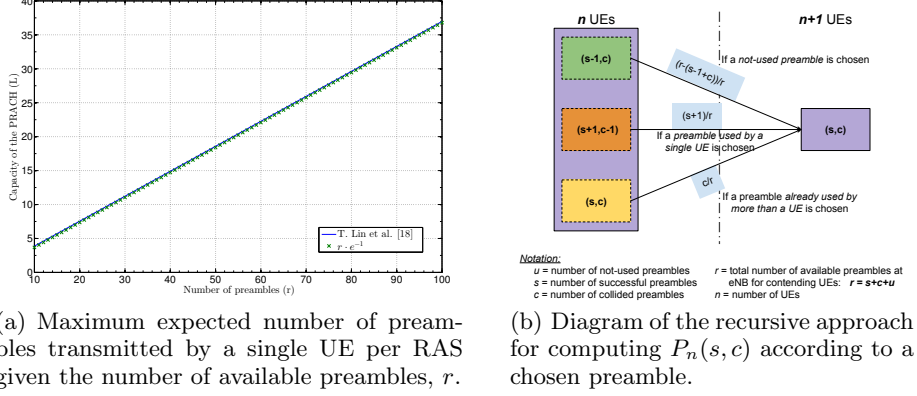


Fig. 1. System capacity and recursive approach diagram for computing $P_n(s, c)$ as the number of UEs, n , increases in the system.

As MTC imply a massive number of UEs accessing the network, it is necessary to find an efficient way to compute the conditional joint probability distribution, $P_n(s, c)$, for a great number of UEs registered in the system. For doing so, in Fig. 1(b) we illustrate what happens at the eNB when a UE performs a new access attempt in a determined RAS; the diagram represents the number of s and c preambles (s, c) detected in the RAS when there are a certain number of users, n , attempting access at the eNB. If the number of UEs increases in the RAS, $n + 1$, a new preamble must be selected so that we can identify three outcomes:

1. There are $(s - 1, c)$ preambles detected at the eNB, and the arriving UE has chosen one of the not-used preambles; in that case, it is a successful access attempt, and the detected preambles becomes (s, c) .
2. There are $(s + 1, c - 1)$ preambles detected at the eNB and the arriving UE has chosen a preamble used by a single UE; in that case, a collision is produced, and the detected preambles becomes (s, c) .
3. There are (s, c) preambles detected at the eNB, and the arriving UE selects a preamble already used by more than one UE; in that case, the eNB does not detect the preamble, remaining the same (s, c) preambles.

As a result of the described observations, we devised a recursive method for computing the conditional joint probability distribution function (PDF) for the number of successful and collided preambles in a RAS as the number of UEs in the system, n , increases. It is computed as follows

$$P_{n+1}(s, c) = \frac{r - (s - 1 + c)}{r} P_n(s - 1, c) + \frac{s + 1}{r} P_n(s + 1, c - 1) + \frac{c}{r} P_n(s, c). \quad (5)$$

Note that this operation is computationally tractable and this problem can be solved offline numerically for a great number of UEs (n). A look-up table is obtained with r rows r columns for different values of s and c , respectively.

Moreover, this table is computed once and can be used throughout the operation of the system.

Finally, we can estimate N by using the maximum likelihood estimation (MLE) from s as

$$\hat{N} \equiv \hat{N}(s) = \arg \max_n P_n(S = s). \quad (6)$$

3.1 Improving the initial estimation

As a baseline, the number of contending UEs in a RAS can be estimated using only the conditional probability that a device can transmit successfully its preamble by (6). Note that the number of the devices with successful preamble transmissions, S , is limited as $S \leq r$, since there are r preambles available in the RAS. Nevertheless, we can also use the number of collided preambles for the estimation of N . Specifically we can use the conditional joint PDF for the number of successful and collided preambles, $P_n(s, c)$, in the MLE as follows

$$\hat{N} \equiv \hat{N}(s, c) = \arg \max_n P_n(S = s, C = c). \quad (7)$$

In the simulations, as will be explained in the following Section 4, it can be observed that we get acceptable results estimating the number of contending UEs by either (6) or (7). Furthermore, we observe that \hat{N} is always less than or equal to the real number of contending UEs. Therefore, based on the latter observation, the MLE can be improved through an iterative approach. For that, we consider that given an initial estimation of the contending UEs, \hat{N} , it can be refined as follows

$$\hat{N}^{(1)} = \frac{\hat{N}^2}{\mathbb{E}_n[\hat{N}]}, \quad (8)$$

where $\mathbb{E}_n[\hat{N}] = \sum_s \hat{N}(s) P_n(s)$ if the initial estimation, \hat{N} , has been computed by (6). On the other hand, if the initial estimation has been computed by (7), $\mathbb{E}_n[\hat{N}] = \sum_{s,c} \hat{N}(s, c) P_n(s, c)$. This refinement allows for an elegant solution that yields accurate results, as demonstrated further.

4 Numerical Results

In this section, we present the simulation results for getting the MLE of N from $P_n(s)$ and $P_n(s, c)$, respectively. We have performed 10000 runs for each value of N to find the mean values of the MLE. In real implementations, the number of UEs within a single cell could be significantly large, so we vary the number of UEs from 1 up to 100 times the capacity of the PRACH, L , computed using (1). For example, when $r = 54$, the PRACH capacity becomes $L = 20.05$; therefore, we vary the number of UEs from 1 up to 2005. We used different values for the number of preambles available in the system for the contention-based random access, $r \in \{1, 2, 3, \dots, 54\}$. But we observed that the results are qualitatively similar for each value of r tested; therefore, next we show and analyze the ones corresponding to the most typical scenario (in conformance to

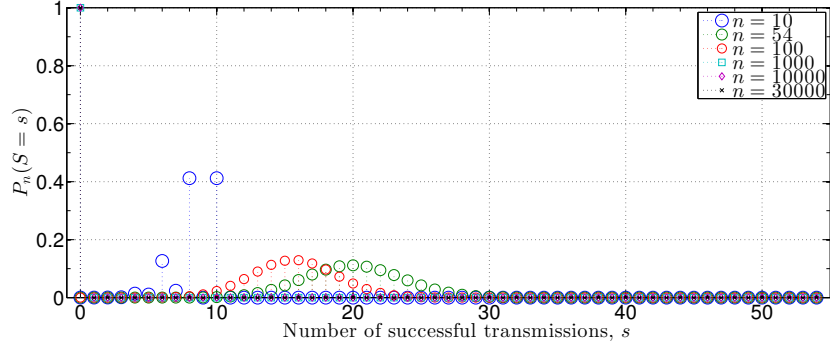


Fig. 2. Distribution of the number of successful transmissions in a given random access slot, $P_n(s)$, $r = 54$.

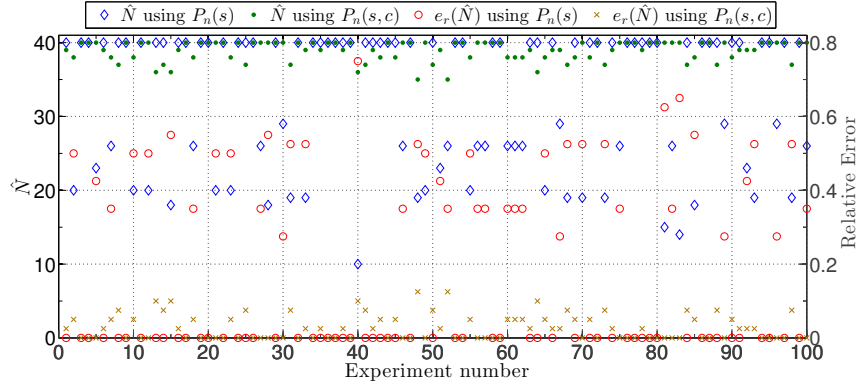


Fig. 3. MLE of contending UEs at eNB when $N = 40$, $r = 54$.

the LTE-A specification [2, 3]) when the eNB have $r = 54$ preambles available for the contention-based random access procedure.

As a baseline, Fig. 2 illustrates the probability distribution of the number of successful transmissions, $P_n(s)$, in a RAS comprising $r = 54$ preambles. As intuitively expected, the distribution is irregular when the number of UEs is low (see the curve of $n = 10$). As n increases, the curves makes a well-defined peak at a determined value of n ; additionally, we see that the probability mass is concentrated around the mean value, as can be seen clearly in the curve of $n = 100$. In essence, Fig. 2 intuitively suggests that for a relatively large r , the PDF of the number of successful transmissions is highly concentrated around its mean value, for either n small and large [21].

Then, we analyze two cases for estimating the number of contending UEs at the eNB. First, we consider that the unique available information at the eNB in a time slot is the number of successful preamble transmissions, s , and find the MLE

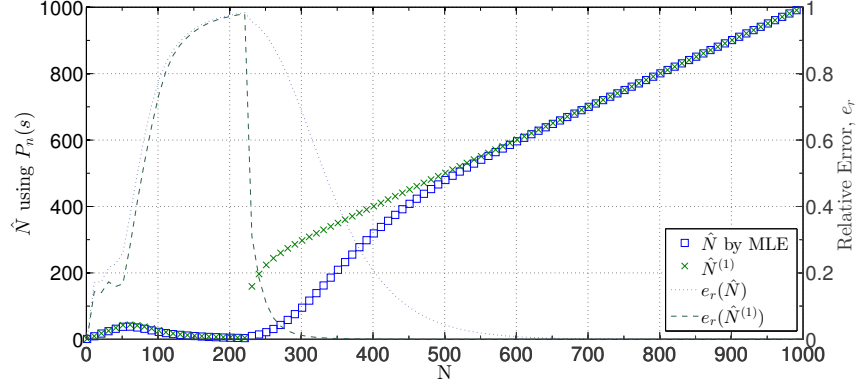


Fig. 4. MLE of contending UEs at eNB using $P_n(s)$, $r = 54$.

of N by (6). We observe that, for a test value $N = 40$ (see Fig. 3), the relative error of the estimated values ranges from 0 to 70% in the worst case. It might be a reasonable estimation; however, it can be improved considering additional information available at the eNB as is explained next. Second, we consider the fact that, at the eNB, the number of successful and collided preambles in a RAS is known. Thus, we can use the joint probability distribution (3) in the MLE of N . By doing so, equation (7) can provide much better results for any value of N , when compared to the ones obtained by (6). For example, see Fig. 3, for $N = 40$, the relative error in the estimations ranges from 0 to 14%.

One of the main observations in the results obtained by simulation, for instance the ones of Fig. 3, is that the error is upper-bounded by the real value of N , i.e., the estimation is always lower than or equal to the real number of contending UEs in a RAS. Taking advantage of this fact, an iterative approach, (8), can be devised. It allows us to increase further the accuracy of the estimation. Figures 4 and 5 illustrate the results following the iterative approach; we varied N from 1 up to 100 times the capacity of the PRACH computed using (1) as a function of the preambles available for contending UEs, $r = 54$.

On the one hand, in Fig. 4 we illustrate the estimation of the number of contending UEs by (6) from $P_n(s)$. Analyzing these results, we observe that in the extremes, $\{1 < N < 50\}$ and $\{N > 500\}$, the estimation is very accurate; whereas, there are values of N , $\{50 < N < 500\}$, in which the MLE is biased and the relative error of the estimation reaches the 98%. Clearly, an enhancement is necessary in these cases; therefore, we use the iterative approach by (8) for refining the initial estimation. This refinement yields a remarkable improvement in the accuracy at $\{N > 222\}$ and we observe that the enhancement especially improves the accuracy in cases where the initial estimation was longer acceptable. Conversely, when the estimation accuracy was not acceptable in many cases, $\{50 < N < 223\}$, it continues without being after the refinement. On the other hand, in Fig. 5 we illustrate the estimation of the number of contending UEs by (7) from $P_n(s, c)$; we observe that the initial results are very accurate, an

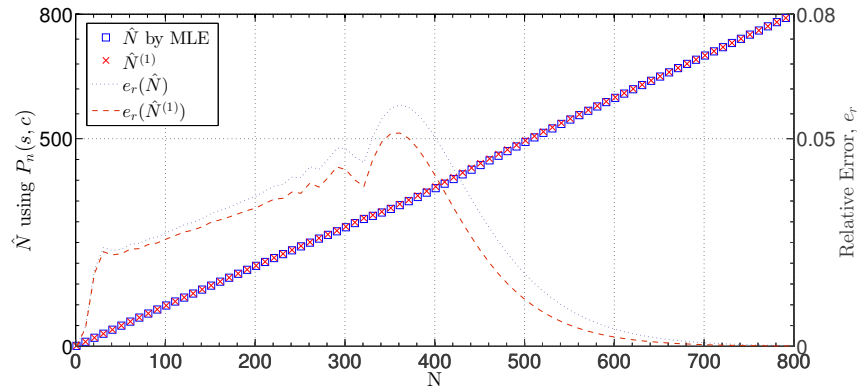


Fig. 5. MLE of contending UEs at eNB using $P_n(s, c)$, $r = 54$.

absolute error of 6% is obtained in the worst case. After refining these results by (8), we obtain an $e_r \leq 5\%$ for all the values of N tested.

5 Conclusions and Future Work

We presented an analytical approach to estimate the number of contending users in one random access slot for an LTE-A network. Based on the preamble information available at the evolved Node B, we find the conditional joint probability distribution function (PDF) for the number of successful and collided preambles within a random access slot. Then, we design a maximum likelihood estimator using this PDF. Numerical results showed that the proposed method can accurately estimate the number of contending MTC devices even for large network loads. This approach can further be extended to design static or dynamic congestion control schemes for alleviating the radio access network overload and optimize the parameter setting of these overload control schemes.

References

1. 3GPP: TR 36.912, Feasibility study for Further Advancements for E-UTRA (Apr 2011)
2. 3GPP: TR 37.868, Study on RAN Improvements for Machine Type Communications (Sept 2011)
3. 3GPP: TS 36.321, Medium Access Control (MAC) Protocol Specification (Sept 2012)
4. 3GPP: TS 36.211, Physical Channels and Modulation (Dec 2014)
5. 3GPP: TS 36.213, Physical layer procedures (Dec 2014)
6. 3GPP: TS 36.331, Radio Resource Control (RRC), Protocol specification (Dec 2014)
7. 3GPP: TS 23.682, Architecture enhancements to facilitate communications with packet data networks and applications (Mar 2016)
8. Arouk, O., Ksentini, A.: General model for RACH procedure performance analysis. IEEE Communications Letters 20(2), 372–375 (Feb 2016)

9. Arouk, O., Ksentini, A.: Multi-channel slotted ALOHA optimization for machine-type-communication. In: Proceedings of the 17th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems. pp. 119–125. ACM (2014)
10. Bertsekas, D.P., Gallager, R.G., Humblet, P.: Data networks, vol. 2. Prentice-Hall International New Jersey (1992)
11. Biral, A., Centenaro, M., Zanella, A., Vangelista, L., Zorzi, M.: The challenges of M2M massive access in wireless cellular networks. *Digital Communications and Networks* 1(1), 1–19 (2015)
12. Cheng, R.G., Chen, J., Chen, D.W., Wei, C.H.: Modeling and analysis of an extended access barring algorithm for machine-type communications in LTE-A networks. *IEEE Transactions on Wireless Communications* 14(6), 2956–2968 (June 2015)
13. Choi, J.: On the adaptive determination of the number of preambles in RACH for MTC. *IEEE Communications Letters* (99), 1–1 (2016)
14. Chu, D.C.: Polyphase codes with good periodic correlation properties. *IEEE Trans. Inform. Theory* 18 (1972)
15. Condoluci, M., Dohler, M., Araniti, G., Molinaro, A., Zheng, K.: Toward 5G densenets: architectural advances for effective machine-type communications over femtocells. *IEEE Communications Magazine* 53(1), 134–141 (January 2015)
16. Ghavimi, F., Chen, H.: M2M communications in 3GPP LTE/LTE-A networks: Architectures, service requirements, challenges, and applications. *Communications Surveys Tutorials, IEEE* 17(2), 525–549 (Secondquarter 2015)
17. Laya, A., Alonso, L., Alonso-Zarate, J.: Is the random access channel of LTE and LTE-A suitable for M2M communications? a survey of alternatives. *IEEE Commun. Surveys Tuts.* 16(1), 4–16 (2014)
18. Lin, T., Lee, C., Cheng, J., Chen, W.: PRADA: prioritized random access with dynamic access barring for MTC in 3GPP LTE-A networks. *IEEE Trans. Veh. Technol.* 63(5), 2467–2472 (2014)
19. Lo, A., Law, Y., Jacobsson, M.: A cellular-centric service architecture for machine-to-machine (M2M) communications. *Wireless Communications, IEEE* 20(5), 143–151 (October 2013)
20. Tello-Oquendo, L., Leyva-Mayorga, I., Pla, V., Martinez-Bauset, J., Casares-Giner, V.: Analysis of LTE-A random access procedure: A foundation to propose mechanisms for managing the M2M massive access in wireless cellular networks. In: Workshop on Innovation on Information and Communication Technologies (ITACA-WIICT 2015). pp. 95–104 (2015)
21. Wei, C., Bianchi, G., Cheng, R.: Modeling and analysis of random access channels with bursty arrivals in OFDMA wireless networks. *IEEE Trans. Wireless Commun.* 14(4), 1940–1953 (2015)
22. Wei, C.H., Cheng, R.G., Tsao, S.L.: Modeling and estimation of one-shot random access for finite-user multichannel slotted ALOHA systems. *Communications Letters, IEEE* 16(8), 1196–1199 (2012)
23. Wei, C.H., Cheng, R.G., Tsao, S.L.: Performance analysis of group paging for machine-type communications in LTE networks. *Vehicular Technology, IEEE Transactions on* 62(7), 3371–3382 (2013)
24. WG2, G.T.R.: Further performance evaluation of EAB information update mechanisms. Meeting N. 77 R2-120270, 3rd Generation Partnership Project (3GPP) (february 2012)
25. Wiriaatmadja, D.T., Choi, K.W.: Hybrid random access and data transmission protocol for machine-to-machine communications in cellular networks. *IEEE Transactions on Wireless Communications* 14(1), 33–46 (Jan 2015)

Drift clock analysis on distributed embedded systems for IoT applications

M. Navia^{1,2}, J.C. Campelo¹, A. Bonastre¹, R. Ors¹, J.V. Capella¹

¹ Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, España
{jcampelo, bonastre, rors, jcapella}@itaca.upv.es

² Carrera de Computación, Escuela Superior Politécnica Agropecuaria de Manabí Manuel Félix López, 10 de Agosto No. 82, 130601 Calceta, Ecuador
mnaviam@espam.edu.ec

Abstract. The use of microcontrollers to build nodes or devices for Internet of Things, commonly briefed as IoT, (and especially for wireless sensor networks WSN) is a common practice. These nodes have constrained resources, due to costs and power consumption restrictions. Many IoT applications, because of its inherent distributed nature, require the nodes to work with a common time base, and thus they have to be synchronized. Therefore, is necessary to evaluate the limitations of the different options offered by microcontroller clocks implementations. This paper analyzes the variability of clock of a STM32F4xx microcontroller, an ARM-based M4 microcontroller, which offers two options as a source for time counting. Through performing tests on several nodes with identical hardware, we have obtained results that characterize the drift caused by any of these time sources, and enhance the need of clock synchronization.

Keywords. Clocks, synchronization, clocks drift, WSN, IoT.

1 Introduction

The Internet of Things (IoT) paradigm has become into a new technological revolution. The possibility to connect many kinds of devices (such sensors, actuators, mobile phones, and others), that interact among them, gives a new dimension to the Internet of the future. Relevant scenarios such as healthcare, smart environments, transports, logistics, etc., are application domains of IoT [1]. In this sense, many researches have been developed, covering most of topics that conforms this paradigm.

The WSN is a fundamental part of IoT. Their main function is to sense environmental aspects, collect data, transport obtained data, and/or perform specific actions. Several approaches for development and deployment of WSN for IoT can found in

literature of Internet. These may originate on citizen-based or industry-based systems, may be standard or not, and may be based on open or closed (proprietary) architectures [2]. Anyway, the devices of these kinds of networks usually are resource constrained, because they are structured by simple and limited hardware. The WSN nodes usually are based on a microcontroller with some input/output and peripherals options. One of the most widely used microcontroller architectures for WSN is the ARM architecture, very popular at last the years due its capabilities.

Many IoT applications require that the involved devices, such as the WSN nodes, use a common base time, even local time (all the nodes synchronize its operation) or external (time base has to be related to real time). In those applications, clocks synchronization is considered a crucial component for IoT operation [3].

Most of IoT systems, such as WSN's are distributed system. The first found reference about synchronization of a distributed system was published at 1978, and it refers how distributed systems can register events in correct order [5]. This work has been used as reference for new proposals, in the same research field, but with novel methods [6].

The challenge of clocks synchronization in WSN's, and many solutions, is briefed in [3]. In [3], a performance analysis of several techniques in this field is included, but it does not analyze how the drift and offset of clocks of different devices can affect this performance. An updated and more complex statistical study is done in [7]. This article is focused on the two-way message exchange mechanisms for synchronization. However, this work does not consider in its equations the variability of clock drift.

A high clock drift may useful in some case. For instance, it can be used as reference for identify nodes, such as the addressing scheme proposed in [10]. Nevertheless, this scheme needs a base station with a clock used as reference skew, but this clock skew may also vary along time.

Although it is not just focused on WSN or IoT approach, a mobile application for measuring of clocks drift in multimedia devices is presented in [8]. The source of clock signal in these devices (crystal oscillators) is the same that the used in WSN nodes. The authors of this tool have tested it on some mobile devices such smartphones or tablets, which usually have more resources than IoT devices. The drifts were found to be different for almost all tested devices, as expected, but authors do not perform tests on several units of the same device.

1.1 Synchronization in IoT

Every microcontroller system requires a clock. Some IoT nodes also include a Real Time Clock (RTC) feature inside them, but must commonly these devices are replaced by software approaches in order to reduce costs. While hardware RTC's have a high-precision crystal oscillator as frequency source for their operation, software approaches may use as time base different clock sources, incorporated in the microcontroller, and supported by different hardware designs. Sometimes, low-quality internal quartz crystals incorporated in the microcontroller are used. In other cases, microcontroller board design includes a more accurate external crystal oscillator. However, these solutions are less exact that hardware RTC's and different devices in the same IoT envi-

ronment may provide different accuracies. Hence, time synchronization of nodes is one of the most important challenges for implementation of IoT [4].

Two error sources in clock operation may be defined.. Clock *drift* (or *skew* in some studies) is the difference between clock periods or frequency. Clock *offset* is the phase difference between clocks. The clock offset varies depending of clock drift, and is different between clocks. The clock drift depends of RTC sources. Most low-cost microcontrollers, as those used on most of sensor and actuator network nodes, usually integrate low-quality crystals oscillators.

This work analyzes the clock drift between microcontrollers, in order to evaluate these differences. Our study demonstrates that, in some cases, the drift may not be negligible, and the use of sophisticated synchronization methods is justified, even taking into account the constrained resources of WSN nodes and most of IoT devices.

2 Testing variability of Drift

2.1 Hardware used for tests

Many architectures, components, and devices have been used, or are suitable to be used, for IoT applications. However, low-cost microcontroller systems are the base of most IoT mobile elements. In these environments, the most used clock source is a crystal oscillator.

As mentioned above, the evaluation of clock drift between devices in the same IoT environment requires some real experiments. This way, as a first step in order to evaluate a complete IoT ecosystem, a reduced three-device experiment was driven. Three identical microcontroller boards were considered for comparison.

For these tests, an ARM-based microcontroller was chosen, a ST32F407VG, incorporated on an evaluation board STM32F4Discovery [9]. ARM architecture was chosen because its popularity when building small embedded devices in many fields, not only in IoT environments, but in others such as smartphones, low-cost devices or single-board computers. Also, although this feature was not used in the tests, these nodes have wireless communication. This device offers two options for RTC source: A Low Speed Internal crystal (LSI) at 32768Hz, and a High Speed External crystal (HSE) at 4MHz. Three identical nodes based on this microcontroller were used for tests.

2.2 Experiments

In order to measure the drift among the three nodes, a common event, generated by a Test PC, was driven into the three boards under evaluation. With the detection of the event, every node recorded its current time and stored it in a non-volatile local memory. The comparison of the recording times on all the three nodes measures the clock drift.

Two different event sources were considered. In first time, the reception of an Ethernet frame sent by the Test PC was used as a event.

Ethernet was chosen because of its great implantation. The use of a hub, which does not store the frame in internal buffers, makes all the nodes to receive the frame at the same time with a very low difference.

To evaluate the jitter introduced by the Ethernet, a second event source was considered. This way, a digital signal was sent to all the three nodes through digital lines, using GPIO (General Purpose Input/Output) pins. This signal was used as interrupt line in all the three boards. Fig. 1 and Fig. 2 show both the connection between nodes and Test Server (through digital lines and Ethernet, respectively).

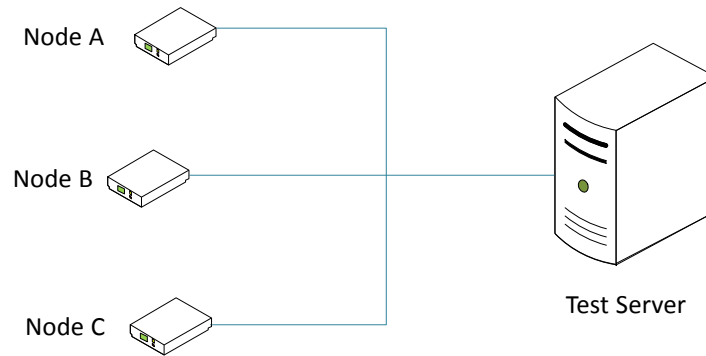


Fig. 1. Digital line connection scheme for experiments

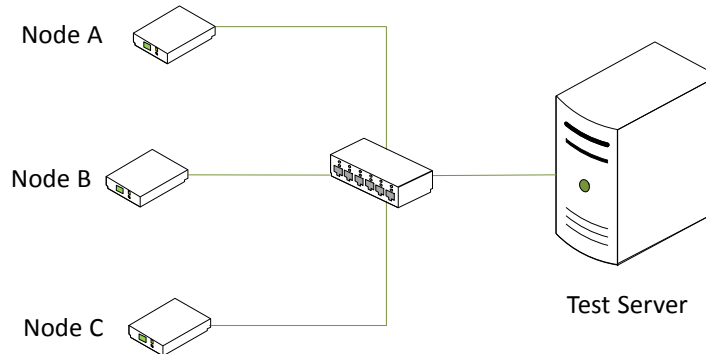


Fig. 2. Ethernet connection scheme for experiments

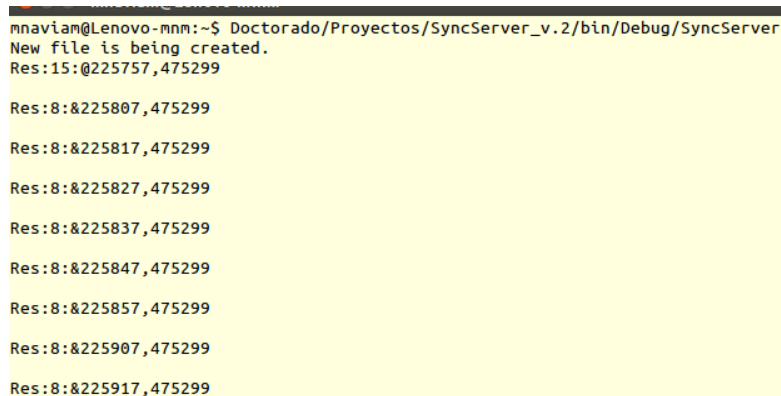
Finally, both (LSI and HSE) options for clock source that are available in the microcontroller board were used. Low Speed Internal crystal (LSI) at 32768Hz is present into the ST32F407VG microcontroller. It is very commonly found in microcontroller architectures because of its simplicity, as it offers low cost and low consumption. On the other hand, the board includes a High Speed External crystal (HSE) at 4MHz. This

crystal offers better performance and higher precision, but increases power consumption.

2.2 Performed tests

Four experiments were designed, combining the two crystal sources (LSI and HSE) with two event generation possibilities (Ethernet and Parallel via GPIO).

In a PC acting as the Test Server a program generates periodically (each 10 seconds) an event to be registered by all three nodes at the same time. Detection of this event causes the nodes to store the local time in a text file for further analysis. Fig. 3 shows a screen capture of the program. The time of each event generation is reflected in *hhmmss,useg* format, which is also used when writing in the text file.



```
mnaviam@Lenovo-mnm:~$ Doctorado/Proyectos/SyncServer_v.2/bin/Debug/SyncServer
New file is being created.
Res:15:@225757,475299

Res:8:&225807,475299

Res:8:&225817,475299

Res:8:&225827,475299

Res:8:&225837,475299

Res:8:&225847,475299

Res:8:&225857,475299

Res:8:&225907,475299

Res:8:&225917,475299
```

Fig. 3. Screen capture of program used for tests. The program generates an event and stores the time in a file

The clock in all the nodes is enabled and configured to be initialized with a predefined time when they boot up. A common reset line is used to synchronize the initial time. Each time that a node receives an event from the Test Server, a timestamp of local time is stored in a non-volatile memory (SD card).

Both Ethernet based events and GPIO based events have been studied. Ethernet events are received through a NIC (Network Interface Controller) added to the F32Discovery. Reception interrupt – generated by the NIC – is used to trigger the time stamp recording procedure. In order to reduce the variance of Ethernet transmission, a hub (with no frame storage-and-forward mechanisms) has been used. This experiment was necessary because it will be the usual procedure to achieve synchronicity among the nodes in future developments.

On the other hand, to verify the correctness of the Ethernet approach, a direct GPIO interrupt-driven event has been also studied. In this event generation mechanism, a digital line triggers a high-priority interrupt in all three nodes. This mechanisms, de-

spite of the fact that is more simple, it is not suitable to be applied when the number of nodes to synchronize increases.

Ten minutes length simulations have been performed, being generated 60 events for each execution. Absolute capture time of each event in each node and recorded time of last event – that summarizes all the measures – were considered.

In order to obtain reliable results, n replications have been performed for each measurement, n being calculated as follows:

The results for each measurement have been considered as random variables (X_1, X_2, \dots, X_n) with a μ mean value. n measurements have been repeated until an estimation of μ has been obtained with a 90 % confidence interval according to the Equation 1, where $t_{n-1,0.95}$ represents the upper limit of the Student's t-distribution on $n-1$ degrees of freedom, and $\bar{X}(n)$ and $S^2(n)$ are the mean and the variance of the results obtained in the different experiments.

$$\bar{X}(n) \pm t_{n-1,0.95} \sqrt{S^2(n)/n} \quad (1)$$

3 Tests Results

After performing the battery of tests, the timestamps stored in each node were downloaded to a computer for the analysis of the obtained data. For each experiment (combination of clock source and event notification mechanism), some of the results are shown in Table 1. The time format in Table 1 is *hh:mm:ss,useg*.

Table 1. Average last event recording time in nodes

CLK source	Trigger	Node A	Node B	Node C	Max difference
LSI	GPIO	00:10:36,914364	00:09:21,148519	00:10:15,517819	00:01:15,765845
HSE	GPIO	00:10:00,041362	00:10:00,033442	00:10:00,041361	00:00:00,007920
LSI	ETH	00:10:27,072259	00:09:11,703967	00:10:05,747093	00:01:15,368292
HSE	ETH	00:10:09,878117	00:10:10,041762	00:10:10,049842	00:00:00,171725

Figure 1 brings many interesting conclusions. It must be remarkable that clock difference among nodes is very notable. In the worst experiment, a 10% clock drift has been detected. These differences may be caused by environmental conditions – such as temperature in microcontroller, electromagnetic interference, etc. – and even caused by imperfections in the oscillators used as clock source.

Looking at rows with the same clock source, it can be seen that in both LSI and HSE clock sources, Ethernet and GPIO generate the same results. This validates the use of Ethernet as event propagation method in further studies.

When comparing LSI and HSE performance as clock source, it must be noted that HSE offers far better performance, but its costs and power consumption is greater than internal LSI. This difference could be expected because manufacturers usually include low-cost crystals inside microcontrollers. Nevertheless, three-magnitude-order differences should make the microcontroller's designers think about the suitability of too

cheap crystals in many applications. Finally, without corrective methods, LSI must be discarded as a primary clock source in most IoT applications.

On the other hand, HSE offers better accuracy, and thus the difference between nodes reaches some milliseconds. For some IoT applications this may be tolerable in short periods, but power consumption has to be taken into account.

In all cases, the performed tests enlighten the necessity of node synchronization beyond the local clock, and thus a synchronization algorithm must be considered when implementing IoT applications.

4 Conclusions

In this paper has been shown that clock sources of embedded system may suffer problems in their correct operation that may prevent the use of local microprocessor time in IoT applications.

Most of current microcontroller-based systems minimize cost and power consumption by the incorporation of a low-cost internal clock source for Real Time Clock implementation. These internal sources must be deeply studied in order to avoid drift and offset problems when applied to IoT applications.

External clock sources, like HSE used in our tests, usually obtain more accuracy than low-cost internal clock sources. It should be analyzed, depending on the application, if the use of a crystal like LSI is precise enough to fulfill the requirements. This solution may use less energy but requires additional synchronism methods that may cause more overhead due to message interchange. Other solutions may rely on the use of an external crystal like HSE, which use more energy but requires less frequent updates, thus reducing the synchronization message overhead but increasing both cost and power consumption.

Acknowledgements

This research was supported by the Universitat Politècnica de València under Research Projects VLC/Campus UPV SP20150050.

References

1. L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Networks*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.
2. A. Kotsev, S. Schade, M. Craglia, M. Gerboles, L. Spinelle, and M. Signorini, "Next Generation Air Quality Platform: Openness and Interoperability for the Internet of Things," *Sensors*, vol. 16, no. 3, p. 403, Mar. 2016.
3. Y.-C. Wu, Q. Chaudhari, and E. Serpedin, "Clock Synchronization of Wireless Sensor Networks," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 124–138, 2011.

4. D. Datla, X. Chen, T. Tsou, S. Raghunandan, S. M. S. Hasan, J. H. Reed, C. B. Dietrich, T. Bose, B. Fette, and J. Kim, "Wireless distributed computing: a survey of research challenges," *Commun. Mag. IEEE*, vol. 50, no. 1, pp. 144–152, 2012.
5. L. Lamport, "Time, clocks, and the ordering of events in a distributed system," *Commun. ACM*, vol. 21, no. 7, pp. 558–565, Jul. 1978.
6. P. Eugster, V. Sundaram, and X. Zhang, "Debugging the Internet of Things: The Case of Wireless Sensor Networks," *IEEE Softw.*, vol. 32, no. 1, pp. 38–49, Jan. 2015.
7. X. Wang, D. Jeske, and E. Serpedin, "An Overview of a Class of Clock Synchronization Algorithms for Wireless Sensor Networks: A Statistical Signal Processing Perspective," *Algorithms*, vol. 8, no. 3, pp. 590–620, Aug. 2015.
8. M. Guggenberger, M. Lux, and L. Böszörményi, "ClockDrift: A Mobile Application for Measuring Drift in Multimedia Devices," in *Proceedings of the ACM International Conference on Multimedia - MM '14*, 2014, pp. 767–768.
9. STMicroelectronics, "STM32F4DISCOVERY Discovery kit with STM32F407VG MCU." [Online]. Available: http://www.st.com/content/st_com/en/products/evaluation-tools/product-evaluation-tools/mcu-eval-tools/stm32-mcu-eval-tools/stm32-mcu-discovery-kits/stm32f4discovery.html. [Accessed: 16-May-2016].
10. M. M. Kassem, H. S. Hamza, and I. A. Saroit, "A Clock Skew Addressing scheme for Internet of Things," in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, 2014, pp. 1553–1557.

Simulation of Telecom Traffic Sources in Hospitals

Yang LI¹, Angel Gomez-Sacristan², Miguel A. Rodriguez-Hernandez²

¹International School, Beijing University of Posts and Telecommunications, 102209 Beijing, China

²ITACA, Universitat Politècnica de València, Camino Vera S/N, 46022 Valencia, Spain

Abstract. The objective of this paper is to simulate the communications of a network with several smart hospitals in order to evaluate the performance and the quality of service (QoS). Several traffic sources have been classified and characterized in this work. Proper design of the traffic sources would lead to a better quality of service and lower costs as well. Human Type Communication and Machine Type Communication will be studied in order to make the design of resources more accurate. Meanwhile a new scenario with 2 hospitals has been designed. Because all the heterogeneous services will share a limited bandwidth, a QoS mechanism will be introduced. After simulation, it is found that the requirements of different services had been achieved by QoS design.

1. Introduction

In modern society, traditional services will have new features with rapid development of technology. Smart Hospital is a specific scenario. It combines the services of hospital and telecommunication, aiming to serve patients better. Fig. 1 shows the general model of the smart hospital scenario. It consists of different services, network and cloud server. This work is based on the hospital scenario, using simulation tool OMNeT++ to evaluate the performance of communication.

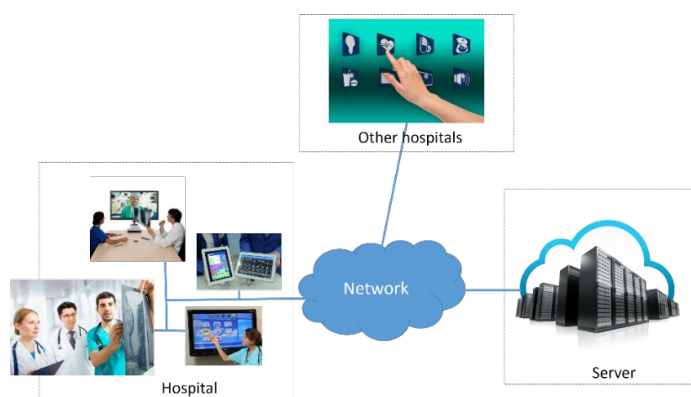


Fig. 1 Smart hospital communications model

In previous work [1], there are some examples of traffic sources in a general scenario, such as schools, financial, etc. There are two main types of communication using in the scenario, human type communication (HTC) and machine type communication (MTC). In HTC, voice and video sources will be characterized. In MTC, text sources and image sources play an important role in the work. After learning examples and studying how to define new services in XML file, new services will be

classified and characterized. Patient information, diagnosis images and sensors for monitoring, all these things could be new traffic sources in the smart hospital network.

According to limited bandwidth and all the services will share the same bandwidth, there exist a balance between cost and quality [2]. If all services are assigned with multimedia which has the highest priority, the costs will be pretty high. Thus, different Quality of Service (QoS) types for different services will be assigned in order to make both efficiency and lower cost. In this work, voice sources will be assigned with multimedia type, video sources and important data with gold type, rest of data sources will be silver type. Two rate three color marker (TR-TCM) algorithm is used to manage this traffic [3]. Traffic belonging to “silver” quality will be discarded in congestion periods.

After added new hospital with new services to the scenario, the network capability and whether QoS mechanism works or not will be observed [4].

This article is structured as follows: Chapter 2 presents characterization of traffic sources, chapter 3 explains simulation scenario, and chapter 4 describes the results of simulation and analysis. Finally, chapter 5 draws a conclusion and presents further works.

