# Accepted Manuscript

Power transmission and workload balancing policies in ehealth mobile cloud computing scenarios

Josué Pagán, Marina Zapater, José Ayala

PII: S0167-739X(17)30213-3
DOI: http://dx.doi.org/10.1016/j.future.2017.02.015
Reference: FUTURE 3333

To appear in: *Future Generation Computer Systems*

Received date : 20 June 2016
Revised date : 1 February 2017
Accepted date : 8 February 2017

Please cite this article as: J. Pagán, et al., Power transmission and workload balancing policies in ehealth mobile cloud computing scenarios, *Future Generation Computer Systems* (2017), http://dx.doi.org/10.1016/j.future.2017.02.015.

# Power Transmission and Workload Balancing Policies in eHealth Mobile Cloud Computing Scenarios
# Highlights

Josué Pagán Marina Zapater José L. Ayala

December 13, 2016

A real IoT eHealth scenario for ambulatory monitorization and prediction of the migraine disease is shown. In this scenario, energy efficiency techniques are applied in every level of the monitorization and prediction network. Low-power techniques are applied in monitorization nodes, and workload balancing policies are carried out in the coordinator nodes and Data Centers. The results quantifies the energetic and economic effects of the energy policies applied.

**Contact Author: Josué Pagán (jpagan@ucm.es)**

# Power Transmission and Workload Balancing Policies in eHealth Mobile Cloud Computing Scenarios

Josué Pagán[1,2,*], Marina Zapater[1,2], José Ayala[1]

**Abstract**

The Internet of Things (IoT) holds big promises for healthcare, especially in proactive personal eHealth. Prediction of symptomatic crises in chronic diseases in the IoT scenario leads to the deployment of ambulatory monitoring systems. These systems place a major concern in the amount of data to be processed and the intelligent management of the energy consumption. The huge amount of data generated for these systems require high computing capabilities only available in Data Centers. This paper presents a real case of prediction in the eHealth scenario, devoted to neurological disorders. The presented case study focuses on the migraine headache, a disease that affects around 15% of the European population. This paper extrapolates results from real data and simulations in a study where migraine patients are monitored using an unobtrusive Wireless Body Sensor Network. Low-power techniques are applied in monitorization nodes. Techniques such us: on-node signal processing and radio policies to make node's autonomy longer and save energy, have been applied. Workload balancing policies are carried out in the coordinator nodes and Data Centers to reduce the computational burden in these facilities and minimize its energy consumption. Our results draw average savings of € 288 million in this eHealth scenario applied only to 2% of European migraine sufferers; in addition to savings of € 1272 million due to the benefits of the migraine prediction.

*Keywords:*
WBSN, migraine prediction, energy optimization, Data Center, balancing workload, economic savings

## 1. Introduction

The Internet of Things (IoT) brings big opportunities for healthcare in unobtrusive monitoring scenarios for proactive personal eHealth [1]. Predictive models in the eHealth scenario using wearable monitoring devices have increased rapidly, from activity recognition [2] to event detection in neurological diseases [3]. These scenarios present big challenges in data acquisition and processing in ambulatory environments. The main goal of this paper is to study an energetically-efficient massive deployment of an ambulatory Wireless Body Sensor Network (WBSN) for the prediction of migraine events in European patients.

The migraine disease is one of the most widespread neurological disorders. Migraine is a socio-economic problem and degrades the quality of life of those who suffer from it (migraineurs). According to Stovner *et al.* [4], it affects almost 15% of the European population and leads to costs (direct and indirect costs) of €1222 per patient per year in Europe [5]. Currently available medications are not effective when the pain is very advanced. Hu *et al.* demonstrate in [6], that if the intake of specific treatments occurs before the pain starts, the pharmacokinetic time of these drugs is enough to abort the pain. Predicting the onset of a migraine will considerably reduce the patient's pain and thus the effects of migraine over the course of their lives. This will also lead to considerable economic savings over time.

At the onset of a migraine event, a sequence of neurological events occurs along with changes in some hemodynamic variables regulated by the autonomic nervous system (ANS) [7, 8]. These variables are, among others: surface skin temperature (TEMP), electrodermal activity (EDA), heart rate (HR), and peripheral capillary oxygen saturation (SpO2). In previous works we faced the problem of prediction of migraines [9, 10] through the ambulatory monitorization of the aforementioned hemodynamic variables.

According to our previous results and calculations in Appendix B, the prediction of migraines for a target of 2% of population of European migraineurs would save more than € 1200 million. To achieve this, we propose the deployment of an unobtrusive ambulatory monitorization scheme of this population through the use of a Wireless Body Sensor Network (WBSN).

In this paper we present the results for three main research goals: i) the energy efficiency for on-body channel transmissions in WBSNs, ii) the predictive modeling of migraines in a real scenario in the context of a clinical study, and iii) the optimization of energy consumption by means of workload balancing in Data Centers in an eHealth scenario.

The proposed scenario allows to scale our results and

*Corresponding author
*Email addresses:* jpagan@ucm.es (Josué Pagán), marina.zapater@ucm.es (Marina Zapater), jayala@ucm.es (José Ayala)
[1]DACYA, Complutense University of Madrid. Madrid 28040, Spain
[2]CCS - Center for Computational Simulation. Technical University of Madrid, Spain

policies for ambulatory prediction of migraines under energy constraints. Each subject is monitored using two sensing nodes wirelessly connected to a coordinator. A coordinator is a computation system with a higher performance than the monitorization nodes (usually a smartphone). The sensing nodes monitor the biomedical variables. The coordinator may perform the predictions or it can send the data to the Cloud and the predictions could be computed there.

The expenses of electricity bills of such deployment are high, as the Data Center is an infrastructure with a high energy consumption and many smaller power consumers (nodes) are also deployed. This eHealth scenario must be economically rewarding to have a smooth establishment in the market. For that purpose, we focus our policies for the energy efficiency at two different levels of the eHealth scenario:

- In the monitorization nodes: the battery consumption of the monitoring devices must be managed in order to increase their lifetime and autonomy. In this paper, this problem is tackled by optimizing the energy consumption of the monitoring devices: i) in the radio communication and, ii) in the computation of the gathered data. The energy consumption required for sampling signals has not be considered in the optimization, but it is detailed in Section 3.2.
- In the high performance computing systems: this 24-hour monitoring study generates vast amounts of data that must be managed in order to create prediction models for the patients. This process requires computing capacities only available at state of the art Data Centers. Once prediction models have been created, the prediction of an incoming event can be computed at two points of this IoT scenario: i) a coordinator node (assuming an intelligent monitoring device) or, ii) a Data Center. The energy consumption of these facilities leads to unsustainable costs. This work also optimizes the energy consumption by balancing the workload between the two points of the monitoring and prediction framework (the coordinator node and Data Center), in order to save money and achieve energy efficiency.

To fulfill the goal of this paper we consider that each part of the eHealth scenario can work in different modes. We describe three different modes corresponding to the amount of energy used for gathering and preprocessing data, and making the predictions. The first mode represents a baseline where data is gathered and transmitted in streaming without saving energy in the monitorization nodes. The second mode applies energy reduction techniques in the monitoring devices and energy aware off-loading techniques between the coordinator and the Data Center. The third one applies Data Center energy minimization policies on top of the previous off-loading techniques.

The rest of is paper is organized as follows: Section 2 discusses the state of the art of the energy efficiency in the eHealth scenario. Section 3 shows the experimental setup deployed for this paper. Section 4 details the experimental scenarios that lead to Section 5, where the results obtained for the goal of this paper are shown. Finally Section 6 draws the conclusions and future work. In addition, calculations about the sizing of the Data Centers that host the computing, and an estimation of the economic benefits for the national health systems are shown in Appendix A and Appendix B respectively.

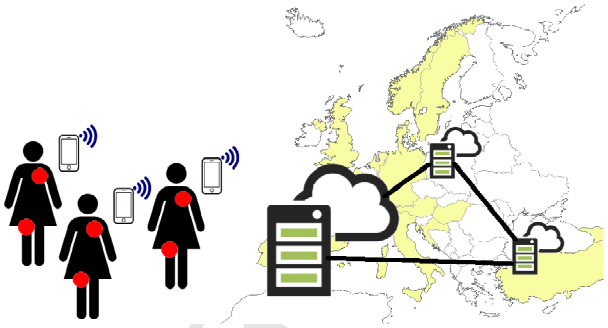

Figure 1: Scheme representing the structure of the system. Coordinator nodes (smartphones) communicate data from monitored migraine patients to a Data Center that belongs to a federation of Data Centers distributed in Europe. Shadowed regions are the countries included in the economic study.

## 2. Related work

As aforementioned, this work targets three main research goals that interact and build a complex scenario. Based on a real application of monitorization for migraine prediction, we study the energetic impact of every element in a Mobile Cloud Computing (MCC) scenario [11]: the WBSN (signal acquisition, signal processing, and radio link), the coordinator node, and the Data Center. Figure 1 depicts a scheme of the whole system. Each patient wears two monitoring devices and a coordinator node (a smartphone) that sends the biomedical data from the monitorization nodes to a Data Center. Data from patients of each country are sent to the Data Center in that country. As the figure represents, Data Centers are understood as a federation of Data Centers distributed in Europe. Shadowed regions in Figure 1 show the countries included in our economic study. In the following lines we describe briefly the state-of-the-art for all the scientific aspects that appear in this work.

### *2.1. The migraine prediction*

This paper focuses on the concrete use case of migraine prediction in a large-scale eHealth scenario. This problem was already tackled by the authors in previuos works as a initial contribution in this field [12]. In this paper we use real data and we base on the N4SID state-space models [13] developed in [9] to predict migraines. Results in our previous studies are currently limited up to 40 min of prediction of the migraine disease, with up to 90% of prediction accuracy and a low rate of false positives. The accuracy of this methodology is investigated in [9, 14].