2. Traffic sources in hospitals

2.1 Characterization of voice source

The voice is part of the HTC. When transmitting voice information through the network, it is necessary to convert the analogue source to digital samples. In this work, G.711 encoder is used to generate a 64Kbit/s stream [5].

In addition to choose encoder, profiles design of voice sources is important as well. The voice sources have active part and in active part. In the time active part, T-on and T-off model is using to represent speaking and silence part during a call [6]. In time inactive part, it won't generate packets. Send interval is the time gap between two consecutive packets, and it is defined by the encoder. Table 1 shows the G.711 codec characteristics.

Table 1 Voice G711 codec characteristics

ITU-T Codec	Codec type	Maximum codec delay	Bitrate (bps)	Packetization interval (ms)	pps	Payload size (bytes)	IP pkt size (bytes) ⁱ	IP bps
G.711	PCM	0.375	64 000	10	100	80	120	96
G.711	PCM	0.375	64 000	20	50	160	200	80

In voice source, there are two sub-types, first is the normal calls happened in the hospital, and the second is call center which is responsible for consultations and customer care.

The duration of normal call is modelled by the exponential distribution and having a maximum 3 minutes and a minimum 1 minutes and a mean 2 minutes. And the silence time between calls is modelled by exponential distribution as well, which has a maximum 120 minutes, a minimum 0 minutes and a mean 60 minutes. Another service is call center, it will response the patients' questions, so the frequency of calls would be higher, which will be reflected by the time inactive distribution. It has a maximum 10 minutes, a minimum 0 minutes and a mean 2 minutes.

2.2 Characterization of video source

Video is another subtype of HTC, and it is more complicated than voice sources. The parameters of video sources are more than voice. In addition to set the duration and inactive time, the movement type need to be set and the frames per second as well. Before designing the profiles of video sources, the first thing is to set the video quality. In this work, H.264 codec is used, and four different qualities are defined: standard, high quality, high definition and high definition plus.

The basic information is shared by all the services, but each service has its own profile. In time active part, the video is transmitted, and in time inactive part, the source dose not generate traffic. Besides, new parameters appear in video profiles, the movement and the FPS (frames per second). Movement parameter indicates the movement level of the video, it varies from 1 to 4, the bigger the number, higher the movement level.

There are three services appears in Smart Hospital scenario, tele-consulting, tele-care and tele-presence. The detailed profile of each service is in Table 2.

Table 2 Video profiles

Service		tele-consulting	tele-care	tele-presence
Duration Type = exponential	Max	15	10	60
	Mean	7	5	30
	Min	3	1	20
Time inactive Type = exponential	Max	20	10	14400
	Mean	5	2	300
	Min	0	0	100
Movement Type = constant		2	1	3
FPS Type = constant		25	25	25

2.3 Characterization of data source

There is another type of communication called MTC, which includes text and image resources. And these sources are different from HTC. When the data sources transmit, TCP could be used instead of UDP, which has connection-oriented feature; it can order the packets and re-order at receiver [7].

These sources can be found in diagnostic, regular checking, monitoring and file transferring. Each service will have its own identifier, number of resources and the resource size [8]. In this work, there are four main data services, and it is necessary to characterize each of them. Table 3 is the summary characterization of data source.

Table 3 Data size

Service	Text resource size			Number of TR		Image resource size			Number of IR	
	Type=exponential			Type=uniform		Type=exponential			Type=uniform	
Chronic patient sensors	Max	Mean	Min	Begin	End	Max	Mean	Min	Begin	end
	10K	5K	100	0	10	500K	300K	100K	0	2
	Type = exponential			Type =uniform		Type = exponential			Type =uniform	
HIS	Max	Mean	Min	Begin	End	Max	Mean	Min	Begin	end
	1M	100K	10K	0	10	500K	300K	100K	0	2
	Type = exponential			Type =uniform		Type = exponential			Type =uniform	
DICOM	Max	Mean	Min	Begin	End	Max	Mean	Min	Begin	end
	3M	1M	100K	0	2	30M	10M	1M	0	2
	Type = exponential			Type =uniform		Type = exponential			Type =uniform	
Administrative services	Max	Mean	Min	Begin	End	Max	Mean	Min	Begin	end
	10M	100K	1M	0	5	100K	200K	500K	0	5

In MTC, four parameters are used to determine how data sources behave in the network: Number of packet, Send interval, Number of burst and Burst interval.

2.4 New services design

New services could be hospital infrastructure, medical equipment, and the diagnosis files transmitting among several departments. In this specific scenario, 10 new services of MTC have been defined in table 4 [9].

Table 4 New services

Services	Text resource size		Number of Text resource			Image resource size		Number of Image resource		
	Type	Value	Type	Value		Type	Value	Type	Value	
Temperature	C	80	U	Begin=1	End=20	C	0	C	1	
Humidity	C	80	U	Begin=1	End=20	C	0	C	1	
Air Quality	C	516	U	Begin=3	End=10	C	0	C	1	
Electrocardiogram	C	132	C	1		U	Begin=933K End=98M	C	1	
Photoacoustic	C	132	C	1		C	3M	C	1	
Bone Scan	C	132	C	1		C	1M	C	1	
Thermal	C	132	C	1		C	368K	C	1	
Breath monitor	C	96	C	0		C	0	C	1	
BMET	C	164	U	Begin=30	End=50	C	0	C	1	
HL7	E	Max Median Min	U	Begin=0	End=10	E	Max 500K Median 200K min 100K	U	Begin=0	End=5

C:Constant, E:Exponential, U: Uniform

3. Experiments description

In this chapter, it will describe the simulation scenario.

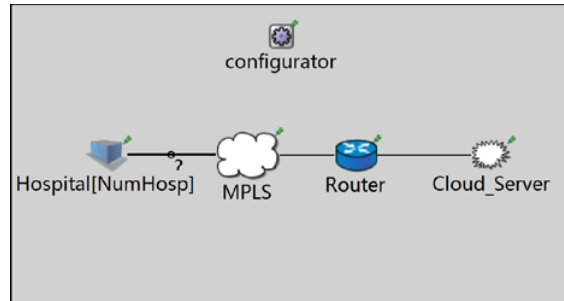


Fig. 2 Simulation scenario

This topology is a general model of the smart hospital communications. Each module has its own function.

- 1) Hospital: it is the host in the network, it has several traffic resources.
- 2) MPLS: it using multi-protocol label switching protocol to do the traffic engineering.
- 3) Router: it is the main router in this network, it connect the hospital and the cloud. The Quality of Service mechanism is built in this part.
- 4) Cloud_Server: it is the remote server in this network.
- 5) Configuration: it refers to the configuration file, which is usually called omnetpp.ini, it consists of all the details of the scenario, like the general setting of the network and definition of each resource, etc.

When the simulator launches, it will first read the NED file to generate network topology, and then read the omnetpp.ini to set all the parameters in the network in order to control how the simulation is executed.

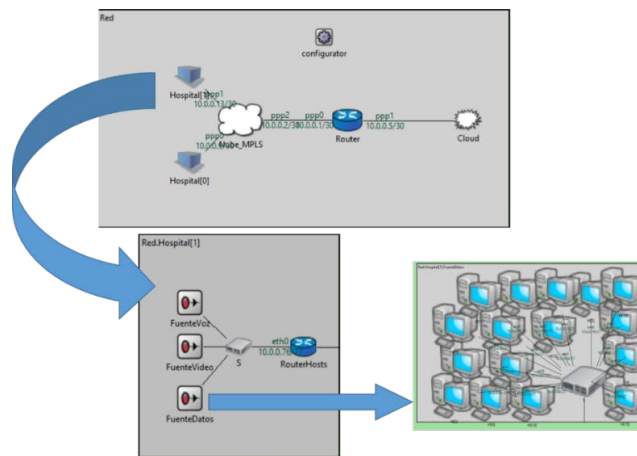


Fig. 3 Topology of new scenario and inner structure of hospital

Fig. 3 shows the new scenario topology of two hospitals. Although the inside structure of each hospital is similar, the number of source, and the type of source can be different. Taking this case as example, there are two hospitals, Hospital (0) and Hospital (1). Both hospitals have three sources: voice, video and data, but in Hospital (1), the number of voice and video sources is 0, it only contains the new data services defined above.

4. Results and discussion

4.1 One hospital scenario simulation

The simple scenario with one hospital has been simulated, the number of each source has been defined in Table 7. Fig. 4 to 8 are parts of result of simulation.

Table 5 Number of sources in the scenario

Name of source	Number of source
VoIP	300
Call center	20
tele-consulting	10
tele-care	20
tele-presence	1
Chronic patient sensors	500
HIS	50
DICOM	5
Administrative services	50

Fig. 4 shows the bandwidth occupied at the network interface of cloud. From the Fig.4, it can be easily found that total usage is less than 350Mb/s.

Fig. 5 describes the packets sent and received of voice sources. Because of multimedia type, the priority of voice sources is highest in all sources, so the packet loss rate is very low; server almost received all the packets that the hospital sent.

Fig. 6 indicates the voice signal jitter. Jitter is an average of deviation from the network mean latency. It clearly shows the variation is between 0.02ms (both positive and negative), and it could be neglected when transmitting, and it indicates that the QoS meets the requirements of voice sources.

Fig. 7 represents the bandwidth consumed by different video services. Blue line indicates the tele-consulting, green one correspond to tele-presence and red one is for tele-care. It can be found that different definition of resources will result to different behavior in the network. For tele-consulting, it used up to 45Mb/s to transmit information, in comparison, tele-care with lower quality keeps the consumption below 10Mb/s. The bandwidth usage for tele-presence will between 8Mb/s and 17Mb/s.

Fig. 8 shows the delay of data traffic sources, because most of data traffic sources are assigned with silver type, the delay is greater than other type of sources. The Fig. 8 clearly shows that the maximum delay is 0.25s.

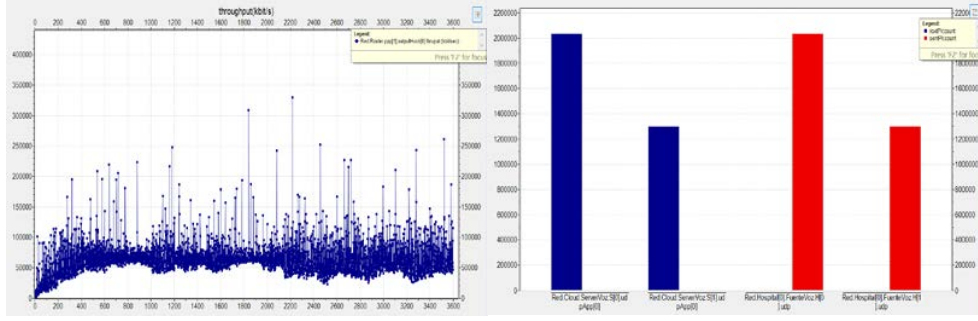


Fig. 4 Bandwidth consume at cloud server interface

Fig. 5 Voice packets

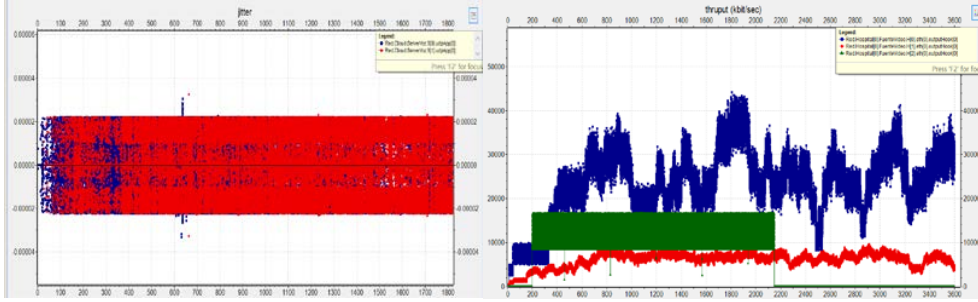


Fig. 6 Jitter of voice

Fig. 7 Bandwidth of video sources

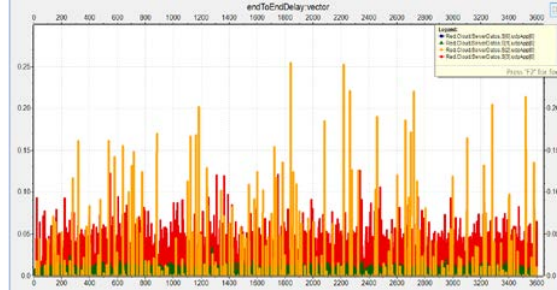


Fig. 8 Delay of data sources

4.2 New scenario simulation

In this scenario, a new hospital with newly defined services is added into the original scenario. In order to get enough information, the simulation time lasts for 2 hours. Fig. 9 to 12 are 4 results of the simulation.

Fig. 9 describes the bandwidth usage of two different hospitals. Blue line represents Hospital (0) and red line represents Hospital (1). It clearly shows that usage of Hospital (0) is greater than Hospital (1), it is still below 350Mb/s and similar to one hospital scenario. For Hospital (1), the usage is smaller than Hospital (0), it is below 150Mb/s.

Fig. 10 shows the end to end delay of all the data resources. It is found that when two hospitals connect to the same network, the delay of data will significantly increase. The maximum delay of a data service is around 8s.

Fig. 11 and 12 are results of information, compared with one hospital scenario, it is found that the delay and jitter are similar to the original results. So the requirements of voice still can be met properly, delay and variation can be neglected again.

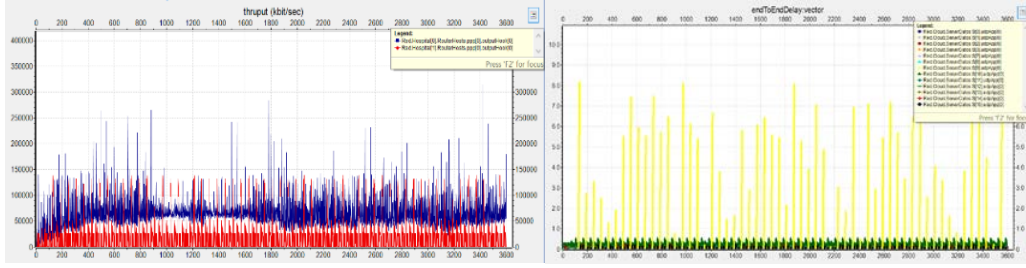


Fig. 9 Bandwidth consumption of two hospitals

Fig. 10 End to end delay of data sources

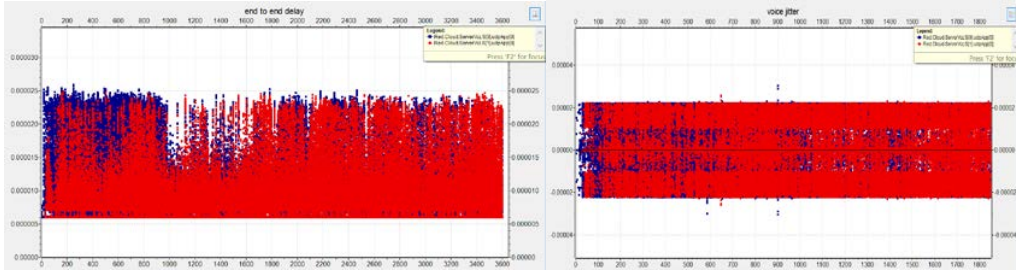


Fig. 11 Delay of voice

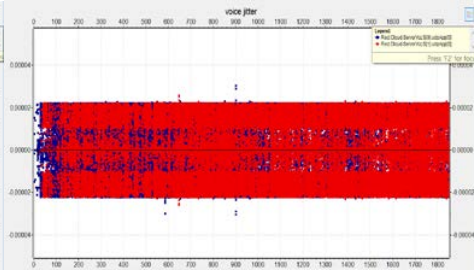


Fig. 12 Voice jitter

5. Conclusions and future work

The properly designing of communication sources will definitely improve the satisfaction degree of patient and stuffs in hospital as well as the efficiency. All the traffic sources will share the same bandwidth, so the different quality of service type will be assigned to different services to make sure the most important information will delivered properly. This part can be adjusted easily in the simulator, the balance between costs and quality can be tested conveniently. Meanwhile, the behavior of HTC and MTC sources can be observed, those sources which assigned to lower quality, will directly lead to big end to end delay. In this work, new services can be easily added to the existed scenario without introducing extra errors.

In future, rapid development of medical equipment will lead to great number of traffic sources appears, characterization of these sources can be added into database for flexible use. Meanwhile, the simulator can be updated as well, making the software more user friendly by designing graph user interface.

Acknowledgement:

This work was partially supported by Spanish Government MEC Project TIN2013-47272-C2-1-R. Thanks to school BUPT for giving Yang LI the chance to do a project in Valencia. Yang LI also want to thanks Polytechnic University of Valencia.

References

- [1] Angel Gomez-Sacristan, Miguel A. Rodriguez-Hernandez, and Victor Sempere, Evaluation of Quality of Service in Smart-Hospital Communications, *Journal of Medical Imaging and Health Informatics*, vol. 5, pp. 1864-1869, 2015.
- [2] John Evans & Clarence Filsfils. Deploying IP and MPLS QOS for Multiservice Networks: theory and practice (ISBN 13: 978-0-12-370549-5)
- [3] J. Heinane, R. Guerin (1999), A Two Rate Three Color Marker, RFC 2698
- [4] David Gomez Cuadrado, Simulation of All-IP access network with configurable traffic sources. Polytechnic University of Valencia, Spain.
- [5] Cisco engineers, Voice over IP - Per Call Bandwidth Consumption, updated: Apr 13, 2016, Document ID: 7934 web site: <http://www.cisco.com/c/en/us/support/docs/voice/voice-quality/7934-bwidth-consume.html>
- [6] A.D. Shaikh, K.J Blow, M.A. Eberhard and S.A. Fowler, Language independent on-off voice over IP source model with lognormal transitions, *IET Communications*, 7 (2013), DOI: 10.1049/iet-com.2012.0686, pp.1449 – 1455
- [7] Rossitza Golevam, Dimitar Atamian, Seferin Mirtchev, Desislava Dimitrova and Lyubina Grigorova, 3G network traffic sources measurement and analysis, *transactions on emerging telecommunications technologies*, 2014; 25:798-814
- [8] Service requirements for Machine-Type Communications (MTC), 3GPP TS 22.368 V1.2.2 (2010-02) web site: <http://www.3gpp.org/DynaReport/22368.htm>
- [9] Angel Gomez-Sacristan, Miguel A. Rodriguez-Hernandez, Victor Sempere Paya, Telecom Services Design in Smart-Hospital Communications, 2016 global medical engineering physics exchanges/pan American health care exchanges (GMEPE/ PAHCE).
- [10] Aeroqual, Portable & Fixed Monitor Gas Sensor Specifications, Series 900 – Fixed Indoor Air Quality Meter. Web site: <http://www.aeroqual.com/product/series-900-fixed-gas-monitor>
- [11] radub85, Vector - Modern Electrocardiogram Monitor Display, 123RF, image ID: 23442573, web site: http://www.123rf.com/photo_23442573_modern-electrocardiogram-monitor-display.html

A dynamic access class barring method to avoid congestion in LTE-A networks with massive M2M traffic

Israel Leyva-Mayorga, Luis Tello-Oquendo, Vicent Pla, Jorge Martinez-Bauset,
and Vicente Casares-Giner

ITACA, Universitat Politècnica de València,
Camino de Vera s/n. 46022, Valencia, Spain.
email: {isleyma,luiteloq,vpla,jmartinez,vcasares}@upv.es

Abstract. In the near future, a massive number of machine-to-machine (M2M) communication devices will provide with ubiquitous information and services. Nevertheless, the actual infrastructure of relaying networks may not be capable of handling such a large number of interconnected user equipments (UEs). This is the case of LTE-A networks, where the random access channel suffers from congestion whenever a bulk of UEs transmit in a highly synchronized manner. As such, the use of congestion control methods, such as access class barring (ACB), is needed. But ACB reduces congestion in exchange of a higher access delay. Hence, maintaining an active ACB induces unnecessary delay to the access of UEs during periods of low congestion. In this paper, we present a novel approach for the dynamic modification of ACB parameters that enhances the UE accesses during periods of high congestion and avoids excessive delay during periods of low congestion.

Keywords: Access class barring; dynamic congestion-control; LTE-A networks; machine-to-machine (M2M) communications.

1 Introduction

Modern society is in the need for ubiquitous device connectivity, where small devices, known as user equipments (UEs), exchange data autonomously to provide continuous access to information and services. Machine-to-machine (M2M) communication stands for the autonomous exchange of data between UEs and is a fundamental component of the Internet of Things (IoT) [2]. But, M2M applications pose important engineering challenges regarding the signaling capabilities of relaying networks. For instance, LTE-A networks present the best option for UE interconnection as its infrastructure has already been largely deployed [7]. Nevertheless, it has been observed that the signaling capabilities of cellular base stations (evolved NodeBs, eNBs, in LTE-A) can be exceeded when a bulk of UEs transmit in a highly synchronized manner (this is a typical behavior in M2M applications) as the random access channel (RACH) of LTE-A was designed to

handle human-to-human communications, where a few UEs (when compared to M2M communications) attempt to access the eNB [5].

The UEs access the eNB by means of the random access procedure (RAP), which comprises a four-message handshake; i.e., preamble transmission (only permitted during random access opportunities, RAOs), random access response (RAR), connection request and contention resolution messages. The RAP will be described in Section 3. The performance of the LTE-A system is measured in terms of several performance indicators, which determine whether the access is being conducted efficiently. These include, among others, access success probability, access delay and the number of access attempts. The degradation of these performance indicators occurs whenever the network suffers from congestion, which is by no means desirable. To enhance the performance of the RACH, several methods have been proposed in the literature.

Access class barring (ACB) is a congestion-control method that redistributes the UE accesses through time. For this, a portion of the UEs delay the transmission of the first message of the RAP according to the barring rate and the mean barring time broadcast by the eNB. As a matter of fact, when ACB is implemented and correctly configured, the congestion in the RACH can be reduced in exchange of a longer wait to access the eNB. But maintaining a constant barring rate and barring time during periods of low congestion leads to a notorious and unnecessary increase in access delay. Building on this, several dynamic ACB methods that periodically modify its parameters have been proposed [12, 9]. But the dynamic selection of ACB parameters and, more importantly, the mechanism to activate/disable these methods is not straightforward.

In this paper, we propose and evaluate a dynamic ACB method (DACB) that relies on the periodic calculation of a load coefficient (LC). The purpose of our DACB method is to prevent congestion in the RACH. By doing so, the access success probability of UEs is maximized during sporadic periods of high congestion, while the access of UEs during periods of low congestion is not significantly affected.

The rest of the paper is organized as follows. The review of the literature of massive M2M communications through LTE-A and congestion-control methods is presented in Section 2. Then, we describe the random access procedure, along with the traditional ACB and our dynamic ACB method (DACB) in Section 3. The basic configuration of the random access channel (RACH) and the selected traffic models are described in Section 4. Results derived from our performance analysis are presented in Section 5. The article concludes with the discussion of results and future work.

2 Related work

Several studies have concluded that the current RA procedure (RAP) of LTE-A is not capable of handling massive M2M communications [1, 8, 10, 13]. As such, several methods have been proposed to enhance the performance of the LTE-A RACH [5, 6]. Among these, access class barring (ACB) methods are part of the

LTE-A specification [3]. The purpose of these methods is to spread the UE accesses through time by restricting the access of certain classes of UEs. ACB is oftentimes modeled as a static method, in which the configuration parameters remain the same throughout its operation, and most studies agree with its efficiency during sporadic periods of congestion [6, 5]. Nevertheless, to the best of our knowledge, no study has evaluated the impact on access delay of a static implementation of ACB during periods of low congestion, where an unnecessary access delay may be induced.

The dynamic modification of ACB parameters has been proposed previously as an efficient solution for enhancing the access success probability with a minimum increase in access delay [11, 9]. This is the case of [12], where the authors propose a dynamic barring method in which the ideal barring factor is calculated at each RAO. To calculate the ideal barring factor, the authors estimate the number of contending UEs based on the probability distribution of successful and failed accesses. But this information may not be known by the eNB. This is recognized in [9], where the authors propose the use of a state transition diagram for the dynamic activation of a barring method. Here, the state of the system depends on the average number of successful preamble transmissions and the dynamic barring method is activated when the system reaches the state of severe congestion. However, the authors do not consider that the maximum number of uplink grants may be lower than the number of successful preambles, which is a relevant limitation of the RACH. In this paper, we propose a simple dynamic ACB method (DACB) that considers the limitations of the RACH.

3 Random access procedure

The UEs access the eNB by means of the RAP, which comprises a four-message handshake. First, the timing configuration of the random access channel (RACH) must be obtained by the UEs through the System Information Blocks, which are periodically broadcast by the eNB. Then:

Msg1: The UEs perform the random selection and transmission of one out of the R available preambles towards the eNB during one of the available random access opportunities (RAOs). These preambles are orthogonal sequences that lack any type of identification field. A collision occurs if the same preamble is transmitted by multiple UEs in the same RAO, i.e., the eNB is unable to detect the transmitted preamble [1].

Msg2: The eNB sends up to N_{UL} uplink grants to the UEs per RAO (up to one for each detected preamble) by means of random access response (RAR) messages. Uplink grants assign time-frequency resources to the UEs for the transmission of *Msg3*. After the transmission of *Msg1*, the UEs wait for a random access response window, W_{RAR} , to receive the uplink grant. If the uplink grant is not received within the W_{RAR} , the UE shall increase its preamble transmission counter, k , ramp up its power, perform backoff and go back to the transmission of *Msg1* if the maximum number of preamble transmissions, *preambleTransMax*, has not been reached. Otherwise, the RAP is terminated.

Msg3: The UEs that received an uplink grant send its connection request message (*Msg3*) during the time-frequency resources specified by the eNB. As such, no collisions can occur during the transmission of this message.

Msg4: Finally, the eNB responds to each *Msg3* transmission with a contention resolution message, *Msg4*. This is the end of the RAP and now the UEs proceed with the transmission of their data packets towards the eNB.

3.1 Access class barring (ACB)

Access class barring (ACB) is a congestion control method that aims to redistribute the first preamble transmission of UEs through time to reduce the number of access requests per RAO. In ACB, every UE that belongs to the normal classes (0 to 9) shall acquire the mean barring time, T_{ACB} , and barring rate, P_{ACB} , from the eNB through the System Information Block Type 2 (SIB2). SIB2 is broadcast periodically by the eNB. Then, the UEs perform ACB checks at the beginning of the RAP; i.e., only before its first preamble transmission. At each ACB check, the UEs transmit its preamble with probability P_{ACB} . Otherwise, the UE waits for a random time,

$$T_b = (0.7 + 0.6 \times U(0, 1)) \times T_{ACB} \text{ seconds.} \quad (1)$$

In highly congested scenarios, the static implementation of ACB has proven to be a valuable congestion control method given an adequate selection of P_{ACB} and T_{ACB} . In the practice, the implementation of a static ACB induces an unnecessary delay in the UE accesses during periods of low or no congestion.

3.2 Dynamic ACB (DACB)

In this section we present a dynamic access class barring method (ACB) and the load coefficient (LC) that is used to calculate the barring rate at the i th RAO, $P_{ACB}[i]$. It is worth noting that in this study we select a constant $T_{ACB} = 1$ s as its dynamic selection is not straightforward. Herein we assume that the eNB updates LC at each RAO and sends this information to the UEs as in [12].

At the end of the i th RAO, the eNB is only aware of the number of successfully decoded preambles, $N_{sp}[i]$, and the number of uplink grants that will be sent within the next W_{RAR} , $N_s[i]$. Other information such as the number of UEs that transmit its k th preamble at the i th RAO, $N[i, k]$, the total number

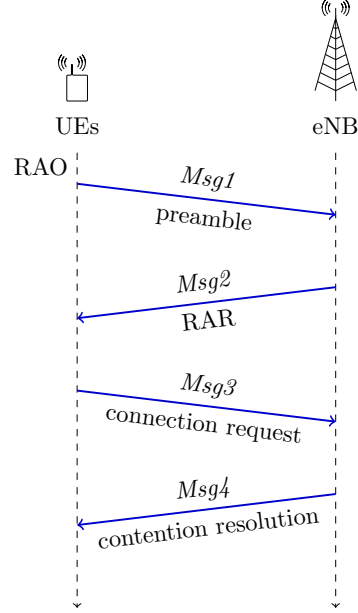


Fig. 1. Four-message handshake in the contention-based random access procedure (RAP) of LTE-A.

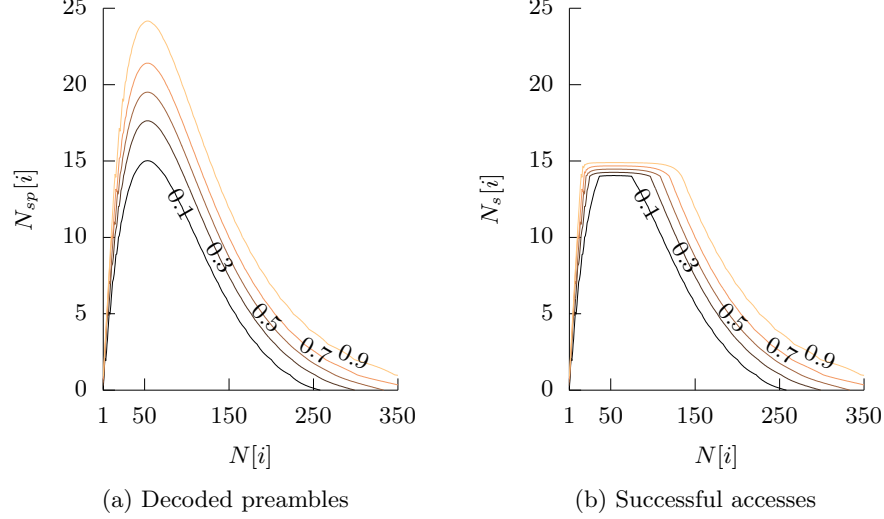


Fig. 2. Cumulative distribution function (CDF) of the decoded preambles, $N_{sp}[i]$, and the successful access, $N_s[i]$, given $N[i] \in \{1, 2, \dots, 350\}$.

of transmitted preambles per RAO and the number of colliding UEs is certainly not known by the eNB. Building on this, the eNB can only measure the traffic load in terms of $N_{sp}[i]$ and $N_s[i]$. Fig. 2a and Fig. 2b show the cumulative distribution function (CDF) of $N_{sp}[i]$ and $N_s[i]$ for $N[i] \in \{1, 2, \dots, 350\}$ in a typical RACH configuration (further explained in Section 4), where the number of available preambles is $R = 54$ and the number of uplink grants per RAO is $N_{UL} = 15$.

Instead of attempting to control the congestion once it has occurred, we adopt a preventive approach towards the use of ACB. For this, we propose the use of a load coefficient, given as

$$\ell[i] = \frac{\max\{N_s[i] - 1, 0\}}{N_{UL} - 1}, \quad (2)$$

to control the barring rate at each RAO as

$$P_{ACB}[i] = 1 - \ell[i]. \quad (3)$$

The block diagram for the implementation of the DACB is shown in Fig. 3.2.

Please observe that the access of UEs will not be affected if $N_s[i] \in \{0, 1\}$ since $\ell[i] = 0$. This is a typical scenario of low congestion in LTE-A, where few preamble transmissions per second occur. Therefore, the UEs would not experience unnecessary access delay when the traffic load is low. $\ell[i]$ will increase with the number of preamble transmissions per RAO, $N[i]$, until a maximum of one when $N_s[i] = N_{UL}$. Given the maximum load coefficient, $\ell[i] = 1$, is obtained,

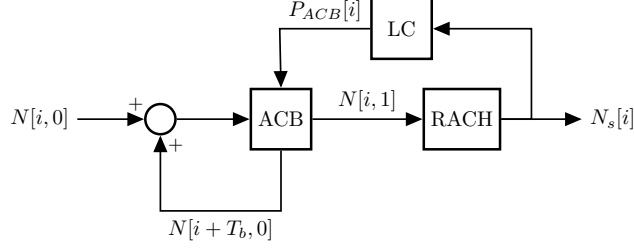


Fig. 3. Block diagram for the implementation of a dynamic ACB (DACB).

Table 1. RACH configuration.

Parameter	Setting
RAO periodicity	$T_{RAO} = 5$ ms
Number of available preambles	$R = 54$
RAR window length	$W_{RAR} = 5$ subframes
Maximum number of uplink grants per subframe	$N_{RAR} = 3$
Maximum number of uplink grants per RAO	$N_{UL} = W_{RAR} \times N_{RAR} = 15$
Maximum number of preamble transmissions	$preambleTransMax = 10$
Preamble detection probability for the k th preamble transmission	$P_d = 1 - 1/e^k$
Backoff Indicator	$B = 20$ ms

$P_{ACB}[i] = 0$, hence the first preamble transmissions will not be permitted until $\ell[i]$ drops. From Fig. 2b, it can be seen that our DACB will react accordingly given $N[i] \leq 130$. Therefore, the main objective of our DACB method is to hinder the growth of $N[i]$ beyond 130.

4 RACH configuration and selected traffic model

Throughout this study we assume the basic configuration of the RACH as suggested in [1]. Table 1 shows the selected RACH configuration.

In this configuration (PRACH Configuration Index 6), the duration of a subframe is 1 ms and RAOs occur every $T_{RAO} = 5$ ms. The number of available preambles, R , determines the probability distribution of successful preambles, $N_{sp}[i]$, in the i th RAO for a given number of preamble transmissions, $N[i]$ (see Fig. 2a). $N_{UL} = W_{RAR} \cdot N_{RAR}$ determines the maximum number of uplink grants that the eNB can transmit within the RAR window (for each RAO). Note that only the UEs that receive an uplink grant may proceed with the transmission of *Msg3* and *Msg4*; therefore, the maximum number of successful UE access per RAO for any given $N[i]$ is $\max\{N_s[i]\} = N_{UL}$ as observed in Fig. 2b. The preambles transmitted by single UEs are detected by the eNB with probability P_d . The backoff time of failed UEs is calculated randomly as

$$T_B = U(0, B), \quad (4)$$

Table 2. M2M traffic models for RACH Evaluation [1]

Characteristics	Traffic model 1	Traffic model 2
Number of M2M UEs (N)	1000, 3000, ..., 180000	30000
Arrival distribution over T	Uniform	$Beta(3, 4)$
Distribution period, T	12000 RAOs	2000 RAOs

where B is the *Backoff Indicator* broadcast by the eNB.

The 3GPP has proposed two traffic models for RACH evaluation in [1]. Traffic model 1 corresponds to a typical scenario, where the UE arrivals are uniformly distributed through time within $T = 12000$ RAOs (60 seconds). While the 3GPP defines the maximum number of M2M UEs in the traffic model 1 as 30000, we use this model to evaluate the effect of ACB on the access delay of UEs during periods of low congestion, when up to 180000 M2M UEs access the eNB. On the other hand, Traffic model 2 corresponds to a highly congested scenario, where $N = 30000$ M2M UEs arrivals follow a $Beta(3, 4)$ distribution within $T = 2000$ RAOs (10 seconds). We use this model to evaluate the effect of ACB on the access success probability, P_s , and the access delay of UEs that access the eNB during periods of high congestion. Table 2 shows the traffic models and the number of M2M UEs considered during this study.

5 Performance analysis

In this section we evaluate the performance of the RACH and compare the efficiency of a static ACB and our DACB method. For this, we perform simulations of the RAP until the cumulative results obtained in the j th simulation differ from those obtained in the $j - 1$ th simulation by less than 1%. First, we evaluate the performance of the RACH given that the UE access the eNB according to traffic model 2, depicted in Table 2. Fig. 5 shows the average number of UE arrivals, $N[i, 0]$ ¹ and the average number of successful accesses per RAO, $N_s[i]$, for three cases in the given scenario. In the first, no ACB is implemented, and it can be clearly seen that the performance of the RACH is degraded when $N[i, 0]$ exceeds $N_{UL} = 15$. Consequently, the access success probability if no ACB is implemented is $P_s = 31.305\%$, which is by no means desirable.

In the second, a static ACB is implemented. We have selected $P_{ACB} = 0.5$ and $T_{ACB} = 4$ s for the static ACB because it leads to the lowest access delay while maintaining an adequate $P_s \geq 95\%$, concretely $P_s = 97.44\%$. In the third, our DACB is implemented and even a higher $P_s = 99.93\%$ is achieved. Such high access success probabilities are achieved because both ACB methods successfully reduce the number of first preamble transmissions, $N[i, 1]$, below N_{UL} . Also, in Fig. 5 we can observe a noticeable difference in the shape of $N_s[i]$ between both ACB methods. For the static ACB, $N_s[i]$ rises rapidly and almost reaches

¹ In case no ACB is implemented, the average number of UE accesses, $N[i, 0]$, is equal to the average number of first preamble transmissions, $N[i, 1]$, in the i th RAO.

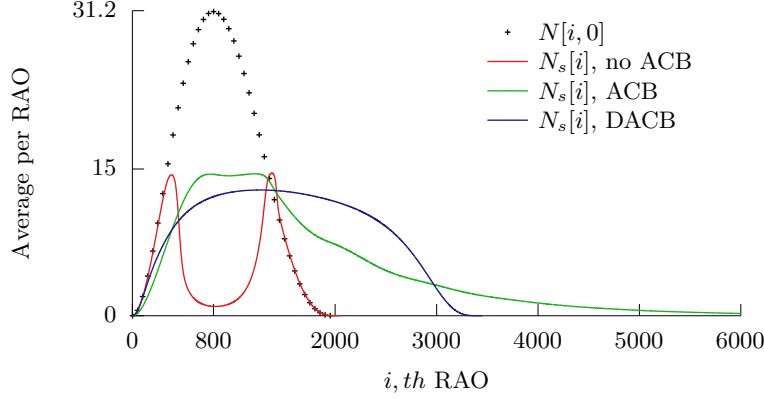


Fig. 4. Average number of UE accesses, $N[i, 0]$, and the average number successful UE accesses, $N_s[i]$, given the implementation of no ACB, a static ($P_{ACB} = 0.5$ and $T_{ACB} = 4$ s) and the dynamic ACB (DACB), traffic model 2.

$N_{UL} = 15$. Then, after $N[i, 0]$ drops below N_{UL} , $N_s[i]$ drops slowly. For the DACB, $N_s[i]$ rises rapidly but stops at around 12.9 successful accesses per RAO. Then, after 2000 RAOs, $N_s[i]$ drops rapidly.

The difference in the shape of $N_s[i]$ between both methods results in a noticeable disparity in the probability distribution of access delay. To closely observe this performance indicator, we have obtained the cumulative distribution function (CDF) of access delay based on the timing values of Table 16.2.1-1 in [4]. In Fig. 5, it can be seen that the initial growth of the CDF of the static backoff is more rapid than that of our DACB method. However, this growth stops just below 0.5 and the CDF of the DACB is higher after 2 s have elapsed. As a result, high percentiles of access delay are much lower for the DACB than that of the static ACB. Hence, our DACB method presents a lower access delay than the best possible combination of P_{ACB} and T_{ACB} for the static ACB in an extremely congested scenario (traffic model 2).