### 2.2. Ambulatory monitoring systems

Large scale population monitoring systems in MCC scenarios are starting to become a reality, specially under the paradigm of Smart Cities. Many different and vague definitions of Smart Cities are described in the literature. According to the authors of the ranking of middle-sized smart European Cities reported for the Austrian Centre of Regional Science [15], a Smart City is understood as a certain ability of a city and not focusing on single aspects and certain characteristics are required for an evaluation. The authors of this study also indicate that the term is not used in a holistic way describing a city with certain attributes, but is used for various aspects, that in another study Deakin *et al.* [16] try to describe. It may said that Smart Cities are based on the application of information and communication technologies (ICT) through IoT, and they merge a set of services and resources in the urban environment to make citizen's life better. The services applied over the health care lead to the eHealth scenario that targets our research work. Training sport systems or healthcare ambulatory monitorization systems foresee a future where data and communication management—between the monitoring systems and the data bases and computing centers—will be crucial to reduce the energetic impact that these already cause in our days.

Among proposed Smart Cities services devoted to monitorization of daily activity and healthcare (e-Health systems), there are several research areas that exhibit a high interest for future applications. As an example, tracking prediction to optimize tactics in team sports [17], or mobile applications for sport training [18]. Further, we can find dozens of applications focused on the medical environment, such as: remote diagnosis, disease alarms generation [19], prediction of atrial fibrillation [20], or arrhythmia detection using mobile devices [21].

In the market, during the last years, dozens of companies have developed a large number of monitoring devices. Most of them are not oriented to the science but mostly to the personal wellbeing. They usually incorporate movement sensors and rarely others such as oxygen saturation, electrodermal activity or skin surface temperature. The accuracy of these devices—generally presented as wristbands—is not enough for medical purposes in a continuous ambulatory study.

Different medical bio-monitors can be found in the state-of-the-art, such as the IMEC's devices *Smart ECG necklace* for ambulatory cardiac monitoring, or the Health Patch[3] developed in collaboration between the Holst Centre[4] and SHINKO[5]. Another example are the medical devices for rehabilitation and muscle strain diagnosis developed by TMG[6].

Wristband Empatica E4[7] is a commercial device that, unlike Surge (FitBit[8]), also commercial, is thought for scientific usage. Empatica E4 integrates several sensors, but the accuracy problem in the wristband format makes it not useful for our work. Due to this is an research study, we are going to focus on the less portable but well known Shimmer devices[9]. These devices allows wireless communication and control of the whole acquisition system to enlarge the battery life. Shimmer nodes are thought for non-invasive biomedical research [22]. Two Shimmer nodes are used to monitor the hemodynamic variables aforementioned in Section 1.

### 2.3. Energy efficiency in Wireless Sensor Networks

Studies such as the presented by Casino [23] show the importance of the estimation of the wireless system operation in the radio link to reduce the energy consumption. In our research, we analyze the energetic impact between the extremes of the radio link, *i.e.*, from the Wireless Body Area Network (WBAN) to the Data Center, as the decisions in one extreme affect to the other.

The energy efficiency in WBANs has been studied in the literature to reduce the battery consumption and enlarge the availability of the monitoring devices. We can distinguish different levels to achieve energy efficiency in WBANs:

- Hardware: low power consumption electronic components.
- Coding: efficient codes, bare metal programming, compiler optimization, *etc*.
- Wireless communication: efficient technologies, optimized spectrum utilization.
- System: modular on/off control of electronic components
- Signal processing: computation on the device such as pre-processing, feature extraction, data removal, *etc*.
- Workload: high level decisions that can include some of the previous levels.

The aim of this work is to reduce the overall energy consumption in the eHealth application, from a holistic perspective. To this end, in this paper focus on: i) minimizing the energy of the monitoring nodes via reducing the radio link power, ii) developing strategies to minimize the impact of signal processing ever the whole system, iii) developing workload off-loading techniques to minimize energy consumption at the Data Center, and iv) developing specific strategies to minimize energy consumption in the Data Center.

#### 2.3.1. Energy efficiency in the radio link

The wireless communication between nodes in a WBAN—typically with a star topology—is called an on-body channel communication. The blocking of the direct line of sight between the sensing nodes and the receiver due to the movement of users dramatically affects the link quality [24,

[3] IMEC devices: http://www2.imec.be/be_en/research/wearable-health-monitoring.html
[4] Holst Centre: https://www.holstcentre.com
[5] SHINKO: http://www.shinko.co.jp/english/index.html
[6] TMG: http://www.tmg-bodyevolution.com/medical
[7] Empatica: https://www.empatica.com
[8] FitBit: https://www.fitbit.com
[9] Shimmer: http://www.shimmersensing.com

25]. Given the dynamic and complex nature of the on-body channels, the challenge is to maintain a good quality of the link between sensor nodes while extending the network lifetime. This leads to the development of power transmission control policies to avoid data loss.

In the context of non-invasive WBSNs both, the human body and the node placement, affect the signal propagation, so that the radio network coverage can be reduced up to 1 meter. In the literature, different authors like Augustine [26], or Smith *et al.* [27] have shown that the quality of the on-body links in a non-invasive WBSN are highly dependent on the body posture, movement, the physical characteristics of the human body and the local environment. The movements of the user causes changes in the quality of the links, affecting their reliability [28, 29]. In this sense, some studies try to detect the body position to select the power transmission, such as in the work by Aulery *et al.* [30].

For this purpose, in this paper we use the techniques developed in our previous work [31] to minimize the radio link power without quality loss. These techniques are reactive policies for power level control based on algorithms using body posture detection. Our techniques are trained with information of the Received Signal Strength Indicator (RSSI) level of the communication between the nodes and the coordinator, and the body posture. This allows to change automatically the power transmission to save battery.

In this work, we combine our previously developed techniques with other simple but effective strategies to reduce radio power consumption. In this sense, we apply strategies to minimize the number of transmissions, managing the trade-offs between reducing the radio link power and increasing the processor power—due to the increase of signal processing.

#### *2.3.2. Energy efficiency through computation in the node*

Several works apply some data processing techniques in the node to achieve energy efficiency applied in the WBSN level. As an example, Mamaganian *et al.* [32] apply compress sensing techniques for low-complexity energy-efficient ECG compression, to reduce the amount of data to transmit and save energy. A low-power biosignal acquisition and classification system for ECG signal is proposed by Lee *et al.* [33].

Tobola *et al.* [34] analyze the power consumption of medical monitoring devices at system level; the same author also studies the impact in the battery life at a signal processing level in [35] by reducing the sampling rate as much as possible, and studying how it affects to state-of-art signal processing algorithms in WBSN.

In this paper we apply an ECG signal processing algorithm to assess the impact on the global energy reduction. We based on the implementation of Rincon *et al.* [36] for the ECG delineation using their single-lead optimization for the Shimmer node. For further implementation details, we encourage the reader to refer to the original work in [36]. The surface skin temperature and the electrodermal activity values do not need to be processed, as well as the SpO2 data that are gathered directly from the OEM-III module (see Section 3.2.1).

### *2.4. Data processing in the Cloud*

Data obtained using the WBSN is communicated to an embedded processing element, *i.e.* a coordinator (usually a smartphone), that forwards the data to the Cloud. In order to predict migraine, huge data sets must be analyzed. To deal efficiently with such computationally intensive tasks, part of the processing and storage will be local to the coordinator, while another part will be communicated and processed in the Cloud, *i.e.* in Data Centers—the extreme end of the network. As the target population is so large, the key challenge in MCC scenario is the definition of strategies to off-load and distribute efficiently the workload between the different elements of the system.

In this regard, the goal of our work is to balance efficiently the workload, reducing the computational burden in the Data Center and minimizing its energy consumption. Data Centers infrastructures consume a huge amount of power and generate a tremendous amount of heat. In year 2010, these facilities represented 1.3% of electricity use in the world [37]. In year 2012 alone, global Data Center power consumption increased to 38GW, and further rise of 17% to 43GW was estimated in 2013 [38]. Data Center power budget is mainly devoted to the energy drawn by servers and the cooling needed to keep IT equipment under safe environmental conditions. In the last years, industry has devoted significant effort to decrease the cooling power, thus decreasing the Data Center Power Usage Effectiveness (PUE)—defined as the ratio between total facility power and IT power. According to a report by the Uptime Institute, average PUE improved from 2.5 in 2007 to 1.89 in 2012, reaching 1.65 in 2013 [39]. PUE values close to 1 are preferred. Moreover, academia has also focused on the development of Data Center optimization strategies to minimize energy from the computational and cooling perspective. However, in a MCC scenario, efficient computation off-loading, combined with tailored strategies to minimize the energy consumption at the Data Center, still represents an important challenge.

Our work devises new strategies to combine these two aspects: i) workload off-loading and ii) Data Center energy efficiency techniques, to further reduce the energy consumption of eHealth applications. This work presents a realistic case study for migraine prediction and shows the overall impact of this strategies over the system.

In the following section, we describe each aspect of the experimental setup relative to every major research goals: the migraine prediction (Section 3.1), the energy efficiency in WBSNs (Section 3.2), and the energy efficiency in Data Centers (Section 3.3).

## 3. Experimental setup

The migraine prediction study starts in the hospital with the patient, the doctors, and the engineers. After a usability learning process, the monitorization period starts. The study comprises two phases: i) the training, or offline phase, and ii) the real-time prediction, or online phase (see Figures 2 and 3).