Finally, we are interested in evaluating the effect of both ACB methods in access delay during periods of low congestion, described by traffic model 1 (see Table 2). Specifically, we assess the access delay in these scenarios in terms of the 95th percentile, t_{95} in seconds; i.e., 95% of the UEs experience an access delay lower than or equal to t_{95} . As Fig. 6 shows, the static ACB causes the exact same t_{95} regardless of the number of UEs that access the eNB per RAO. In other words, if a static ACB is implemented, 5% of the total UE accesses will be delayed, at least, 15.6 seconds in any scenario, which is by no means desirable. On the other hand, the use of our DACB method sharply reduces t_{95} . In fact, the lowest $t_{95} = 0.057$ s is achieved when $N[i] = 1$ and increases with $N[i]$, but is always lower than the t_{95} achieved by the static ACB.

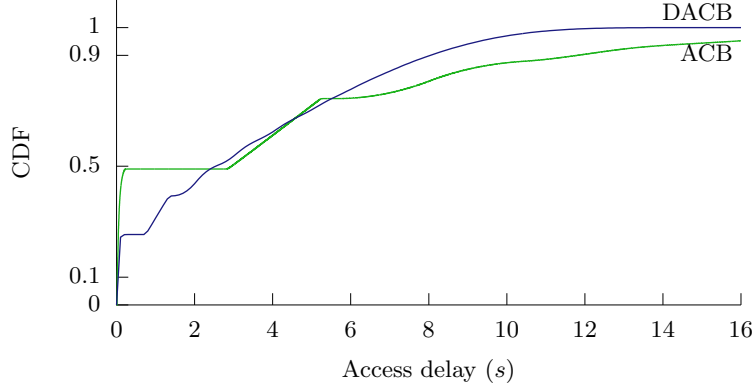


Fig. 5. Cumulative distribution function (CDF) of access delay for the static ($P_{\text{ACB}} = 0.5$ and $T_{\text{ACB}} = 4$ s) and for the DACB, traffic model 2.

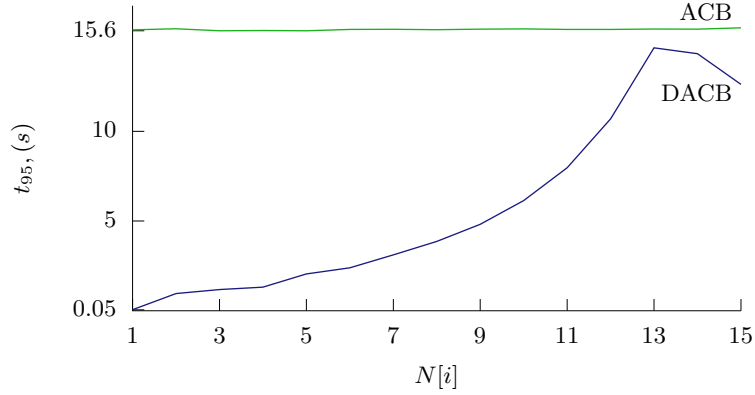


Fig. 6. 95th percentile of access delay, t_{95} , for the static ($P_{\text{ACB}} = 0.5$ and $T_{\text{ACB}} = 4$ s) and the dynamic ACB, traffic model 1.

6 Conclusion

The current random access procedure of LTE-A is not efficient at handling massive M2M communications. As such, congestion control methods must be implemented. Without a doubt, static ACB methods sharply enhance the performance of LTE-A during periods of extreme congestion. However, these methods greatly affect the access of UEs during periods of low or no congestion. By comparing our DACB method with the best possible implementation of a static ACB, we have observed that the former leads to a higher access success probability and lower access delay in extremely congested scenarios. Our DACB method also maintains an acceptable access delay in scenarios where a few UEs access the eNB, whereas implementing a static ACB sharply increases this parameter. Future work includes the performance analysis of our DACB method given that the calculation of the LC and the update of the barring rate at the UEs can only be

performed at certain intervals. In addition, the use of LCs from previous RAOs for the calculation of the barring rate may increase the reliability of the DACB.

Acknowledgment

This work has been supported by the Ministry of Economy and Competitiveness of Spain through the projects TIN2013-47272-C2-1-R and TEC2015-71932-REDT. The research of I. Leyva-Mayorga was partially funded by grant 383936 CONACYT-Gobierno del Estado de México 2014. The research of L. Tello-Quendo was supported in part by Programa de Ayudas de Investigación y Desarrollo (PAID) of the Universitat Politècnica de València.

References

1. 3GPP: TR 37.868, Study on RAN Improvements for Machine Type Communications (2011)
2. 3GPP: TS 22.368, Service Requirements for Machine-Type Communications (Dec 2014)
3. 3GPP: TS 36.331, Radio Resource Control (RRC), Protocol specification (2014)
4. 3GPP: TR 36.912 Feasibility study for Further Advancements for E-UTRA (2016)
5. Biral, A., Centenaro, M., Zanella, A., Vangelista, L., Zorzi, M.: The challenges of M2M massive access in wireless cellular networks. *Digit. Commun. Netw.* 1(1), 1–19 (2015)
6. Ferdouse, L., Anpalagan, A., Misra, S.: Congestion and overload control techniques in massive M2M systems: a survey. *Trans. Emerg. Telecommun. Technol.* (2015)
7. Ghavimi, F., Chen, H.H.: M2M Communications in 3GPP LTE/LTE-A Networks: Architectures, Service Requirements, Challenges, and Applications 17(2), 525–549 (May 2015)
8. Lin, G.Y., Chang, S.R., Wei, H.Y.: Estimation and Adaptation for Bursty LTE Random Access. *IEEE Trans. Veh. Technol.* 65(4), 2560–2577 (2016)
9. Lin, T.M., Lee, C.H., Cheng, J.P., Chen, W.T.: PRADA: Prioritized random access with dynamic access barring for MTC in 3GPP LTE-A networks. *IEEE Trans. Veh. Technol.* 63(5), 2467–2472 (2014)
10. Osti, P., Lassila, P., Aalto, S., Larmo, A., Tirronen, T.: Analysis of PDCCH Performance for M2M Traffic in LTE. *IEEE Trans. Veh. Technol.* 63(9), 4357–4371 (2014)
11. Suyang Duan, Shah-Mansouri, V., Wong, V.W.S.: Dynamic access class barring for M2M communications in LTE networks. In: *Proc. IEEE Global Communications Conference (GLOBECOM)*. pp. 4747–4752 (2013)
12. Tavana, M., Shah-Mansouri, V., Wong, V.W.S.: Congestion control for bursty M2M traffic in LTE networks. In: *Proc. IEEE International Conference on Communications (ICC)*. pp. 5815–5820 (2015)
13. Wei, C.H., Bianchi, G., Cheng, R.G.: Modeling and analysis of random access channels with bursty arrivals in OFDMA wireless networks. *IEEE Trans. Wirel. Commun.* 14(4), 1940–1953 (2015)

Minimal Configuration of Body Surface Potential Mapping for Discrimination of Left and Right Atrial Drivers during Atrial Fibrillation

Miguel Rodrigo¹, Andreu M Climent², Alejandro Liberos², Francisco Fernández-Avilés^{2,3}, Felipe Atienza^{2,3}, María S Guillem¹ and Omer Berenfeld⁴

¹ Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politècnica de València, Valencia, Camino de Vera s/n, 46022 Valencia, España
{mirodbor, mguisan}@upv.es
<http://www.itaca.upv.es>

² Cardiology Department, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain. Calle Dr Esquerdo 46 28007

³ Facultad de Medicina, Universidad Complutense de Madrid, Pza. Ramón y Cajal, s/n, Ciudad Universitaria, 28040, Madrid, Spain

⁴ Center for Arrhythmia Research, University of Michigan, Ann Arbor, USA. 2800 Plymouth Road, MI 48109, USA

Abstract. Ablation of drivers maintaining Atrial Fibrillation (AF) has been demonstrated as an effective therapy. Drivers in the form of highest frequency rotors can be non-invasively localized to either left or right atria (LA, RA) with Body Surface Potential Mapping (BSPM) systems. This study quantifies the surface localization of high dominant frequency rotors in order to assess the minimal configuration of a reduced-leads BSPM system for ablation guidance. Nine uniformly distributed lead sets from 8 to 66 electrodes were evaluated. BSPM signals were registered simultaneously to intracardiac electrocardiograms (EGMs) in 14 AF patients. Rotor activity was analyzed on the surface electrical activity for the 9 leads configurations, and the non-invasive measures were compared with the EGM recordings. The surface rotor distribution assessed with 32 or more electrodes presented a high correlation with the full-resolution BSPM system of 66 leads. The surface reentrant activity at the torso areas corresponding to the LA and RA electrical activity presented significant differences between LA and RA fastest patients for lead sets with 24 or more leads. Reduced-leads BSPM systems enable non-invasive discrimination between LA or RA atrial drivers with similar results as higher-resolution 66-leads system. Our findings demonstrate the possible incorporation of simplified BSPM systems into clinical planning procedures for AF ablation.

1 Introduction

Several studies have demonstrated the maintenance of atrial fibrillation (AF) by discrete atrial drivers whose ablation is associated with a favorable outcome [1-3], mak-

ing the identification and localization of these drivers the goal for AF mapping and ablation [2,4]. Emerging technologies have used Body Surface Potential Mapping (BSPM) systems to non-invasively identify those atrial drivers, either by locating reentrant patterns [3,5,6] or Dominant Frequencies (DF) [7]. Although these preliminary results seem promising, the main disadvantage to extend the use of BSPM systems is the complexity of the recording approach with multiple electrodes.

The aim of this study is to determine the minimal configuration of body surface leads necessary to non-invasively identify the atrial rotors that presumably drive AF and to facilitate the incorporation of the non-invasive mapping into the clinical practice. Previous studies have already investigated how the selection of different leads' sets affects the ECG waveform [8-10] and have concluded that for AF signals, systems with more than 32 electrodes do not increase the signal reconstruction accuracy [11]. However, the ability of these lead sets to characterize the nature and position of possible AF drivers in forms of rotors has not been studied. The present study demonstrates that BSPM with an approximate uniform distribution of 32 or more electrodes over the torso surface is capable of rotor identification and localization in the left and right atria.

2 Methods

2.1 Patients

We included 14 patients admitted for ablation of drug-refractory paroxysmal and persistent AF [6]. All patients gave informed consent and the protocol was approved by the Institutional Ethics Committee.

2.2 Electrophysiological study and EGM recordings

The electrophysiological study was performed under general anesthesia and periodic heparin bolus administration. Intracardiac electrograms (EGM) were obtained from both atria during the procedure by using the following catheters introduced via the right femoral vein: (1) a standard tetrapolar catheter in the right atrial (RA) appendage; (2) a deflectable 4-mm mapping catheter (Marinr; Medtronic Inc., Minneapolis, MN) in the distal coronary sinus; (3) a decapolar circular mapping Lasso catheter (Biosense-Webster, Diamond Bar, CA) used to map the PV-left atrial (LA) junctions; and an ablation catheter (Navistar cooled-tip catheter, Biosense Webster, Inc, Diamond Bar, CA) in the LA. In patients arriving in sinus rhythm, AF was induced by burst pacing. AF episode durations longer than 5 minutes were required for inclusion in the study.

2.3 Body surface potential recordings

Surface electrocardiograms (ECGs) were simultaneously recorded with the EGM recordings by using a vest grid of 66 electrodes distributed as follows: 30 electrodes on the anterior, 34 on the posterior and 2 on each lateral side of the torso (Figure 1A). An additional left leg electrode was used in order to obtain a Wilson Central Terminal. The signals were recorded using a commercial system (Active Two, Biosemi, The Netherlands) at a sampling frequency of 2048 Hz and were stored for off-line analysis. Ventricular activation was removed by administration of a central venous bolus of adenosine (12-18 mg) and 4-seconds segments of surface ECGs surrounding the longest RR interval were used for the analysis. In cases with ventricular activation pauses shorter than 4 seconds, the QRSTs were cancelled [6]. Baseline was estimated by decimation to 51.2 Hz and filtering with a Butterworth 10th-order low-pass filter with a cut-off frequency of 2 Hz. This signal was interpolated to 2048 Hz and subtracted from the original signal. Leads presenting >0.5 % of their spectral content at 50 Hz were filtered with a second-order infinite impulse response notch filter centered at 50 Hz. Potential signals were then low-pass filtered with a 10th-order Butterworth filter with a cut-off frequency of 30 Hz [11].

2.4 Lead distribution on torso and 3D model

To analyze the effect of lead configuration on AF mapping, different combinations sets of the BSPM electrodes were defined. Those sets included uniform distributions of 8, 12, 18, 24, 32, 42 and 54 electrodes as well as a non-uniform distribution of 8 leads matching the standard ECG derivation distribution (see Figure 1A). The surface electrical activity was then projected on the surface of a 3D torso model with 771 nodes and 1538 faces (see Figure 1B), in which the surface potentials were reconstructed using the ECG signals provided by the different lead sets. The electrical potentials in areas between the nodes selected as electrodes for a particular lead configuration were interpolated as previously described.

Two areas on the surface of the 3D torso model were defined as being dominantly influenced by either left or right atrial electrical activity (Figure 1B): an LA area (red) covering the left, upper back and top side; and an RA area (green) covering the right, lower back and front side [6,7]. Seven smaller surface areas were also defined at the front, upper back, lower back, left, right, top and floor regions for quantification of the rotor distribution (Figure 1C).

2.5 Signal analysis and pattern recognition

Power Spectral Density (PSD) of EGMs and ECGs was computed by using Welch periodogram (2-second Hamming window with an 8192 point Fast Fourier transform per window and 50% overlap) to determine the DFs. For the reentrant pattern identification, reconstructed surface potentials were resampled at 128 Hz and then filtered at the highest DF (HDF) found on the EGM by using a cascade of a high-pass elliptic filter with cut-off frequency equal to $\text{HDF}-1$ Hz and a low-pass elliptic filter with a

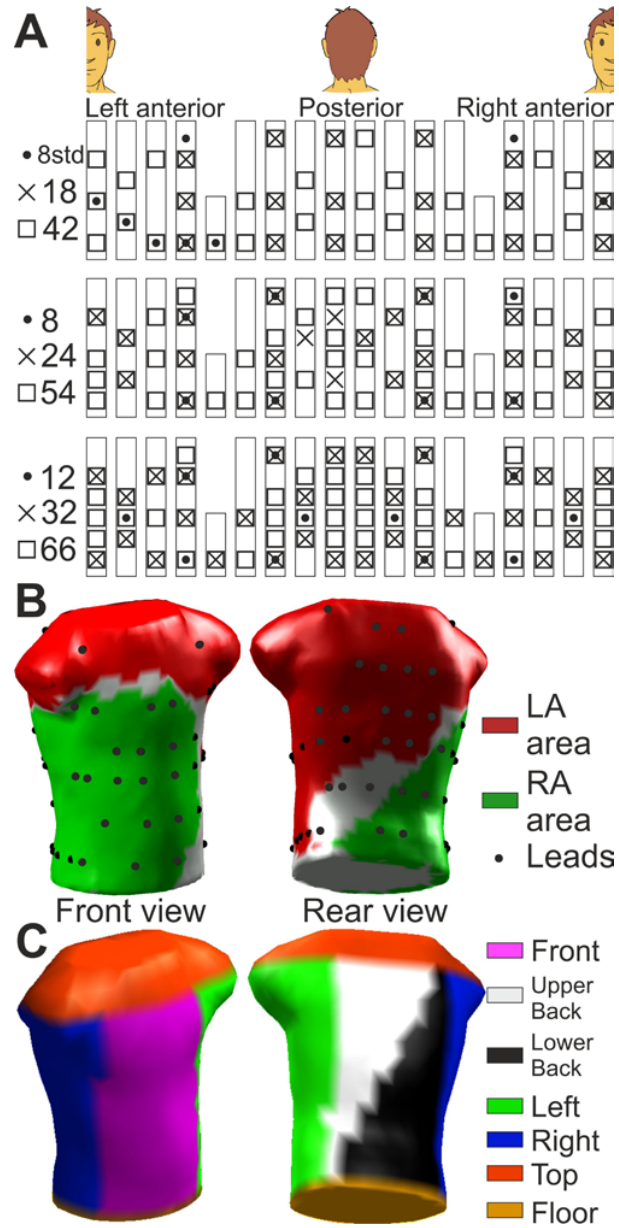


Fig. 1. Lead and area distribution on the torso. (A) Distribution of the lead sets. (B) 3D torso with the 66 leads (black dots) and the LA and RA dominant influence areas (red and green, respectively). (C) Locations and names of the 7 torso areas used to assess the stability of the surface rotors

cut-off frequency equal to HDF+1 Hz [6]. Phase maps on the torso surface were obtained from the potential phase signal of each node by the Hilbert transform. A Singularity Point (SP) was defined as any point in a phase map which is surrounded by phases monotonically increasing from 0 to 2π . Only those SPs that were present for the duration of at least two full rotations were considered as rotors [6].

2.6 Statistical analysis

The Student t-test was used to evaluate the statistical significance between continuous paired or unpaired variables, and statistical significance was considered for $p < 0.05$. All data are reported in mean \pm SD.

3 Results

3.1 Surface distribution of AF drivers for reduced-leads BSPM

With the aim of analyzing the number of leads required to detect AF maintenance mechanisms, we studied the reproducibility of the surface distribution of reentrant patterns for different lead arrangements. Figure 2 shows a qualitative example of the phase maps and the rotor distribution obtained for one of the patients using different BSPM lead's distribution. The singularity point that can be observed at the back of that patient is at the same place for 66, 32 and 8 leads distribution, although the lower resolution of the latter also shows a SP in the front that is not seen at higher resolutions (in the higher resolution of 66 and 32 leads a SP is at the floor of the torso and is not seen in the panels). As shown by the cumulative SP presence maps in the right side panels, the rotor distribution is relatively similar for the 66 and 32 leads, but it considerably altered for the 8 leads configuration.

To quantify such effect of the leads configuration on rotor localization on the surface, the number of SPs and rotors (i.e., SPs lasting more than a full rotation) at 7 distinct torso surface areas, as defined in Figure 1C and following HDF band-pass filtering, for the different lead distribution was counted. The number of SPs and rotors at each region for each lead configuration was correlated with the distribution obtained with highest resolution configuration of 66 leads. In Figure 3 it is shown that, as expected, the low-resolution leads' configuration showed low correlation with the full resolution distribution, both for SPs (0.05 ± 0.23) and rotors (-0.23 ± 0.22). The correlation with the full resolution distribution increased with the number of leads, reaching a constant maximum value for 32 or more leads. The SPs distribution reached a maximum correlation of 0.95, while the rotor distribution reached a correlation coefficient around 0.8 for 32 leads configuration or more.

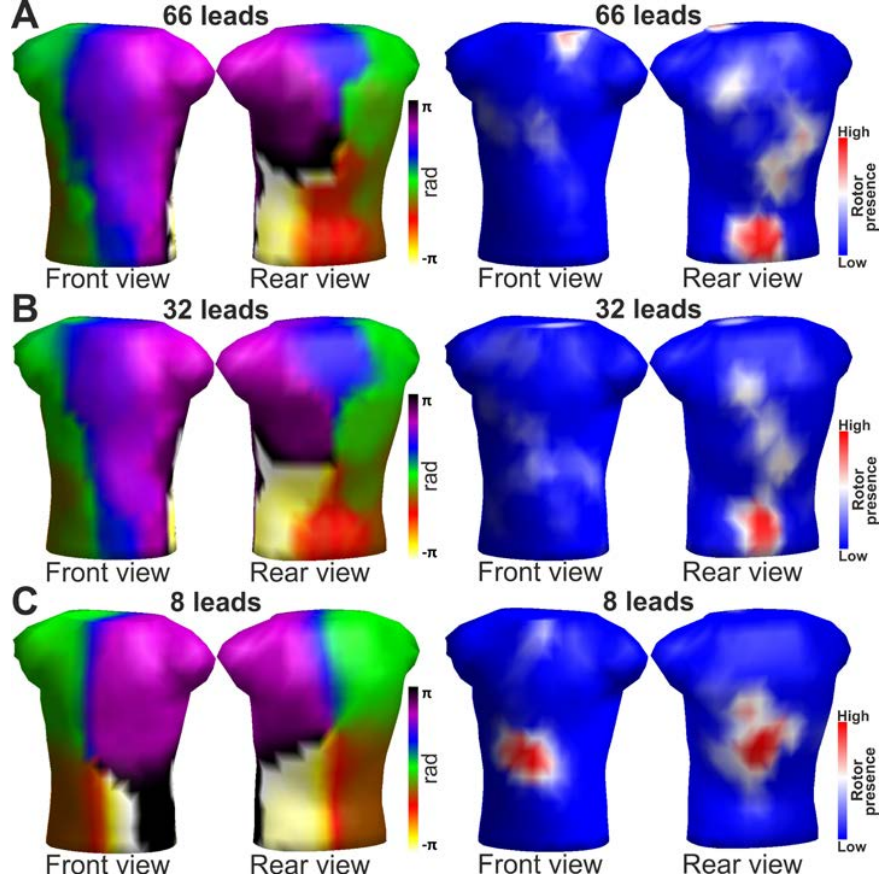


Fig. 2. Sample surface phase maps (left) and surface rotor number distributions (right) for different leads BSPM configuration: (A) 66 leads. (B) 32 leads. (C) 8 leads uniformly distributed.

3.2 Diagnostic value of the non-invasive atrial activity

Once the accuracy of noninvasive measures was evaluated for different leads configurations, we assessed the statistical ability of these configurations to provide drivers localization in AF patients. We measured the percentage of time with surface rotors after HDF band-pass filtering in the LA and RA surface areas (Figure 1B) for the different lead configurations. In Figure 4 the patients were separated into LA-fastest patients (bottom bars) and RA-fastest patients (top bars) groups based on whether the DFs at the EGM recordings when the LA-DF was higher than the RA-DF or vice versa, respectively. Here we predict that the surface area corresponding to the atrium

with the fastest activation rate would show the highest number of rotors. Indeed, as expected, cumulative data in Figure 4 shows the RA-fastest patients present more time with rotors in the RA surface area than in the LA area, and conversely, the LA-fastest patients present more time with rotors in LA surface area than in RA area. However, this trend is only observable for mid- to high-resolution leads configurations; statistical significance of $p < 0.05$ of the distinct rotor presence was obtained only with 18 or more surface electrodes. Specifically, there were significant differences between the number of LA and RA area rotors for 18, 24, 32, 42, 54 and 66 leads for RA-fastest patients, and for 32 leads between LA and RA area rotors for the LA-fastest patients. Comparing the rotor presence in the same area between patients' groups, there was significant difference of rotor presence in the LA area between RA-fastest and LA-fastest patients for 24, 32, 42, 54 and 66 leads; and in the RA area for 32 leads. These results show that BSPM with an approximate uniform distribution of 24 or more electrodes is capable of the rotor localization between the LA and RA similar to differentiation with 66 leads configurations.

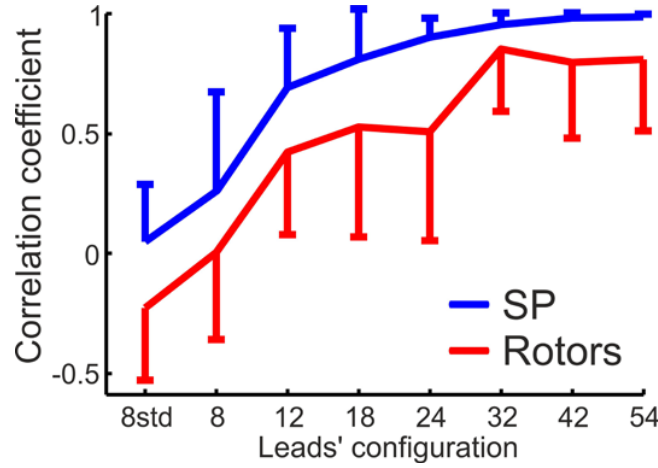


Fig. 3. Cross correlation between the amount of SPs and rotors localized to each of the 7 surface regions (defined in Figure 1), as a function of the leads' configuration, for the 14 patients.

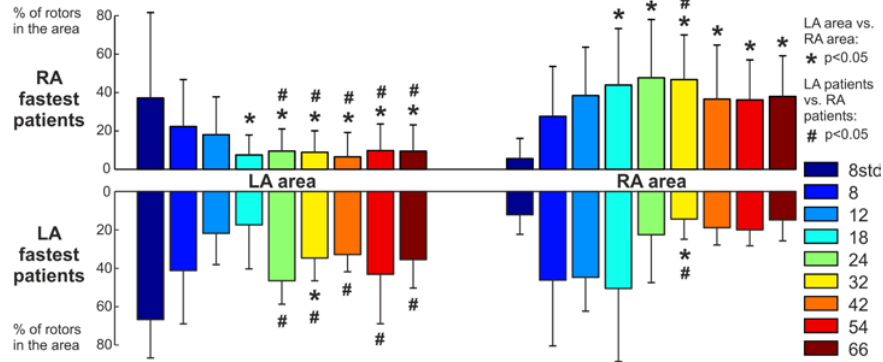


Fig. 4. Summary of surface distribution of rotors for the 14 patients. Bars represent the number of rotors localized to the LA area (left) and RA area (right) for RA-fastest patients (Top, N=10) and LA-fastest patients (Bottom, N=4). Each color represents a different leads' configuration. Statistical comparisons for the 14 patients were made between rotors presence in LA vs RA areas (left and right) for each leads' configuration (*), and between RA-fastest and LA-fastest patients (top and bottom) for each leads' configuration (#).

4 Discussion

4.1 Major findings

The present study demonstrates that surface rotor localization based on surface reentrant activity identification at the torso areas enables similar differentiation between LA and RA fastest patients for lead sets with 32 or more leads than with full-resolution BSPM systems. Thus, the increase of the number of uniformly distributed leads above 32 does not increase the BSPM diagnosis value for LA versus RA drivers in AF patients. Therefore, reduced-leads BSPM systems enable non-invasive discrimination between LA or RA atrial drivers with similar results as higher-resolution 66-leads systems.

4.2 Mechanisms of AF maintenance and their non-invasive characterization

The mechanisms underlying AF in patients are not always clear. Recent studies utilizing new mapping methods have supported the existence of a hierarchical AF organization of activity whereby identifiable sources of fast activity [1,4] or in the form ectopic or rotor activation patterns maintain the arrhythmia [2,3], but controversies still exist [12-13]. Arguably, better and easier identification of localized AF sources would render better targets for ablation thus efforts to improve technologies for localization of AF drivers is ongoing [5-7]. Although intra-cardiac recordings are considered the

most reliable method for AF characterization [2], they are not free of difficulties [14] and the need for panoramic mapping led to the development of noninvasive AF mapping methods. Noninvasive mapping methods not only would allow for relatively simpler deployment than the usage of multiple intra-cardiac electrodes, but will also enable studying patients outside of the clinical EP laboratory and pre-planning individual ablation procedures. Some of the limitations of non-invasive technologies are their reduced sensitivity to the atrial activity during fibrillation. We have recently studied the effect of the body torso on body surface mapping and found that a significant amount of atrial information is retained on the body surface recordings [6-7,15]. These studies showed that utilizing the appropriate frequency and phase domain processing, with a band-pass filtering at the highest DF and SP detection allows localization of the surface reentrant activity originating at either the LE or the RA [6]. Thus, non-invasive analysis of the AF reentrant activity on the torso surface enables rotor drivers localization at a dominant atrium to facilitate ablation procedures [1,4].

4.3 Lead distribution in body surface electrocardiography

The number of body surface leads and their position have been broadly acknowledged as affecting arrhythmia diagnostic accuracy, but in clinical practice the standard ECG leads configuration largely prevails. Indeed, the standard ECG lead set does not accurately respond to all diagnostic cardiology needs, and for this reason body surface potential mapping (BSPM) systems with a broad distribution of leads and higher spatial resolution were developed. However, today there is no standard evaluation or recommendation for the surface electrode configuration for the different cardiac diseases and AF in particular.

The number of electrodes necessary for BSPM utilization has been mostly evaluated for ventricular activity based on the quality of signals [8-11]. Although some studies argue that in the ECG signal there are about 10 statistically independent components [16], Lux et al. observed that for the study of healthy ECG, 30 leads allowed to reconstruct the set of torso signals with a correlation of 98% and an error less than 3% with the original signals [8]. The position of these 30 electrodes was not uniform across the torso surface, being the front side of the torso the densest area. More recently, a study by Finlay et al in a population of healthy and myocardial infarction patients observed that 32 electrodes were enough to provide an acceptable level of reconstruction, but in this case different configuration of the electrodes yielded little differences in the ECG signals [9]. In agreement to this, we conducted a study in patients with myocardial infarction, bundle-branch block, and ventricular hypertrophy patients and found that increasing the lead number above 30 electrodes had little effect on the ECG signal and the leads configuration had little importance on the signal reconstruction, which was more dependent on the population under study [11].

Only few studies investigated the role of the number and distribution of BSPM leads in studying atrial electrophysiology. Although the morphology and variability of the standard ECG signal in AF has been widely studied, only few authors have considered that the standard ECG system is sensing mostly the ventricles and have proposed the use of a different arrangement of the 9 standard leads in order to increase the effec-

tiveness in AF characterization [17]. The broader electrode settings of BSPM has been used to study atrial arrhythmias [5,18-20], but without an evaluation of accuracy for AF mechanisms characterization. In a recent study, Guillem et al studied the number of leads necessary to properly reconstruct the torso potentials during in AF signals [11]. This study showed that during atrial fibrillatory rhythms a set of 34 electrodes was able to reconstruct the surface potential with the same accuracy than systems with more leads, since the reconstruction error reaches the noise level at the recording electrodes. The latter is consistent with our results here: we propose that mapping with more than 32 uniformly distributed surface leads does not increase the diagnostic information of inter-atrial rotor localization. For the DF inter-atrial distribution, even less leads could be enough. In spite of the disorganized AF signal nature, the BSPM systems for the study of the AF seem to need the same amount of electrodes (about 32) than the BSPM systems for the study of ventricular non-fibrillation arrhythmias [8-11]. This observation is also consistent with the natural smoothening filtering effect of the torso volume conductor that projects on the torso surface only the strongest characteristics of the atrial waves, with reduced sensitivity for the small electrical waves, no matter how many leads are used for mapping [6].

4.4 Limitations

The present study was conducted on a small number of 14 patients, however its results are consistent with previous findings [18-20]. Moreover, as the optimal lead configuration on the torso depends on the population under study, we submit that the use of our heterogeneous AF patients cohort, with paroxysmal, persistent and different number of ablation procedures [6-7,15] enhances the credence of our conclusions on AF inter-atrial characterization. Additionally, other studies have demonstrated correlations of 0.98 or higher between ECG signal waveform [8-11] which are higher than our correlations for DF and rotors. This is likely the result of propagating errors of secondary calculated characteristics of the electrical activity. Notwithstanding, the physiological meaning of the inter-atrial DF and rotor distribution as ablation targets is still evident.

5 Conclusions

For coarse AF diagnosis, BSPM systems with about 32 uniformly distributed leads can achieve the same results as systems with larger number of leads. Consistent with the literature, those leads would be sufficient to properly reconstruct the surface electrical potentials for inter-atrial localization of AF drivers in the form of highest DF or rotors. Medium-resolution BSPM systems retain the relevant clinical information that enables discrimination between RA- or LA-fastest patients, highly relevant for increased ablation effectiveness. The results of this work can be the basis for the simplification and lower-cost non-invasive classification systems for AF patients based on the use of BSPM.

References

1. Atienza, F., Almendral, J., Ormaetxe, J.M., Moya, A., Martínez-Alday, J.D., Hernández-Madrid, A., et al. RADAR-AF Investigators. Comparison of radiofrequency catheter ablation of drivers and circumferential pulmonary vein isolation in atrial fibrillation: a noninferiority randomized multicenter RADAR-AF trial. *J. Am. Coll. Cardiol.* (2014) 2455-67
2. Narayan, S.M., Krummen, D.E., Clopton, P., Shivkumar, K., Miller, J.M. Direct or coincidental elimination of stable rotors or focal sources may explain successful atrial fibrillation ablation: on-treatment analysis of the CONFIRM (CONventional ablation for AF with or without Focal Impulse and Rotor Modulation) Trial. *J. Am. Coll. Cardiol.* (2013) 137-147
3. Haissaguerre, M., Hocini, M., Denis, A., Shah, A.J., Komatsu, Y., Yamashita, S., et al. Driver domains in persistent atrial fibrillation. *Circulation* (2014) 530-538
4. Atienza, F., Almendral, J., Jalife, J., Zlochiver, S., Ploutz-Snyder, R., Torrecilla, E.G., et al. Real-time dominant frequency mapping and ablation of dominant frequency sites in atrial fibrillation with left-to-right frequency gradients predicts long-term maintenance of sinus rhythm. *Heart Rhythm* (2009) 33-40
5. Lim, H.S., Zellerhoff, S., Derval, N., Denis, A., Yamashita, S., Berte, B., et al. Noninvasive mapping to guide atrial fibrillation ablation. *Card. Electrophysiol. Clin.* (2015) 89-98
6. Rodrigo, M., Guillem, M.S., Climent, A.M., Pedrón-Torrecilla, J., Liberos, A., Millet, J., et al. Body surface localization of left and right atrial high-frequency rotors in atrial fibrillation patients: a clinical-computational study. *Heart Rhythm* (2014) 1584-1591
7. Guillem, M.S., Climent, A.M., Millet, J., Arenal, A., Fernández-Avilés, F., Jalife, J., et al. Noninvasive localization of maximal frequency sites of atrial fibrillation by body surface potential mapping. *Circ. Arrhythm. Electrophysiol.* (2013) 294-301
8. Lux, R.L., Smith, C.R., Wyatt, R.F., Abildskov, J.A. Limited lead selection for estimation of body surface potential maps in electrocardiography. *IEEE Trans. Biomed. Eng.* (1978) 270-276
9. Finlay, D.D., Nugent, C.D., Donnelly, M.P., Black, N.D. Selection of optimal recording sites for limited lead body surface potential mapping in myocardial infarction and left ventricular hypertrophy. *J. Electrocardiol.* (2008) 264-271
10. Guillem, M.S., Castells, F., Climent, A.M., Bodí, V., Chorro, F.J., Millet, J. Evaluation of lead selection methods for optimal reconstruction of body surface potentials. *J. Electrocardiol.* (2008) 26-34
11. Guillem, M.S., Bollmann, A., Climent, A.M., Husser, D., Millet-Roig, J., Castells, F. How many leads are necessary for a reliable reconstruction of surface potentials during atrial fibrillation? *IEEE Trans. Inf. Technol. Biomed.* (2009) 330-40
12. Narayan, S.M., Jalife, J. CrossTalk proposal: Rotors have been demonstrated to drive human atrial fibrillation. *J. Physiol.* (2014) 3163-3166
13. Allessie, M., de Groot, N. CrossTalk opposing view: Rotors have not been demonstrated to be the drivers of atrial fibrillation. *J. Physiol.* (2014) 3167-3170
14. Berenfeld, O., Oral, H. The quest for rotors in atrial fibrillation: different nets catch different fishes. *Heart Rhythm* (2012) 1440-1441
15. Pedrón-Torrecilla, J., Rodrigo, M., Climent, A.M., Liberos, A., Pérez-David, E., Bermejo, J., et al. Noninvasive Estimation of Epicardial Dominant High-Frequency Regions during Atrial Fibrillation. *J. Cardiovasc. Electrophysiol.* (2016) 10.1111/jce.12931
16. Hoekema, R., Uijen, G., van Oosterom, A. The number of independent signals in body surface maps. *Methods Inf. Med.* (1999) 119-124
17. Ihara, Z., van Oosterom, A., Jacquemet, V., Hoekema, R. Adaptation of the standard 12-lead electrocardiogram system dedicated to the analysis of atrial fibrillation. *J. Electrocardiol.* (2007) 68.e1-e8

18. Gerstenfeld, E.P., SippensGroenewegen, A., Lux, R.L., Lesh, M.D. Derivation of an optimal lead set for measuring ectopic atrial activation from the pulmonary veins by using body surface mapping. *J. Electrocardiol.* (2000) 179-185
19. Lian, J., Li, G., Cheng, J., Avitall, B., He, B. Body surface Laplacian mapping of atrial depolarization in healthy human subjects. *Med. Biol. Eng. Comput.* (2002) 650-659
20. Sippensgroenewegen, A., Peeters, H.A., Jessurun, E.R., Linnenbank, A.C., Robles deMedina, E.O., Lesh, M.D., et al. Body surface mapping during pacing at multiple sites in the human atrium: P-wave morphology of ectopic right atrial activation. *Circulation* (1998) 369-380

Towards Customizable Assisted Living with friendly environment using intelligent agents: A Survey

Sandra Lopez C.¹, Rosario Baltazar F.¹, Arnulfo Alanis², Lenin Lemus³,
Guillermo Méndez Z.¹

¹ Instituto Tecnológico de León, Guanajuato, México
jclsandra@gmail.com, r.baltazar@ieee.org, guillermomendez06@gmail.com

² Instituto Tecnológico de Tijuana, Baja California, México
alanis@tectijuana.edu.mx

³ Universitat Politècnica de València
lemus@upv.es

Abstract. The elderly people or with physically problems, always seek independence and performance their daily activities without having to depend on others, so the purpose of a AAL system is prolong the life of these people while living at home independently using information technology and computers. (Pavlina Koleva, 2015). Customizing a system is critical to user acceptance and to achieve assistance for all the persons. A key point to provide customization is using learning techniques. In this work we made a survey about the state of art of work in Ambient Assisted Living and work about learning of people. Also we proposed a trend about the Customizable Assisted Living and some future directions.

1 Introduction

Recent advances in health technologies along with improved medical care have led to a steady increase in life expectancy in recent decades. As a result, there is significant growth in the number of older people in all the world. In recent years, study to get a comfortable environment for the elderly has gained much attention.

By leveraging information technology and communication, the focus Ambient Assisted Living (AAL) shows great promise in meeting many needs of elderly people and allows them to live safe, healthy and Independent. In recent years, several AAL systems based in wireless networking technologies have been proposed to improve the quality of life of elderly (D. He & S. Zeadally, 2015).

The elderly people or with physically problems, always seek independence and performance their daily activities without having to depend on others, so the purpose of a

AAL system is prolong the life of these people while living at home independently using information technology and computers. (Pavlina Koleva, 2015).

A fundamental problem in people with motor disabilities, people in adulthood or handicapped people is that often they are home alone. This fact affects the user mentally and emotionally. Due this, as system must have friendly interaction, social communication and must providing company.

It appears that happiness does not predict longevity in sick populations, but that it does predict longevity among healthy populations. So, happiness does not cure illness but it does protect against becoming ill. The effect of happiness on longevity in healthy populations is remarkably strong. The size of the effect is comparable to that of smoking or not. If so, public health can also be promoted by policies that aim at greater happiness of a greater number. That can be done by strengthening individual life-abilities and by improving the livability of the social environment (R. Veenhoven, 2008).

A customized system for the assistance not only must consider confort and help to do activities. It is necessary also to consider the quality of life by promoting social interactions, especially for those elderly who live alone. Internet could play a role in relieving the loneliness of elderly, in order to address its social fragility. But the design of appropriate services and ensure its adoption by older adults remain open questions (Alaoui & Lewkowicz, 2015).

Customizing a system is critical to user acceptance and to achieve assistance for all the persons. A key point to provide customization is using learning techniques. The theme of learning activities or user behavior has been widely investigated because it is considered very important, relevant, and also a subject with difficulty (Cook & Krishnan, 2014). Today this problem has not been fully resolved, as many of the investigations are made in a individual way and for the needs of the application (Aztiria, Izaguirre, & Augusto, 2010). So researchers are trying to create a comprehensive model. This model could be based in the opening of combination of different techniques of artificial intelligence.

It can be seen that user support personalized way is a very complex work. Therefore it is proposed as a solution strategy, make use of intelligent agent technology to make individual and global tasks. In this way, an intelligent System with a multi-agent architecture hierarchically seems to be a promising solution. This survey provides a collection of works related user assistance and it is presented a strategy for personal assistance solution.

2 A study of current works

Over the last decade there have been various investigations about learning users since several points of view.

2.1 Focused works in learning users

A very important job where they began to implement fuzzy logic for learning and also applied in a real intelligent environment was conducted by (Hagras, et al, 2004). In an experiment of five years and a half, a user occupied the iDorm, testing their ability to learn user behavior and adapt to the needs of this. Built discreetly agent iDorm controlled according to user preferences.

The technique they used (Hagras, et al, 2004) was to use a set of fuzzy logic controllers in parallel (FLC, fuzzy logic controllers), where each forms a user behavior. Hagras et al. (2004) justified the use of FLCs. They said that these (FLC) are useful when processes are too complex for conventional quantitative analysis techniques. The FLC are also used when available sources of information are interpreted qualitatively, imprecisely or inaccurately. They also used a hierarchical assembly fuzzy controllers to reduce the number of required rules. The reason of this is because there were many entrances and exits of the agent and control user behavior was complex. In this way Hagras et al. (2004) divided the behaviors available in embedded agents of iDorm in fixed and dynamic sets. The dynamic behaviors are learned from the behavior of the person who is in the room. Fixed behaviors are preprogrammed.

The authors used for each of the FLC a singleton fuzzifier, triangular membership functions, inference products, max-product composition and defuzzification with maximum membership. The reason of this was because of the computational simplicity and real-time considerations. In summary, the project had a set of fuzzy rules a priori controls of each behavior that takes place in the room (fuzzy set of fixed controls). The system changed the rules to suit the user's wishes with interactuación-user system. In this way new fuzzy controls (dynamic fuzzy controls) were obtained.

Another work was made by (Doctor, Hagras, & Callaghan, 2005). In this work was used fuzzy logic for learning and it was used the intelligent room iDorm of Hagras et al. (2004) for experimentation. This work differs from (Hagras, et al, 2004) in the fact that they generated their membership functions and fuzzy rules from the database collected by monitoring user. That is, the knowledge base was obtained from the pattern recognition. The authors used a technique of artificial intelligence of dual clustering. The first cluster grouped feature vectors from minimum distance, setting a priori the number of clusters that wished to obtain.

Then (Doctor, Hagras, & Callaghan, 2005) took the center of each cluster and merged into a new database. In the new database was applied the second hierarchical cluster, but now each dataset considered a single feature as a group (each column as a group). Therefore the cluster number was equal to the number of features. The elements of

each cluster were ordered sequentially. In the same time, the number of elements in each iteration was reduced by the fusion of two element subsequent more like. This is repeated until the number of elements is the number of membership functions desired for each feature.

It is noteworthy that the feature vector at work (Doctor, Hagra, & Callaghan, 2005), was formed by the values of sensors and actuators intelligent environment. Therefore, the number of features or dimension of the vector is equal to the sum of the number of sensors and actuators in the existing space or environment.

Doctor et al. (2005) proposed that the fuzzy rules were obtained from the combination of sensor values with operators AND. The result of the conditional values was the learned actuators database. So fuzzy control was obtained, whereas a change of habit or change of user could be a new rule to control, as long as this changed action was repeated a number of times.

Subsequently, in (Hagra, Doctor, Callaghan, & Lopez, 2007) the authors used the proposed learning model by Doctor et al. (2005), but they used fuzzy logic type 2, showing satisfactory results. In a study made by (Aztiria, Izaguirre, & Augusto, Learning patterns in ambient intelligence environments: a survey, 2010) say that the work done by (Doctor, Hagra, & Callaghan, 2005) is robust when sensors are continuous (temperature, humidity, time, etc.).

Of course fuzzy logic has not been the only technique for learning behavior. There are research in neural networks such as the control system (Mozer, et al, 1995) installed on "The adaptive house". The goal of this research was to design an adaptive control that considers the lifestyle and consumption of energy habitants. They used a feed-forward neural network to develop a prediction of occupation and anticipation of area. In this way, it was predicted where the user would be in the next seconds to control (on or off) lighting.

Also, (Campo, Bonhomme, Chan, & Esteve, 2006) developed a system that calculates the probability of occupation for each area of the house and systematically compared the probability with the current situation, using neural networks.

There has been research in the area of discovery sequence. The group of researchers who have more actively working on this area are those who have developed the "Houses" and "MavHome" project (Cook & Das, 2007).

Naive Bayes classifier has been used with promising results for the recognition of the activity in smart places, as in the case of (Brdiczka, L.Crowley, & Reignier, 2009), (Maurer, Smailagic, Siewiorek, & Deisher, 2006) (Kasteren & B. Krose, 2007) and (Nazerfard & Cook., 2013). The authors model a probability function that can be used to identify the most likely given the feature set of sensor activity. They use sensors dependence on labels activities. These classifiers are robust when large amounts of sample data are used. However, Naive Bayes classifiers not explicitly model every

time information that generally is considered important in the recognition of the activity (Cook & Krishnan, Mining the home environment, 2014).

(Aztiria, Izaguirre, & Augusto, 2010) conducted a study on the different techniques used in state of the art, until 2010. The authors commented that the analysis of different techniques reveals that different groups have chosen techniques based on the specific needs of their environments or applications.

They stated that there is still no global or integrated approach that allows the system to learn, to understand with ease patterns, to represent human behavior by the meaning of subsequent actions, to relate actions and adapt other patterns. Aztiria et al. (in the study of the different techniques) propose that the combination of artificial intelligence techniques may be a promising way.

A job that has begun to combine artificial intelligence techniques for learning activities is carried out by (Ros, Delgado, Vila, Hagra, & Bilgin, 2012). They make use of various techniques such as fuzzy logic, the Apriori algorithm, and an algorithm for optimal convergence of learned action to pick the most likely action in each defined instant "k".

Article Learning Frequent Behaviors of the Users in Intelligent Environments of (Aztiria, Augusto, Basagoiti, Izaguirre, & Cook, November 2013), the authors talk about a system that learns from frequent patterns of behavior of system users. The system discovers frequent user behavior taking into account the specific characteristics of ambient intelligence. The authors divide learning three stages, conversion layer, layer learning and application layer.

Another work of great importance is (Krishnan & Cook, 2014), in which mention that activity recognition approaches use a script or sequential segments sensor events related to the activities. Therefore they propose a sliding windowing time database approach.

In the previous research of works related with learning, we can see that the theme of learning activities or user behaviors has been widely investigated because it is considered very important, relevant, and also a subject with difficulty (Cook & Krishnan, 2014). Today this problem has not been fully resolved, as many of the investigations are made to private individuals and to needs of the application (Aztiria, Izaguirre, & Augusto, 2010).

2.1 Focused works in Ambient Assisted Living (AAL)

There is an association called "Ambient Assisted Living" (AAL) was founded on September 19, 2007 in Brussels, Belgium.

The aim of the association is to improve the conditions of life, autonomy, participation in social life, competition, employability, reducing health costs and social care of older adults, so it also leads to strengthening industrial opportunities in Europe through the use of information and communication technologies (Ambient Assisted Living Joint Programme, 2012).

The AAL Association, has developed many projects, among them is "DOME0" which focuses on the development of robotic platform for integration and adaptation of care services custom home, which include voice recognition and synthesis of dialogue, cloud service tele-presence, medical and nonmedical sensors, interface with infrastructure home.

In (Pavlina Koleva, 2015), describe the practical application of an architecture of a system AAL, which aims to provide a monitoring system of personal health combined with positive reinforcement, medical reminders, connecting to social networks for improving of social relations, in addition to home security and alerts for medical personnel in case of emergency.

In (A. J. Jara, 2013), this paper presents an interconnection framework for mobile health (mHealth) based on the Internet of things. This makes a continuous and remote monitoring of vital signs and introduces technological innovations and devices with Internet capabilities that enable patient monitoring and remote monitoring centers and personal, such as tablets platforms.

Lloret, et al (2015) uses the information obtained from sensors, data traffic patterns and behavior of a person (habits) to make decisions and send alarms.

The system proposed by Lloret, et al (2015), includes the following features:

- Independence communication technology
- Supports different handlers on the same network
- Adaptable to different needs of the elderly advanced
- Suitable for different environments
- Low power consumption to avoid large load periods
- Low consumption of bandwidth to allow many wireless technologies are saturated
- Easy to use
- Scalability in order to allow new sensors are added or removed as needed
- Small size and not easily detectable in order to avoid visibility
- Allow and control the movement and displacement of older people

The authors have applied an intelligent monitoring system that learns from user decisions and caregivers and / or medical for future cases. Intelligence techniques used were some patterns classifiers, however they do not express the performance study them. They also presented the design of some algorithms and representation of the knowledge acquired environmental information management and wireless network.

2 Strategy for personalized assistance

In a user support must be met multiple objectives. These objectives can be help to perform daily activities, environmental comfort, appointment reminders and messages, communication through social networks, home security or environment, anomaly detection and risk situations, monitoring user and even developing therapies or Healthcare. In a user support must be met multiple objectives. These objectives can be help to perform daily activities, environmental comfort, appointment reminders and messages, communication through social networks, home security or environment, anomaly detection and risk situations, monitoring user and even developing therapies or Healthcare. Considering the above you could say they have to fulfill particular tasks or even tasks could relate to each other to meet their objectives.

Therefore you can take advantage of the characteristics of multi-agent systems (MAS) to design a strategy for user support. Some characteristics of MAS presented below Sycara (1998):

- A global control not exist in the system. Individual agents can cooperate with other agents in the MAS to achieve individual goals or they can coordinated to maximize global profits.
- Data and environment are decentralized, all the agents can generate changes in the environment with their own “influences spheres”
- The computing is asynchronous, agents can take their own tasks independently, without having to wait for a signal of controller center.
- Each agent has incomplete information to solve a problem. Agents can achieve global goals through competition, collaboration or other interactions.

Whereas in user support, some tasks will be more important than others, the architecture of MAS can be considered hierarchically. This type of architecture is characterized by some agents who have authority over the actions of other agents. In the state of art architectures have been proposed over two or three hierarchies, considering that you can be more (Kantamneni, 2015).

In the multi-agent system is essential to consider an agent responsible for user interaction, which specializes in providing a friendly assistance and also provide natural communication, it is proposed to call this agent: Virtual Assistant by voice output. No doubt it is still in research communication through natural language between machine-human. However in early applications and usability studies you can consider issuing recordings of comments, reminders or explanations of the actions performed by the system in the environment using a horn in the environment. This way you can see the user-system interaction by issuing voice with particular styles that make the user will not feel alone.

Other indispensable agents in the multiagent system should be in charge of learning patterns, habits or user trends that will get the personalization of care. This also should

consider de-protection agents anomalies, agents sensors, actuators and monitoring agents.

This strategy allows research on the interaction of different actors to integrate different techniques artificial intelligence, learning models and technologies. This approach is a new area in the implementation of MAS with different deliberative agents and artificial intelligence (see figure 1).

The elderly people or with physically problems, always seek independence and performance their daily activities without having to depend on others, so the purpose of a AAL system is prolong the life of these people while living at home independently using information technology and computers.

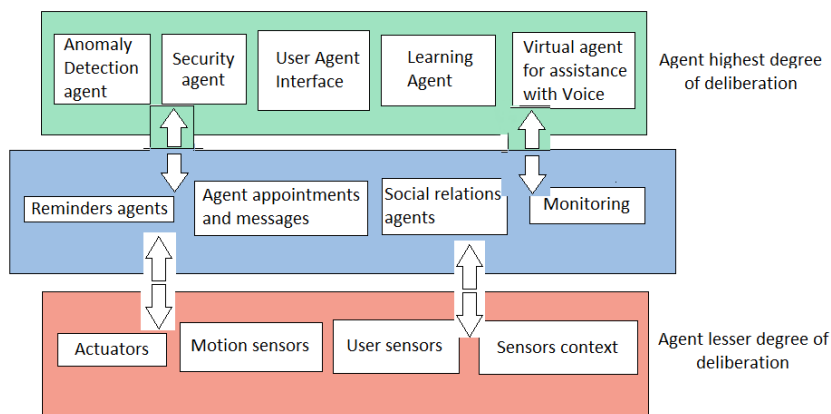


Fig. 1. Model of Multiagent System

Future directions

The Multi-agent model is proposed based on the study of works in Ambient Assisted Living and learning users. With the multi-agent system is intended to facilitate the daily lives of these people but aspects that improve the quality of life they are also considered. The research focuses on the design of the model of care, where you can conduct research for coordination and communication strategies to manage the information necessary to achieve the desired assistance. Future work will be performed as studies of therapies to increase mood, health therapies, usability studies, etc. The tendency for assistance user is increasing the quality of life considering the social, emo-

tional and about the feeling of loneliness needs. Preventing disease and improving users due to an appropriate mood. As a result is expected that the adults and / or persons with disabilities have more quality life time at home.

References

1. Baldonado, M., Chang, C.-C.K., Gravano, L., Paepcke, A.: The Stanford Digital Library Metadata Architecture. *Int. J. Digit. Libr.* 1 (1997) 108–121
- A. J. Jara, M. A.-I. (2013). Interconnection Framework for mHealth and Remote Monitoring Based on the Internet of Things. *IEEE Journal on Selected Areas in Communications*, 31(9), 47-65.
- aal-europe.eu. (2012). Recuperado el 12 de Enero de 2015, de <http://www.aal-europe.eu/contacts/>
- Asier Aztiria, J. C. (noviembre 2013). Learning Frequent Behaviors of the Users in Intelligent Environments. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS*.
- Aztiria, A., Augusto, J. C., Basagoiti, R., Izaguirre, A., & Cook., D. J. (noviembre 2013). Learning Frequent Behaviors of the Users in Intelligent Environments. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS*.
- Aztiria, A., Izaguirre, A., & Augusto, J. (2010). Learning patterns in ambient intelligence environments: a survey. *Springer*, 34, 35-51.
- Aztiria, A., Izaguirre, A., & Augusto, J. C. (2010). Learning patterns in ambient intelligence environments: a survey. *Journal Article. Artificial Intelligence Review. Springer Netherlands*, 34(1). doi:10.1007/s10462-010-9160-3
- Brdiczka, O., L.Crowley, J., & Reignier, P. (Febrero de 2009). “Learning situation models in a smart home,” *IEEE. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(1), 56-63. doi:10.1109/TSMCB.2008.923526
- Campo, E., Bonhomme, S., Chan, M., & Esteve, D. (2006). Learning life habits and practices: an issue to the smart home. *International conference on smart Homes and health telematic*, 355-358.
- Cook, D. J., & Das, S. K. (Marzo de 2007). How smart are our environments An updated look at the state of the art. *Pervasive and Mobile Computing*, 3(2), 53-73. doi:10.1016/j.pmcj.2006.12.001
- Cook, D. J., & Krishnan, N. (22 de Octubre de 2014). Mining the home environment. *Journal of Intelligent Information Systems*, 503-519. doi:10.1007/s10844-014-0341-4
- Doctor, F., Hagraas, H., & Callaghan, V. (Junio de 2005). An Intelligent Fuzzy Agent Approach for Realising Ambient Intelligence in Intelligent Inhabited Environments. *IEEE Trans. Syst., Man, Cybern*, 35(1), 55-65. doi:10.1109/TSMCA.2004.838488

- Hagras, H., Callaghan, V., Colley, M., Clarke, G., A., P.-C., & Duman, H. (Nov-Dic de 2004). Creating an ambient-intelligence environment using embedded agents. *Intelligent Systems, IEEE*, 19(6), 12-20. doi:10.1109/MIS.2004.61
- Hagras, H., Doctor, F., Callaghan, V., & Lopez, A. (Febrero de 2007). An Incremental Adaptive Life Long Learning Approach for Type-2 Fuzzy Embedded Agents . *Fuzzy Systems, IEEE Transactions on Ambient Intelligent Environments*, 15(1), 41-55.
- I. Bisio, F. L. (January de 2015). Smartphone-centric ambient assisted living platform for patients suffering from co-morbidities monitoring. *IEEE Communications Magazine*, 53(1), 34-41.
- Juan Carlos Augusto, P. M. (2007). *Ambient Intelligence: Concepts and Applications*. ComSIS
- Kasteren, T. v., & B. Krose. (2007). Bayesian activity recognition in residence for elders. *Proceedings of the IET International Conference on Intelligent Environments*, 209–212.
- Krishnan, N. C., & Cook, D. J. (Julio de 2014). Activity recognition on streaming sensor data. *Pervasive and Mobile Computing. El Sevier*, 10, 138–154. doi:10.1016/j.pmcj.2012.07.003
- Mozer, M., Dodier, R., Anderson, M., Vidmar, L., Cruickshank, R., & Miller, D. (1995). The neural network house: an overview. . Erlbaum. *Current trends in connectionism*, 371–380.
- Nazerfard, E., & Cook., D. (2013). Using Bayesian networks for daily activity prediction. *Proceedings of the AAAI Workshop on Plan, Activity and Intent Recognition Workshop* .
- Pavlina Koleva, K. T. (2015). Challenges in Designing and Implementation of an Effective Ambient Assisted Living System. *Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS)*, 2015 12th International Conference on, 305-308. doi:10.1109/TELSKS.2015.7357793
- Romero-Rodríguez Wendoly J. Gpe., Z. R.-F. (2011). Comparative Study of BSO and GA for Optimizing Energy in Ambient Intelligence. *Lecture Notes in Computer, Science* 7095/2011(DOI: 10.1007/978-3-642-25330-0_16), 177-188.
- Ros, M. (2014). A Fuzzy Logic Approach for Learning Daily Human Activities in an Ambient Intelligent Environment. *IEEE Transaction Power Electronics*.
- Thathachar, M. A., & Sastry, P. S. (2002). Varieties of learning automata: an overview. *IEEE Transactions on Systems, Man, and Cybernetics, Part B.*, 32(6), 711-722.
- R. Veenhoven: Healthy happiness: effects of happiness on physical health and the consequences for preventive health care. *Journal of Happiness Studies*. September 2008, Volume 9, Issue 3, pp 449-469. DOI: 10.1007/s10902-006-9042-1

Process Choreography in Wireless Sensor and Actuator Networks: a proposal to achieve Network Virtualization

Sara Blanc, Jose Luis Bayo-Monton, Jose Carlos Campelo, Carlos Fernandez-Llatas

Instituto Universitario de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas. Universitat Politècnica de València. Camino de Vera s/n. 46022 Valencia, España.

{sablaca, jobamon, jcampelo, cfillatas}@itaca.upv.es

Abstract. Wireless Sensor and Actuator Networks (WSANs) can be considered as one of the basic constituents of Internet of Things (IoT). Although during last years WSANs has been designed for a wide range of applications, they have been task oriented and domain specific. Emerging research advocates for reusing WSAN deployments enabling their usage by multiple users. This can be achieved by means of the virtualization concept: it allows the abstraction of computing resources into logical units. Several approaches to follow this concept have been proposed at different levels. This paper proposes the use of Process Choreography, a paradigm based on Service Oriented Architecture (SOA), combined with WSANs to obtain WSAN virtualization and to allow the use of WSANs in IoT in an effective way.

1 Introduction

Internet-of-Things (IoT), Web-of-Things (WoT) or Sensor Cloud(SC) [1,2] are examples of emerging concepts that are considered to be the next technological revolution. During last decade, Wireless Sensor and Actuator Networks (WSAN) have been useful for a wide range of applications [3,4,5,6]. These deployments have been domain-specific and task-oriented. So WSAN designs have been tailored for particular applications and with no possibility of using them in other applications.

To mitigate previous drawbacks, new trends have emerged in this context: (i) the possibility of running different applications on heterogeneous nodes in different WSANs in order to better exploit deployed networks infrastructure and (ii) the capacity of accessing sensor node data from the web to achieve the IoT, WoT or SC concepts.

Virtualization [3,4] is a key technique for the realization of the future Internet, and its use in the WSAN context must be researched. Virtualization makes possible to present physical computing resources by abstracting them into logical units, enabling their usage by multiple users (or applications). Additionally, it allows the deployment of applications that were not even envisioned during initial WSAN deployments.

So, virtualizing WSNs can offer many interesting benefits: new applications can run over existing WSN implementations, the elimination of tight coupling between WSN services/applications and WSN deployments and WSN applications and services can be used by third-party applications.

This kind of network virtualizations require the creation of distributed architecture, an architecture capable of manage the data of multiple sensors and petitions of heterogeneous applications. The architecture needs to be highly configurable and dynamic to achieve the virtualization necessities.

To tackle this architecture, the patter of Service Oriented Architecture (SOA) [7] is a good proposal. In this pattern the system is composed of a self-contained units of functionality called services. Each of this services communicates and interoperates with other through a common communication protocol, to provide the whole functionality of the system.

Due to the nature of a WSN the architecture needs to deal with highly coupled processes and to represent the different states of the process. This kind of necessities are better covered by the paradigm based on SOA known as Process Choreography [8,9]. In this paradigm the process is defined as a composition a services and the decision to change from one state to other is taken in a distributed way.

2 Materials and Methods

2.1 WSN Virtualization

Existing research about WSN virtualization is classified in one of the next three groups: i) Node Level virtualization, ii) Network Level Virtualization and iii) Hybrid solutions.

Research in **Node level virtualization** [10] are focused on how to execute different applications in each sensor node. Different solutions can be found in the bibliography: a) OS-based solutions, b) Middleware-based solutions and c) Virtual-machine solutions. In OS-based solutions, virtualization is part of the OS. In Middleware and Virtual-machine based solutions a specialized component running on top of the OS is in charge of virtualization. Several proposals can be referenced: some of them supports the execution of concurrent applications in resource-constrained nodes; others focus on the event-drive or thread-based programming models; the stack memory and how logical addresses are managed are studied in other references. Benchmarking studies of those proposals can be found [11,12,13,14].

Nevertheless, Node level virtualization must be extended with the actual SoCs designs, taking into account the re-programmability of programmable devices (i.e FPGAs) included and recent advances in sensor hardware technology. So, this advances have made it possible to produce sensor nodes that have more processing power, memory and prolonged battery life evolving towards Smart Sensors. New generation of IoT sensors focus on three aspects: Internet connectivity, security and context-aware resources, which are already a reality to both, microcontroller and microprocessor based platforms. While microprocessors are assumed in all the three related as-

Context-aware resources: Dynamic runtime application change and reconfiguration are two complementary modes of providing context-aware capability. While the traditional method is supported by embedded operating system dispatchers, ISP (In-System Programming) and In-Application Remapping are now additional resources to faster reconfiguration. The In-System Programming (ISP) system service can be used for field upgrades of the hardware (a SoC includes an FPGA component) and for application code reprogramming. The application program running on the Cortex-M3 controls the ISP operations. Such application program is allocated in a non-volatile-memory slot and dormant during normal operation. An in-application remapping internal change activates this application program that must retrieve the programming file from a communication port and feed it to the System Controller through the ISP system service. The System Controller performs the actual programming of the FPGA and the Non-Volatile-Memory (application code memory) using the programming data it receives from the Cortex-M3 as part of the ISP system service. This scheme provides intrinsic node virtualization capabilities in an interesting way.

Research in **Network level virtualization** [15] are focused on two lines: virtual network/overlay-based solutions and cluster—based solutions. Virtual network/overlay are logical networks created on top of physical network. In cluster-based solutions, the nodes in a physical network are grouped to work together in connected groups (clusters). Figure 2 shows the network-level virtualization types. Proposals of network-level virtualization define the networks formation mechanisms (based on application overlays, virtual links, embedded agents or cluster trees) and the protocols used (JXTA, AODV, Hierarchical clustering by example). Nevertheless, there has been very few implementations and only simulation results are showed in papers [16,17,18,19,20,21].

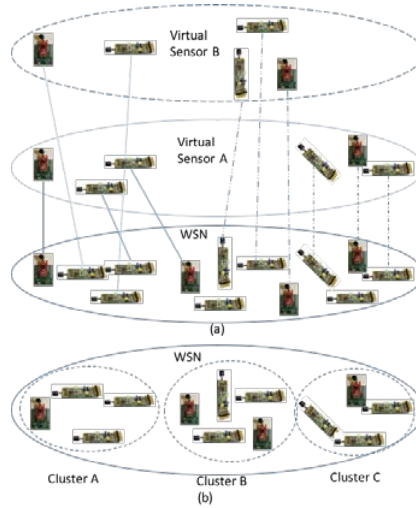


Fig. 2. Network level virtualization solutions: a) Virtual network-based solution, b) Cluster-based solutions

Research in **Hybrid Network Virtualization** [22, 23] combines both node and network-level virtualization mechanisms. Basically, three types of hybrid solutions can be found: middleware and cluster-based solutions, middleware and virtual networks-based solutions and virtual machine and dynamic grouping-based solutions. In the first one a middleware handles node-level virtualization while network-level virtualization is achieved by grouping sensor nodes into clusters. In the second one, a middleware handles node-level virtualization while network-level virtualization is achieved using virtual networks/overlays. Finally, in the third one, virtualization is achieved using a virtual machine and a tailored sensor node grouping scheme is used for network-level virtualization.

From the previous desired objectives to be achieved by means of virtualization several requirements can be enumerated to generate an effective solution:

1. Availability of node level virtualization: sensor nodes must support the concurrent execution of multiple applications.
2. Network-level virtualization: the ability of sensor nodes to dynamically form groups to perform the isolated and transparent execution of multiple tasks in such a way that each group belongs to a different application.
3. Support for application/service priority: mission-critical applications or service must have prioritized execution mechanisms.
4. WSN virtualization should be platform-independent.
5. A resource discovery mechanism for neighbor discovery and service discovery must be supported.
6. Although today's sensor nodes have enough resources, solutions must be also applied to resource-constrained sensor nodes.
7. Virtualization solutions should be applicable to a variety of WSN platforms (heterogeneity).

2.2 Process Choreography

In the actual context of business processes, an architecture that provides flexibility and adaptation capabilities to the changing environment is essential and the paradigm of Service Oriented Architecture (SOA) has been proposed as a solution for various organizations [24,25,26].

The OASIS organization defines SOA as “*paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains. It provides a uniform means to offer, discover, interact with and use capabilities to produce desired effects consistent with measurable preconditions and expectations. The SOA-RM specification bases its definition of SOA around the concept of “needs and capabilities”, where SOA provides a mechanism for matching needs of service consumers with capabilities provided by service providers.*” and Service is defined as “*Mechanism to enable access to one or more capabilities, where the access is provided using a prescribed interface and is exercised consistent with constraints and policies as specified by the service description.*” [27].

A more concrete definition is provided by The Open Group who describes SOA as *“an architectural style than supports service-orientation”*. Being service-orientation a *“way of thinking in terms of services and service-based development and the outcomes of services”*. And ending with the definition of a Service as a *“Logical representation of a repeatable business activity that has a specified outcome. It is self-contained, may be composed of other services and is a black box to consumers of the service”* [28].

Recapping all the definitions, a system than implements a SOA Architecture is composed by services, each service is self-contained and provides a unit of functionality. The systems must provide the mechanism to communicate the services, mechanism by means of the services can be composed and interoperate to create the functionality for a whole process.

SOA provides multiple benefits to the system, being some of them: flexibility, reduce costs by service re-use, adaptability, rapid develop of new functionality based on service composition and scalability of the system [28].

At the SOA generic architectures the Services tend to be decoupled by definition, but the processes than run in a WSN need to be coupled due to the necessity of coordinate sensors or configurations of specific hardware. Furthermore, SOA Services typically are stateless and in a WSN manage the state of the process is required.

A specialization of SOA called Process Choreography [8,9] is suggested as a better approach. This paradigm uses a top to down design where a unique process is defined in a distributed way, the functionality of a process is provided for a composition of services, in a composition each of the services involved knows when to execute and with whom to interact. The management of the state and the decision to change it is taken in a distributed way, the combination of the different flows of the system provides the state.

Moreover, the Choreography model avoids the SOA overheads derived from the service discovery or data adaptation between services.

In order to test the proposed architecture of this paper, a Choreography engine implementation developed by the authors and presented in previous works is used [8, 29, 30]. This engine provides a methodology to register services inside them and protocols to send messages between services.

The engine is complemented by a set of connector services that provides extra functionalities, like the one that interconnect, through TCP, engines running on different machines, abstracting to services of where are running, providing the tools to create a distributed and scalable architecture without rewriting original services. Among other, they have been developed connectors than provides REST communication, Web Service communication, Message Queuing communication, Service Orchestration, execution of R scripts or execution of Matlab code.

The specific message protocol interchanged between services is called XMSG [8] and is based on well-known protocols like FIPA [31] and SOAP [32]. This protocol allows broadcast, multicast and Peer to Peer message management using wild card symbols and patters in the destiny and reception address.

3 Results

Based on the capabilities of the described Choreography engine, the authors propose an architecture for WSN based on the Virtualization Network Level. As we can see at the Fig.3, the engine is positioned at the center of the architecture, used as a middleware and integrator system.

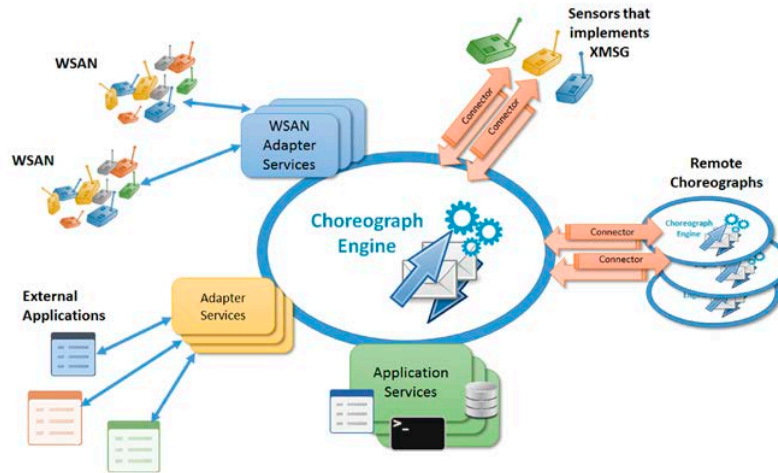


Fig 3. Proposal of Architecture for WSN based on Choreography.

The physical network architecture is virtualized to the sensors and applications by the Choreograph, it is the responsible to manage the distribution of choreography messages regardless of where are located, different Choreographs situated on different locations can interchange the messages using the communication connectors, furthermore, connectors can be configured to create different networks between Choreographs and different topologies of networks.

From the point of view of the different elements integrated in the architecture, each of them use different approaches.

- Sensors with enough resources to implement XMSG protocol to send or receive choreography messages, can be directly connected to choreograph through a communication connector.
- Sensors without enough resources to implement XMSG protocol or legacy WSN can be virtualized developing adaptor services than translate between protocols.
- Legacy External Applications are integrated on the virtual network by the creation of adaptor services for the translation of protocols.
- New Applications can be developed by creating new services and using composition of services. It provides distributed, scalable applications with a reduction on costs.

The virtualization of a WSN give it the capacity of being discovered, retrieved, interpreted and managed. Virtualization of WSN lead it to be promoted as a service

provider. Thus, legacy or new applications subscribe to the WSN service to both, requesting for information and retrieving information.

As initial works to test the proposal of architecture, a specific implementation of a real scenario is in development. This will produce experimental data to evaluate the architecture.

The test scenario proposes a WSN with legacy or resource-constrained sensor that will be virtualized. Besides, some services will be created to complete the choreography that defines the work flow to solve, as a result the needed applications and functionality will be obtained.

For a potential subscriber services, WSN discovering and requesting processes depend on the semantic interpretation of metadata gathered from the WSN. Thus, the sensor metadata model must be a useful image of the WSN performing context to any application to know when data sent by the WSN are or not the type of data that the application really needs.

Although sensors package their measures in a standard format, for example SensorML or O&M models, it enabled interoperability at the syntactic level, but the semantic understanding of WSN activities need a specific component in the network virtualization.

Moreover, because sensors are constrained in resources, actions done by a sensor per time slot are also limited. However, smart sensors are assumed to be reconfigurable in runtime changing actions done by the sensor oriented to a specific network activity: the context change.

Managing context changes and semantic interoperability can be done in the local network or at cloud level. However, the integration of the WSN in a choreograph middleware makes sensible to derivate both semantic model interoperability adaption and WSN context change managing to choreograph.

Such derivation means the design at the choreograph level of new strategies such as:

- Knowledge of network aspects of functionality, availability, reliability, security, etc.
- Knowledge about different aspects of sensor data, from the description of the sensor itself to the action carried out by the sensor node.
- Knowledge in syntactic aspects of data retrieved.
- Control, filtering and upgrading in WSN message traffic.
- Decision making processes in WSN context change.

To overcome these challenges, we propose a WSN management service in charge of WSN discovery, data retrieve and context management. Semantic operability is virtualized in this service while syntactic interoperability is done at node level by using standard formats.

Letting out of the scope of this paper both discovery and context management, the work starting point has been focus on data retrieve. Three possible solutions are under study. First, a solution based on Identification Masques and events. Second, a Central Event Retrieve Service; and third, implementing Message Type Oriented Services.

Identification Masques

Figure 4 shows a data retrieve process with three parties: the WSA, a manager service and an application subscriber. Events represent messages sent by the WSA and identified by type. It is assumed that information contained in data includes sensor specific metadata such as timestamp and geo-position.

The application/service subscriber requests a subscription to the manager service whom returns an acknowledgement in case of current availability of the type of message requested. Since this moment, events are directly received by the application subscriber without the manager intervention. Thus, the subscriber needs the ID that identifies the type of message expected by the application and a dynamic update of its own ID into the choreograph software. The new application ID will be an extension of its unique identifier plus the ID of messages type A (in the example). The choreography is able to resolve message dispatch by consulting.

Central Event Retrieve Service

Fig 5 represents the message flow interchange between the WSA manager service and several application/service subscribers. Events received in the choreography pass through the manager service whom re-directs them to subscribers. No update in the application ID is needed while the manager service maintains a message redirection table crossing subscribes IDs and message type IDs.

Message Type Oriented Services

This proposal is an extension of the event retrieve service into several dedicated services to any of possible WSA message types.

The use of identification masques for WSA message types is a decentralized solution which offers advantages such as:

- Reduction of choreograph internal traffic by reducing event retransmission.
- Increases subscriber's autonomy by stopping and restarting retrieve without supervision.

However, focusing on context decision making, the solution unguaranteed an efficient configuration of the WSA. While contextual changes are based on managing subscription requests the solution lacks of control about un-subscriptions. Even if un-subscription messages will be added to the control flow were not guarantee of real sending by non-penalty.

On the other hand, an Event Retrieve Service is a centralized solution which increases both the complexity of the data retrieve process and the internal choreograph traffic. Context decision making increases in efficiency, but a high rise in subscriptions will turns out in a detriment to reception delays. Thus, the third solution, the Message Type Oriented services, will be analyzed in scenarios of high subscriptions variability.

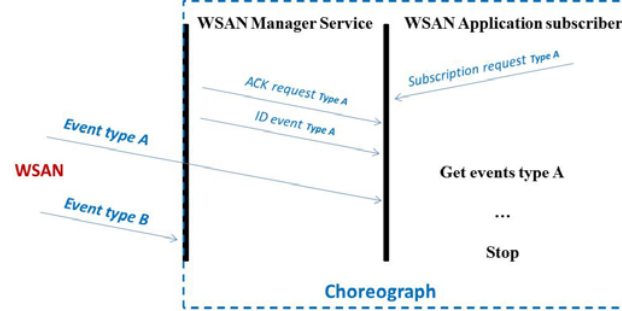


Fig 4. Control message flow of an Identification Masques data retrieve solution

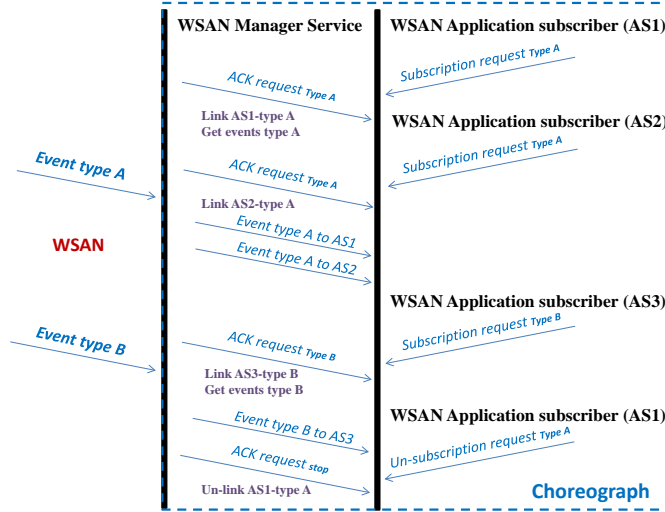


Fig 5. Control message flow of a Central Event Retrieve Service solution

4 Conclusions

At this paper the authors expose a proposal of architecture for Network Virtualization of a WSAN. Previous works on WSAN have been domain-specific, task-oriented and tailored to solve specific problems. This provides isolated networks, with no possibilities to extend to other fields.

To tackle some of the requirements to generate effective network virtualization of WSAN, the Process Choreography paradigm is introduced as virtualization middleware at the center of the architecture. Also, an implementation of Choreography engine based in previous works by the authors are suggested to test de architecture.

The first of the requirements to accomplish, is the virtualization at node level, although the majority of nodes in a WSAN tend to be resource-constrained, some of

them have enough capabilities to run an operating system and run an implementation of a Choreograph engine, having the possibility of replicate the architecture at node level and providing the desired virtualization.

As a second requirement the network virtualization is required, this is tackled through the connection of the different elements of the WSAN to Choreographer and his redefinition as a Services, all the communication is based on choreography messages and elements/services are independent of his physical localization. Furthermore the engine provides connection capabilities between choreographs running on different machines, providing high flexibility and scalability. Thanks to this different work flows related to different applications can function in cooperation or in aisle mode, and elements can be grouped virtually by changing network topology or using defined patterns in service identifiers.

The third element defined as an application/service priority support, at this moment it is not supported by the actual implementation of the proposed choreography engine, but the architecture is not a limitation element at this point and the development of the engine is in continuous process, as a future work this functionality may be an option.

To generate a platform-independent architecture, requested as a fourth requirement, there are three ways to get it. We can have elements that implement XMSG protocol and can be integrated directly by using connector services. Next, for elements that do not implement XMSG, adapter services can be developed. And finally, implementations of the choreograph engine for different architectures can be developed.

A resource discovery mechanism, presented as a fifth requirement, is not tackled by the architecture itself, this mechanism can be provided by intermediary services, at results section has been explained a work in progress to test this by three different points of view: Identification Masques, Central Event Retrieve Service, Message Type Oriented Services. The results of this test will be explained by the authors in future publications.

The sixth requirement is accomplished by the virtualization of entire legacy WSANs and developing specific adapter services to connect them to choreographer.

The seventh requirement, of be applicable to a variety of WSAN, is tackled by the possibility of develop specific adapters for every WSAN or the capacity of the sensor to implement XMSG protocol.