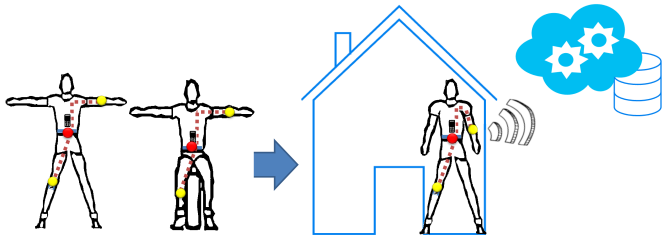

Figure 2: Sensor placement and monitorization. Two sensors are wirelessly connected in a star topology to a coordinator (a smartphone in the waist). The sensor node S1, placed in the right arm, communicates with the coordinator via the L1 link. The sensor node S2, in the knee, connects via the L2 link. On left side of the figure, the patient performs a sequence of movements to calculate the on-body channel transmission parameters. On right side, data are transmitted wirelessly to the Data Center.

In the offline phase, and still in the clinic, the study of the on-body transmission parameters is carried out. The offline phase of the study lasts between 2 weeks and one month [9]. During this time the patient's migraines are recorded and the prediction models are developed. In the online phase, patients are continuously monitored and a runtime prediction of the migraine is performed. As we show in following sections, each of these two phases has a different impact on the global consumption of the application.

In this section we go through the details of all the elements of the network. Moreover, the most important characteristics of the models—on-body channel transmission, migraine prediction and workload balancing—are shown.

### *3.1. Overview of the experiments*

Figure 3 summarizes the experiments undertaken and described in the following sections.

During the training or offline phase, in the hospital, the sensing nodes are placed on patients' body describing a star topology. The coordinator (a smartphone) is placed in the waist (just over the navel), the sensor node S1 in the right arm (link L1) and the sensor node S2 in the right knee (link L2). The sensors are wirelessly connected to the coordinator. Sensor node S1 senses the ECG signal on the chest using one derivation, EDA in the arm using two electrodes, and TEMP near the armpit using an NTC thermistor. Sensor node S2 senses the SpO2 in the capillarity zone near the groin using the 8000R SpO2 sensor and the OEM-III module (both from NONIN[10]).

#### *3.1.1. On-body channel transmission parameters*

Still in the hospital, the on-body channel transmission parameters for each patient are calculated. These parameters are required to feed the transmission models and apply the radio techniques in [31] to reduce the energy consumption.

To develop the models some physiological measurements of each patient are required, such as: the arm circumference, body fat mass, bone mass, muscle mass, *etc*. [40]. Based on our previous work, a proactive technique to control the transmission power in the communication between the nodes and the coordinator is carried out [41]. This proactive technique adapts dynamically the transmission power based on the variations in the RSSI index. These models are patient-dependent and use an ANFIS Link Quality Estimator (A-LQE) model—based on ANFIS networks—to estimate the quality of the radio links. RSSI levels and minimum transmission power affordable are related through values stored in a Look Up Table (LUT). To calculate the parameters of the models, the patients have to perform a sequence of movements standing up and sitting down [40].

The exercises are a sequence of movements that simultaneously combine different positions $P_i$, $i = 1,2,...,N$, of arm and knee. The number of exercises $N = 5$ for sequence 1 of movements, and $N = 4$ for sequence 2. The positions of arm and knee are named as: L1P1+L2P4, L1/P2+L2/P3, L1/P3+L2/P1, L1/P4+L2/P2 and L1P5+L2P1. For each link, the size of the LUT in the corresponding node matches the number of exercises considered in that link. Each one of the positions is described below:

- Sequence 1: the subject sits on a chair and performs five movements of the arms (Link 1, L1): i) hands on thighs, denoted as L1/P1; ii) arms crossed, L1/P2; iii) arms extended forward, L1/P3; iv) arms extended up, L1/P4; and v) arms extended to both sides, L1/P5.
- Sequence 2: the subject sits on a chair performs four movements of the legs (Link 2, L2): 1) leg in 90º angle with the body, L2/P1; 2) left leg crossed over the right knee, L2/P2; 3) right leg crossed over left knee, L2/P3; and 4) leg extended forward, L2/P4.

#### *3.1.2. Ambulatory monitorization: the training phase*

In the following lines we describe the training phase of the predictive modeling.

The ambulatory monitorization starts when the patient leaves the hospital. During a period that comprises from two weeks to one month approximately (day D in Figure 3), data are transmitted by the coordinator to the Data Center via a 3G link. In this training phase, the patients are inquired to record in a smartphone (the coordinator) their subjective pain. Using these data points, a normalized curve is generated through the adjustment of two semi-Gaussian curves [9].

In the *pre-processing* block in Figure 3, after HR calculation from the ECG signal, data from all sensors are synchronized to each other. After synchronization, data losses are regenerated using a Gaussian Process Machine Learning (GPML[11]) [42]. This GPML process fits the data with an average precision of 75.4% for HR, 85.7% for TEMP, 93.2% for EDA and 73.4% for SpO2.

When a migraine is recorded, a migraine model is generated using the previous data and the data during the migraine. Due to the high computational burden of these calculations when done

[10]NONIN: `http://www.nonin.com`

[11]GPMLTool: `http://gaussianprocess.org/gpml/code`

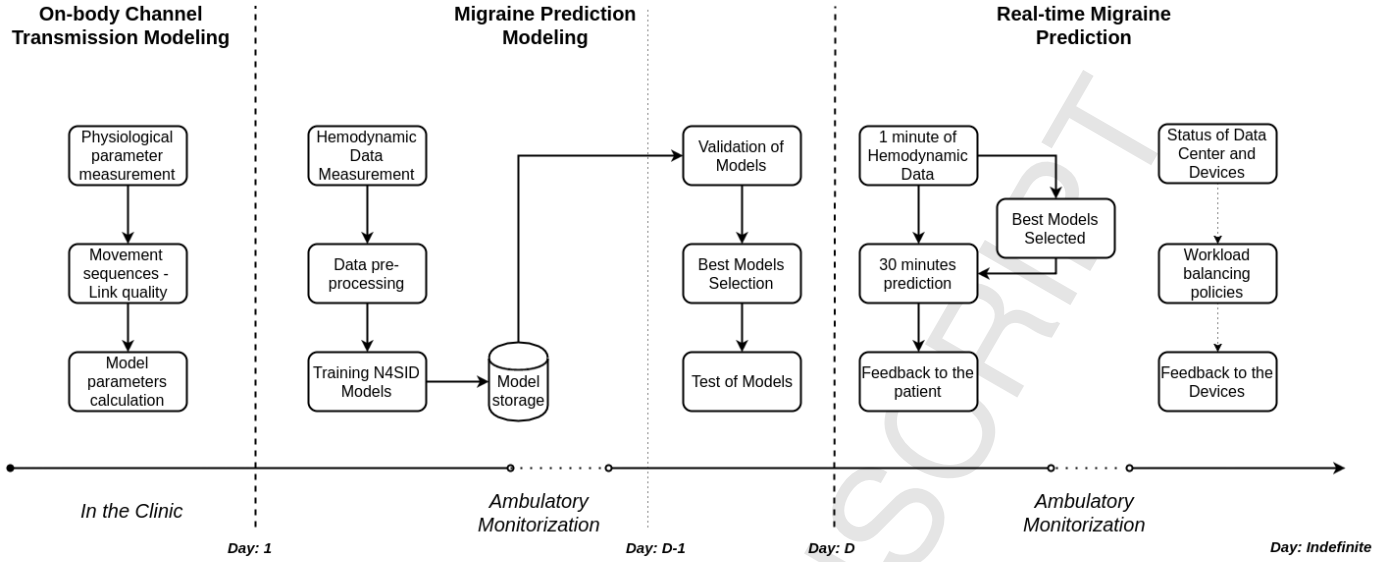


Figure 3: Basic scheme of the monitorization study. At the beginning of the study the parameters of the models for the on-body channel transmission are calculated in the hospital. After that, the study continues: first the migraine prediction models are calculated. For the remaining of the study the ambulatory monitorization will continue. Data are always transmitted from the two monitoring nodes to the coordinator, and then to the Data Center.

for a large population, these models need to be generated in a Data Center.

All models we use in this paper were created for a prediction horizon of 30 minutes. A simple feature selection analysis was undertaken and different models were generated for each combination of three hemodynamic variables. Based on our previous work, five models are stored for each combination of sensors. Changing between models according to the availability of sensors or quality of the gathered signals makes the system more robust (see Appendix B). Data from two female migraine patients (Patient A and Patient B) are used to train and test the models. 15 migraines are used to train and select the best models for Patient A; 8 for Patient B. Test is performed with 5 and 4 migraines for Patient A and Patient B. The accuracy (and selectivity too) depends on the set of sensors used. The best case is different for each patient. Predictivity is always 100% according to our previous results, what means that the number of false positives is zero. The True Positive Rate (TPR) ranges from 67% to 90% for Patient A and patient B respectively, leading to F-scores (accuracy) of 80% and 95% respectively.

When the number of migraines is large enough, the ambulatory monitorization finishes. During day D-1 (Figure 3) models are validated and tested, and the more accurate ones are selected. The final prediction is the average of the prediction of several models in order to avoid overfitting or loss of accuracy.

The N4SID algorithm is a state-space model. These models define immeasurable states to describe difference equations that calculate the current and future outputs from past and current inputs. N4SID is defined as a combined deterministic-stochastic model with the following parameters in Eq. 1:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k \\ y_k &= Cx_k + Du_k + v_k \end{aligned} \tag{1}$$

$u_k$ are the $U = 4$ inputs representing the sensors and $y_k$ is the predicted value of pain at time $k$. $A$ is the state transition matrix, $B$ relates the states at time $k$ ($x_k$) with the inputs, $C$ is the state to output matrix, and $D$ equals zero in our case. $v_k$ and $w_k$ are immeasurable white noises. The N4SID order $nx$ (size of the square matrix $A$) has a maximum size of 10 in our implementation.

The training, validation and test codes have been developed using the System Identification Toolbox of the MATLAB software [43]. To increase the speed-up of our algorithms and reduce the computational burden in the servers of the Data Center, we use the MATLAB compiled version (from MATLAB Compiler for standalone applications).

### *3.1.3. Migraine prediction: the online phase*

During online phase the ambulatory monitorization system is still used, but now the patients do not have to record their pain symptoms. Each minute, data from all sensors are sent to the coordinator. Data can be preprocessed in the sensing nodes, in the coordinator or even in the Data Center. Online prediction is much less computationally intensive than the offline phase, thus, the coordinator (*i.e.*, the smartphone) can either perform the prediction itself, or forward the computation to the Data Center. Each prediction is calculated 30 minutes forward and announced to the patient. The output of the prediction is the probability of pain occurrence.

Various scenarios representing the most representative combinations of the working modes of each element are presented in Section 4. Moreover, we prove later how the most efficient solutions consist on off-loading the computation to the sensing nodes.

The N4SID matrices calculated in the training phase have been implemented in C-code to run them in the servers of the Data Center and in the coordinator node. In parallel to the prediction calculation, the workload balancing policies are

carried out, as can be seen in the right side of Figure 3. Depending on the number of patients being monitored, the computational burden of the prediction, the Data Center utilization, and the battery status of the coordinator nodes, the off-loading policy decides during runtime where the prediction takes place: in the Data Center or in the coordinator node.

### *3.2. The WBSN*

As aforementioned, the WBSN is composed by two Shimmer nodes and the coordinator (a smartphone) wirelessly connected in a star topology. Two different working modes for the sensor-coordinator communication are analyzed: i) streaming mode—where data are gathered and transmitted without preprocessing, and ii) preprocessing—where data is preprocessed in the nodes before being transmitted.

#### *3.2.1. The sensor nodes*

The Shimmer devices are based on the MSP430F1611 16-bit microcontroller with a maximum frequency of 8 MHz. The RAM size of the microcontroller is 10 kB, and 48 kB the flash memory. The CC2420 chip radio performs the radio interface, and implements the IEEE 802.15.4 radio standard. Shimmer uses the TinyOS[12] operating system for embedded devices. However, FreeRTOS[13] has been used instead. FreeRTOS is portable, open source, and it has a hard real-time mini kernel that includes support for the microcontroller and the IEEE 802.15.4-compliant radio chip used by Shimmer.

Table 1 shows the configuration of the two Shimmer nodes and summarizes the amount of gathered data per day. In case of preprocessing in node S1, the HR is calculated every 1.25 seconds using 20 seconds of ECG data. This rate ensures the detection of peaks at 48 bmp for normal resting rates described in [44]. In case of using node S2 for data processing, SpO2 is only computed once a second, according to the OEM-III module bitrate.

The power consumption required for the sampling of the signal has not been considered in the optimization as it cannot be modified in some of the sensors, or it would require the application of compress sensing techniques. As shown in Table 1, the sampling rate of the TEMP and EDA sensors is negligible. The commercial digital OEM III module bitrate, 3 kbps, cannot be modified by software. The sampling of the ECG sensor is 250 Hz. This is the standard quality level in most of the commercial ECG monitoring devices, thus we do not contemplate a reduction of the ECG sampling frequency. It could be possible if we apply compressed sensing techniques, but that is out of the scope of our research.

Battery discharge or sensor disconnections occur frequently. Table 2 shows empirical results of the probabilities of disconnection or unavailability for our sensors. Probabilities have been computed from 8 migraines lasting 7.7 hours on average. TEMP, EDA and HR sensors have the same probability of disconnection, 6.2%. This is due to the battery discharge rate of node S1. However, the SpO2 sensor has a higher probability of disconnection, 29.7%, mainly due to the misplacement of the sensor in the groin. On average, almost 1 of 3 packets of SpO2 are lost. Note that whenever a packet is lost, we run a the GPML algorithm as part of the data preprocessing to recover the signal and be able to obtain a prediction. Thus, data loss increases the computational burden, and needs to be considered during runtime prediction.

Table 1: The acquisition parameters in the Shimmer nodes. Processed data information shown in parenthesis.

| | Placement | Sampling rate (Hz) | Bits | Precision | Data-24h (kB) |
|---|---|---|---|---|---|
| TEMP* | Armpit | 1/60 | 12 | 0.0223 °C | 2.1 |
| EDA* | Arm | 1/60 | 12 | 0.0062 $\mu S$ | 2.1 |
| ECG (HR) | Breast | 250 (0.8) | 12 (8) | 4 ms (1 bpm) | 31640.6 (67.5) |
| PPG (SpO2) | Groin | 75** (1***) | 16 (8) | 50.5 $\mu V$(1%) | 31640.6 (84.4) |
| Total (MB) | | | | | 61.8 (0.15) |

* EDA and TEMP are considered negligible for energy consumption.
** Only PPG is needed but OEM III module captures and sends 375 Bytes.
*** SpO2 is internally calculated by the OEM-III.

Table 2: Probability of sensor data loss.

| | TEMP | EDA | HR | SpO2 |
|---|---|---|---|---|
| # breaks/hour | 0.24 | 0.24 | 0.24 | 0.57 |
| Average duration (min/break) | 15.5 | 15.5 | 15.5 | 31.3 |
| Probability (%) | 6.2 | 6.2 | 6.2 | 29.7 |

Regarding to the radio, for the 802.15.4 standard, a 104 Bytes data payload has been used, with 250 kbps transmission rate. One transmission packet (including headers) is 128 Bytes long. Table 3 shows the transmission rate $TX_r$, the amount of data sent $D$ (including headers), the transmission time cost $TX_t$, and the duty cycle (the ratio of transmission time to transmission rate) for nodes S1 and S2 in each working mode.

Figures 4a through 4d complement values in Table 3. These figures are an interpretation of the energy levels of the tasks and their timestamps. In the streaming working mode for node S1 (in Figure 4a), when 104 Bytes are recorded (69 samples of ECG), the acquisition stops and data flow from the buffer in the microcontroller to the buffer in the CC2420 chip to be transmitted. Thus, every 277 ms the radio wakes up and transmits. During the acquisition time, the radio is in low power mode. The transmission takes 4.1 ms, as Table 3 shows. Every 60 seconds, when ECG is processed to reduce the amount of transmitted data, 48 samples (1 Byte each) of HR are sent.

Regarding the working modes in node S2 (Figures 4b and 4d), we can see some peculiarities. The data to be sent are always larger than the radio buffer size of 128 Bytes. With this, data have to be truncated and sent in burst mode. 375 Bytes of data have to be sent in the streaming

Table 3: Data transmission properties for Shimmer nodes.

| | Streaming mode | | Processing mode | |
|---|---|---|---|---|
| | Node S1 | Node S2 | Node S1 | Node S2 |
| $TX_r$ (ms) | 277 | 1000 | 60000 | 60000 |
| D (Bytes) | 128 | 471 | 72 | 84 |
| $TX_t$ (ms) | 4.1 | 15.1 | 2.3 | 2.7 |
| Duty cycle (%) | 1.48 | 1.51 | 0.004 | 0.005 |

[12] TinyOS: `http://tinyos.stanford.edu/tinyos-wiki/index.php`
[13] FreeRTOS: `http://www.freertos.org`

mode; this leads to transmit 3 packets of 104 Bytes of payload and an extra packet of 63 Bytes. Thus, 4 headers of 24 Bytes are sent. In this way, we achieve the 471 Bytes in Table 3. The radio wakes up once a second and sends the data in 15.1 ms. In the second working mode, this node extract only the SpO2 information once a second, and transmits data every minute. The SpO2 data extracted are 60 Bytes (one Byte per second) and 24 Bytes of header are added, leading to the 84 Bytes that appear in Table 3.

The beacon time of the communication protocol and the packet transmission time from the microcontroller to the radio chip in Figures 4a through 4d are based on previous work by Rincon *et al.* [36].

### *3.2.2. The coordinator node*

The coordinator node in the real experiments is an smartphone. To simplify the energy characterization and isolate the contribution of the smartphone's processor from other components, we perform power characterization in a BeagleBone Black platform[14]. This platform uses the same processor as the Samsung Galaxy S smartphone, and can be easily instrumented to measure power. The processor is an ARM Cortex-A8 at 1 GHz with 512 MB DDR3 RAM. Data are stored in an external 2 GB SD memory card. In order to isolate the energy measurements, as radio device, the same CC2420 as in the Shimmer devices is supposed to be used in the smartphone, and this interface has been attached to the BeagleBone platform. As we envision a case study to apply energy-aware off-loading policies that balance computation between the elements in the WBAN and the Data Center, the coordinator node—in addition to transmitting an receiving data—may perform two main actions: i) data preprocessing (off-loading computation from sensors), and ii) online prediction (off-loading computation from the Data Center).

### *3.2.3. Energy costs of the WBSN*

The energy consumption of the microcontroller and the external devices has been measured using a high precision digital amperimeter, whereas, the consumption of the radio has been simulated, due to the complexity in measurement of energy consumption in the real scenario. Simulation has been carried out with the open source simulator Castalia[15], which is designed specifically for WBSN networks and it includes a channel model based on data measured empirically, and a model of radio CC2420, used for our experiments.

In order to guarantee reliability, most approaches transmit at a fixed transmission power that ensures no packet loss. However, these techniques lead to a waste of energy. Thus, transmission reactive policies based on knowledge of body posture have been implemented. These accelerometry-based transmission power control policies allow to regulate the transmission power levels according to changing link conditions, and minimize the power consumption of the radio chip, while maintaining service quality. As aforementioned, the algorithm requires the characterization of the patient in terms of the RSSI and the Packet Error Rate (PER) metrics with respect to positions in Section 3.1. Once the communication link has been characterized properly, the calculation of the transmit power optimal level is done offline using the experimental data, and as result, a LUT with the best power levels for each body position is generated. Controlling the transmission power is done online; the optimum level of power required to ensure the quality of the link is dynamically selected from the values stored in the LUT by using motion detection—based on accelerometry with low complexity and low overhead.