The work in progress explained at previous section and future works to test the qualities of this architecture need to be tackled, but a priori, the defined architecture seems a good point of start to virtualize networks for WSAN

References

1. Madria, A., Kumar, V., Dalvi, R. Sensor Cloud: A Cloud of Virtual Sensors. IEEE Software (2014) 70-77
2. Phan, D.H., Suzuki, J., Omura, S., Oba, K. Toward Sensor-Cloud Integration as a Service: Optimizing Three-tier Communication in Cloud-integrated Sensor Networks. BODYNETS, September 30, Octobre 02, (2014), Boston, USA

3. Khan, I., Belqasmi, F., Glitho, R., Crespi, N., Morrow, M., Polakos, P. Wireless Sensor Network Virtualization: A Survey. *IEEE Communication Surveys and Tutorials*, vol. 18, n° 1, (2016) 553-576
4. Khan, I., Belqasmi, F., Glitho, R., Crespi, N., Morrow, M., Polakos, P. Wireless Sensor Network Virtualization: Early Architecture and Research Perspectives. *IEEE Network*, May/June 2015, 104-112
5. Lia, W., Delicatob, F.C., Piresb, F., Leea, Y. C., Zomayaa, A.Y., Micelib, C., Pirmez, L. Efficient allocation of resources in multiple heterogeneous Wireless Sensor Networks. *Journal of Parallel and Distributed Computing*, Volume 74, Issue 1, January 2014, 1775–1788
6. Feki, M. A., Kawsar, F., Boussard, M., Trappeniers, L. The Internet of Things: The next technological revolution, *Computer*, vol. 46 (2013), no. 2, 24–25.
7. Erl Thomas. *Service-Oriented Architecture (SOA): Concepts, Technology and Design*, Prentice Hall, 2005. ISBN: 0-13-185858-0
8. C. Fernández-Llatas, J. B. Mocholí, A. Moyano, and T. Meneu. Semantic Process Choreography for Distributed Sensor Management, in *SSW*, 2010, pp. 32–37.
9. Zimmermann, O., Doubrovski, V., Grundler, J., & Hogg, K. Service-oriented architecture and business process choreography in an order management scenario: rationale, concepts, lessons learned. In *Companion to the 20th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications*, 2005 pp. 301-312 ACM
10. Yu, Y., Rittle, L.J., Bhandari, V., LeBrun, J.B. Supporting concurrent applications in wireless sensor networks in *Proc. 4th Int. Conf. Embedded Netw. Sensor Syst.*, New York, NY, USA, (2006), pp. 139–152.
11. Dargie, W., Poellabauer, C., *Fundamentals of Wireless Sensor Networks: Theory and Practice*. Hoboken, NJ, USA: Wiley (2010).
12. Chu, R., Gu, L., Liu, Y., Li, M., Lu, X. SenSmart: adaptive stack management for multi-tasking sensor networks, *IEEE Trans. Comput.*, vol. 61, no. 2, pp. 137–150, Jan. 2012.
13. Nath, S., Gibbons, P.B., Seshan, S., Anderson, A., Synopsis diffusion for robust aggregation in sensor networks, *ACM Trans. Sen. Netw.*, vol. 4, no. 2, pp. 7:1–7:40, Apr. 2008.
14. Baccelli, E., Hahm, O., Gunes, M., Wahlisch, M., Schmidt, T.C., RIOT OS: Towards an OS for the Internet of Things in *Proc. 32nd IEEE INFOCOM Poster*, 2013, pp. 79–80.
15. Jayasumana, A.P., Han, Q., Illangasekare, T.H. Virtual sensor networks—A resource efficient approach for concurrent applications, in *Proc. 11th IEEE/ACM Int. Conf. on Sensor Networks*, 2007, pp. 115–124.
16. Khan, I., Belqasmi, F., Glitho, R., Crespi, N., A multi-layer architecture for wireless sensor network virtualization, in *Proc. 6th Joint IFIP WMNC*, Dubai, UAE, 2013, pp. 1–4.
17. Hoebeke, J., Poorter, E.D., Bouckaert, S., Moerman, I., Demeester, P., Managed ecosystems of networked objects, *Wireless Pers. Commun.*, vol. 58, no. 1, pp. 125–143, May 2011.
18. Ishaq, I., Hoebeke, J., Moerman, I., Demeester, P., Internet of things virtual networks: Bringing network virtualization to resource-constrained devices, in *Proc. IEEE Int. Conf. GreenCom*, 2012, pp. 293–300.
19. De Poorter, E., Troubleyn, E., Moerman, I., Demeester, P., IDRA: A flexible system architecture for next generation wireless sensor networks, *Wireless Netw.*, vol. 17, no. 6, pp. 1423–1440, Aug. 2011.
20. Tynan, R., O'Hare, G.M.P., O'Grady, M.J., Muldoon, C., Virtual sensor networks: An embedded agent approach, in *Proc. ISPA*, (2008), pp. 926–932.
21. Bandara, H.M., Jayasumana, P., Illangasekare, T.H., Cluster tree based self organization of virtual sensor networks, in *Proc. IEEE GLOBECOM Workshops*, (2008), pp. 1–6
22. Haghighi, M., Cliff, D., Multi-agent support for multiple concurrent applications and dynamic data-gathering in wireless sensor networks, in *Proc. 7th Int. Conf. IMIS*, 2013, pp. 320–325.

23. Smith, R.B., SPOTWorld and the sun SPOT, in Proc. 6th Int. Conf. Inf. Process. Sensor Netw., New York, NY, USA, 2007, pp. 565–566.
24. The Open Group, The SOA Work Group,
<http://www.opengroup.org/getinvolved/workgroups/soa>
25. OASIS, Advancing Open Standards for the Information Society, OASIS SOA Reference Model Technical Committee,
https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=soa-rm
26. OMG, Object Management Group, www.omg.org
27. OASIS, Reference Architecture Foundation for Service Oriented Architecture Version 1.0 Committee Specification 01,
<http://docs.oasis-open.org/soa-rm/soa-ra/v1.0/cs01/soa-ra-v1.0-cs01.html>
28. The Open Group, The SOA Source Book,
<https://www.opengroup.org/soa/source-book/intro/index.htm>
29. C. Fernández-Llatas, J. B. Mocholi, C. Sanchez, P. Sala, and J. C. Naranjo, Process choreography for Interaction simulation in Ambient Assisted Living environments, in Proceedings of XII Mediterranean Conference on Medical and Biological Engineering and Computing, 2010.
30. A. Martinez-Millana, C. Fernandez-Llatas, L. Sacchi, D. Segagni, S. Guillen, R. Bellazzi, and V. Traver, From data to the decision: A software architecture to integrate predictive modelling in clinical settings, in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, pp. 8161–8164.
31. P. D. O'Brien and R. C. Nicol. "Fipa towards a standard for software agents". BT Technology Journal, vol. 3, pp. 51-59, Jul. 1998.
32. Henrik F. Nielsen, Noah Mendelsohn, Jean J. Moreau, Martin Gudgin and Marc Hadley. "SOAP version 1.2 part 1: Messaging framework" W3C recommendation, Jun. 2003.

Assessment of Ad Hoc Routing Protocols for Network Deployments in Disaster Scenarios

Miquel Martínez¹, Yusheng Ji², Juan-Carlos Ruiz¹, and David de Andrés¹

¹ ITACA, Universitat Politècnica de València, Valencia, Spain
{mimarra, jcruizg, ddandres}@disca.upv.es

² National Institute of Informatics, Tokyo, Japan
kei@nii.ac.jp

Abstract. The use of ad hoc networks to recover the communications infrastructure after a natural disaster situation has been deeply studied in the literature. In these kind of situations, the current infrastructure that provides access to the Internet can be damaged, which turns out in a complete loss of service in some areas. This problem has been approached by creating an ad hoc network conformed by the remains of the damaged infrastructure and additional relays. Thus, providing service to isolated areas. However, the harsh conditions of these deployments and the unstable conditions of the environment can affect its performance. In this work, the deployments done in previous works using real data over the city of Tokyo, are evaluated in presence of different ad hoc routing protocols to determine the performance of the network. These results are analysed using *multi-criteria decision-making* methods in order to determine under which protocol the network performs the best.

Research area: Fault Tolerant Systems

1 Introduction

The frequency of earthquakes in Japan is extremely high compared to other countries of the world. Around 30% of the world's earthquake every year take place in Japan. When such a disaster occurs, keeping the population informed is of primary importance, and for that to happen communication infrastructures must keep working.

Nowadays people have quick access to the information through their smartphones or laptops via the Internet. Online news and social networks let the users be aware of the current situation of the disaster, the status of evacuation areas and shelters, or stay in touch with their loved ones.

Communication infrastructures, however, can be damaged during a disaster leaving people in some affected areas isolated from the rest. This lack of communication in some areas is mainly caused by the malfunctioning of the base stations that operate in such areas providing Internet access for the people. But after a disaster, one of the main reasons for these base stations to fail is power

outage. Although base stations can be powered with additional sources to keep them running, like power generators or batteries, these are temporary solutions and have a short lifetime.

The use of wireless and infrastructureless technologies, like ad hoc networks, is a common approach in these works. Some of them are focused on determining the kind of ad hoc network to deploy (MANET, VANET, MESH, etc), and the wireless technology to use for the nodes. Others, like [7], approach the connectivity problem by studying how to deploy static and mobile relays to create an ad hoc network to provide Internet access to isolated areas in a disaster scenario in the city of Tokyo (Japan). To do so, this work applies an algorithm developed in [3] and adapt it to study to what extent the number of nodes can be reduced while providing network service to all nodes.

There are various factors that will impact the performance of this network. As stated in works like [4] [5], an important factor that should be considered when deploying an ad hoc network is the routing protocol used. Thus, being able to determine which routing protocol improves the performance of the network in such harsh conditions can make a difference.

Given that, the work done in this paper focuses on benchmarking the performance of the network deployments done in [7] for different ad hoc routing protocols. These deployments are simulated for a post-disaster scenario in the city of Tokyo. Thus, the benchmarking process will determine which routing protocol improves the performance of the ad hoc network deployed.

Assuming the context for these deployments, a post-disaster scenario, the main purpose of the network is to keep people informed, so having a low ratio of *packet loss* will be more important than having a high *throughput*. This requirements must be reflected in the analysis of results in order to provide meaningful and context-aware conclusions. To cope with this kind of analysis, previous works have shown the feasibility of *multi-criteria decision-making* (MCDM) [6] methods to consider evaluator's requirements in the analysis. So in this work, MCDM methods are used to assist the analysis of the results obtained from the experiments, and provide the reader with our conclusions.

The rest of the paper is structured as follows. Section 2 presents a brief description of the ad hoc routing protocols that are evaluated in this work. The experiments performed in this work are described in Section 3, while the MCDM method used to analyze the experiments' results is presented in Section 4 together with the analysis done. Finally, Section 5 discuss the impact of the work done and concludes the paper.

2 Routing Protocols

Ad hoc network deployments in disaster scenarios are expected to have the best possible performance, as people lives may rely on this network to work. Thus, this work is a first approach to study the improvement in the network's performance that one routing protocol has over another. However, the number of existing

routing protocols is quite large, so as a first approach, among all of them three well known routing protocols have been considered for this work.

OLSR The *Optimized Link State Routing* protocol ([14]) is a proactive protocol that makes use of link-state information to determine the best route from source to destination among those available. Its constant transmission of *hello* and *topology control* packets let the nodes discover and propagate the information about the status of the network topology in real time. Thus when information has to be forwarded towards a given node, the route is already known. Of course, the constant exchange of information introduces some overhead in the network traffic.

AODV Unlike OLSR, Ad hoc On-Demand Distance Vector Routing ([11]) has a reactive behavior. Route discovery only takes place when information has to be forwarded, not before. When a node has to send information to another node, probe-like packets are sent through the network to find a route for the information to be delivered. These packets are called *Route Request* (RREQ) and *Route Reply* (RREP). This last packet is the response from the destination node, and contains all the necessary information about the route that the information has to follow to reach it. As these packets are only transmitted “on-demand”, the overhead introduced in the network traffic is expected to be lower than in OLSR.

DSDV Destination-Sequenced Distance-Vector Routing ([12]) follows a table driven routing scheme based on the Bellman-Ford algorithm. Thus it has a proactive behavior. Each node maintains routing information for all known destinations which is updated periodically. Nodes use a sequence number pattern to announce its routing information to its neighbors. When links are present, even numbers are used, odd numbers otherwise. This way, nodes use this sequence numbers to keep their own routing tables updated and avoid the appearance of routing loops.

In order to evaluate and compare these three protocols it is necessary to define a set of repeatable and reproducible experiments. The conditions of each experiment should be reproduced so the experiment can be repeated using every one of the routing protocols. Next section describes the different aspects that were taken into consideration when defining the experiments.

3 Experimental Set Up

The experiments done in this work are based on the deployments studied in [7]. Here, authors study the feasibility of their approach to create an ad hoc network by studying multiple base scenarios. Each scenario differs from the rest in their initial assumption of which areas are disconnected and which ones remain connected. Thus, different initial states led to different deployments of nodes to provide isolated areas with access to the Internet. Given the amount of nodes, all these scenarios were simulated using a widely used tool, the *Network Simulator 3* (NS3) [2].

The upcoming sections provide a description of the process followed to design the experiments and determine the measures that would be used for its later analysis.

3.1 Target of evaluation

During a disaster people is encouraged to go to their nearest evacuation site or shelter. The information of their location is available online through governmental sites [1]. These locations are interpreted as areas in Tokyo that may have or not access to the Internet. Thus, they are considered as nodes of the network, meaning that additional nodes must be deployed between them so that every area has access to the Internet, hence creating the ad hoc network. For the experiments, an area of 35 km^2 in the center of Tokyo that encompass a total of 67 sites - nodes, from now on - is considered. The locations of the initial nodes can be seen in Figure 1.

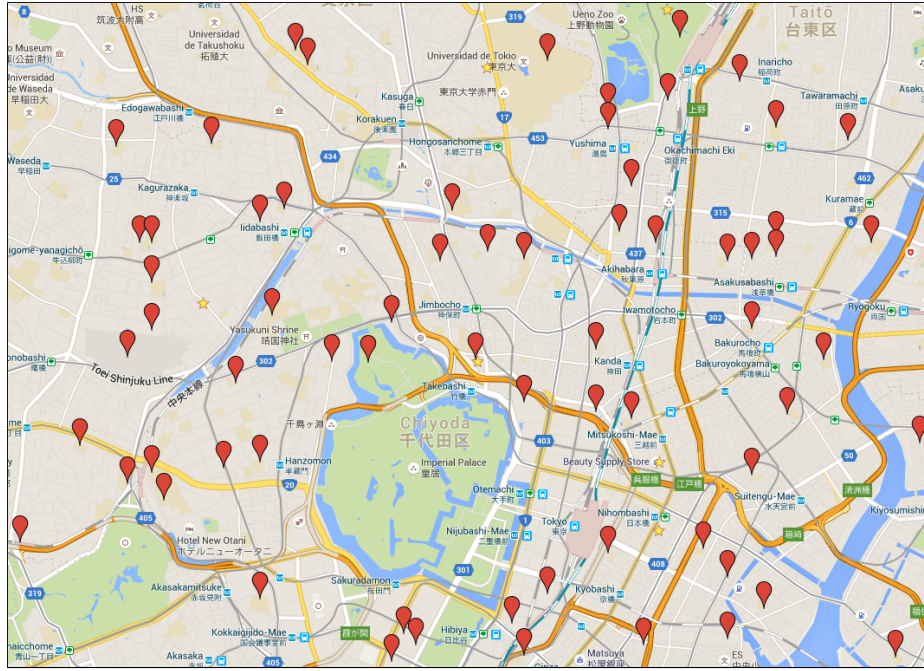


Fig. 1. Evacuation areas and shelters in the city center of Tokyo.

3.2 Definition of the experiments

Determining which of these nodes are initially out of service is done randomly based on the probability that each area has to have access to the Internet after

a disaster occurs. This information, as it is described in [7], it is extracted from real data used to create maps of Tokyo that indicate the probability of service in each area in four different time slots (6h, 12h, 18h and 24h) after the disaster. So the four maps must be considered for the experiments, as the probability of having service in an area will drop dramatically between 6 hours and 24 hours after the disaster.

In addition to these four possible initial scenarios, three different workflows are defined to evaluate the performance of the network. Each workflow is defined as a different *Constant Bit Rate* (CBR) that is generated in the disconnected nodes towards the connected ones. The CBRs considered are the following: i) 0.5 Mbps, ii) 1 Mbps and iii) 3 Mbps.

The total number of experiments performed depends also on how many initial scenarios are created. For every map the disconnected nodes are randomly selected, leading to different topologies each time. Then, the more number of initial scenarios are considered, the more number of experiments will be performed. In this work a total of 144 experiments were performed. This number of experiments is the result of considering four initial scenarios for each map, thus leaving 16 topologies to evaluate (4 initial scenarios x 4 maps). Each of this scenarios is evaluated for every CBR defined, which gives a total of 48 different experiments (16 scenarios x 3 CBR). The 48 experiments must be performed considering all three routing protocols defined in previous section, making a total of 144 experiments.

3.3 Performance measures

The purpose of this work is to evaluate the performance of the network with different routing protocols. Keeping that in mind, the measures provided as the outcome of an experiment were selected considering the impact they would have on the network performance. The performance of the whole network must be assessed by analysing all the traffic flows generated in the experiment. So, every experiment provides the following information for each data flow in the network: i) The percentage of *Packet Loss*, ii) the *Throughput* perceived by the receiving node, iii) the average *Delay* of the packets, and iv) the *Availability* of a route. This last measure indicates if a route between a disconnected node and an Internet gateway was available when a node needed to send information.

It is obvious that the number of data flows in an experiment can be very high. In order to compare the three routing protocols, the final measures used in the analysis will be the average for all the data flows generated. Table 1 shows the final measures obtained for an experiment after its execution with the three routing protocols.

As aforementioned in this paper, when analyzing the results of the experiments it is necessary to do it according to the application context of the network deployment, a disaster scenario. The methodology used to analyze the results from the experiments, as well as the conclusions driven from the analysis are presented in next section.

Table 1. Average results obtained for an experiment performed using the 12 hours probability map and a CBR of 0.5 Mbps

	Throughput (Kbps)	Delay (ms)	Packet Loss (%)	Unavailable routes (%)
OLSR	418.4	2.3346	8.65	0.3
AODV	418.3	7.661	4.26	0.1
DSDV	404.5	39.686	9.45	0.4

4 Analysis of results

In the benchmarking literature, it is common to find that the analysis of results is done using methods that consider all measures equally relevant, like the *arithmetic mean*, the *geometric mean* or the *kiviat diagrams*. In contrast, some later works in this field ([8][9]) have proved the feasibility of MCDM methods to analyze results considering different levels of importance for each measure, as it happens in real situations. Then, in this work, a widely used MCDM method in different fields of research has been chosen to perform the analysis, the *Analytic Hierarchy Process* (AHP) [13].

The AHP allows the evaluators to model their analysis requirements into a hierarchical structure, where some criteria represent the result of other sub-criteria. The contribution of each criterion to its immediate upper criterion must be quantified, which can be understood as the weight that the value of one criterion has to calculate the value of its upper one. For evaluators to quantify their requirements, pairwise comparisons of the criteria at the same level must be done. Thus, with the use of a numeric scale, evaluators compare the criteria two-by-two to quantify the importance that one criterion has with respect to another to achieve the immediate upper criterion. This process is repeated until every criterion has been compared with all the others. These comparisons are stored in a matrix form, and the final priority of each criterion is calculated by applying the principal right eigenvector of the matrix.

After comparing the four measures considering the application context of the deployment, the priority (or weight) of each criterion (measure) to calculate the global score for each protocol are as follow (in [0,1] scale): 0.09 for *Throughput*, 0.11 for *Delay*, 0.52 for *Packet Loss* and 0.28 for *Unavailable routes*. Figure 2 depicts the hierarchical model of the requirements used in this work to compare the routing protocols and the calculated weights for each one.

When the priorities for all criteria are calculated, the results obtained from all routing protocol must be pairwise compared too according to each criterion. This will provide a value for each protocol relative to the others in each criterion, so a global score can be calculated for each one of them, thus allowing to rank the protocols from best to worst. However, this process can be very tedious given the fact that 48 different types of experiments were done. To automate this process, the *Assisted Pairwise Comparison Approach* (APCA) (introduced in [10]) was used to compute all the comparisons.

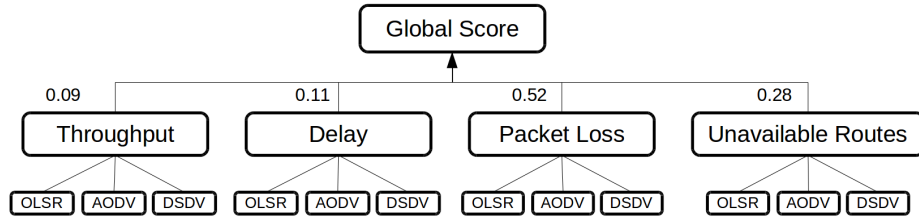


Fig. 2. Hierarchical model of the requirements applied to analyze the results.

After analysing and comparing the results obtained from performing the 48 types of experiments with the three routing protocols, we found ourselves with 48 rankings. These rankings represent under which protocol the network presented a better performance, so routing protocols are ranked from best to worst.

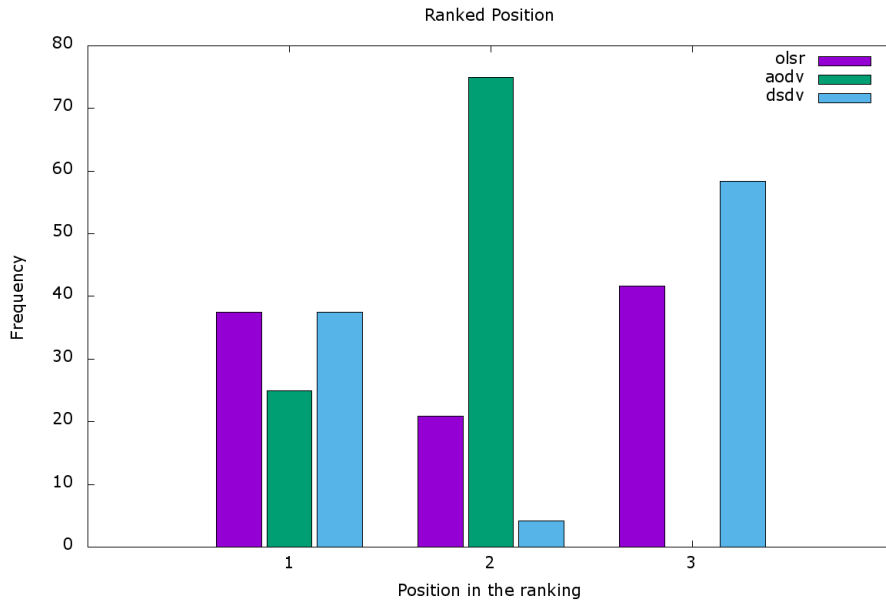


Fig. 3. Frequency of each routing protocol classified as first, second or third among all experiments

To determine which of them was the best candidate to be used for a network deployment in this scenario, the accumulated frequency of their ranked positions was studied. These frequencies are depicted in Figure 3. From the figure, it can be seen that even though OLSR and DSDV are ranked in the first position more times than AODV, AODV is ranked around 75% of the time second and

the other 25% first. But unlike OLSR and DSDV, AODV is never ranked third. Based on these results, it can be stated that in average, the network will perform better running AODV than with OLSR or DSDV, as these two would present the worst performance in around 40% and 55% of the scenarios, respectively.

5 Conclusions

Natural disasters seem to have become more common during the last decades. In Japan, where earthquakes represent a high threat to population, a lot of measures have been taken and actuation plans have been developed to improve people's safety. However, there are currently no implemented solutions to prevent communication's infrastructure to keep working after a disaster occurs, leaving people without service or Internet access.

Authors in [7] have approached the problem of communication failures through the deployment of an ad hoc network using static and mobile relays. They make use of an algorithm to determine, based on the isolated areas, which is the best way to deploy the network nodes so every isolated area has service again.

Starting from this approach, this work has been focused on evaluating the performance of the ad hoc network deployed with three well known routing protocols, OLSR [14], AODV [11] and DSDV [12], thus find out which one is more suitable for the network. Every protocols was evaluated for a total of 48 different possible deployments generated through the methodology described in [7].

Although traditional approaches like *arithmetic* or *geometric* mean are used in benchmarking studies to analyze the results, the post-disaster scenario of these deployments required a different approach. As people's life may rely on these networks to work properly, the analysis of the experiment's results has been carried out using *multi-criteria decision-methods*, that let evaluators apply specific requirements to the analysis. More concretely, the *Analytic Hierarchy Process*.

This study has shown that the selection of the ad hoc routing protocol can make a significant difference in the performance of the network. From the considered protocols, the network has shown on average a better performance when using AODV than with the other two.

Nevertheless, this study is only a first approach to contribute to the improvement of the performance in this kind of deployments. The next step in this work involves not only to evaluate more routing protocols, but also to compare different deployment techniques. Thus, being able to determine which kind of deployment and ad hoc routing protocol would be better to cope with failures in the communications infrastructures in this type of situations.

Acknowledgment

This work has been partially supported by the “Programa de Ayudas de Investigación y Desarrollo” (PAID) de la Universitat Politècnica de València, and the international internship program of NII.

References

1. Evacuation sites, areas and shelters in central tokyo. <http://www.realestate-tokyo.com/news/evacuation-sites-in-central-tokyo/>
2. The network simulator ns-3. <https://www.nsnam.org>
3. Bredin, J.L., Demaine, E.D., Hajiaghayi, M.T., Rus, D.: Deploying sensor networks with guaranteed fault tolerance. *IEEE/ACM Transactions on Networking* 18(1), 216–228 (Feb 2010)
4. Friginal, J., de Andrés, D., Ruiz, J.C., Martínez, M.: A survey of evaluation platforms for ad hoc routing protocols: A resilience perspective. *Computer Networks* 75, 395–413 (2014)
5. Friginal, J., de Andrés, D., Ruiz, J.C., Martínez, M.: Refrahn: A resilience evaluation framework for ad hoc routing protocols. *Computer Networks* 82, 114–134 (2015)
6. Ishizaka, A., Nemery, P.: *Multi-criteria Decision Analysis: Methods and Software*. Wiley (2013)
7. Król, M., Ji, Y., Yamada, S., Borcea, C., Zhong, L., Takano, K.: Extending network coverage by using static and mobile relays during natural disasters. In: 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA). pp. 681–686 (March 2016)
8. Martínez, M., de Andrés, D., Ruiz, J.C.: Gaining Confidence on Dependability Benchmarks’ Conclusions through “Back-to-Back” Testing. In: 2014 Tenth European Dependable Computing Conference. pp. 130–137 (2014)
9. Martínez, M., de Andrés, D., Ruiz, J.C., Friginal, J.: Analysis of results in dependability benchmarking: Can we do better? *International Workshop on Measurements and Networking* pp. 127–131 (2013)
10. Martínez, M., de Andrés, D., Ruiz, J.C., Friginal, J.: From Measures to Conclusions Using Analytic Hierarchy Process in Dependability Benchmarking. *Instrumentation and Measurement, IEEE Transactions on* 63(11), 2548–2556 (2014)
11. Perkins, C.E., Royer, E.M.: Ad-hoc on-demand distance vector routing. In: *Mobile Computing Systems and Applications, 1999. Proceedings. WMCSA '99. Second IEEE Workshop on*. pp. 90–100 (Feb 1999)
12. Perkins, C.E., Bhagwat, P.: Highly dynamic destination-sequenced distance-vector routing (dsdv) for mobile computers. In: *Proceedings of the Conference on Communications Architectures, Protocols and Applications*. pp. 234–244. SIGCOMM '94, ACM, New York, NY, USA (1994), <http://doi.acm.org/10.1145/190314.190336>
13. Saaty, T.: Decision making with the analytic hierarchy process. *International Journal of Services Sciences* 1(1), 83–98 (2008)
14. T. Clausen and P. Jacquet: Optimized Link State Routing Protocol(OLSR). RFC 3626 (2003)

Optical Disdrometer Based on a Universal Frequency-to-Digital Converter

R. Barrales-Guadarrama¹, A. Mocholí-Salcedo², and F. Mocholí-Belenguer²

¹ Universidad Autónoma Metropolitana—Azcapotzalco.
02200, Mexico City, Mexico
rbg@correo.azc.uam.mx

² Instituto ITACA, Universitat Politècnica de València. Camino de Vera s/n.
46022 Valencia, España
amocholi@eln.upv.es; fermocbe@upv.es

Abstract

In this paper, it is presented an optical disdrometer that uses a quasi-digital photodetector of our own and a universal frequency-to-digital converter instead of a conventional analog-to-digital converter as data acquisition system. This feature has allowed the design of a low cost, robust and simple sensor-to-microcontroller interface as demanded by Intelligent Transportation Systems (ITS) applications. This instrument is capable of detecting drops of water greater than or equal to three millimeters in diameter simply by knowing two data and warning of adverse weather conditions. No calibration was performed due to the limited capacity of the measuring device.

I. INTRODUCTION

Because of the convenience and price, using the car to move as mean of transport between two relatively nearby places is becoming more common. However, the continued urban growth and the increasing number of vehicles are important factors that often cause traffic congestion problems in cities. At the same time, it is observed a relationship between an increase in traffic accidents and adverse weather conditions, being the rain the most harmful because it is the most habitual. These problems are observed both in developing countries and in developed countries. Therefore, a reasonable measure to prevent accidents closely related to environmental factors would be the development and implant of new intelligent traffic systems that warn the driver of potential dangerous situations.

The precipitation that reaches the surface of the earth as water droplets greater than or equal to 0.5 mm in diameter is identified as rain. Traffic accidents due to rain occur when vehicles skid on wet pavement, traffic signs are not visible, there are shorts by runoffs in electrical systems, visibility is reduced and several others factors. Therefore, it has been studied that the probability of having an accident grows rapidly to over 15 millimeters per hour ($\frac{mm}{h}$) of rainfall [1]. This amount belongs to a measure called precipitation intensity (PI) and can be characterized as follows:

Precipitation intensity	Accumulation in one hour
Weak	Less than 2 mm
Moderate	Between 2 and 15 mm
Heavy	Between 15 and 30 mm
Very heavy	Between 30 and 60 mm
Torrential	Over 60 mm

Table 1. Relationship between accumulation of water in one hour and precipitation intensity [1]

Rain monitoring must be done with reliable tools which are easy to install on the motorway and maintenance-free. An instrument that meets these specifications is the optical disdrometer presented in this article. It should be pointed out that this instrument is not a catching instrument as some types of rain gauges, but lends itself perfectly to the ITS requirements. 114

II. OPERATING PRINCIPLES

A. Measurement Principle of Rainfall

To measure the precipitation intensity there are different instruments that are classified as catching instruments and no-catching instruments [2]. The first ones catch precipitation through a well defined hole size and measure the equivalent mass or volume of water that has been accumulated in a given time interval. On the contrary, no-catching instruments calculate the precipitation intensity (PI) by mathematical integration of all particles which have passed through a cross section in a given time.

The optical disdrometer uses one or two thin layers of light to detect particles which pass through it. Each particle within the beam blocks the intensity of transmitted light, reducing it to a quantity proportional to its diameter. The range of measure for the diameter covers typically from 0.3 mm to 8 mm. The precipitation intensity can be calculated directly by integrating the number of the detected particles and their size on a period of time from fifteen seconds to one minute. The resolution in the precipitation intensity is typically $0.01 \frac{mm}{h}$, although there is a possible source of error when several drops coincide and they are detected as a giant drop. This produces a poor estimate of the volume of water, being necessary a statistical correction. Hence, the upper limit of the measuring range is restricted typically to $250 \frac{mm}{h}$.



Figure 1. Optical disdrometer

To measure the PI, it is assumed that a certain number of drops fall through the volume (V) of the disdrometer in a sampling period (T). The duration of the pulse signal generated by the sensor signal conditioner is equivalent to the time (t_i) that the drop takes to cross the volume. Illigworth and Stevens (1987) [3] demonstrated the validity of the following equation:

$$N = \frac{\sum t_i}{VT} \quad (1)$$

Where N is the density of raindrops in the sensing volume. Therefore, this parameter is measured in m^{-3} . This calculation is valid as long as a large number of raindrops are evenly distributed in space and if the sampling period T is large compared to the transition time of each raindrop. On the other hand, the final speed of raindrops is given by another equation (2) [4].

As the optical disdrometer must be oriented to ITS applications and considering that high accuracy in measuring rainfall rate is not required because it is taken as the absolute measure the final velocity of falling raindrops without regard to factors such as wind speed or a non-spherical shape droplet, it is possible to change the equation to express the speed depending on the diameter of the raindrop.

$$v_{fall} = 17,67 (2r)^{0,67} = 17,67 (\phi)^{0,67} \quad (2)$$

This parameter is measured in $\frac{m}{s}$. ϕ is the diameter of the raindrop and r is the radius, assuming always spherical shape. Although it will be seen in the next section in more detail, it is clear that there is proportionality between the peak value V_p of the signal from the optical disdrometer conditioner, which represents the degree of concealment of the beam due to the obstruction of the raindrop, and size drop. It is shown an image to clarify (Figure 2).

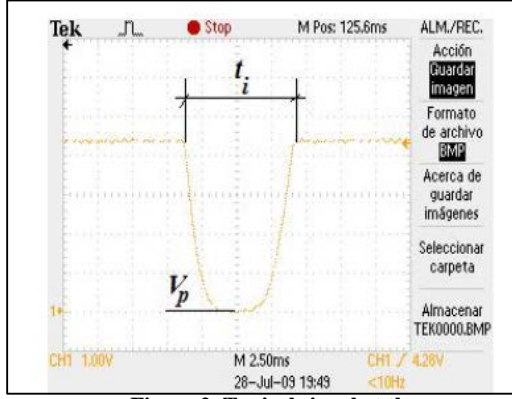


Figure 2. Typical signal at the output of the disdrometer

This degree of proportionality can be expressed as follows:

$$V_p = K \cdot \phi^2 \quad (3)$$

Where V_p is the peak value of the signal from the optical disdrometre conditioner measured in volts as it shown in Figure 2, and K is a proportionality constant. Therefore, combining this new equation with the last one, the resulting equation is:

$$v_{fall}(\phi) = 17,67 \left(\frac{V_p}{k} \right)^{0,67/2} \quad (4)$$

On the other hand, it is necessary to know the amount of water that has fallen as precipitation for calculating the PI. Then, if $m(M_{H_2O})$ is the mass of the equivalent amount of raindrops in the sampling period T , V_{H_2O} is the volume and ρ_{H_2O} is the density of water, it is known approximately that:

$$\rho_{H_2O} = 1000 \left[\frac{kg}{m^3} \right] \quad (5) \quad \text{And as all density} \quad \rho_{H_2O} = \frac{M_{H_2O}}{V_{H_2O}} \quad (6)$$

As it is assumed that raindrops are spherical and it is known that the diameter (ϕ) is twice the radius (r), we can express the volume as:

$$V_{H_2O} = \frac{4}{3} \pi r^3 = \frac{1}{6} \pi \phi^3 \quad (7)$$

Lastly, it is possible to reach the following equation by combining formulas 3, 5, 6 and 7.

$$m = \rho_{H_2O} \cdot V_{H_2O} = 1000 \left(\frac{\pi}{6} \right) \phi^3 = 1000 \left(\frac{\pi}{6} \right) \left(\frac{V_p}{k} \right)^{3/2} \quad (8)$$

Therefore, using the above equations, the precipitation intensity (PI) [4] can be defined as:

$$PI = m \cdot N \cdot v_{fall}(\phi) = 1000 \cdot (17,67) \left(\frac{\pi}{6} \right) \left(\frac{V_p}{k} \right)^{3/2} \left(\frac{1}{VT} \right) \sum_{i=1}^n t_i \left(\frac{V_p}{k} \right)^{0,67/2} \quad (9)$$

And in a sampling period T , the total PI for a given number of drops would be:

$$PI = 1000 \cdot (17,67) \left(\frac{\pi}{6} \right) \left(\frac{1}{VT} \right) k^{-3,67/2} \sum_{i=1}^n (V_p)^{3,67/2} t_i \quad (10)$$

Thus, it is observed that the number before the sum is a constant and the other part depends on t_i . In this way, this method for measuring the intensity of rainfall will require measurement of the peak value (V_p) of the signal produced by an optical disdrometre and the duration of this signal width (t_i). Conventionally, the peak value could be measured with an ADC (Analog-to-Digital Converter) and time could also be measured by comparing a clock signal on a microcontroller, but as mentioned above, this optical design must meet the objective of using a UFDC (Universal Frequency-to-Digital Converter) and thereby making a contribution to the development of environmental instruments for ITS. In the following sections is explained how the problem of measuring parameters (V_p , t_i) by a UFDC has been solved.

B. Instrument Implementation

To verify the performance of the optical disdrometer, it has been necessary to construct a model of the same characteristics that are presented below. The emitter in the optical system of the disdrometer consists of an IR LED whose wavelength is in the near infrared (NIR) spectra. The selection of this LED will depend on the opening angle that allows creating a cylindrical collimated beam with a certain size of diameter (consistent with sufficient dimensions to obtain a representative sample of the concentration of water droplets for estimating the PI), a sufficient intensity of emission, and the range of wavelengths available on the market.

The detector in the optical system will be a high sensitivity photodetector to the emission wavelength. Occlusion produced by the raindrop crossing the collimated beam will be turned into an electric pulse by the detector whose characteristics will contain information about the droplet size and speed. In the figure below (Figure 3) it is shown the block diagram of a basic optical disdrometer system.

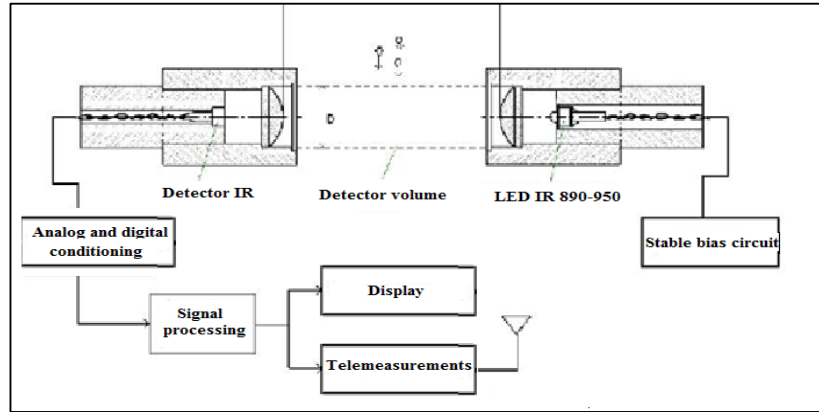


Figure 3. Block diagram of an optical disdrometer

The concept of the optical disdrometer of Figure 3 has been the basis for advanced developments of this instrument, such as the Löffler-Mang and Joss' disdrometer (2000) [5] which uses a horizontal light sheet produced with a laser beam and the Denby's disdrometer [6] with horizontal double sheet.

In this case it is necessary to extract from the sensor signal the parameters involved in estimating the PI (V_p and t_i). Clearly, there is not a quasi-digital sensor with two frequency outputs which represents values of these parameters. Therefore, it is essential to design a material solution that generates the signals needed for the UFDC. Figure 4 shows the concept of an optical disdrometer operating this way.