It is assumed for our experiments that the radio switches on for every transmission, going off after finishing. Simulations have been executed at three different power transmission levels available in the radio CC2420 chip: $-15dBm$, $-10dBm$ and $0dBm$; and then the reactive policies have been applied. A re-transmission rate has been simulated as random re-transmissions along the time, simulating the channel effects of medium access, using a 802.15.4 MAC with two transmission attempts. The simulation adds a model with temporal variation to recreate the dynamics of the path-loss fluctuations and without collisions.

The configuration setup of the general parameters in simulation has been as follow:

- The energy consumption model: the power consumption in Castalia has two components i) radio consumption and ii) baseline consumption. The default baseline consumption value has been modified, thus, the energy consumption of a mote when the radio is off and the microcontroller is active, will be 0 mW. In this way, the baseline value does not affect the real values of energy consumption measured due to processing of the microcontroller. Finally, only the radio consumption is taken into account.
- Duty Cycle: differences in energy occur because the radio is active for different periods of time in each node. The parameters, beacon order (BO) and superframe order (SO), define the duty cycle between the active and inactive periods. In Castalia, BO equals six and SO equals four, while a duty cycle of 25% is the default value for MAC protocol 802.15.4.
- Collision Model: in Castalia, the radio collision model is configured according to the *InterfModel* parameter, which can take three different levels: Level 0, 1 and 2. In Level 0, the simulator assumes no collision at all. We have used for this study Level 0.

Table 4 summarizes the configuration parameters of in Castalia for sensor and coordinator nodes in the two working modes. Notice that the values of the duty cycle in Table 3 and in simulation are different. This is due to the minimum duty cycle as parameter for simulation is 1.56%. For the simulation of the nodes working in streaming mode, this will lead to be an overestimation of the consumption.

[14] Beagle: `http://beagleboard.org/BLACK`

[15] Castalia: `https://castalia.forge.nicta.com.au/index.php/en`

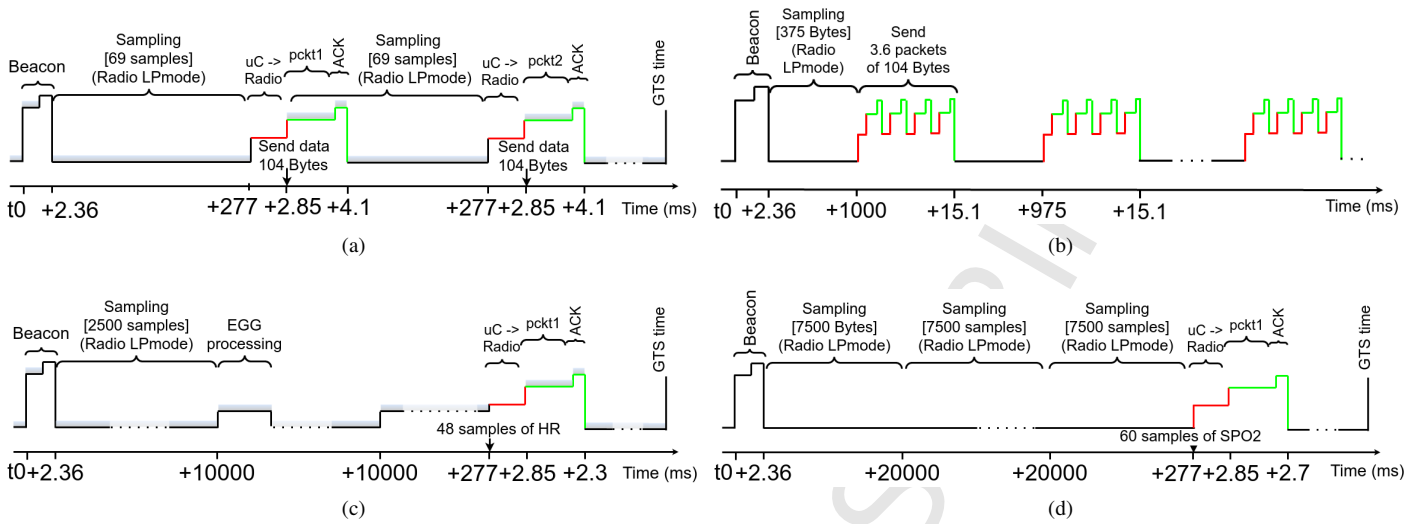


Figure 4: Diagram of the streaming and processing working modes in the two sensor nodes. The y-axis (not in scale) of these figures represent the energy level of each task. The x-axis (not in scale) marks the most important events. (**a**) ECG streaming; (**b**) All NONIN data packet streaming; (**c**) HR processed; (**d**) SpO2 signal extracted. Shaded areas in Figures 4a and 4c represent the data acquisition.

Table 4: Parameters in the Castalia radio simulator for the Shimmer nodes and the coordinator.

| | Streaming mode | | | Processing mode | | |
|---|---|---|---|---|---|---|
| | Node S1 | Node S2 | Coordinator | Node S1 | Node S2 | Coordinator |
| $Pkt_r$ (pkt/s) | 3.610 | 3.596 | 1 | 1/60 | 1/60 | 1 |
| Payload (Bytes) | 104 | 104 | 20 | 48 | 60 | 20 |
| Headers (Bytes) | 18 | 18 | 18 | 18 | 18 | 18 |
| MAC (Bytes) | 6 | 6 | 6 | 6 | 6 | 6 |
| Duty cycle (%) | 1.56% | 1.56% | 1.56% | 1.56% | 1.56% | 1.56% |

### *3.3. The Data Center*

The Data Center setup comprises two clusters: i) a High-Performance Computing (HPC) cluster to train and validate the models, as these are CPU and memory intensive tasks, and ii) a virtualized Cloud computing cluster for online prediction. The tasks of model training and validation are CPU and memory intensive, and require of a resource of computation during the whole execution of the tasks, that use a large amount of data. Therefore, they require an HPC cluster. On the other hand, the execution of the model for prediction is a task driven by requests, that can be executed on a virtualized machine to release the resources after that. Therefore, it is executed on a Cloud cluster.

The HPC cluster is composed of Quad-core Intel Xeon RX300 servers with 16GB of RAM, and the cluster is managed using the SLURM[16] resource manager. The virtualized Cloud computing cluster consists on Intel SandyBridge S2600GZ servers with six cores and 32 GB of RAM. The SandyBridge servers are virtualized using KVM, and the OpenStack software is deployed to manage the deployment and allocation of Virtual Machines (VMs). Due to the intensiveness of prediction of models, we do not oversubscribe VMs, *i.e.*, we only deploy 1 VM per core, and assign 2GB of RAM per VM. In order to characterize power and performance, servers are monitored via IPMI software and using current clamps, allowing us to obtain power values with an error of ±10 W.

We characterize in terms of power and performance all the offline tasks in the Intel Xeon server: i) the data preprocessing (GPML), ii) training, and iii) validation. We also characterize all the online tasks in the SandyBridge server: i) the data preprocessing (GPML), and ii) online prediction. We run stress and performance tests on the VMs of the virtualized cluster to obtain the maximum number of GPMLs and predictions that can be performed in each VM without degrading the throughput.

The characteristics of the HPC and the virtualized Cloud cluster to compute predictions are next outlined. For a detailed explanation on how these parameters are obtained, the reader is referred to Appendix A:

- The preprocessing, training and validation phase of 1 patient runs for approximately 3.5 hours in a dedicated server, and degrades up to 3.76 hours when the server is fully utilized, with a power consumption that varies from 157 W to 173 W. The HPC cluster consists on 2275 server, providing up to 9100 cores. Preprocessing, training and validation are performed in a simulated HPC SLURM cluster [45]. We consider that, in average, models need to be re-trained once a month per patient, to stay within allowable error margins.

- In the online phase predictions are computed every minute. Runtime prediction consists on a data preprocessing stage (*i.e.*, data recovery via GPML in case there was a packet data loss), and migraine prediction 30 minutes forward, *i.e.* a migraine is predicted 30 minutes in advance. The virtualized cluster consists on 1638 servers that can run up to 9828 VMs. Up to 230 preprocessing instances and 250 prediction instances can be packed together in the same VM, without degrading performance. When

[16] `https://slurm.schedmd.com`

a VM reaches this limit, OpenStack is responsible for automatically scaling up the resources, launching as many VMs as needed to handle the workload.

The previous setup allows the HPC and Cloud clusters to handle in the offline and online phases up to 1,393,649 migraine patients. This represents a 2% of the migraine sufferers in Europe (see Appendix A and Appendix B), which is the target population in our case study.

It needs to be considered that runtime prediction is a light-weight process that can be performed either in the coordinator node or in the Data Center. In this sense, if the coordinators off-load computation from the Data Center, OpenStack will reduce the amount of VMs used. This, together with the usage of server turn-off policies, can drastically reduce the power usage of the Data Center, as shown in Section 5.

Regarding cooling, we assume that the whole cluster is cooled via traditional raised-floor air-cooling techniques. As a baseline, we assume a room temperature of 22 $^{\circ}C$ and a PUE of 1.65, which are the most common value for room temperature and the world average PUE value respectively [39]. Cooling optimization policies will be applied to reduce cooling costs while keeping servers under save environmental conditions.

To simplify the deployment, we have divided patients into four large groups, according to their nationalities. Patients are included in the system gradually, as shown in Figure 5. The inclusion rates are: 50%, 25%, 15% and 10%, and the total evaluation period is 10 weeks. The gradual inclusion is an exercise of evaluation of the methodology proposed, and it is considered that all patients in a group are included at the same time. First, when the HPC cluster is still empty, 50% of the population is included (patients from Turkey, Germany and United Kingdom). Every two weeks, the remaining groups come into the system. The second group are two countries that represent 25% of European migraineurs (France and Italy). The third group comprises three countries (Spain, Netherlands and Switzerland) with 15% of patients, and finally the remaining 10% patients are distributed in eight countries. For further details, the reader is referred to Appendix B.

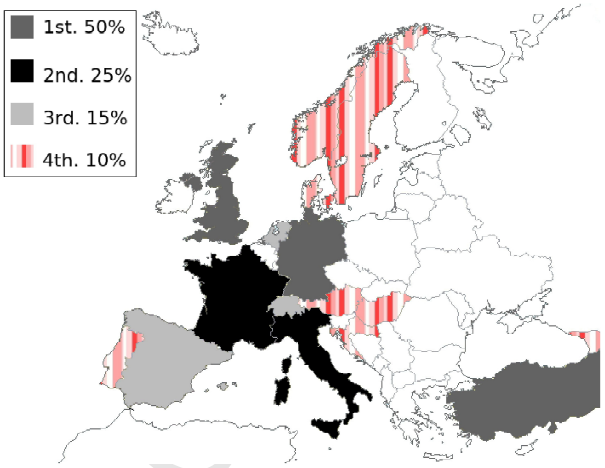


Figure 5: Gradual inclusion of migraine sufferers in the system. 50% of this population is firstly included in the study, when the HPC cluster in Data Center is still empty.

Table 5: Tasks that each level of the network is able to perform.

| | Tasks |
|---|---|
| Sensor device | Collect and transmit data, process data (limited) |
| Coordinator | Receive, transmit and process data, perform predictions |
| Data center | Process data, perform predictions |

## 4. MCC eHealth optimization scenarios

As mentioned in Section 2.3, in this work we tackle the energy efficiency challenge from the workload allocation perspective, including the processing and wireless communication level. From the signal processing level we perform very light preprocessing techniques in the sensor nodes—due to their limited processing capabilities. From the wireless communication level, we focus on the on-body channel communication between the monitoring devices and the central node or coordinator.

Each one of the three network elements can operate in various modes. Table 5 summarizes the tasks that each level of the network can perform when working cooperatively.

The energy efficiency policies take these possibilities into account in order to minimize the power consumption of the whole system. For the sake of clarity, among all combinations of scenarios, we choose the five than we understand to be more significant from the energy perspective. These scenarios are shown in Table 6. Our claim is that using the WBSN and coordinator to off-load computation minimizes overall energy consumption. This technique, together with tailored energy optimization strategies at the Data Center level can drastically reduce the impact of the application deployment.

According to Table 6, the sensor nodes only have two different working modes:

- Streaming mode: in scenarios SC1-SC2-SC3. Node S1 collects and transmits data from ECG immediately. Node S2 collects and transmits data from the NONIN devices every second.
- Processing mode: in scenarios SC4-SC5. Node S1, after collecting ECG data, calculates the HR and transmits it every minute. Node S2 extract SpO2 data from the whole frame of 375 Bytes of the OEM-III device and transmit it once a minute as well.

Regardless of the working mode, partial or total data loss is usual in ambulatory studies. As mentioned before, these can occur due to two factors: i) discharge of the node battery, or ii) sensor disconnection. When a disruption occurs, a GPML is executed to recover the signals. This process can be executed both in the coordinator as in the Data Center. For the purpose of this paper, the GPML is going to be executed with the

Table 6: Five scenarios for the workload balancing policies.

| | Sensor device | Coordinator | Data Center |
|---|---|---|---|
| SC1 | Collect + transmit data | Receive + transmit data | Process data + perform predictions |
| SC2 | Collect + transmit data | Receive + process + transmit data | Perform predictions |
| SC3 | Collect + transmit data | Receive + process data + perform prediction + transmit data | - |
| SC4 | Collect + process + transmit data | Receive + transmit data | Perform predictions |
| SC5 | Collect + process + transmit data | Receive data + perform predictions + transmit data | - |

probabilities shown in Table 2, which have been experimentally obtained. The coordinator and the Data Center can also process raw signal from sensor nodes (in streaming mode). The online phase can also take place either in the coordinator or in the Cloud Data Center, whereas the offline phase—due to its computational burden—can only be carried out in the HPC Data Center.

## 5. Results

In this section we present the results of the strategies applied at the node, coordinator and Data Center level. In addition, a study of the economic impact of the energy savings is performed.

Results have been extrapolated from real data acquired in the clinic and in the laboratory. The data used correspond to the migraines of two young, middle-sized and middleweight female migraineurs; one of them with treatment. Power and performance characterization for the coordinators and Data Center corresponds to the description in Section 3.

### *5.1. Energy results in WBSN*

The energy consumption of the microcontroller for node S1 in both working modes has been obtained from [36]. The consumption in node S2 has been measured with a HAMEG HM8012 digital multimeter. Radio consumption has been calculated in simulation as described in Section 3 for both of the nodes.

Simulation results have been calculated as the average of the results of 10 simulations of 60 seconds each. For the CC2420 radio chip, the default values provided by the simulator have been chosen.

Table 7 shows the energy consumption of the Shimmer nodes—microcontroller and radio for node S1, and microcontroller, 8000R oximetry sensor and OEM-III module and radio for node S2—for both working modes: streaming and processing. Energy has been calculated for an execution time of 1 minute and only for 1 patient.

When working in streaming mode, the nodes collect data and transmit them immediately without saving energy. These situations are shown in Table 7 for 0 dBm of radio power transmission in the streaming mode.

Radio switches on and switches off in every transmission. This explains the high consumption of node S1 in streaming mode. In processing mode, node S1 switches on just one time per minute, while in streaming mode it does for 217 times (see Table 4). A similar situation occurs with node S2; where in streaming mode radio switches on 60 times more than in processing mode.

Energy savings when computing HR in node S1, applying reactive techniques and transmitting once per minute, reach 96.9% (from 13767 mJ to 432 mJ). Energy savings in node S2 after the extraction of the SpO2 data, applying reactive techniques and transmitting once per minute, reach 50.3% in node S2 (from 7235 mJ to 3637 mJ). Notice that reactive policies in the radio lead to 1.1% energy savings in streaming mode for node S1 (from 13526 mJ to 13371 mJ); almost negligible. This is because most energy (99.5%) is used during the initialization process, and barely goes to the energy transmission $\epsilon_{TX}$. These policies reach 80% of energy saving (from 9 mJ to 1.8 mJ) in streaming mode for the same node. For node S2, energy savings in the radio link lead to 33.7% and 79.0% for streaming and processing modes respectively.

Even though, the energy reduction in the radio link is significant. The absolute savings are masked due to the high power consumption of node computation. To increase the impact of radio energy savings, the computation of nodes S1 and S2 should be reduced. In this sense, more efficient HR calculation techniques—or sampling rate reduction techniques—would reduce the power consumption. For node S2 the strategy would be changing the sensor node and its proprietary processing module OEM-III.

According to our results, processing data in nodes is more convenient than streaming them, reaching a power consumption of 432 mJ for node S1 and 3637 mJ for node S2. Thus, for the remaining of the paper, we use this setup, an apply our workload balancing policies in scenarios SC4 and SC5 in Table 6, to further reduce the energy consumption.

### *5.2. Workload off-loading policies*

In this section we present the results for the workload off-loading policies, corresponding to scenarios SC4 and SC5. Moreover, in this section we show the benefits of combining workload off-loading techniques (*i.e.*, moving workload from the Data Center to the coordinators) together with energy optimization strategies in the HPC and Cloud Data Centers.

The energy optimization strategies are called: i) server turn-off policies, ii) workload consolidation, and iii) cooling power reduction. This yields the following scenarios:

- SC4 (baseline): the original scenario considers no workload off-loading. The coordinator nodes simply forward computation to the Data Center. The off-line phase is performed in the HPC cluster, and the online phase in the virtualized cluster.

Table 7: Energy consumption for nodes without intelligence in transmission and after applying reactive transmission policies.

| | | Streaming mode | | | | Total | Processing mode | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Computing | Radio | | | streaming (mJ) | Computing* | Radio | | | processing (mJ) |
| | Radio power$_{TX}$ | uC (mJ) | $\epsilon_{total}$ (mJ) | $\epsilon_{TX}$ (mJ) | RSSI (dB) | | uC + NONIN (mJ) | $\epsilon_{total}$ (mJ) | $\epsilon_{TX}$ (mJ) | RSSI (dB) | |
| Node S1 | 0 dBm | 396 | 13526 | 61 | -59 | 13922 | 430 | 9.0 | 0.3 | -57 | 439 |
| | -10 dBm | | 13371 | 38 | -68 | 13767 | | - | - | - | - |
| | -15 dBm | | - | - | - | - | | 1.8 | 0.2 | -70 | **432** |
| Node S2 | 0 dBm | 3635 | 3600 | 6 | -63 | 7235 | 3635 | 10.5 | 0.3 | -63 | 3646 |
| | -15 dBm | | 2388 | 4 | -75 | 6023 | | 2.2 | 0.2 | -75 | **3637** |

*uC consumption includes the consumption of NONIN 8000R SpO2 sensor and the consumption of OEM-III module (predominant).

- SC4 (optimized): in this scenario we apply Data Center energy minimization techniques, but without any workload off-loading policy.

- SC5, 100% prediction: coordinator nodes perform data preprocessing (*i.e.*, GPML), but the virtualized Cloud takes care of 100% of the prediction. The off-line phase is performed in the HPC cluster.

- SC5, 30% prediction: in this case, coordinator nodes execute both GPML and 70% of the predictions (*i.e.*, when battery power allows it). The remaining 30% are computed in the virtualized cluster.

All scenarios except for the first SC4 (baseline) apply the three energy minimization strategies aforementioned. The SC4 (baseline) is used to calculate the energy savings of the three other experiments.