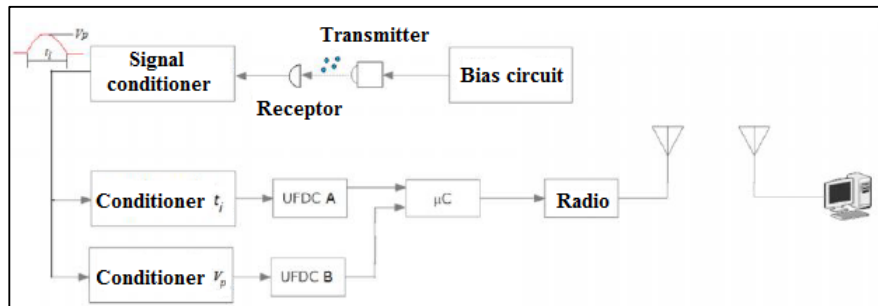


Figure 4. Block diagram of the optical disdrometer

The Universal Frequency-to-Digital Converter (UFDC) [7] is a universal two-channel high precision multifunctional converter based on novel frequency-time conversion method. It is perfectly suited to any applications where frequency-time parameters have to be measured with highest resolution and programmable accuracy.

In this way, regarding the conditioners parameters, it is clear that a significant trouble considered here is to process two parameters of the same signal to the respective UFDC. This requires having a quasi-digital representation of the peak value of the signal receiver, as this parameter is fundamental in calculating the PI, and having a quasi-digital representation of the time that the drop of water takes to cross the sensing volume. According to this, the UFDC A (Figure 4) will make the conversion to digital code of the quasi-digital signal that contains the parameter t_i and UFDC B will do the same with the quasi-digital signal that contains information about parameter V_p . Later, the microcontroller will be the main responsible for controlling and calculating. Its main features would consist of processing the information (V_p , t_i) to obtain the estimated value of the PI and send this information to a light panel of the highway by using any technology.

Therefore, once the emitter, receptor, bias circuit, signal conditioner, UFDB and their respective problems have been analyzed, it is time to introduce them in more detail. The selected LED has the characteristics indicated in the following table:

Characteristic	Value
Model	HE8404SG
Brand	Opnext
Typical peak wavelength	820 nm
Minimum output optical power	40 mW

Table 2. Table with the characteristics of the emitter

This component [8] has been chosen because of its high output power (it can radiate up to 60 mW) and its emission pattern is nearly uniform across a wide field of view (less than 5 % variation over an angle of 80°). In Figure 5 is shown the mechanical scheme of the frame transmitter. A holographic diffuser lens (Figure 6) model # 54-506 of Edmund [9] has been added to this frame allowing a uniform distribution of the LED light whose irradiance, after passing through the lens, is independent of the angle [10].

According to the guide Edmund's applications, the diffusion angles for homogenization applications of short light distances are between 60° and 80° . For this work, the lens that has been selected has the following characteristics:

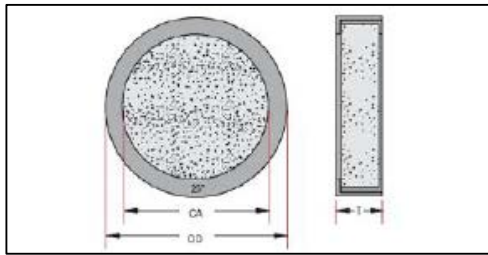


Figure 5. Mechanical scheme of the optical emission system

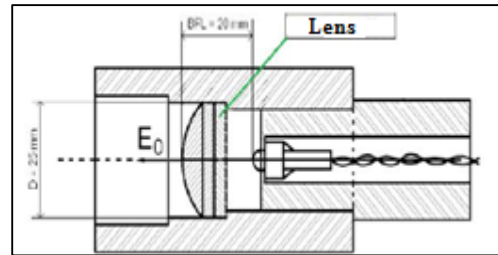


Figure 6. Mechanical scheme of the diffusing lens

The detector was chosen by analyzing the market options in search of some device that has a good sensitivity to the wavelength of the IR LED and has an enough area to ensure the light occlusion detection at a wide LED angle. Finally, a PIN photodiode made of silicon was taken. The device found in the market is the TEMD5010X01 of Vishay [11] because is very useful in the NIR region. Figure 7 shows the responsivity curve of the PIN diode selected (extracted from the component data sheet). By observing this curve, it is possible to confirm that the wavelength of the emitter ($\lambda_{\text{EMITTER}} = 820 \text{ nm}$) corresponds to a receiver responsivity of about 0.9, which guarantees a very good detection of the occlusion light signal by raindrops.

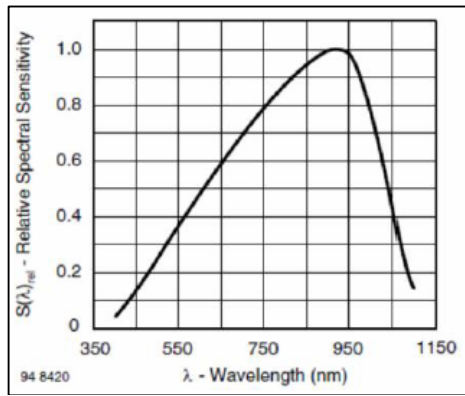


Figure 7. Graph which shows the responsivity curve

III. EXPERIMENTAL SETUPS

A. Polarization of the Emitter LED

The disdrometer emitter must produce a beam of maximum intensity. Therefore, it is necessary to investigate the I-V function and choose an appropriate point of operation (V_F , I_F), considering the maximum current of the device of 250 mA specified by the manufacturer.

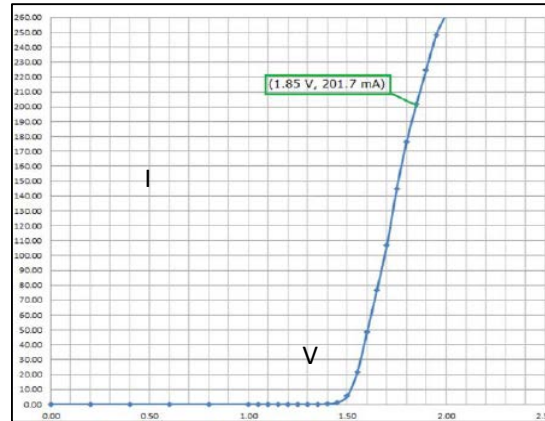


Figure 8. Graph which shows the experimental curve characteristic of the IR LED HE8404SG

In the figure above (Figure 8) it is observed that the proper operating point for maximum emission intensity without risk of overheating the LED is $(V_{F0}, I_{F0}) = (1.85 \text{ V}, 201.7 \text{ mA})$. All this could be implemented by a current source (200 mA) constructed with a transistor and resistors, or more precisely, by a LM317 voltage regulator configured as current regulator. The circuit to polarize the IR LED for the selected operating point is shown in Figure 9.

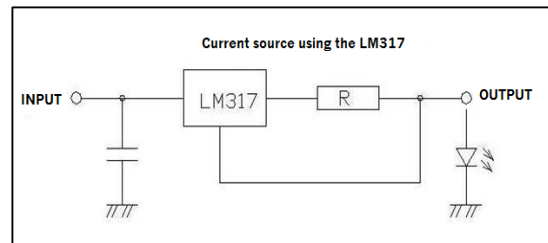


Figure 9. Circuit diagram polarizer

B. Receiver conditioning

Once the PIN polarization was done, it was taken into account that the PIN photodiode output current is proportional to the input irradiance. The measurements carried out showed that its absolute value is probably always less than $50 \mu\text{A}$ and it is known that the value of this signal should not be degraded by an overload of the PIN photodiode. In addition, PIN photodiode is prone to receive unwanted radiation

that not only interferes with the signal of interest, but also it can cause saturation. For these reasons and because of the intrinsic noise from the signal conditioner PIN photodiode (current / voltage converter), it was studied which electrical circuit could solve these problems, and the best solution was a transimpedance amplifier. The problem of saturation of the photodiode has been controlled in the laboratory implicitly by accommodating the receiver in a closed frame for preventing lighting interference and lining up with the emitter. The paradigm of data acquisition using frequency conversion to digital code requires minimizing or eliminating any kind of interference by unconventional means.

The first thought design did not consider the use of a capacitor, but in order to avoid the possibility that the circuit oscillates, it was necessary to connect one to ensure that there is a pole before the unity-gain frequency response of the open loop ($f_p < f_{BW}$) but without degrading the bandwidth of the closed loop amplifier (which is shown in Figure 10).

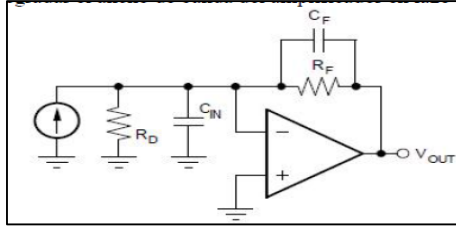


Figure 10. Solution to the problem

Value	Meaning
R_D	IR LED equivalent resistance
C_D	IR LED equivalent capacitance
C_{AO}	Input capacitance of operational amplifier
C_{IN}	$C_{IN} = C_D // C_{AO}$
R_F	Feedback resistor
C_F	$\frac{1}{4\pi R_F (GBW)} [1 + \sqrt{1 + 8\pi R_F C_{IN} (GBW)}]$ [12]

Table 4. Explanation of the components

Where GBW ($GBW = G \cdot BW$) is the gain-bandwidth product of the operational amplifier (8 MHz in this case). In this design, it was selected the operational amplifier OP27G [13] for its excellent features as a precision amplifier. From the table and IR LED data [14], it was selected R_F as 10 kΩ and C_F was computed as 8.64 pF. Finally, a standard value of 10 pF was selected. The final circuit is shown in Figure 11.

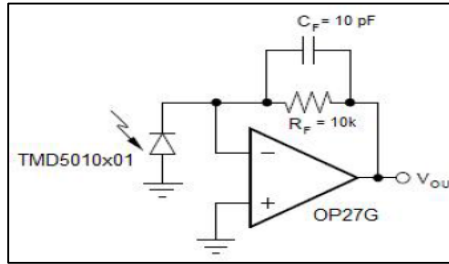


Figure 11. Final circuit

Figure 12 shows the aspect of the output signal corresponding to a drop of water passing through the detection volume. The signal has an offset and due to the type of feedback amplifier, the signal has negative values. The necessity to eliminate this offset and represent a signal with positive values to be processed accordingly to the proper UFDC functionality has required interconnecting an averaging filter and a precision inverter. Finally, in Figure 13 the processed final signal is presented.



Figure 12. Oscillogram for a detected drop

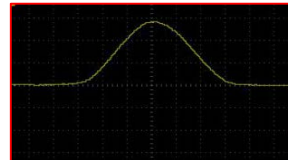


Figure 13. Oscillogram for a drop detected after passing through the filter

IV. EXPERIMENTAL RESULTS

A. Parameters extraction

The parameter t_i is the time that the drop takes to traverse the detection volume and implies its speed of fall. This measurement must be taken by a UFDC-A as a digital value so that the microcontroller can process it directly. The parameter V_p is the peak value of the signal produced when the drop traverse the detection volume and also reflects the size of it. But before designing an electronic system which produces a TTL (Transistor-Transistor logic) signal with a proportional frequency to V_p , it must be obtained the function that relates this peak value with the droplet size to test the hypothesis of formula 3.

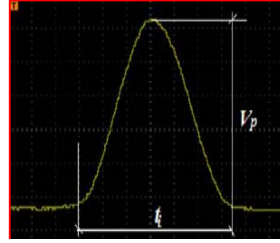


Figure 14. Values represented in the oscillogram

How to obtain t_i

The measurement of t_i is possible if the UFDC-A is set in mode 3 (time interval between start- and stop-drives) [15]. However, according to the device technical information, it is necessary to begin the measurement with a pulse (start) and stop it with another pulse (stop). Therefore, the design of an electronic system that detects the start and end of the sensor signal was necessary. This system is composed of a LM311 voltage comparator, which will compare the magnitude of the sensor signal against a V_{offset} voltage level set just above the zero signal, whereby the output of the comparator switches quickly from 5 V to 0 when the sensor signal V_s meets $V_s < V_{offset}$. The response of the monostable pulse is called V_o , whose width is defined by the relation $t_w = 1,1R_3C_1$.

In the case described, the circuit would generate the pulse which indicates that the raindrop has left the detection volume and the measurement of t_i must be suspended. The width of this pulse is set to about 100 microseconds, which corresponds to $R_2 = R_3 = 10000 \Omega$ and $C_3 = 10 \text{ nF}$. Therefore, it is evident that a similar circuit is required to generate the pulse for starting the measurement of t_i . Thus, Figure 15 shows the complete circuit with the complementary part of the circuit described above.

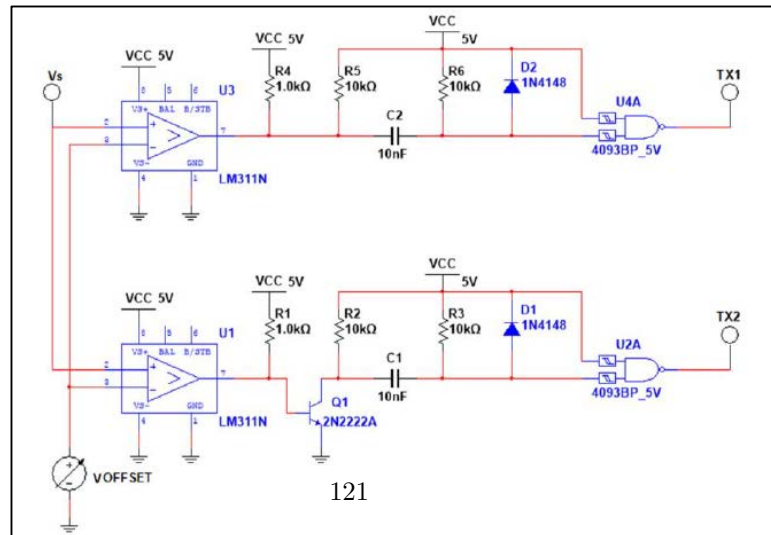


Figure 15. Wiring diagram of the electronic system to generate the start pulse TX1 and the stop pulse TX2 which are necessary to measure t_i of the sensor signal with the UFDC-A.

Verification of the quadratic relationship $V_p = f(\Phi^2)$

Equation 3 expresses an intuitively quadratic relationship between V_p and Φ . In order to test this hypothesis, raindrops were simulated with steel pellets with known diameters. The peak voltage V_p values were measured for 20 drop trials of each available pellet at three different volume lengths near the emitter ($l = 0$), at half the detection volume ($l = 35$ mm) and close to the receiver ($l = 70$ mm).

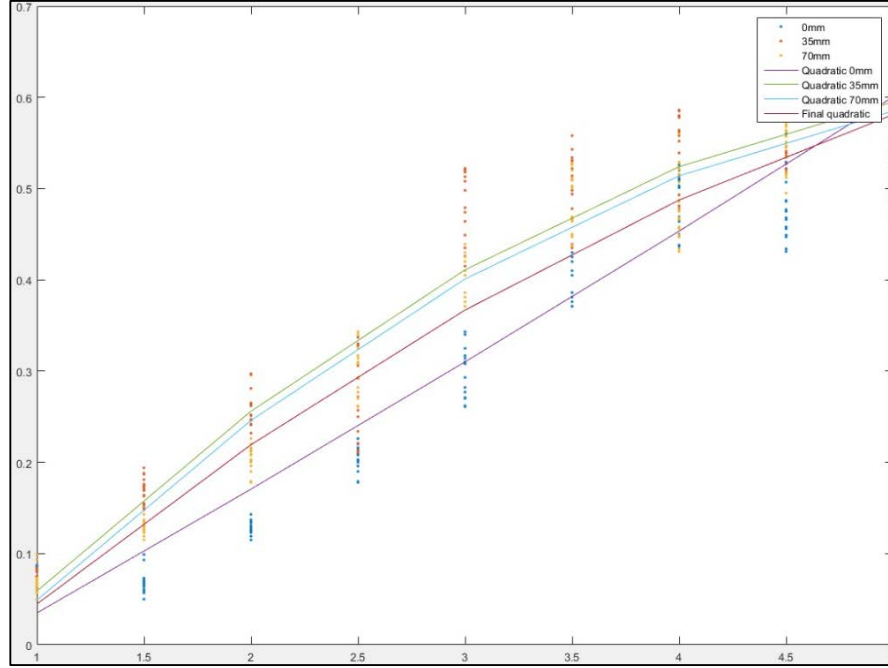


Figure 16. Quadratic settings and tolerance limits when a) $l = 0$; b) $l = 35$ mm and c) $l = 70$ mm.

In Figure 16 it can be seen that substantially all of the tests are within the tolerance limits of the quadratic fit. Therefore, considering an average of each of the settings, the quadratic function related to the peak value V_p of the signal drop with diameter ϕ assuming a spherical shape, is:

$$V_p = -0,0135\phi^2 + 0,2151\phi - 0.157 \quad (11)$$

Whereupon it proves to be necessary to obtain the peak value of the signal sensor, convert this value to a representation in the frequency domain by UFDC-B and, using this value, obtain an estimate of the PI.

How to obtain V_p

The measurement of the parameter V_p is possible if it is set the UFDC-B mode 0 [15] (Frequency-Channel 1, FX1) which is the natural mode of the frequency converter to digital code. Therefore, it is necessary to obtain firstly the peak value of the sensor signal and, secondly, make a voltage to frequency conversion of this value. The peak detector used for this purpose is well known in the analog signal processing [16, page 446] and once calibrated, the system used allows linearity of 0.01% between the detected value (V_p) and the corresponding frequency, according to the manufacturer.

Results obtained for t_i and V_p

After all the process, Figure 17 shows an oscillogram which records the start pulse signal used to start the measurement of t_i parameter (signal TX1 in the UFDC-A set in mode 3). Figure 18 shows the pulse used to finalize that measurement, and finally Figure 19 shows an oscillogram which records the output of the peak detector with the V_p value of the sensor signal.

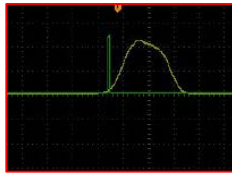


Figure 17. Start pulse

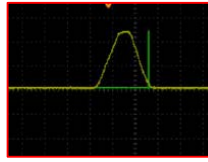


Figure 18. Stop pulse

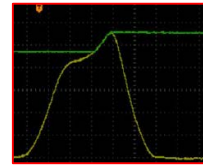


Figure 19. Peak value V_p

The UFDC-A will make the measurement of the time interval t_i of the sensor signal and UFDC-B will do the same for the representation in frequency of the voltage V_p . It is necessary to recall that the cumulative measurement of both parameters will allow to obtain an estimation of the PI. This implies that the microcontroller must acquire both data and accumulate them in the system over a period T . Finally, the microcontroller will be programmed with the necessary instructions to extract the parameters as discussed above.

V. CONCLUSIONS AND FUTURE WORK

It has been demonstrated the feasibility of using the frequency conversion to digital code in the design and construction of an optical disdrometer oriented ITS applications as parameter information. In the absence of a quasi-digital commercial sensor for measuring parameters related to the rainfall intensity, it has become necessary to design the conditioning necessary for signal input frequency in UFDC which has been demonstrated to be more powerful than other technologies. In addition, due to the nature of the measurement, it has not been necessary to take measures to suppress possible interferences of the signal in the sensor of the instrument. In general, it is to be noted that the ultimate purpose of the ITS sensors does not involve an exact measurement, but an estimate that allows to take the proper safety measures for the driver on motorways. From this point of view, the inherent uncertainties encountered during the characterizations of the instrument are not significant.

Finally, as future researches it would be advisable to improve the design proposed in this paper as well as continue researching in this area. It would also be a good project to continue working on the implementation of other environmental sensors ITS such as snow detector, measuring pollutants, hail impact and other factors.

This work has been funded by the Spanish Ministerio de Economía y Competitividad: Progama Promoción de Empleo Joven e Implantación de la Garantía Juvenil 2014 (PEJ-2014-A-18839).

REFERENCES

1. National Institute of Meteorology of Spain
2. *Guía de Instrumentos y Métodos de Observación Meteorológicos*/Organización Meteorológica Mundial, Ginebra, 1990, Serie: OMM 8, ISBN: 9263300089.
3. A. J. Illingworth, C. J. Stevens, *An Optical Disdrometer for the Measurement of Raindrop Size Spectra in Windy Conditions*, Journal of Atmospheric and Oceanic Technology 1987; 4: 411-421.
4. Martin Grossklaus, Klaus Uhlig, Lutz Hasse, *An Optical Disdrometer for Use in High Wind Speeds*, Journal of Atmospheric and Oceanic Technology 1998; 15: 1051-1059.
5. Martin Löffler-Mang, Jürg Joss, *An Optical Disdrometer for Measuring Size and Velocity of Hydrometeors*, Journal of Atmospheric and Oceanic Technology 2000; 17: 130-139.
6. Denby B., Prevotet J.C., Garda P., Granado B., Barthes L., Gole P., Lavergnat J., Delahaye J.Y., *Combining Signal Processing and Machine Learning Techniques for Real Time Measurement of Raindrops*, IEEE Transactions on Instrumentation and Measurement, Vol. 50, No. 6, Dec. 2001, pp. 1717-1724, Digital Object Identifier: 10.1109/19.982973.
7. http://www.sensorsportal.com/DOWNLOADS/UFDC_1.pdf
8. Web page: <http://pdf1.alldatasheet.es/datasheet-pdf/view/153394/OPNEXT/HE8404SG.html>
9. Web page: <http://www.edmundoptics.com/optics/windows-diffusers/optical-diffusers/holographicdiffusers/1363/>
10. Web page: <http://www.edmundoptics.com/technical-resources-center/optics/diffuser-selection-guide/>.
11. Web page: <http://www.vishay.com/docs/84679/temd501.pdf>
12. Web page: <http://www.ti.com/lit/an/sboa055a/sboa055a.pdf?keyMatch=compensate%20transimpedance%20amplifiers%20intuitively&tisearch=Search-EN-TechDocs>
13. <http://www.analog.com/media/en/technical-documentation/data-sheets/OP27.pdf>
14. <http://www.vishay.com/docs/84679/temd501.pdf>
15. Table 3. http://www.sensorsportal.com/DOWNLOADS/UFDC_1.pdf
16. Sergio Franco, *Diseño con amplificadores operacionales y circuitos integrados analógicos*, McGraw-Hill Interamericana, México, 2005.

Visual Interfaces for Clinical Archetype Editing

Christian Ponce, José Alberto Maldonado, Diego Boscá, Montserrat Robles

Instituto ITACA. Universitat Politècnica de València, Camino de Vera s/n, 46022
Valencia, España, chponrui, jamaldo, diebosto, mrobles @upv.es

Abstract. Nowadays is still difficult to use interfaces for developing standardized detailed clinical information models. Modern clinical standardized models let us define data structures by using archetypes. In this paper we present two applications developed to improve the way in which the medical professionals interact with clinical data models, making the editing process simpler, easier, and interoperable with the goal of optimizing this kind of applications. These applications use the information of the chosen archetype to generate automatically the interface. Our approach has been tested for CDA, openEHR, and ISO-13606 clinical models.

1 Introduction

Nowadays there is clearly a need of interoperability in the Electronic Health Record Systems (EHRs)[1]. The use of EHRs standards is required to reach interoperability. Archetypes, also called Detailed Clinical Models, are defined in the modern reference models ISO-13606[2], and openEHR[3]. They are used to constraint the structure of EHRs information fulfilling the restrictions of standardized models. LinkEHR methodology developed by Biomedical Informatics Group of ITACA institute, allows to standardize the EHRs by making use of archetypes[4]. Archetype editor LinkEHR originally shows the archetype as a tree following the Archetype structure, which needs knowledge of the underlying reference model to its fully comprehension.

In this paper we show new visual interfaces which eases the realization and comprehension of the archetype. These interfaces are mindmap archetype viewer and clinical forms viewer. The development of this application interfaces in a optimized way eases the work for people who are not too acquainted with computers. Interfaces have to meet some features, such as:

- Applications must be usable, interactive, and adaptable.
- They must be able to exchange and understand the data.
- They must have the capacity to represent a large number of data.
- Software libraries must have support for a large period of time.

Relevant examples in the biomedical field are Cytoscape[5], Ingenuity Pathway Analysis[6], PHINCH[7], Omix[8], and BiMS Viewer[9].

2 Materials and methods

In this section we describe the applications and the methods developed to represent the mindmaps and clinical forms, as well as the languages and libraries used to carry out the development.

2.1 LinkEHR Editor

LinkEHR is a powerful archetype editing framework capable of handling multiple reference models of EHRs and oriented towards the semantic description and standardization of legacy data[10]. This application is the base framework which includes the developed applications. The current version of LinkEHR mindmap representation is based on Freemind using a flash viewer[11]. Freemind is an application written to allow the user to edit a hierarchical set of ideas around a central concept. Freemind uses its own format file to draw a mindmap. Mindmap can be exported from linkEHR. Flash technology, which is proprietary and in particular the Freemind flash viewer is somewhat obsolete. For this reasons we have decided to implement an upgraded and more supported ones.

Our main goals with mindmap viewer are to develop an application in HTML5 to create a tree representation with the Data-Driven Documents library(D3)[12] and JavaScript. This library works with JSON data files to represent data, so the JSON file of the archetype was created.

In addition to mindmap viewer, we have also decided to work on a clinical form representation tool. Our goal is to use a widget developed by Lister Hill National Center for Biomedical Communications[13]. It is a open-source widget that creates input forms for Web-based applications. In the current version of LinkEHR the archetype is transformed to create an XML file, and XSLT is applied to the XML to create the HTML form. This transformation is functional only for ISO-13606 standard model. However, as the new implementation works with a well-defined and easy to learn structure, our goal is to extend easily this application to any other standard model such as CDA or openEHR.

2.2 Mindmap Visualization

Mindmap viewer is responsible for carrying out the tree representation of the archetype. It uses both HTML and JavaScript languages. JavaScript is in charge of performing the operations of the visualization part. It makes use of the Data-Driven Documents (D3) library[14]. D3 library is a JavaScript library for producing dynamic, interactive data visualizations in web browsers. D3 library makes use of the widely implemented SVG, HTML5, and CSS standards. D3 library facilitates the manipulation of web documents. This library is generic, for this reason we have made some changes for deal with certain problems which have arisen during our development. The first problem we found was related with the data file. We had to adapt the library to retrieve the JSON data as we wanted, fitting correctly with the structure coming from the archetype. Furthermore we had to add some methods to give more functionality to the mindmap, such as:

- Adding nodes to the tree
- Expand and collapse all children nodes from its parent
- Fit the tree to the screen, which allows to represent large mindmaps
- Export SVG view or data into a file

Another main feature we have implemented is Drag & Drop. Drag & Drop allows the edition of the mindmap view of an archetype.

2.3 Medical Forms Visualization

Medical forms representation application eases the comprehension of an archetype coming from a Electronic Health Record System (EHR) or Personal Health Record System (PHR). This widget uses LForms library from Lister Hill National Center for Biomedical Communications[15]. This library is coded in angularJS language, designed for building new apps. LForms library implements a widget that creates input forms for Web-based medical applications. Lforms data is entered with JSON file, which allows us to extend this application to adapt it to any EHR standard. LForms library has the following features:

- Data types.
- Cardinality.
- Units of measure.
- Selections in list (one choice/multiple choice).
- Default value settings.
- Validation.
- Export data to JSON file.

3 Results

Fig. 1 and Fig. 2 show an example of existing and new mindmap visualizations of an archetype. Nodes with one or more children have a blue color (except root node) and nodes that have not child nodes are white. Occurrences are also shown inside the node, following the name of each node. Class to each node belongs is shown both as an icon and as a tooltip which can be accessed by hovering the mouse over the icon. Tooltip shows:

- | | |
|---------------------------------|-----------------------------|
| – Node description. | – Node comments. |
| – If the attribute is required. | – Parent attributes. |
| – The type of the node. | – Allowed values or ranges. |

Each node is connected with its children with a line. When the node could have more than one occurrence, it will have an * to denote the {0..*} or {1..*} relationship. This application also allows mouse events and zoom. The zoom can be done with the mouse wheel. When a left click over a node is received, the node will be toggled showing or hiding its children. When a right click over a node is detected, shows a menu with the following three options: a) Expand all

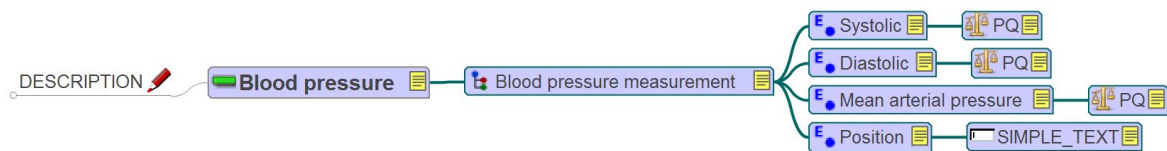


Fig. 1. Existing Mindmap view of a "Blood Pressure" ISO-13606 archetype

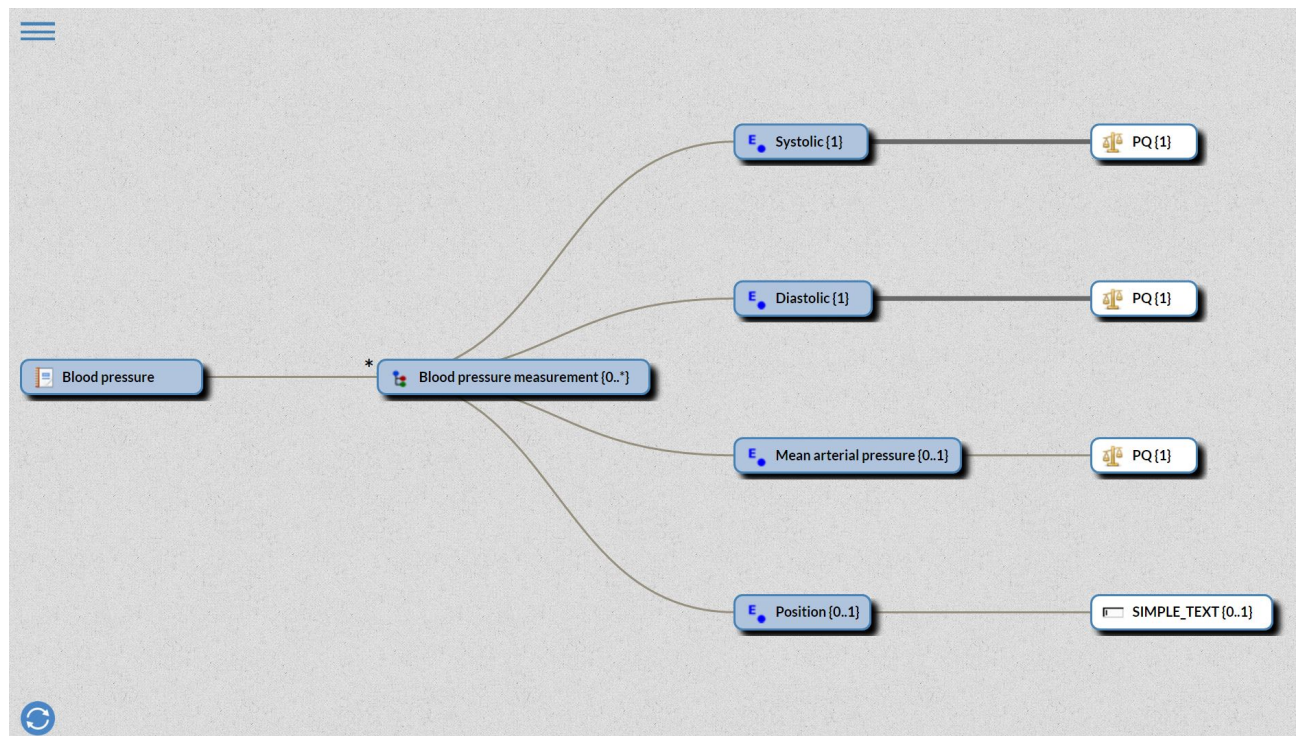


Fig. 2. New Mindmap view of a "Blood Pressure" ISO-13606 archetype

nodes from the actual node, b) Collapse all nodes from the actual node, and c) Add a node. This last option shows a window which lets the user to specify a name and to choose between all compatible types for the actual node. This application also supports Drag & Drop. We are able to drag the tree and any node from the tree individually to drop it to any compatible node. By doing so, this node will become child of the chosen node. We have implemented a tools menu. This tool menu let us to collapse all nodes, expand all nodes, fit the tree to the screen and export data and view in JSON or SVG format respectively.

Medical forms representation Fig. 3 shows us dynamically the variables of the selected archetype. It is designed in a horizontal layout where each item is divided in 3 columns. The first one shows the name, archetype node identifier (if activated) and description (if activated). The second column shows the section to input a value of the item. Value format depends on the data type. E.g. a Date data type is shown as a calendar, in case of String data type, a list of values will be shown. This column also has the capacity of checking if a value is required or not. In the last column we can choose between the available units for the item. When an item has more than one occurrences, an additional button with a label "Add another ..." will be shown. It is also possible to get the form in JSON format by pressing the corresponding button.

CEN-EN13606-ENTRY.Blood_pressure.v1			
Date Done	Time Done	Where Done	
DD/MM/YYYY	Type a value	Select or type a value	
Name	Value	Units	
└ Blood pressure			
└ Blood pressure measurement ?			
└└ Systolic ?	Type a number between [0.0..1000.0]	mm[Hg]	
└└ Diastolic ?	Type a number between [0.0..1000.0]	mm[Hg]	
└└ Mean arterial pressure ?	Type a number between [0.0..750.0]	mm[Hg]	
└└ Position ?	Select one		
Add another 'Blood pressure measurement'			
Get JSON instance		1: Standing	
		2: Sitting	
		3: Reclining	
		4: Lying	

Fig. 3. Medical form representation of a "Blood pressure" ISO-13606 archetype

In addition to the above implementations, we have created REST services and forms for both tools to easily test the presented applications^{1,2}. This can be done by providing an archetype.

4 Conclusions

As we discussed before, is needed to improve visual interfaces for archetype editing. The applications developed help to the user to define concepts in an easier way. We have developed these applications to improve the usability, interoperability and the capacity of represent different amount of data. Regarding the improvements related to the developed applications are:

- The applications are easy to use and can be adapted to other fields of work.
- The implementations are based on new technologies as HTML5, angularJS or D3 have been used to develop the applications.
- The libraries used are standard and modern libraries.

In spite of this improvements, there are some things that can improve this applications and could be considered as future work, as well as:

- Transform mindmap viewer into a full editor, giving features as deleting or editing nodes.
- Make usability tests to evaluate the applications, this will help us to know if we have to make any changes in the interface of the applications.
- Improve clinical forms for being the real data input interfaces.

5 Acknowledgements

This work has been funded by the Spanish *Ministerio de Economía y Competitividad: Programa Promoción de Empleo Joven e Implantación de la Garantía Juvenil 2014*: PEJ-2014-A-18839.

References

1. HealthIT.gov | the official site for Health IT information, <https://www.healthit.gov/>, accessed on 2016-05-24
2. ISO 13606-1:2008 - Health informatics – Electronic health record communication – Part 1: Reference model, http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=40784, accessed on 2016-05-24
3. openEHR Architecture Overview, http://openehr.org/releases/BASE/latest/docs/architecture_overview/architecture_overview.html, accessed on 2016-05-09
4. ISO/CD 13606-2 - Health informatics – Electronic health record communication – Part 2: Archetype interchange specification, http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=62305, accessed on 2016-05-09

¹<http://tiny.cc/MMViewer>

²<http://tiny.cc/FormRepresentation>

5. Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization, <http://www.cytoscape.org/>, accessed on 2016-05-09
6. Ingenuity IPA - Integrate and understand complex 'omics data, <http://www.ingenuity.com/products/ipa>, accessed on 2016-05-09
7. PHINCH - A framework for visualizing bio data, <http://phinch.org/>, accessed on 2016-05-09
8. Omix Visualization, <https://www.omix-visualization.com/#sthash.d5kORAGp.dpbs>, accessed on 2016-05-09
9. P. Díaz Ruata, A. Perera Lluna: Biomedical Multi-Scale Viewer (BiMS Viewer): visualizador interactivo de datos biomédicos. In: CASEIB, pp. 217–220 (2015)
10. Maldonado, J.A., Moner, D., Boscá, D., Fernández-Breis, J.T., Angulo, C., Robles, M.: LinkEHR-Ed: A multi-reference model archetype editor based on formal semantics. *International Journal of Medical Informatics* 78(8), 559–570 (Aug 2009), <http://www.sciencedirect.com/science/article/pii/S1386505609000513>
11. FreeMind, <https://sourceforge.net/projects/freemind/>, accessed on 2016-05-09
12. Bostock, M.: D3.js - Data-Driven Documents, <https://d3js.org/>, accessed on 2016-05-09
13. U.S. National Library of Medicine | Lister Hill National Center for Biomedical Communications, <https://lhncbc.nlm.nih.gov/>, accessed on 2016-05-09
14. Murray, S.: Interactive Data Visualization for the Web. In: *Interactive Data Visualization for the Web*, pp. 7–8. "O'Reilly Media, Inc." (Mar 2013)
15. LForms | U.S. National Library of Medicine, <https://lhncbc.nlm.nih.gov/project/lforms>, accessed on 2016-05-09

Multiscale technical systems (in particular, totality of cell telephones) reaffirms the existence of a fractal system of tripled periods

E. V. Kolombet¹, V. N. Lesnykh¹, O. Yu. Seraya¹, V. A. Kolombet¹, G. Verdú^{2,3},
V. Milian-Sanchez², A. Mocholí-Salcedo⁴

¹ Institute of Theoretical and Experimental Biophysics of Russian Academy of Sciences (ITEB RAS), Institutskaya str, bld. 3. Pushchino, Moscow region, 142290, Russian Federation
kvision@mail.ru, 1253vadim@mail.ru, olga-seraya@yandex.ru, v.kolombet@rambler.ru
<http://web.iteb.psn.ru>

² Institute for Industrial, Radiophysical and Environmental Safety, Universitat Politècnica de València ,
Camí de Vera, s/n. Valencia 46021, Spain; vicmisan@iqn.upv.es

³ Chemical and Nuclear Engineering Department, Universitat Politècnica de València ,
Camí de Vera, s/n. Valencia 46021, Spain; gverdu@iqn.upv.es

⁴ Traffic Control Systems Group, Instituto ITACA, Universitat Politècnica de València,
Camino de Vera, s/n, Valencia, Spain; amocholi@eln.upv.es

Abstract. The distribution of periods of various processes (astronomical, geophysical, biophysical and biological) leads to the discovery of a fundamental tendency to fractal organization. Proves of the existence of a fractal system of the tripled periods are also observed in technical devices. For example, progress in increasing the speed of modern CPUs ..., 1.12 GHz, 3.36 GHz is hindered because of the need to jump to the next tripled clock frequency, that is, to 10 GHz. The fundamental frequencies of the fractal are described by the formula $f_{n,k} = f_0 \cdot n / 3^k$, where $f_0 = 1/114.6$ years, $n = 1$ or 2 and k is an integer in the range from -40 till $+15$. In this report, we describe the related difficulties in today's mobile telephony. The frequencies received by mobile phones from a wide variety of businesses providers appear to avoid the frequencies of the fundamental fractal.