Server turn-off policies are applied in both the HPC and Cloud clusters when the Data Center is not fully utilized. In this sense, workload is packed into the minimum number of servers possible and the remaining servers are turned-off. In order to guarantee the absorbance of the sudden workload peaks, we always leave 20% of unused servers on. Workload consolidation is applied only in the virtualized Cloud cluster, and consists on packing together as many instances of either GPML or prediction as possible, until the server reaches a per-core utilization of 100%. As shown in Appendix A and explained in Section 3, up to 230 and 250 instances of GPML and prediction can be packed together, respectively. These two techniques are particularly useful during the initial deployment of the application. As both the HPC and the Cloud cluster have been sized according to the maximum workload, during the first weeks of the migraine prediction deployment, the clusters are heavily underutilized.

Figure 6 shows the evolution of the HPC cluster utilization for a period of 10 weeks as new models that need to be trained arrive to the system, due to the gradual inclusion of patients (blue dots). The utilization of the HPC cluster raises to absorb the peaks. We also observe how model re-training increase (red line) as more migraines are being predicted. Finally, we see how utilization stabilizes around 80%, the design point of the HPC cluster.

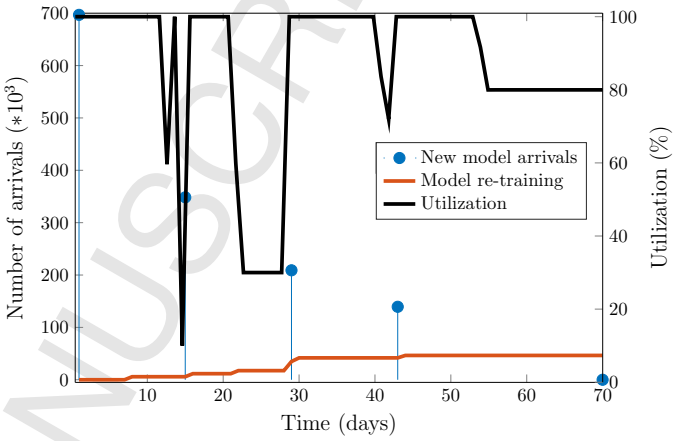


Figure 6: Utilization of the HPC Data Center. Blue dots correspond to gradual inclusion of patients (50%, 25%, 15% and 10%). Red line corresponds to models that need to be re-trained.

On the other hand, Figure 7 shows how the utilization of the virtualized cluster decreases as we off-load computation to the coordinators. This, by itself, does not decrease energy. However, when combined with consolidation and server turn-off policies, drastically minimizes power consumption.

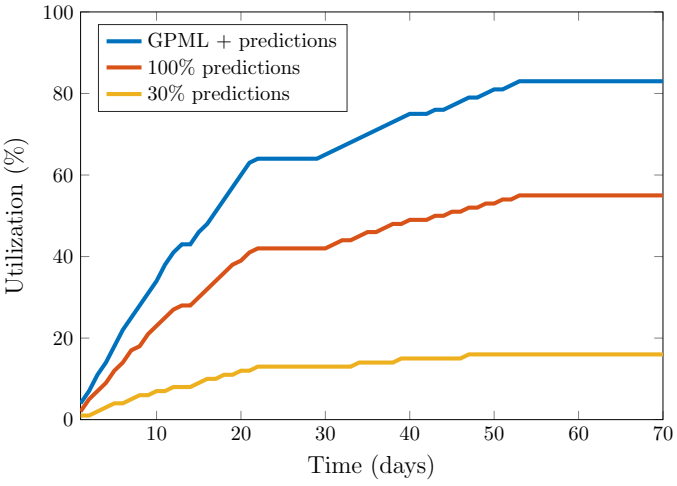


Figure 7: Utilization of the Cloud cluster.

This behavior can be observed in Table 8, which shows the power consumption breakdown for the HPC and Cloud Data Centers for the whole period of 10 weeks. We achieve cooling power reduction by increasing the room temperature of both clusters. To do so, we run experiments at increased inlet temperature for both the SandyBridge and the Intel Xeon

Table 8: Energy consumption breakdown in HPC and Cloud Data Center for 10 weeks and various policies (MWh).

| | HPC IT | HPC IT+Cool. | Cloud IT | Cloud IT+Cool. | Total |
|---|---|---|---|---|---|
| SC4 (base.) | 643.5 | 1061.2 | 282.4 | 466.1 | 1527.3 (-) |
| SC4 (opt.) | 584.7 | 907.5 | 219.1 | 340.0 | 1247.5(18.3%) |
| SC5 (100%) | 584.7 | 907.5 | 159.1 | 247.0 | 1154.5 (24.4%) |
| SC5 (30%) | 584.7 | 907.5 | 78.3 | 121.5 | 1029 (32.6%) |

Table 9: Energy consumption for various workload off-loading scenarios for 1 week in the stationary state (MWh).

| | Coord. | HPC DC | Cloud DC | Total | Savings |
|---|---|---|---|---|---|
| SC4 (baseline) | 170.05 | 104.9 | 49.9 | 324.8 | - |
| SC4 (optimized) | 170.05 | 85.4 | 42.4 | 297.8 | 8.3% |
| SC5 (100% pred.) | 170.12 | 85.4 | 32.2 | 287.7 | 11.4% |
| SC5 (30% pred.) | 170.32 | 85.4 | 14.7 | 270.4 | 16.7% |

servers, and heuristically observed that room temperature can be raised up to 26 $^{\circ}C$ without thermal redlining. Assuming an initial PUE equal to the world average PUE of 1.65, and a reduction of 4% [39] in cooling power for every degree of increase in ambient temperature, we compute the savings in cooling power. As can be seen, the proposed techniques achieve significant energy reductions, of up to 32.6% in overall data center power when compared with the non-optimized scenario SC4 (baseline).

Table 9 shows the energy consumption for the overall application framework of migraine monitoring and prediction for all scenarios. Results are aggregated per device, *i.e.* each column shows the aggregated energy value for all 1,393,649 patients under test. As can be seen, in a scenario with so large number of sensors and coordinators, the impact of these devices on the overall energy consumption of the application is large. In this sense, reducing the energy consumed by the sensor devices implies significant savings. Data Centers, however, are still the most important contributors to power consumption, and off-loading data to coordinators always yields important benefits, reducing overall energy consumption in the steady-state by up to 16.7%.

Finally, Table 9 also summarizes the maximum energy savings achieved for sensor nodes and the offloading scenario between the coordinator and the Data Center. This results are used in the next subsection to compute the economic benefits of our approach.

### *5.3. Economic benefits*

To compute the economic benefits expected by our approach, we compute in Appendix B the total amount of economic savings that can be achieved by migraine prediction in Europe. These savings are of € 1271.8 million with an efficiency in prediction of 76%. Given the amount of energy consumed by the nodes, coordinators, and Data Centers, together with the price of household and industrial electricity, we can provide some insight on the economic impact of our approach. The goal of this result is simply to provide a general overview of the dimension of this problem in particular, and of Mobile Cloud Computing applications in general.

Table 10: Total energy savings in nodes and Data Centers and expenses in coordinator nodes.

| Saving in nodes S1* (kWh) | Saving in nodes S2* (kWh) | Daily saving in Data Centers (kWh) | Daily expenses in coordinators (kWh) |
|---|---|---|---|
| $3.7 * 10^{-6}$ | $1.0 * 10^{-6}$ | 7819.3 | 51.2 |

*1 minute simulation.

Table 10 shows the energy savings for each element of the monitorization system. After applying processing techniques and low-power radio policies in node S1 for HR measurement, we achieve a daily saving of $5.32 * 10^{-3}$ kWh; $1.44 * 10^{-3}$ kWh doing the appropriate in node S2. These values are shown in Table 10 for 1 minute of simulation of the radio channel in Castalia. Bigger than these values are the energy savings in Data Centers. Energy saving in these facilities brings an issue: part of the execution is done in coordinators leading to daily expenses of 51.2 kWh. As shown in Table 11, it worth to balance the workload. On average, economic savings in a year of execution in Data Centers lead to € 287.7 ± 104.1 million. This average calculation understands the HPC and Cloud clusters as a federation of Data Centers located in Europe (see Appendix A), and only the variable part of the industrial and household energy prices is taken into account. Economic savings in nodes in Europe lead to € 692,400. This calculation is the sum of all the savings in Europe taking into account the household energy prices in each country. Finally, the total amount of saving is € 288.4 ± 104.1 million.

## 6. Conclusions

The new paradigm of the Internet of Things poses an important challenge because of the large volumes of data that need to be gathered and analyzed. Among IoT applications, population monitoring applications in general and eHealth in particular, impose important restrictions, and require the deployment of energy minimization strategies that allow the deployment of healthcare solutions in massive population deployments. In this paper we propose multi-level techniques to reduce the energy consumption of nodes, coordinators and Data Center in a eHealth application for migraine prediction. We describe a realistic case study and we test our policies with realistic data, reducing the energy consumption of nodes, and off-loading workload from the Data Center to coordinators, achieving savings of up to 32.6% in the Data Center for the whole deployment of the system. From the economical perspective, our solution yields average savings of € 288 million when applied only to 2% of European migraine sufferers; in addition to savings of € 1272 million due to the benefits of the migraine prediction.