1 Introduction

A paper of Stephen J. Puetz et al, published in April 2014 [1], presented the proof of the global footprint of a tripled periods fractal system (TPFS) that attracted the attention of scientific community all over the world. On the basis of experimental data analysis of the geological, geophysical, biological and astronomical processes, the authors demonstrated the existence of some special meta-resonance of periodic processes; the periods differ from each other in 3^k , where $k = 1, 2, \dots, 15$. In other words, a time range from ~ 50 years to ~ 1.5 billion years, in fact, the cosmological

time scale, includes a lot of tripled periods [1], cf. [2].

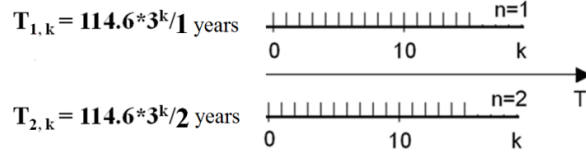


Fig. 1. Schematic representation of TPFS periods described in the article [1]. The figure uses a logarithmic scale of the timeline providing a user-friendly way to show periods of the multiscale system. It covers the range from ~50 years ($n = 2, k = 0$) till ~1.5 billion years ($n = 1, k = 15$).

The result of [1] was the disclosure of at least two systems of periods: the first one extends upward from the period ... 114.6 / 1 year, and the second one extends upward from the period 114.6 ... / 2 years. For simplicity, they are marked below as the system with $n = 1$ and the system with $n = 2$, respectively. Each of them is a 16-period value. (Fig. 1). We decided to explore other values of the parameter k .

2 Extension of the detected values of k

The possibility of further expansion of TPFS in the longer period region is unlikely because the horizon - the boundary of the known universe - is located at 13.8 billion light years and the process has to do, during this time, at least a couple of oscillations. The maximum value of known TPFS period is 1.64 billion years [1], which is approximately 3^2 times smaller than the time life of the universe limited by the horizon. Certainly, one can look for candidates in the astronomical processes with evaluated periods $T_{n,k} = 2.5, 4.9$ and 7.4 billion years [3], but the chance of luck is small because above 100 Mpc scale, i.e. across 326 million light-years, the universe is considered to be homogeneous. Nevertheless, the size of recently found the Laniakea cluster is 500 million light-years and one could think about larger structures of this kind.

We also investigated the possibility of FSTP-states expansion due to the shift of the range the lower limit of the periods (see $k = 0$ in Fig. 1) in the region of shorter periods, i.e. in the region $k < 0$. So, we considered the periodic (in the logarithmic scale) deterioration / improvement of human memory (the so-called Bondar's law [4-6]), the phenomenology of human hearing [6, 7], as well as the preferred frequency in modern multiscale technical devices. It turns out, for example, that a person's memory of the events not only weakens with time, but also oscillates with naturally expanding period.

Regarding the information supplied for memorizing, the recalled ("memory enlightening"), and the almost completely forgotten information ("memory dimming"), it has been experimentally obtained (see [5-7]), that the period of enlightenment / dimming memory is tripled (Fig. 2, 3).

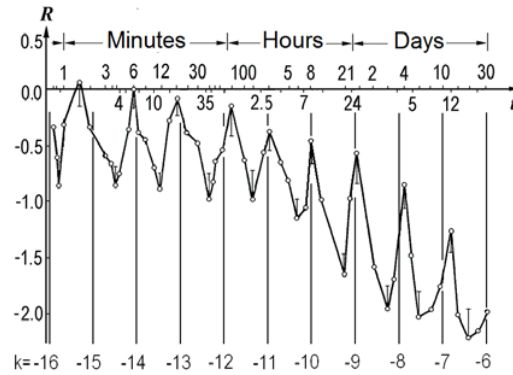


Fig. 2. Periodicity (in logarithmic scale) of human ability to recall events. Peaks in the curve correspond to the periods $T_{n,k} = 114.6 * 3^k / n$ years, where $n = 2$, $k = -15 \div -6$.

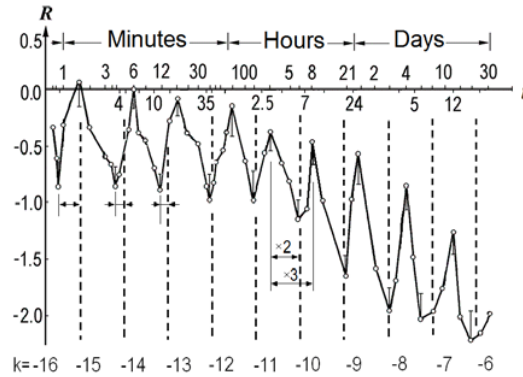


Fig. 3. Periodicity (in logarithmic scale) of the human ability to forget an event. Gaps in the curve correspond to the period $T_{n,k} = 114.6 * 3^k / n$ years, where $n = 1$, $k = -14 \div -6$.

In a logarithmic scale, the successive moments of enlightenment memory and the successive moments of its darkening appear almost at regular intervals. In many of the 51 time slots, studied by A. T. Bondar, the playback efficiency was significantly ($p < 0.01$) increased (see peaks in Fig. 2), and in other cases was significantly ($p < 0.01$) reduced (see minima points in Fig. 3). During the month, there were about 10 oscillations.

It appeared also that the system of depressions (Fig. 3) is also not accidental. It corresponds to twice the period of neighboring peaks located to the left of these depressions. Periods of TPFS with parameter $n = 1$ are situated in these places. It can be assumed that the existence of periodic depressions associated with the processes occurring in the human brain, obstructs the memory retrieval process, since recording

may, for example, occur at this time in memory.

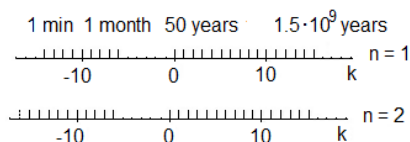


Fig. 4. Extending the triples periods schema in various native processes as a result of Bondar's experiments with human memory.

How, then, appears the adjustment to TPFS periods at different scales of the universe in various periodic processes? It is possible to try to follow the example of evolutionary adjustment to TPFS of human auditory apparatus [7].

The frequency range of 20 Hz - 20 kHz, distinguishable by the human ear, includes seven triples of TPFS frequency (Figure 5.). As one can see, this is the very thing that results just in seven white keys of octave - they get into the middles of pairs of TPFS frequencies. So division and multiplication on dyad leads to the replication of the seven found particular frequencies into all octaves of range of human hearing. Another five triples are obtained by continuing of recurrent procedure of the frequency tripling. The black keys of all octaves appear after divisions on the respective powers of two. This choice of five frequencies is a winning convention [8].

It can be assumed that the emergence of singing is a result of permanent TPFS influence on the human brain in the course of long biological evolution. A kind of singing is known even among great apes, such as gibbons have. Music has appeared on this same basis as an instrumental replacement of singing. Music based on the Seven notes - F, C, G, D, A, E, B - is perfectly matched to the wide range of hearing.

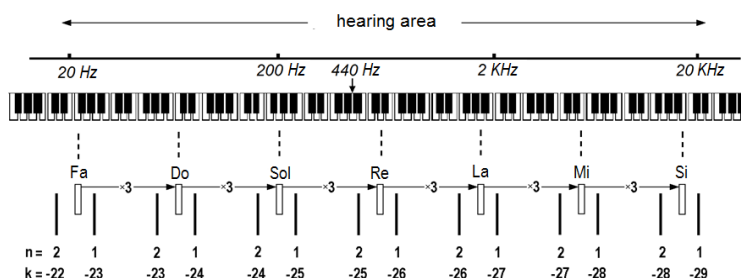


Fig. 5. The seven "basic white keys" are adopted to the TPFS almost as close as possible; they hit the centers of pairs of "fundamental periods" $T_{n,k} = 114.6 \cdot 3^k / n$ years; ($n = 1, 2$ and $k = -22, -23, \dots -29$). However, there is a slight shift to the right from the pair centers provided by the modern tune to frequency 440 Hz; perhaps this is due to different levels of severity of TPFS with $n = 1$ and with $n = 2$. Really, according to [1], TPFS system with $n = 1$ is the dominated one: it is manifested more clearly. The frequencies of the seven "basic white keys" are shifted it is to this (right) side.

The seven "basic notes" roughly correspond to European music centers of TPFS pairs of frequencies but do not correspond to 14 TPFS frequencies themselves (Fig. 5). Imperfection, over-simplicity of European 7-note music in comparison with the "universal music" based on 14 TPFS frequencies is evident.

So, the nontrivial gravity of the seven "basic white keys" (Fig. 5) to the centers of TPFS pairs expands the set of TPFS periods (Fig. 6) by bringing the parameter k to value -29.

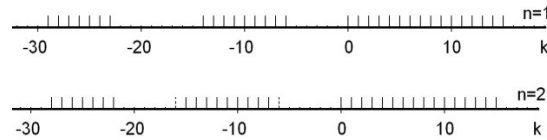


Fig. 6. Further expansion of the TPFS periods into high frequency area as a consequence of the music phenomenon leads to $k = -29$.

It should be noted that the displacement of the lower border of TPFS periods from fifty years to $1 / (20 \text{ kHz})$ does not lead to deviation from the predicted values of TPFS periods. This suggests the prospect of extrapolation of TPFS to even more short periods, which aims to find the characteristic TPFS frequencies in the area, which has mastered by modern electronic technical devices. To do this, we first have analyzed the clock frequencies of multiscale devices like computer processors. It turned out that the frequencies of processors are clearly adjusted by the manufacturers to the frequencies of FSTP.

It is known that the clock frequency characterizes the performance of the processor, that is, the number of operations per second.

For a long time, the main processor performance has increased by the growth of clock frequency. Nevertheless, near the 2005 the growth stopped in front of 4 GHz. Over the past decade this limitation has not been overcome [9, 10].

The emerging constraint could have a fundamental origin.

In support one can consider a smoothed histogram for frequencies of modern computer processors [11].

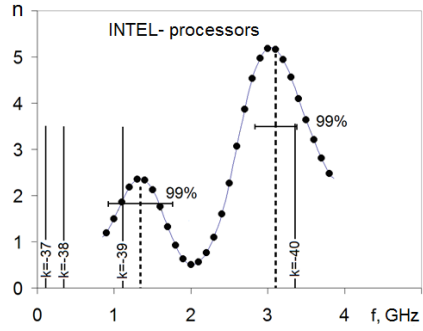


Fig. 7. A smoothed histogram for clock frequencies of Intel computer processors. Basic data for the construction of this figure are taken from Wikipedia [11]. The centers of the peaks (dashed lines) actually coincide with $1 / T_{1,-39}$ and $1 / T_{1,-40}$ (solid lines). $n = 1$

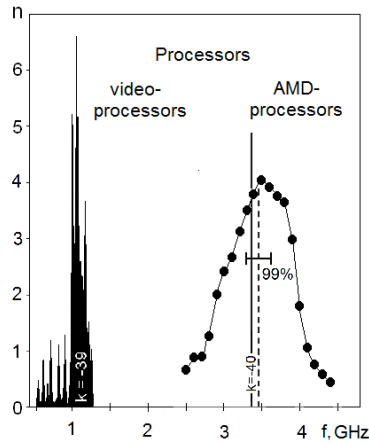


Fig. 8. Smoothed histogram of clock frequencies of AMD processors and processors of graphics cards collectively repeat the smoothed histogram shown in Fig. 7. $n = 1$

The proximity of mean clock frequencies (i.e. of peaks observed in Fig. 7) to the frequencies $1/T_{1,-39}$ and $1/T_{1,-40}$ is evident. Statistical evaluation leads in both cases to hit into the boundary of 99% confidence interval [12]. Distributions of the processor clock speed for graphics cards [13] and for AMD processors [14] show the same pattern (Fig. 8).

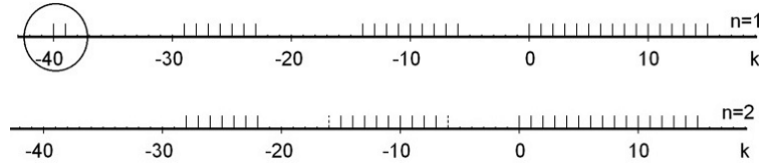


Fig. 9. New extension of the detected TPFS frequencies to 1 - 4 GHz after analyzing the clock frequency of modern processors.

From Fig. 9, and summarizing the achievements, one can assume that the new technological findings will lead to a new processor with the clock frequency of 10 GHz. It is possible to expect a sharp jump from already mastered area at 3-4 GHz straight to clock frequency in vicinity of 10 GHz. Intermediate frequencies are hopeless.

Apparently, a new generation of processors will work on the clock frequency close to $1/T_{1,-41}$.

2 Features of the area of parameter k overlapping with the area of mobile phone frequencies

The previous example of processors shows how unexpected manifestations of the frequencies $1/T_{n,k}$ can be. For instance, use of electromagnetic waves of the UHF range in household appliances intensified dramatically in recent decades. These are mainly microwaves and cell phones. Among other things, this raises the question of the protection of human health from the potentially harmful microwave radiation.

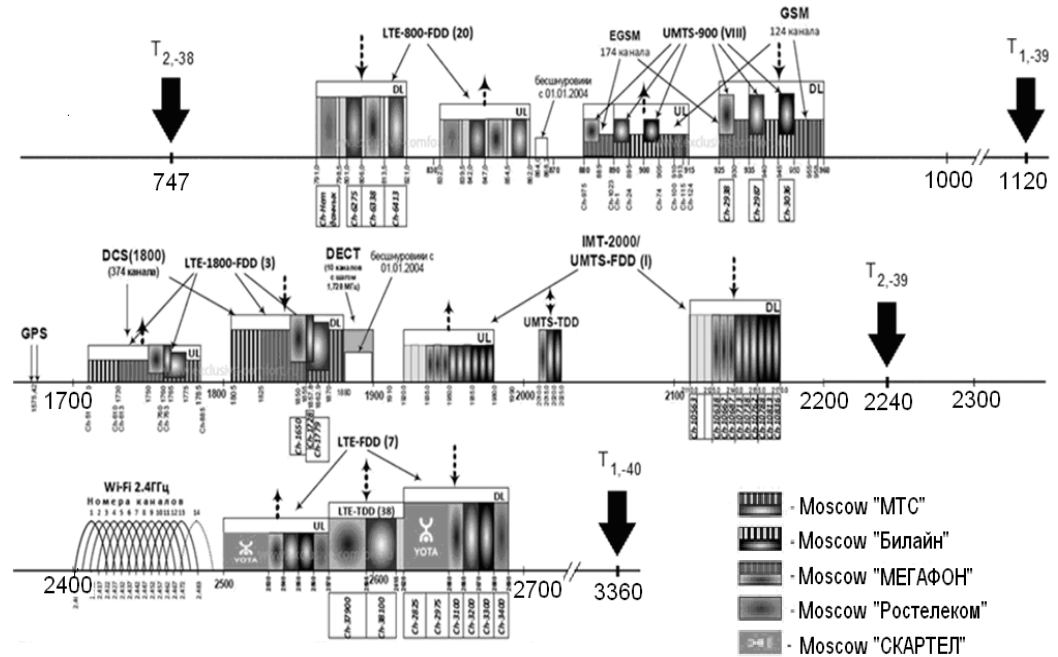


Fig. 10. Areas of operating frequencies (MHz) for mobile phones lie far from the frequencies corresponding to frequencies of TPFS [15]. (Do they repulse each other?)

However, it is known that the biological action of not very large doses of the microwave radiation of different wavelengths is small; the observed effects are mainly thermal. Nevertheless, the detected feature frequencies $1/T_{n,k}$ can be considered as a possible source of acute alternative for this tranquil picture. We can imagine that after a rigorous study of biological action of frequencies $1/T_{n,k}$, they appear to be channels for the effective "resonance" physical therapy and for other "resonance" influences.

Standard microwave frequency of 2,450 MHz (and 2,495 MHz), being selected to ensure rapid heating of water, are far enough from the nearest TPFS frequencies $1/T_{2,-39} = 2,240$ MHz and $1/T_{1,-40} = 3,360$ MHz. "Resonance" effects are not expected here.

The cell phones work at frequencies arranged into series of bands in the 800–2,700 MHz range; but both the 960 – 1,700 MHz range encompassing $1/T_{1,-39} = 1,120$ MHz and the 2,170–2,400 MHz range encompassing $1/T_{2,-39} = 2,240$ MHz are free of cell phone frequencies (Fig. 10). "Resonance" effects are not expected here

too.

3 Results

So, some new prospects associated with the emergence of ideas about clearly determined discrete spectrum of special frequencies $1/T_{n,k}$ can appear in the field of medicine, biophysics, technology, etc. At these frequencies, one can search a variety of new manifestations of "resonance" effects which could be analogous to "attraction" to the frequencies of $1/T_{n,k}$ in processors and to "repulsion" from the frequencies $1/T_{n,k}$ in mobile telephony, and also they could be analogous to the "tune" to the frequencies $1/T_{n,k}$ of mechanisms of human's hearing and memory action.

References

1. Stephen J. Puetz, Andreas Prokoph, Borchardt Glenn, Edward W. Mason. Evidence of synchronous, decadal to billion year cycles in geological, genetic, and astronomical events // Chaos, Solitons & Fractals. 2014. Vol. 62-63. P. 55–75.
2. H. Muller. Fractal scaling models of resonant oscillations in chain systems of harmonic oscillators // Progress in physics. 2009. Vol. 2. P. 72-76.
3. Yu. V. Baryshev and P. Teerikorpi, The Fractal Structure of the Universe: An Essay on the Development of Cosmology (Siberian Branch, Russ. Acad. Sci., Novosibirsk, 2005) [in Russian]
4. A. T. Bondar', DAN SSSR 236 (6), 1503 (1977) [in Russian].
5. A. T. Bondar', in Memory and Trace Processes (ONTI NCBI, Pushchino, 1979) [in Russian].
6. V. Kolombet, Hypnographic Psychoanalysis. An Introduction. (OOO Mailer, Moscow, 2008) [in Russian]
7. A. T. Bondar', M.V. Fedorov, V.A. Kolombet. Biophysics, 2015, Vol. 60, No. 6, pp. 1006–1012
8. Z. Tukhmanova, Starinnaya Muzyka 3–4, 23 (2005) [in Russian]
9. <http://www.electrosad.ru/Processor/ProcTech3.htm>
10. <http://habrahabr.ru/company/intel/blog/194836/>
11. https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%80%D0%BA%D0%B8%D1%80%D0%BE%D0%B2%D0%BA%D0%B0_%D0%BF%D1%80%D0%BE%D1%86%D0%B5%D1%81%D1%81%D0%BE%D1%80%D0%BE%D0%B2_%D1%84%D0%B8%D1%80%D0%BC%D1%8B_Intel
12. V.A. Kolombet, E.V. Kolombet, V.N. Lesnykh. Izvestiya Instituta inzhenerniy fiziki 40 (2) 55-58 (2016). [in Russian]
13. <https://market.yandex.ru/catalog/55314/list?hid=91031&exc=1&how=dpop&page=1>
14. http://monitor.ykt.ru/index.php?show_aux_page=6
15. <http://www.exclusive-comfort.ru/udata/image/mobile-frequency/new/>

Multi-core and Cluster-based Approaches to Enable Exhaustive Fault Injection on Highly Complex HDL Models

Ilya Tuzov, David de Andrés, and Juan Carlos Ruiz

ITACA, Universitat Politècnica de València
Camino de Vera s/n, 46022 Valencia, Spain
{tuil, ddandres, jcrui zg}@disca.upv.es

Abstract. Increasing integration scales are also increasing the sensitivity of integrated circuits to different kinds of faults. Nevertheless, although existing simulation-based techniques can be more or less easily integrated into common semicustom design flows, the simulation time required to estimate the robustness of a circuit described at the implementation-level is prohibitive. This paper takes benefit of different computing platform available nowadays, like multi-core processors and cluster-based systems, to parallelise the execution of fault injection experiments and enable exhaustive fault injection campaigns.

Research Area: Fault Tolerant Systems

1 Introduction

Electronic Design Automation (EDA) tools are nowadays sophisticated enough to achieve a close-to-optimum implementation of integrated circuits with designer assistance [3]. Hence, designers can reason in terms of functionality and the EDA tools take care of the implementation details. On the other hand, increasing integration scales are making VLSI integrated circuits more sensitive to external single event effects (SEE) [7], burn-out and wear-out mechanisms, and soft errors. Accordingly, it is necessary to integrate fault injection techniques into common EDA design flows to estimate the actual impact of faults into the designed circuit.

Existing simulation-based fault injection techniques usually work with models described at the behavioural or register-transfer levels, and inject faults using either simulator commands, mutants, or saboteurs [2]. These techniques can be fairly easily integrated into common EDA design flows, by adapting each injection approach to the commands and options available in the selected simulator. Furthermore, the closer to the implementation-level models are, the more representative of the behaviour of the final circuit become the related simulations. Nevertheless, this representativeness comes at the cost of prohibitive simulation times, even for relatively simple models [6]. The size of the model,

understood as the number of signals (or logical primitives) used to describe that model, is directly related to those costs. The more detailed (lower level) and thus larger the model, the longer the simulator takes to handle and keep track of all the signals of the model. Accordingly, exhaustive fault injection campaigns on implementation-level HDL models is unaffordable.

This paper addresses this problem by taking advantage of existing multi-core and cluster-based computing platforms to execute the fault injection experiments in parallel, thus reducing the total cost of execution of the whole fault injection campaign. Section 2 describes the common simulation-based fault injection procedure implemented and how it has been integrated into the semicustom design flow. The considered computing platforms integrated into the current prototype of the tool implementing that fault injection procedure are presented in Section 3. Section 4 details the considered case study and estimates the speed-up attained by the proposed approaches. Finally, Section 5 summarises the main conclusions of the paper.

2 Simulation-based Fault Injection

As any other fault injection technique, simulation-based approaches consist of three general stages: i) an experimental set-up, which defines parameters and files to configure and control the rest of the procedure; ii) execution of experiments, which carries out experiments in absence (Golden Run) and in presence of faults; and iii) observation and analysis of results, which collects the required measurements to compute the expected measures.

When dealing with highly complex (implementation-level) HDL models all those stages are tightly coupled with state of the art back-end tools for circuits verification and implementation. The latest version of the prototype tool that implements the simulation-based fault injection procedure depicted in Figure 1, relies on Xilinx’s ISE Design Suite [9] to implement the target HDL design on Field-Programmable Gate Arrays (FPGAs), and Mentor Graphics’ ModelSim [4] to simulate and inject faults into the resulting implementation-level HDL model. Other tools could be used by adjusting existing scripts to use the available commands and parse the provided logs and intermediate files.

2.1 Experimental Set-Up

Xilinx’s common design flow for Field programmable Gate Arrays (FPGAs) is used to obtain the implementation-level model of the target circuit from the provided higher level (behavioural or RTL) model. *Register-Transfer Level (RTL) models* describe digital circuits in terms of the flow of (data and control) signals between hardware registers, and the logical operations performed on those signals. Afterwards, this model is further refined following a *synthesis* process, which generates a more detailed description of the circuit (*gate-level netlist*) in terms of its basic logic elements and their interconnection. A *technology mapping* process translates these basic building blocks into those available in the selected

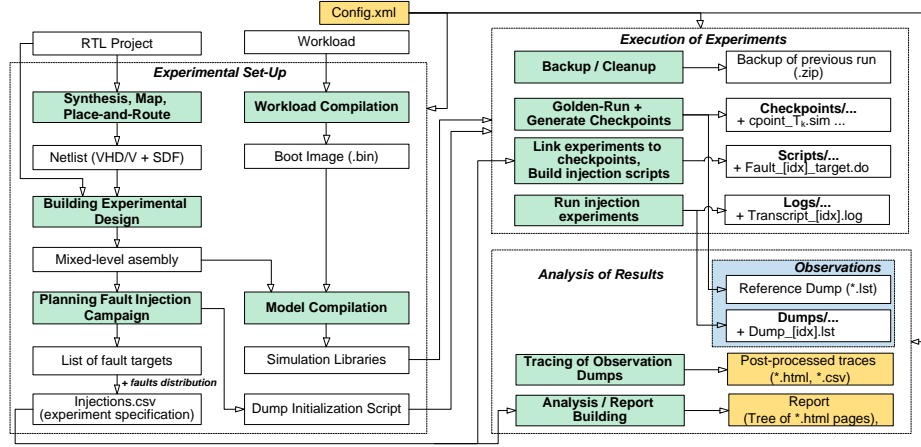


Fig. 1. Experimental procedure: main stages and files

implementation technology (FPGAs in our case). This netlist can be subsequently placed and routed to produce the final *implementation-level model*, also known as *post place and route model*.

The implementation-level model of the target circuit is parsed, according to the naming convention of the selected synthesis tool (Xilinx’s XST in this case), to identify existing fault injection and observation points according to the implementation-level primitives defined in the vendor library. As considering FPGAs as implementation targets, Flip-Flops (FFs) and Look-Up Tables (LUTs) are generally selected as suitable fault injection points, and just FFs as observation points, although other components like embedded RAM block (BlockRAM) units could be also taken into account.

The different parameters required to properly configure all the stages of the experimental procedure are defined by means of a custom configuration file (*Config.xml*). A fault injection TCL script file is generated for each single experiment, defining, among other parameters, the selected fault model, the target fault injection point, the injection time and duration of the fault.

2.2 Execution of Experiments

This stage is in charge of executing the selected workload on the design both in absence and in presence of faults. In such a way, results could be compared to determine the impact of those faults in the system behaviour. The execution of each experiment is controlled by the previously generated TCL script files that include all required simulator commands.

Taking into account that implementation-level model may present hundreds of thousands of different fault injection points the execution of exhaustive fault injection campaigns, even for small circuits running simple workloads, may require a prohibitive execution time. Luckily, experiments are independent from

one another, thus enabling a relatively straightforward parallel processing. Next section will present the devised solution to optimise the execution of experiments according to available computing resources.

2.3 Observation and Analysis of Results

The dump initialization script (TCL file), executed at the beginning of the observation interval, provides precise instructions to the simulator (Mentor Graphics' ModelSim in our case) for capturing the state of the system. By default, all internal registers are selected for observation, and interfaces and memory areas can be added by means of the configuration file.

These dump files include, among other things, a simulation timestamp, any number of custom defined auxiliary signals, an two array of values for selected *internal and external* signals.

3 Parallel Execution of Fault Injection Experiments

The execution of implementation-level models is, by itself, a compute-intensive task. Hence, the time required for performing an exhaustive fault injection campaign targeting all possible fault injection points in a circuit, i.e. hundreds of thousands of model executions, is unaffordable.

As all these runs are independent from one another they can be executed in parallel in a fairly straightforward way. So, this section details how to take advantage of available computing resources (multi-core processors and cluster-based systems) to make feasible the execution of exhaustive fault injection campaigns on implementation-level models.

3.1 Multi-core processor-based approach

In general, back-end tools for the verification of HDL models run on a single thread and, thus, on a single core. However, there is nothing preventing the distribution of those experiments among all available cores but the number of available licenses. Accordingly, the current prototype of the proposed fault injection tool provides the required facilities to benefit from the computing power of all the available cores in multi-core processors.

The PC-based version of the tool launches the previously generated fault injection scripts in a managed sub process shell provided by the *python* basic distribution. The user must define the maximum number of simultaneous processes (*MaxProc*) in the configuration file. Then, as shown in Figure 2, the internal list of processes is periodically checked, by calling the *poll* method for each descriptor in the list, to identify the number of active (not finished) processes. Once the number of active processes is lower than the user-defined *MaxProc* constant, a new process is spawned for the next fault injection script, and its descriptor is appended to the list. This procedure is iteratively applied until all fault injection scripts have been executed.

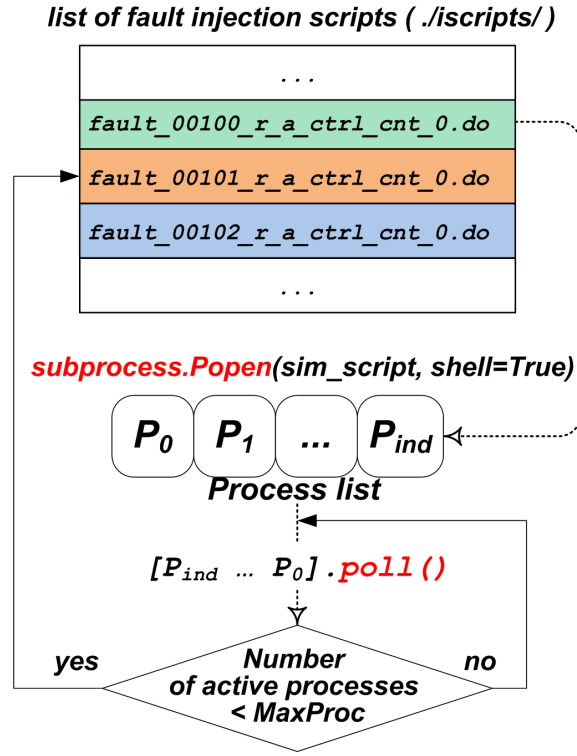


Fig. 2. Multi-core-based execution flow

3.2 Cluster/Grid-based approach

As in the case of the multi-core processor-based approach, the large number of computers comprising a cluster or grid can be used to distribute in parallel the execution of independent fault injection experiments. Hence, our prototype fault injection tool have been extended to take benefit of existing cluster/grid resources.

The main element of a grid system (see the execution flow depicted in Figure 4) is the queue that contains all the tasks (jobs) submitted by users for execution. Those tasks are submitted by means of the qsub command, including the corresponding shell script as parameter. This shell script contains job descriptive information, as the task name, shell interpreter, and the definition of resource requirements (CPU time and memory, among others), and the set of scripts to be executed by the grid.

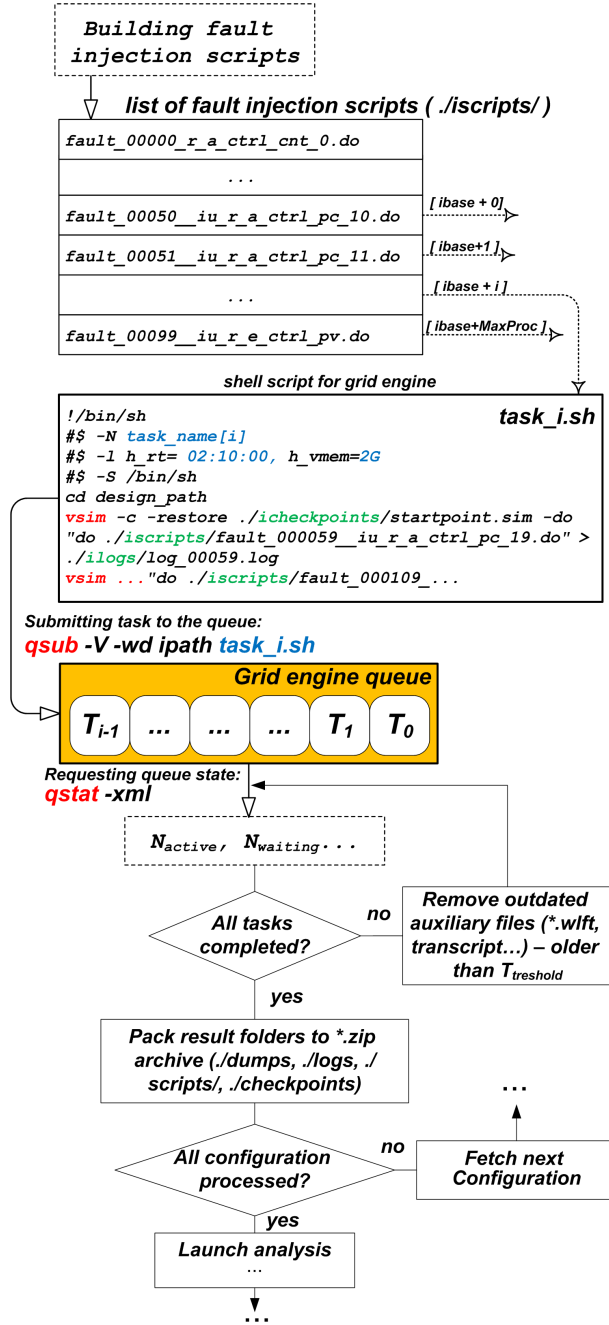


Fig. 3. Grid engine-based execution flow

Tasks are fetched for execution from the queue depending on requested resources, available compute nodes, user quotas, etc. As in the case of the multi-core processor-based approach, the *MaxProc* parameter from the configuration file determines the maximum number of tasks executing in parallel.

The whole set of fault injection scripts is distributed between N tasks as follows. Each task executes the set of simulator scripts identified by indexes $ibase + task_id$, where $ibase$ is incremented by N , and $task_id$ uniquely identifies each task in the range $[0: N-1]$. This procedure enables the gradual execution of experiments without any gap between experiments indexes (given that there is no suspended task), allowing to stop and resume the execution from a given experiment index. The state of submitted tasks (active, waiting, etc.) is monitored by means of repeated *qstat* calls to the grid engine.

3.3 Estimated Speed-up

The maximum attainable speed-up for executing fault injection experiments in parallel the speed-up factor can be estimated by Equation 1.

$$Speed - up = \frac{N \cdot T_{execution}}{\frac{N}{MaxProc} \cdot T_{execution}} = MaxProc \quad (1)$$

Where N is the number of fault injection experiments to be executed, *MaxProc* is the maximum number of processes executed simultaneously, and $T_{execution}$ is the time required to execute a single experiment.

In the case of cluster/grid-based approaches, the attainable speed-up depends on several different factors, like the grid scheduler policies, available resources, user quotas, etc. Only in an ideal case, where there are no suspended/waiting tasks and all computing nodes have the same performance, then the attainable speed-up could be close to that estimated by Equation 1.

Unlike cluster/grid engines, the multi-core processor-based approach launches tasks whenever there are vacant computing facilities in terms of operating system virtual processors. This provides a stable processor load which causes the attainable speed-up to be very close to that estimated by Equation 1 (if *MaxProc*, is lower than or equal to the number of physical cores).

4 Experiments and Results

The developed prototype tool has been used to inject faults into the LEON3 processor running an integer matrix multiplication workload. This section describes the experimental set-up and the real speed-up obtained when applying the two proposed approaches.

4.1 Experimental Set-Up

LEON is a 32-bit soft core processor compliant with the SPARC V8 architecture, whose fault-tolerant version is known to be used as the European Space Agency

(ESA) standard microprocessor. A general-purpose single-core assembly of the 3rd generation of this processor, LEON3 [1], has been selected as case study. This assembly includes a single LEON3 unit (7-stage pipelined unit, separate data and instruction cache controllers, memory management unit, register file, and cache memory), a memory unit, AHB controller and a APB/AHB bridge, all connected to AMBA bus. Peripheral units are connected through the APB bridge, the main PROM stores the boot image, and a UART is used as a debugging interface.

This assembly has been synthesized with Xilinx ISE Design Suite (14.7) for the Virtex-6 target device (6vcx240tff784-2). To understand the complexity of the resulting implementation-level model, it must be noted that it consists of 173 .vhd and 88 .sdf files, comprising a total of 3747 FFs and 7012 LUTs.

A matrix multiplication program implemented in C, adapted from the *basimath* automotive benchmark from the MiBench benchmark suite[], has been selected as workload. Accordingly, only those FFs and LUTs from the integer unit (833 FFs and 1816 LUTs) have been considered as suitable fault injection points. Permanent faults (*stuck-at-0* and *stuck-at-1*) have been injected for both FFs and LUTs (one experiment per fault model per fault injection point), whereas transient ones (*bit-flip*) have only been injected into FFs (three experiments per fault injection point). This leads to a grand total of 7797 experiments.

4.2 Available Computing Resources

Experiments were carried out using the two proposed computing platforms: a multi-core PC and a cluster based on a grid architecture.

The PC featured an Intel Core i5-4670 processor (4 cores/4 threads, 3.4 GHz) and 8 GB DDR3 RAM, running a CentOS 6.7 x64 operating system. The prototype fault injection tool does not require any additional libraries and is executed by means of the basic python distribution provided by the operating system.

Rigel cluster [8] from UPV includes 72 Fujitsu BX920S3 nodes installed on chassis BX900S2. Each node includes two Intel Xeon E5-2450 processors (8 cores/16 threads, 2.12.5 GHz), and 64 GB DDR3 RAM. Nodes are linked by two 10 GB Ethernet interfaces. The cluster runs a CentOS 6 operating system, and its load is managed by Sun Grid Engine [5].

In both environments, Mentor Graphics ModelSim simulator [4] has been used for experimentation.

4.3 Actual Speed-Up

After executing the fault injection campaigns on the different computing platforms, the actual speed-up of using each one can be estimated by comparing with its execution on a single-core machine. Due to the long time required for the execution of the fault injection campaign on the single-core machine, the expected values have been estimated after the execution of 500 experiments.

The average execution time for a single experiment is of 46 s for the PC and 74 s for the cluster. This can be easily explained by the higher performance per

core/thread offered by the desktop processor (cores running at 3.4 GHz vs. 2.1 GHz) and other hardware factors.

Nevertheless, the main goal is to reduce the total execution time for the whole fault injection campaign. In this case, the minimum execution times obtained are of 100.63 hours for the single-core machine, 35.08 hours for the multi-core processor ($MaxProc = 3$), and 3.72 hours for the cluster ($MaxProc = 100$). As can be seen, the speed-up attained by the multi-core processor-based approach is close the best possible (2.87 vs. 3), whereas the cluster speed-up, although better, is far from the best possible (27.05 vs. 100). As previously explained, this difference is related to the particular policies of the scheduler and available resources at the experimentation time.

4.4 Robustness Analysis

Finally, and just as a sanity check, it is necessary to verify that the selected experimental platform does not impact in any way the results obtained. As expected, no difference was detected among the results provided by the three different execution platforms, which validates the correctness of the proposed approach.

Just as an example of the type of results provided by the prototype tool, Figure ?? shows the distribution of the failure modes obtained (*masked* – no failure, no error; *latent* – no failure, incorrect internal state; *signalled failure* – detected failure; *silent data corruption* – undetected failure) for the injection of bit-flips into the target FFs.

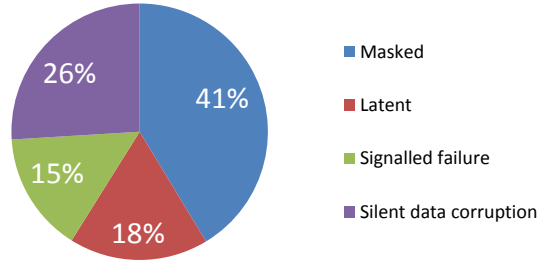


Fig. 4. Sensitivity analysis for bit-flips targeting the FFs of the integer unit of the LEON3 processor

5 Conclusions

Despite existing simulation-based fault injection techniques can be integrated into the common semicustom design flow to analyse the impact of faults into the considered circuit, highly complex implementation-level models still pose

a serious problem for performing extensive fault injection campaigns. This paper proposes the use of existing computing platforms, like multi-core processors and cluster-based systems, to parallelise the execution of fault injection experiments and, thus, reduce the overall time required for exhaustive fault injection campaigns. The selected case study shows the feasibility of the proposal and estimates the attainable speed-up of each platform.

Future work will focus on optimising the execution on the cluster-based platform to further increase the attainable speed-up, and considering a larger set of fault models representative of new integration scales.

Acknowledgment

This work has been partially supported by the “Programa de Ayudas de Investigación y Desarrollo” (PAID) de la Universitat Politècnica de València.

References

1. Cobham Gaisler AB: LEON3 processor product sheet (2016), http://www.gaisler.com/doc/leon3_product_sheet.pdf
2. Gil, D., Gracia, J., Baraza, J.C., Gil, P.J.: Study, comparison and application of different vhdl-based fault injection techniques for the experimental validation of a fault-tolerant system. *Journal of Systems Architecture* 34(1), 41–51 (2003)
3. Kaeslin, H.: *Digital Integrated Circuit Design: From VLSI Architectures to CMOS Fabrication*. Cambridge University Press (2008)
4. Mentor Graphics: ModelSim (2016), <https://www.mentor.com/products/fv/modelsim/>
5. Oracle: Oracle Grid Engine, User Guide, Release 6.2 (2012), http://www.upv.es/entidades/ASIC/menu_urlc.html?entidades/ASIC/catalogo/SGEUserGuide.pdf
6. Palnitkar, S.: *Verilog®Hdl: A Guide to Digital Design and Synthesis*, Second Edition. Prentice Hall Press, Upper Saddle River, NJ, USA, second edn. (2003)
7. Ubar, R., Raik, J., Vierhaus, H.: *Design and Test Technology for Dependable Systems-on-chip*. Premier reference source, IGI Global (2010)
8. Universitat Politècnica de València: Rigel Cluster Description (2012), <https://www.upv.es/entidades/ASIC/catalogo/857857normalc.html>
9. Xilinx Inc.: ISE Design Suite (2013), <http://www.xilinx.com/products/design-tools/ise-design-suite.html>

A very fast Single Error Correction-Double Error Detection Code for short data words

J. Gracia-Morán¹, L.J. Saiz-Adalid¹, D. Gil-Tomás¹, J.C. Baraza¹, P.J. Gil-Vicente¹

¹ Instituto ITACA. Universitat Politècnica de València. Camino de Vera s/n.
46022 Valencia, España.
{jgracia, ljsaiz, dgil, jbaraza, pgil}@itaca.upv.es

Area of interest: Fault Tolerant Systems

Abstract. Due to technology scaling, register protection against soft errors remains a major concern for deep sub-micron systems. Error Correction Codes (ECC) improve protection at the price of data redundancy. In memories, trying to reduce this redundancy is a really important issue. However, this fact is less important in the case of registers. A major requirement for register protection is to keep codes' encoding and decoding latencies as short as possible. This paper proposes a very fast Single Error Correction-Double Error Detection (SEC-DED) code for short data words that are specially designed for register files. This code can be implemented using 2-gate delay encoders and 4-gate delay decoders.

1 Introduction

As technology scaling continues, the increasing variety of physical mechanisms that may perturb the information stored in key computer elements, such as registers and memories, is growing [1], [2]. As a result, error correction codes (ECC) must be also in continuous evolution [3]. In this way, in the bibliography there is a large range of alternatives of single error correction codes, such as Low-Density Parity Check [4], Turbo [5] or Flexible Unequal Error Control [6] codes.

Delay, area and power consumption of computer-based systems using ECC increases as reliability does. On one hand, certain delay is systematically introduced by data encoding and decoding circuitry. On the other hand, such circuitry additionally induces some area and power overhead, as the additional storage and/or interconnection lines in circuits also do in order to introduce the necessary degree of information redundancy. The challenge is thus to design an ECC that, attending to system and application protection requirements, provides a good balance between area and delay, while minimizing power consumption.

When designing an ECC for memory protection, the main goal is to reduce the necessary redundancy, which refers to the number of (redundant) bits added by the code to each word stored in memory [7], [8]. The problem is that the spatial overhead

scales with memory size, being the minimization of the spatial overhead a key criterion in the design of ECC applied to memories.

In the case of register files, area becomes less relevant than delay. In this case, it is more important to minimize the delay introduced by the ECC, because register files are the most frequently accessed part of any microprocessor, while their overall related area is small. As a result, when protecting a register file with an ECC, encoding and decoding operations should be as fast as possible to minimize their impact in the critical path of the instruction execution unit of the processor. Clock cycle increases when the speed of the encoding and decoding circuitry is not fast enough, provoking that the global system frequency decreases, which negatively influences the overall system performance. The same happens when ECC are applied to other types of processor registers, such as inter-stage latches in the case of pipelined computers.

Attending to the importance of registers, and their increasing sensitivity to faults [9], speeding up their protection mechanisms becomes the main concern of this paper. Register file presents a high access rate, so corrupted data in register files can quickly spread to other parts of the system. Although technology scaling enables the manufacturing of high-speed (faster) and high-density (smaller) VLSI processors, it also provokes registers be much more sensitive to soft errors, mainly to transient faults caused by high-energy particles. This problem becomes more important when considering the ever-decreasing low supply voltage level used to operate such registers [10].

Another problem provoked by the ever smaller VLSI processors derives from the fact that several of the bits in a register can be affected by the very same error, similarly to memory cells. As technology scales up, Multiple Cell Upsets (MCUs) have become more common, being currently a significant percentage of the soft errors that affects information in memory [11]. Measurements from 65nm and 45nm Static Random Access Memories (SRAMs) were reported in [12]. As observed in Fig. 1, MCUs affecting two bits (i.e. having a length of two), are the most frequent ones in SRAMs. As stated in [13], such two errors use to occur in the same stored word, and with high probability, in adjacent bits of the word. Although these investigations mainly refer to SRAMs, it must be noted that the basic memory cell is similar to the one of register files. In other words, it is of upmost importance to protect registers against double adjacent errors. The relevance of coping with this challenge increases as soft error rate does, which is something predicted in [14].

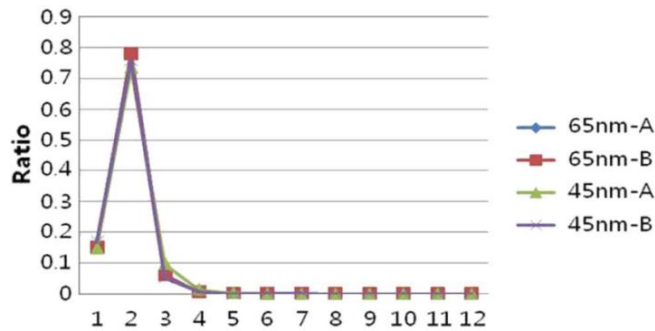


Fig. 1. Ratio of length for MCU events [12]

Extended Hamming codes [7] allow single error correction and double error detection with low redundancy and simple and fast encoding and decoding operations. Nevertheless, as VLSI systems increase their working frequency, reducing the delays introduced by the encoding and decoding circuits becomes of paramount importance. For instance, in [15] we proposed a new type of codes for Single Error Correction for registers (called Ultrafast SEC codes), able to correct a single error with a very short correction latency.

This paper works on such idea and presents a design of an ECC that is able to correct a single error, and to detect a double error for short data words. This ECC is designed especially for the register file. Although its redundancy is very high (for k data bits, k code bits are needed), this code is very fast: 2-gate delay for encoders and 4-gate delay for decoders, assuming the use of 2-input XOR gates to implement the encoding circuit and the syndrome calculation in the decoding circuit.

This paper is organized as follows. Section II introduces the previous work on Ultrafast Error Correction Codes. Section III presents the fundamentals of the Ultrafast Error Correction and Detection Codes. Section IV describes the new code designed. Finally, Section V presents some conclusions and ideas for future work.

2 Previous work on Ultrafast Error Correction Codes

We proposed in [15] a family of SEC (Single Error Correction) codes, called Ultrafast SEC codes, for fast encoding and decoding operations. These codes present a higher redundancy than the equivalent Hamming codes, but they can be implemented using 1-gate delay encoders and 4-gate delay decoders. In addition, these delays are independent of the word length. Ultrafast SEC codes are especially useful in registers, where the encoding and decoding latency may influence the clock cycle duration as well as the processor frequency. At the same time, as the number of registers in a processor is small, a high redundancy is a minor problem. The requirements to construct Ultrafast SEC codes are the following:

- Each column has to be different and nonzero.
- Each column assigned to parity bits must have Hamming weight¹ $w = 1$.
- Each column assigned to data bits must have Hamming weight $w = 2$.
- Each row must have Hamming weight $w = 3$.

The last two requirements allow the design of simple and fast hardware-implemented decoders. However, the last requirement increases substantially the number of parity bits required in the Ultrafast SEC codes. In fact, the number of parity bits is the same as the number of data bits. In addition, only the correction of data bits is performed, as an error in a parity bit does not lead to a wrong decoding.

As an example, this is the parity check matrix for the (8, 4) Ultrafast SEC code, being 8 the number of code bits + data bits, and 4 the number of data bits:

¹ The Hamming weight is the number of 1s in a column.

$$\mathbf{H} = \begin{bmatrix} 1000 & 1100 \\ 0100 & 1010 \\ 0010 & 0101 \\ 0001 & 0011 \end{bmatrix} \quad (1)$$

Following, we will reformulate Ultrafast SEC codes to allow single error correction and double error detection, maintaining the circuit latencies.

3 Fundamentals on Ultrafast Error Correction/Detection Codes

As just commented, the objective of Ultrafast SEC codes proposed in [15] is to reduce the latency introduced in the system by the encoder and decoder circuits. In this paper, some of the restrictions are modified to propose a new code with additional error detection features, while maintaining the encoder and decoder latencies as low as possible.

As stated before, the parity check matrix for Ultrafast SEC codes must fulfill this rule: *each column assigned to data bits must have Hamming weight $w = 2$* . With this rule it is possible to reduce the Hamming weight of the whole matrix, but it complicates the management of double errors.

Following this approach, the new requirements to construct Ultrafast Error Correction/Detection codes are the following:

- Each column has to be different and nonzero: this allows correcting single errors.
- Each column assigned to parity bits must have Hamming weight $w = 1$: this allows minimizing the Hamming weight in those columns; even more, it permits a code be separable (i.e. parity and data bits can be distinguished in an encoded word).
- Each column assigned to data bits must have Hamming weight $w = 3$: although this condition does not minimize the Hamming weight of the parity check matrix, it allows simplifying the final part of the decoding, if only data bits correction is required.
- Each row must have Hamming weight $w = 4$: this requirement is key for our proposal. The result is that each parity bit is obtained during the encoding by XORing only three data bits, and each syndrome bit is calculated during the first part of the decoding by XORing four bits.

The last two requirements allow the design of simple and fast hardware-implemented decoders. However, the last requirement increases substantially the number of parity bits required. As explained before, if redundancy is a major issue, these codes are unaffordable. Nevertheless, when redundancy is a minor problem, these codes will result in an interesting option.

4 New Ultrafast SEC-DED Codes

As stated before, there are two conditions to build a parity check matrix for new ultrafast codes that determine the Hamming weights for their columns:

- Each column assigned to parity bits must have $w = 1$.
- Each column assigned to data bits must have $w = 3$.

As a result, all new Ultrafast codes have odd weight in all the columns of their parity check matrices, and the Hamming distance for these codes is four and, hence, they can detect 2-bit errors.

Let us consider an (8, 4) code, whose parity check matrix is presented in (2).

$$\mathbf{H} = \begin{bmatrix} 1000 & 0111 \\ 0100 & 1011 \\ 0010 & 1101 \\ 0001 & 1110 \end{bmatrix} \quad (2)$$

The syndrome bits are calculated as usual:

r_0	r_1	r_2	r_3	r_4	r_5	r_6	r_7	<i>Syndrome bits</i>
				u_0	u_1	u_2	u_3	
1	0	0	0	0	1	1	1	$s_0 = r_0 \oplus r_5 \oplus r_6 \oplus r_7$
0	1	0	0	1	0	1	1	$s_1 = r_1 \oplus r_4 \oplus r_6 \oplus r_7$
0	0	1	0	1	1	0	1	$s_2 = r_2 \oplus r_4 \oplus r_5 \oplus r_7$
0	0	0	1	1	1	1	0	$s_3 = r_3 \oplus r_4 \oplus r_5 \oplus r_6$

When calculating the syndrome bits, if a single error occurs, the parity bit is nonzero and the syndrome bits point to the bit in error (if all syndrome bits are zero, the error is in the parity bit). Finally, if a 2-bit error occurs, some of the syndrome bits are nonzero, but the parity bit is zero. Under these conditions, a non-recoverable error (NRE) is detected but cannot be corrected. The expression for the NRE signal must include all even nonzero syndromes:

$$\begin{aligned} NRE = & \overline{s_0 s_1 s_2 s_3} + \overline{s_0 s_1 s_2 s_3} + \overline{s_0 s_2 s_1 s_3} + \overline{s_0 s_3 s_1 s_2} + \\ & + \overline{s_1 s_2 s_0 s_3} + \overline{s_1 s_3 s_0 s_2} + \overline{s_2 s_3 s_0 s_1} \end{aligned} \quad (3)$$

Both, the positive and the inverted values, are required for each syndrome bit. Fig. 3 shows the structure commonly employed to implement the decoder circuits. As it can be seen, an additional inverter is required to generate the inverted value, provoking an increase in the logic depth. Nevertheless, other designs are possible. The use of a XOR-XNOR tree is proposed in our design, as shown in Fig. 4. This proposal reduces the logic depth while maintaining the number of logic gates, and its use can be generalized to any decoder circuit.

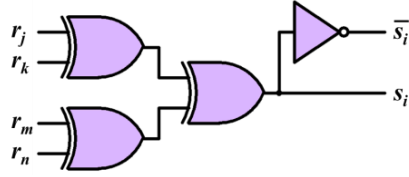


Fig. 3. XOR tree with additional inverter

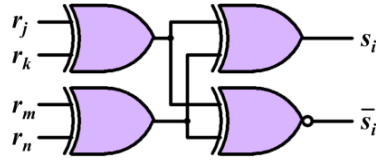


Fig. 4. XOR-XNOR tree

In this way, considering the use of XOR-XNOR trees, the decoder circuit for the (8, 4) Ultrafast SEC-DED code is presented in Fig. 6. The darker gates correspond to the detection logic added to the only-corrector circuit.

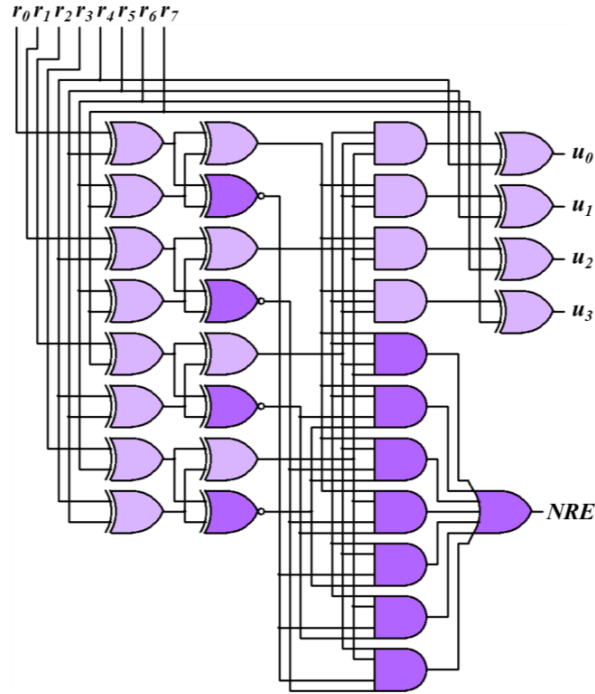


Fig. 6. Ultrafast SEC-DED (8, 4) decoder circuit

On the other hand, the encoder circuit is shown in Figure 7.

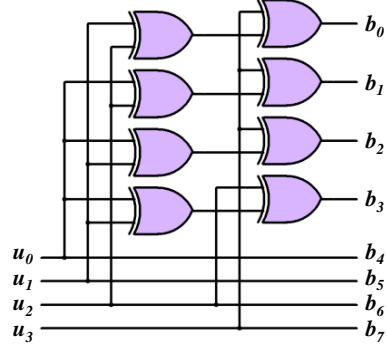


Fig. 7. Ultrafast SEC-DED (8, 4) encoder circuit

The 4-gate delay decoder (as its logic depth is four) shown in Fig. 6 can correct single errors and detect double errors. The correction is performed only in data bits, although it could be easily modified to correct additionally the parity bits, if necessary.

The main concern about these codes is that the detection logic grows exponentially with the word length. For a general $(2k, k)$ Ultrafast SEC-DED code, the number of 4-input AND gates required for detection is $2^{k-1}-1$. In addition, this is the number of inputs of the last OR gate. Depending on the implementation, it may require an OR tree, increasing the logic depth of the detection logic. Nevertheless, if this decoder is employed in the output of the register file in a pipelined processor with retry cycle or checkpoints for recovery, the data output can be employed assuming a right decoding if the NRE signal is processed before the affected instruction was committed.

5 Conclusion and Future Work

In this paper, a very fast Single Error Correction/Double Error Detection code for short data words has been presented. This code is based on Ultrafast SEC codes, a previous version of these codes that only allowed the correction of single errors.

This code has reduced the logic depth of the decoder circuit in order to diminish its delay. However, this reduction is accomplished by an increment of the redundancy. The added redundant bits allow us to find parity check matrices with a lower number of 1s in their rows and columns.

This code is an interesting option for designs where redundancy is not the main concern, data words are short, and fast encoding and decoding is important.

Microprocessor registers, for example, require fast encoding and decoding operations, as an excessive delay may affect the duration of the clock cycle. They use a small silicon area, so probably the spatial overhead introduced by an ECC in this case will not be significant. Another interesting target of the Error Correction Code pre-

sented in this paper can be the protection of high-speed memories or caches, if the redundancy introduced is affordable.

As part of the future work, we want to develop codes for longer data words by reducing the number of logical gates of the code presented here. Also, we want to model the encoder and decoder circuits using a hardware description language. The idea is to synthesize them in a library for VLSI integration. This will allow accurate area and delay measurements. Even more, the modeled circuits will be included in the model of a microprocessor to perform simulated fault injection and study the behavior of the system using this approach.

Furthermore, studying codes with additional error coverage while maintaining a low delay is also foreseen.

ACKNOWLEDGMENT

This work has been partially funded by the Universitat Politècnica de València under the Research Project DesTT (SP20120806), and the Spanish Government under the Research Project ARENES (TIN2012-38308-C02-01).

References

1. P. Shivakumar, M. Kistler, S. Keckler, D. Burger and L. Alvisi, "Modeling the Effect of Technology Trends on Soft Error Rate of Combinational Logic," in Proceedings 2002 International Conference on Dependable Systems and Networks (DSN-2002), pp. 389–402, Bethesda (Maryland-USA), June 2002.
2. D. Gil-Tomás, J. Gracia-Morán, J. C. Baraza-Calvo, L.J. Saiz-Adalid, and P.J. Gil-Vicente, "Studying the effects of intermittent faults on a microcontroller," *Microelectronics Reliability*, vol. 52, no. 11, pp. 2837–2846, November 2012.
3. E. Fujiwara, *Code Design for Dependable Systems*, John Wiley & Sons, 2006.
4. R.G. Gallager, *Low Density Parity Check Codes*, Cambridge, MA: M.I.T. Press, 1963.
5. C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: turbo-codes," in Proceedings ICC 1993, Geneva (Switzerland), pp. 1064–1070, May 1993.
6. L.J. Saiz-Adalid, P.J. Gil-Vicente, J.C. Ruiz, D. Gil-Tomás, J.C. Baraza, and J. Gracia-Morán, "Flexible Unequal Error Control Codes with Selectable Error Detection and Correction Levels," in Proceedings of International Conference on Computer Safety, Reliability and Security (SAFECOMP 2013), Toulouse (France), pp. 178-189, September 2013.
7. R.W. Hamming, "Error detecting and error correcting codes," *Bell System Technical Journal*, vol. 29, pp. 147 – 160, 1950.
8. L.J. Saiz-Adalid, P. Reviriego, P. Gil, S. Pontarelli, and J.A. Maestro, "MCU Tolerance in SRAMs Through Low-Redundancy Triple Adjacent Error Correction," *IEEE Transactions on Very Large Scale of Integration (VLSI) Systems*, vol. 23, no. 10, pp. 2332-2336, Oct. 2015.
9. J.A. Blome, S. Gupta, S. Feng, S. Mahlke and D. Bradley, "Cost-efficient soft error protection for embedded microprocessors," in International Conference on Compilers, Architec-

- ture and Synthesis for Embedded Systems (CASES 2006), Seoul (Republic of Korea), pp. 421–431, October 2006.
10. R. Baumann, “Soft errors in advanced computer systems,” *IEEE Design and Test of Computers*, vol. 22, no. 3, pp. 258–266, May-June 2005.
 11. E. Ibe, H. Taniguchi, Y. Yahagi, K. Shimbo and T. Toba, “Impact of scaling on neutron-induced soft error rate in SRAMs From a 250 nm to a 22 nm Design Rule”, *IEEE Trans. on Electron Devices*, vol. 57, no. 7, pp. 1527-1538, July 2010.
 12. S. Baeg, S. Wen and R. Wong, “SRAM Interleaving Distance Selection with a Soft Error Failure Model,” *IEEE Trans. on Nuclear Science*, vol.56, no.4, pp. 2111-2118, Aug. 2009.
 13. R.K. Lawrence, and A.T. Kelly, “Single Event Effect Induced Multiple-Cell Upsets in a Commercial 90nm CMOS Digital Technology,” *IEEE Trans. on Nuclear Science*, vol.55, no.6, pp. 3367-3374, Dec. 2008.
 14. D. Teixeira-Franco, J.F. Naviner and L. Naviner, “Yield and reliability issues in nanoelectronic technologies”, *Annals of Telecommunications*, vol. 61, issue 11-12, pp. 1422–1457, December 2006.
 15. L.J. Saiz-Adalid, P. Gil, J. Gracia-Morán, D. Gil-Tomás, J.C. Baraza-Calvo, “Ultrafast Single Error Correction Codes for Protecting Processor Registers”, 2015 11th European Dependable Computing Conference (EDCC 2015), pp. 144-154, France, Sept. 2015.

High Power Monocycle Pulse Transmitter for Ultra Wideband Radar Sensor

Younes Ahajjam^{1,2}, José M. Catalá-Civera², Felipe Peñaranda²
And Abdellah Driouach¹

¹ Dept of Physics, Faculty of Science, Abdelmalek Essaadi University 2121, Tetouan, Morocco.

² Instituto ITACA. Universitat Politècnica de València. Camino de Vera s/n. 46022 Valencia, España.

{Ahajjam Younes, youah@itaca.upv.es
José M Catalá-Civera, jmcatala@itaca.upv.es
Felipe Peñaranda Foix, fpenaran@itaca.upv.es
Abdellah Driouach, adriouach@hotmail.com}

Abstract. In this paper, a high power sub-nanosecond pulse transmitter for Ultra-wideband radar sensor is presented. The basis of the generator is considered as a step recovery diode and unique pulse forming the circuit, which gives rise to an ultra-wide band Gaussian pulse. Both transistor driver and transmission line pulse forming network are investigated to transform on one hand, a TTL trigger signal to a driving pulse with the timing and the amplitude parameters required by the SRD and on the other hand, to form an output monocycle pulse respectively. For a good range radar, a high amplitude pulse are indispensable, especially when penetrating thick Lossy material. In order to increase the output amplitude of the transmitter, the outputs of two identical pulse generators have been connected in parallel. This technique leads to a simple and useful solution.

The pulse transmitter circuit is completely fabricated using micro-strip structure technology characteristics. Waveforms of the generated monocycle pulses over 15V in amplitude have been obtained. Good agreement has been achieved between the simulation and the measurement results.

1 Introduction.

Recently, According to Ultra-wideband technology development, increased demands on wireless personal area networks as well as on high-resolution positioning facility is required. Therefore, attracted attentions by many researchers have been concentrated more and more on sub-nanosecond pulses [1, 3]; because Ultra-short pulses with well-defined output waveform plays a crucial role in the UWB transmitter design. A sub-nanosecond pulse transmitter is the fundamental part of any ultra-wideband (UWB) radar sensor [4]. The wide bandwidth of UWB pulses ensures fine range resolution and high penetration capability. Lower frequencies up to 3.5 Ghz are preferred for the application radar in lossy material, because losses of common construction materials go up to unacceptable levels in the higher frequency bands. The corresponding pulse width of the generated baseband pulses is around 200 ps. There are various techniques to generate UWB pulses in the order of nanosecond. The basis of a classical UWB short pulse generator is a driver, a pulse sharpener and a pulse forming network. The driver converts the step-like waveform in Gaussian pulses, and the pulse sharpener converts slow rise time waveform edge to a faster one. However the pulse forming network forms the output pulse to the adequate and favorite pattern: Gaussian or monocycle. Special semiconductor devices are usually utilized as pulse sharpener: avalanche transistor, bipolar diode [5], tunnel diode, Step Recovery Diode (SRD), and FETs [6]. Avalanche diodes are advantageous as high power sharpener but have the disadvantage in the limitation of the maximum pulse repetition frequency, because of the power dissipation in the transistor, Tunnel diodes offers the fastest transition time (sub-picosecond) at very low power (about mW). The SRDs are a compromise alternative for these devices and offer ultra-short transition time (around 100 ps) at moderate power levels (changing from hundreds of mW to tens watt) and with high repetition frequency. This makes them very appropriate to be used in current UWB generator. Normally, an SRD works as a charge controlled switch, which can alter from low impedance to high impedance state much quickly. This qualification of the SRD is used to sharpen the slow waveform edges. The more details of theory and application of SRDs as pulse sharpening is well described in [7].

In this paper, we present a new circuit design of an powerful UWB transmitter. The prototype circuit model has been manufactured using microstrip line technology. The basis of the generated circuit Gaussian pulse generator consists of a simple transistor driver and an SRD sharpener with a pulse-forming circuit. The aim of this paper is to generate high amplitude pulses capable to penetrate thick and lossy materials. To fulfill this task, the outputs of two identical pulse generators were connected in parallel. But, a bad connection of the output generated a ringing into the output waveform, then an additional ringing suppression techniques are necessary to overcome this problem. To resolve this problem, an accurate and complete assembling of two identical generator units on a single board with parallel outputs connected has been implemented using a mini-circuit "power splitter ZN2PD2". Results shows that, the output of the transmitter power has been increased and the peak power of summed pulses is nearly twice of the peak power of the single one. In

addition, adding a PFN, gives a monocycle pulses with 9.7 V in amplitude and 600 ps in total width. All results proves the high performance of the system

2 General Structure of the transmitter.

A block diagramme of the proposed monocycle pulse transmitter is shown in Fig.1. The transmitter consists of two identical Gaussian pulse generators and monocycle forming network (MFN). Both generators are synchronized and trigged by one timing source.

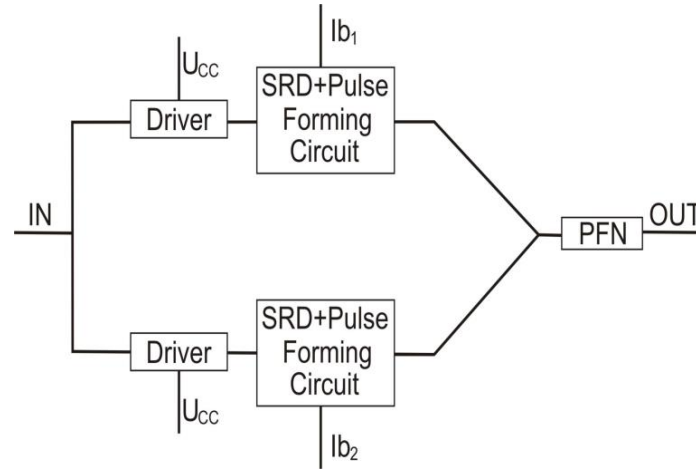


Fig. 1. Block diagram of the proposed radar transmitter

The main parts of the Gaussian pulse generator are a driver and a SRD pulser. The edge-triggered driver generates a well-defined pulse with sufficient power and speed to drive an SRD. This pulse is independent of the input waveform amplitude and duty-cycle. The pulse width is set to a few nanoseconds in order to minimize the current consumption of the circuit. The following stage of the pulse generator, the SRD pulser consists of two main parts. The purpose of an SRD pulse sharpener is to sharpen the leading falling edge of the driving waveform. The sharpened step-like pulse is then processed in a pulse-forming circuit to produce a Gaussian-like pulse. When no input driving pulse is present, the SRD is forward biased by an adjustable constant current source I_b . The outputs of two identical pulse generators are combined in order to obtain higher output pulse amplitude. The resulting Gaussian pulse is then converted to a monocycle pulse by an additional monocycle PFN. Monocycle pulses are of special interest, because their spectrum does not contain low frequency components and the PFN is simple to implement.

1.1 Step recovery diode pulse generator.

The circuit schematic of the pulse generator is shown in Fig 2. Two parts, the avalanche transistor circuit and SRD pulse sharpener circuit construct it.

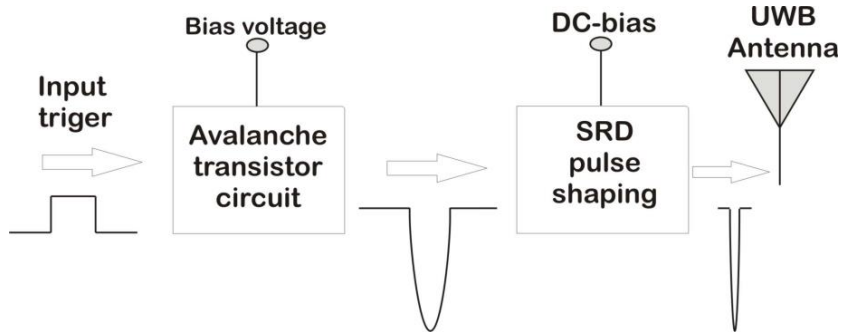


Fig. 2. Block diagram of the gaussian pulse generator

The pulse generator has been designed and simulated by the AWR Microwave office and Multisim transient simulator respectively [7]. The corresponding detailed circuit diagram of the system is shown in Fig.3.

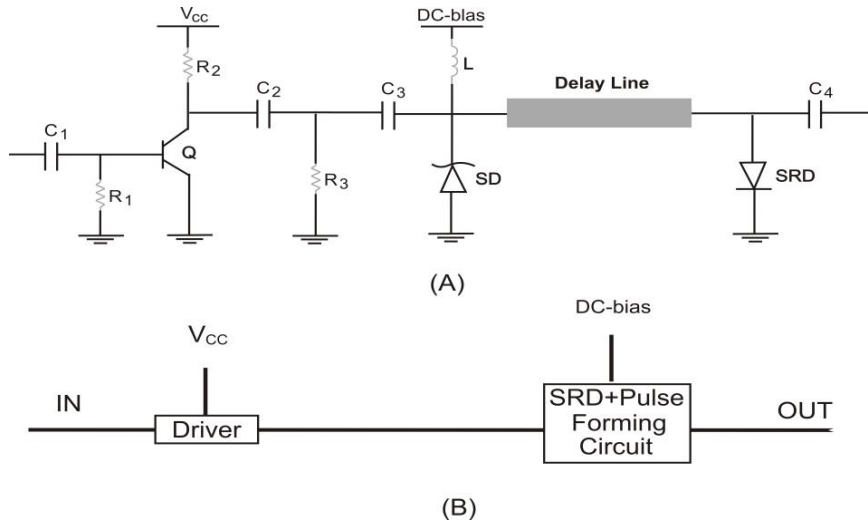


Fig. 3. Circuit (A) and diagram (B) of the Gaussian pulse generator.

The first part of the generator it's called Avalanche transistor circuit, which consists of Q, C₁, R₁, C₂ and R₂ as show in Fig 3. The transistor Q operates as common-

emitter switch, and generates a negative voltage pulse. The driving waveform passes via a coupling capacitor C_2 to the SRD pulser.

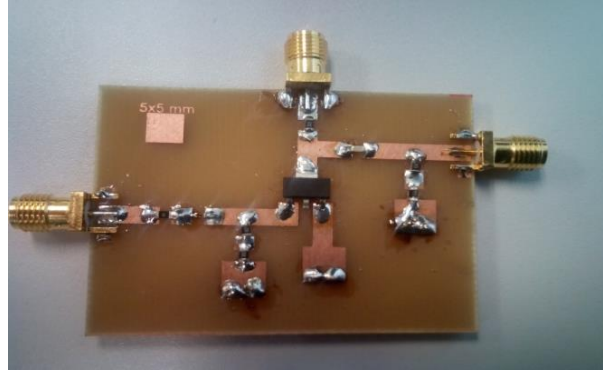


Fig. 4. Photo of Avalanche transistor circuit

As can be shown in Fig. 3. The discharge current of the avalanche transistor circuit flows from the capacitor C_2 through the avalanche transistor and resistor R_3 . The photo of the prototype circuit of the avalanche transistor is shown in Fig.4.

The second part of the generator is called “SRD pulse sharpening” which consists of a SRD sharpener and a Gaussian pulse forming network. The SRD, connected in parallel with a transition line, operates as a falling edge sharpener.

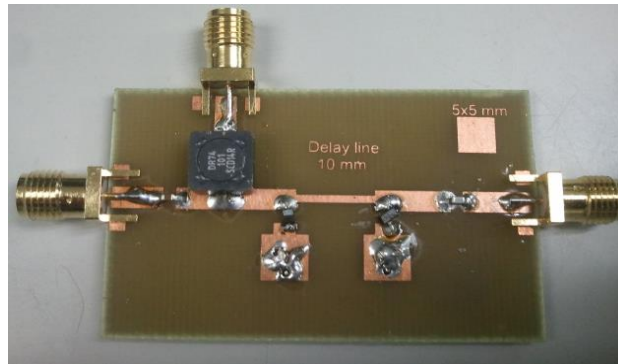


Fig. 5. SRD pulse sharpening Circuit

A photo of the manufactured circuit of the avalanche transistor is shown in Fig.5. The sharpener pulses are processed in a pulse-forming-network circuit, which consists of Schottky diodes and delay lines to produce a Gaussian-like pulse. During the steady

state (forward bias condition), a large amount of charge is injected into the diode, the device appears as a low impedance. After applying the negative driving pulse (reverse biasing condition), the diode continuous as low impedance until all charge is totally removed, at the point where the diode rapidly switches from the low a high impedance [4, 9]. This ability of the SRD to store charge and change its impedance level rapidly is used to sharpen the slow waveform edges. The pulse generator was implemented on an ARLON AD450 substrate of 0,7262 mm in thickness. Waveform was measured using an agilent 86100 C sampling Oscilloscope at 50 Ω load.

1.2 Combining UWB pulses.

For long-range detection or in measurement when penetrating thick lossy obstacles a high amplitude pulses are much desired. However, high power UWB pulse generators are expensive, and are in most cases not compatible with modern planar technology. To overcome this challenge, there are many ways to increase the output power. One of them is to combine the output waveforms from two or more sources. But, combining ultra-wideband Gaussian pulses is a difficult task. Conventional power combining structures, e.g. the Wilkinson power divider, are basically narrowband and distort UWB waveforms. Applying a short UWB pulse results in unwished reflections and ringing at other ports. Some dividers/combiners using planar multilayer techniques have also been described in the literature [5, 6]. Regrettably those dividers/combiners show proper UWB performance, but unfortunately are suitable primarily for the higher UWB band ranging between 3.1–10.6 GHz.

The goal in our design is to obtain a large amplitude waveform with a minimum insertion loss. This requirement is well achieved using a Mini-circuit called “power splitter ZN2PD2” which works in the band 500-5000 MHz. To test the whole functionality of the whole system, we collected two identical pulse generators on a single board. To obtain the sum of the pulses without any disturbance, it's important that both generators have to provide an output pulse exactly at the same time. The pulse amplitude is now over 10V compared with 4.8 V of a single Gaussian pulse generator. The pulse peak power has getting to 1.5 W, which is quasi the double of a stand-alone pulse generator.



Fig. 6. Output waveform of the proposed radar transmitter

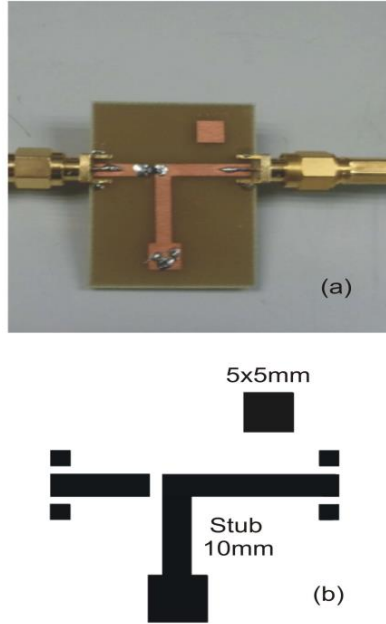


Fig. 7. (a) Circuit and (b) Layout of Monocycle forming network

Finally, a Monocycle Forming Network (MFN) inspired by [10] was connected to the output of the transmitter in order to form monocycle pulses. The photo of the fabricated circuit and layout are shown in Fig. 7. It consists of a shortened stub 10 mm in length and a Schottky diode (BAT15), which reduces output ringing. The impedance of all microstrip lines is fixed at $50\ \Omega$.

3 Conclusion.

In this paper, a new circuit solution of an UWB transmitter has been presented. In order to increase the output power of the transmitter, we assembled two identical generator units on a single board with parallel output connected. The simulation results show that, the peak power of summed pulses is nearly double of the peak power of the single one. With an additional output MFN, the monocycle pulses have a 9 V in amplitude and 600 ps in total width. All simulated and measurement results show very high performance of the designed system

References

- [1] Dieter Genschow; Johannes Kloas “ Evaluation of a UWB radar interface for low power radar sensors” Radar Conference (EuRAD), 2015 European, page (s) 321-324, 2015
- [2] Yu Guo; Guofu Zhu; Zhimin Zhou “ A novel design and implementation of sub-nanosecond smpling pulse generator for ultra-wideband equivalent sampling receiver”, Computing, Communication and Networking Technologies (ICCCNT), 2014 International Conference on page (s): 1-5
- [3] Ahmed abbas Hussein Ameri. “Long-Range Ultra-wideband Radar Sensor for Industrial Applications” Doctoral Diss, Kassel University, November 2012
- [4] Ahajjam Younes, Jose M Catala-Civera, Felipe Peñaranda Foix, Abdellah Driouach “Bibliography, Background and overview of UWB radar sensor”, Journal of engineeering Research and appliactions. Vol,4, Issue 11 (version-5), november 2014, pp. 37-40
- [5] -----,,” Design of ultra-wideband three-way arbitrary power dividers,” IEEE Trans. Microwave Therory Tech, vol. 56, no. 1 pp.194- 201, 2008
- [6] A.ouslimani, G, Vernet, H, Hardallah, and R, Adde, “ Large amplitude picosecond step generation with FETs,” Electrons Lett., Vol. 26, no. 19, pp. 1563-1564, 1990.
- [7] “Hewlett-Packard application note AN918: Pulse and waveform generation wih step recovery diodes,” Hewlett-Packard, Tech. Rep.,

The ITACA-WIICT'16 is a meeting forum for scientifics, technicians and other professionals who are dedicated to Information and communication technologies study and research. Its fundamental scope is to promote the contact among scientific and professionals, improving the cooperation as well as the technological transfer among professionals.