## Acknowledgments

Research by Josué Pagán has been funded by the EU (FEDER) under Research grant PI15/01976 and the Spanish Ministry of Economy and Competitiveness under Research Grants TIN 2015-65277-R and TEC2012-33892. The authors gratefully thank to Monica Vallejo (Department of Electric

Table 11: Average saving for a federation of Data Centers in Europe. Savings calculated for the variable electricity prices. Savings in nodes and Data Centers have repercussion in consumption and electricity costs in coordinator nodes.

| Country | Target population | Household [46] energy price (€/kWh) | Industrial [46] energy price (€/kWh) | Saving in nodes (k€) | Expenses in coordinators (k€) | Saving in Data Center (k€) | Total energy saving (M€) |
|---|---|---|---|---|---|---|---|
| Turkey | 305,138 | 0.122 | 0.070 | 93.0 | 0.50 | 200.4 | 478.9 |
| Germany | 184,150 | 0.295 | 0.149 | 135.2 | 0.73 | 426.1 | 594.3 |
| UK | 183,922 | 0.218 | 0.152 | 100.1 | 0.54 | 433.8 | 601.8 |
| France | 147,472 | 0.243 | 0.160 | 89.3 | 0.48 | 455.8 | 590.5 |
| Italy | 182,348 | 0.168 | 0.095 | 76.1 | 0.41 | 270.8 | 437.3 |
| Spain | 117,211 | 0.237 | 0.113 | 69.2 | 0.37 | 323.4 | 430.4 |
| Netherlands | 78,088 | 0.183 | 0.084 | 35.7 | 0.19 | 238.3 | 309.6 |
| Switzerland [47] | 40,047 | 0.192 | 0.142 | 19.2 | 0.10 | 405.3 | 441.8 |
| Sweden | 25,462 | 0.187 | 0.059 | 11.9 | 0.06 | 168.4 | 191.6 |
| Norway | 23,497 | 0.143 | 0.069 | 8.4 | 0.05 | 195.5 | 217.0 |
| Denmark | 21,496 | 0.304 | 0.091 | 16.3 | 0.09 | 258.6 | 278.2 |
| Hungary | 18,965 | 0.115 | 0.087 | 5.4 | 0.03 | 248.3 | 265.6 |
| Portugal | 18,352 | 0.229 | 0.115 | 10.5 | 0.06 | 329.4 | 346.1 |
| Austria | 17,354 | 0.198 | 0.105 | 8.6 | 0.05 | 298.8 | 314.7 |
| Croatia | 16,138 | 0.131 | 0.093 | 5.3 | 0.03 | 264.9 | 279.6 |
| Georgia [48] | 14,010 | 0.050 | 0.030 | 1.7 | 0.01 | 85.6 | 98.4 |
| Total | 1,393,649 | | | **685.8** | | | **287.7±104.1** |

Energy and Automation, National University of Colombia) for her help in this work.

## References


[1] Y.-L. Zheng, X.-R. Ding, C. C. Y. Poon, B. P. L. Lo, H. Zhang, X.-L. Zhou, G.-Z. Yang, N. Zhao, Y.-T. Zhang, Unobtrusive sensing and wearable devices for health informatics, Biomedical Engineering, IEEE Transactions on 61 (5) (2014) 1538–1554.

[2] E. Dorschky, D. Schuldhaus, H. Koerger, B. Eskofier, A Framework for Early Event Detection for Wearable Systems, in: ACM (Ed.), Proceedings of the 2015 ACM International Symposium on Wearable Computers, 2015, pp. 109–112.

[3] B. E. Heldberg, T. Kautz, H. Leutheuser, R. Hopfengartner, B. S. Kasper, B. M. Eskofier, Using wearable sensors for semiology-independent seizure detection-towards ambulatory monitoring of epilepsy, in: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, IEEE, 2015, pp. 5593–5596.

[4] L. J. Stovner, C. Andree, Prevalence of headache in europe: a review for the eurolight project, The journal of headache and pain 11 (4) (2010) 289–299.

[5] M. Linde, A. Gustavsson, L. Stovner, T. Steiner, J. Barré, Z. Katsarava, J. Lainez, C. Lampl, M. Lantéri-Minet, D. Rastenyte, et al., The cost of headache disorders in europe: the eurolight project, European journal of neurology 19 (5) (2012) 703–711.

[6] X. H. Hu, N. H. Raskin, R. Cowan, L. E. Markson, M. L. Berger, U. S. M. S. P. U. Group, et al., Treatment of migraine with rizatriptan: when to take the medication, Headache: The Journal of Head and Face Pain 42 (1) (2002) 16–20.

[7] C. M. Ordás, M. L. Cuadrado, A. B. Rodríguez-Cambrón, J. Casas-Limón, N. del Prado, J. Porta-Etessam, Increase in body temperature during migraine attacks, Pain Medicine 14 (8) (2013) 1260–1264.

[8] J. Porta-Etessam, M. L. Cuadrado, O. Rodríguez-Gómez, C. Valencia, S. García-Ptacek, Hypothermia during migraine attacks, Cephalalgia 30 (11) (2010) 1406–1407.

[9] J. Pagán, M. I. De Orbe, A. Gago, M. Sobrado, J. L. Risco-Martín, J. V. Mora, J. M. Moya, J. L. Ayala, Robust and accurate modeling approaches for migraine per-patient prediction from ambulatory data, Sensors 15 (7) (2015) 15419.

[10] J. Pagán, , J. L. Risco-Martín, J. M. Moya, J. L. Ayala, Grammatical evolutionary techniques for prompt migraine prediction, in: Genetic and Evolutionary Computation Conference (GECCO), 2016.

[11] A. N. Khan, M. M. Kiah, S. U. Khan, S. A. Madani, Towards secure mobile cloud computing: A survey, Future Generation Computer Systems 29 (5) (2013) 1278–1299.

[12] M. Zapater, P. Arroba, J. L. Ayala, J. M. Moya, K. Olcoz, A novel energy-driven computing paradigm for e-health scenarios, Future Generation Computer Systems 34 (2014) 138–154.

[13] P. Van Overschee, B. De Moor, N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems, Automatica 30 (1) (1994) 75–93.

[14] J. Pagán, J. L. Risco-Martín, J. M. Moya, J. L. Ayala, Modeling methodology for the accurate and prompt prediction of symptomatic events in chronic diseases, Journal of biomedical informatics.

[15] U. Housing, M. S. OTB, Smart cities ranking of european medium-sized cities.

[16] M. Deakin, H. Al Waer, From intelligent to smart cities, Intelligent Buildings International 3 (3) (2011) 140–152.

[17] B. Groh, S. Reinfelder, M. Streicher, A. Taraben, B. Eskofier, Movement prediction in rowing using a Dynamic Time Warping based stroke detection, in: IEEE (Ed.), IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP 2014), 2014, pp. 1–6.

[18] P. Kugler, D. Schuldhaus, U. Jensen, B. Eskofier, Mobile recording system for sport applications, in: Proceedings of the 8th international symposium on computer science in sport (IACSS 2011), Liverpool, 2011, pp. 67–70.

[19] H. Alemdar, C. Ersoy, Wireless sensor networks for healthcare: A survey, Computer Networks 54 (15) (2010) 2688–2710.

[20] J. Milosevic, A. Dittrich, A. Ferrante, M. Malek, C. R. Quiros, R. Braojos, G. Ansaloni, D. Atienza, Risk assessment of atrial fibrillation: a failure prediction approach, in: Computing in Cardiology Conference (CinC), 2014, IEEE, 2014, pp. 801–804.

[21] H. Leutheuser, S. Gradl, P. Kugler, L. Anneken, M. Arnold, S. Achenbach, B. M. Eskofier, Comparison of real-time classification systems for arrhythmia detection on android-based mobile devices, in: Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE, IEEE, 2014, pp. 2690–2693.

[22] A. Burns, B. R. Greene, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer, F. Stroiescu, V. Cionca, Shimmer$^{TM}$–a wireless sensor platform for noninvasive biomedical research, Sensors Journal, IEEE 10 (9) (2010) 1527–1534.

[23] F. Casino, P. Lopez-Iturri, E. Aguirre, L. Azpilicueta, A. Solanas, F. Falcone, Dense wireless sensor network design for the implementation of smart health environments, in: Electromagnetics in Advanced Applications (ICEAA), 2015 International Conference on, IEEE, 2015, pp. 752–754.

[24] A. Dhamdhere, V. Sivaraman, V. Mathur, S. Xiao, Algorithms for

transmission power control in biomedical wireless sensor networks, in: Asia-Pacific Services Computing Conference, 2008. APSCC'08. IEEE, IEEE, 2008, pp. 1114–1119.

[25] S. Xiao, A. Dhamdhere, V. Sivaraman, A. Burdett, Transmission power control in body area sensor networks for healthcare monitoring, Selected Areas in Communications, IEEE Journal on 27 (1) (2009) 37–48.

[26] R. Augustine, Electromagnetic modelling of human tissues and its application on the interaction between antenna and human body in the ban context, Ph.D. thesis, Université Paris-Est (2009).

[27] D. B. Smith, D. Miniutti, L. W. Hanlen, Characterization of the body-area propagation channel for monitoring a subject sleeping, Antennas and Propagation, IEEE Transactions on 59 (11) (2011) 4388–4392.

[28] N. E. Roberts, S. Oh, D. D. Wentzloff, Exploiting channel periodicity in body sensor networks, Emerging and Selected Topics in Circuits and Systems, IEEE Journal on 2 (1) (2012) 4–13.

[29] Q. Tang, N. Tummala, S. K. S. Gupta, L. Schwiebert, Communication scheduling to minimize thermal effects of implanted biosensor networks in homogeneous tissue, Biomedical Engineering, IEEE Transactions on 52 (7) (2005) 1285–1294.

[30] A. Aulery, J.-P. Diguet, C. Roland, O. Sentieys, Low-complexity energy proportional posture/gesture recognition based on wbsn, in: Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference on, IEEE, 2015, pp. 1–6.

[31] M. Vallejo, J. Recas, P. G. del Valle, J. L. Ayala, Accurate human tissue characterization for energy-efficient wireless on-body communications, Sensors 13 (6) (2013) 7546–7569.

[32] H. Mamaghanian, N. Khaled, D. Atienza, P. Vandergheynst, Compressed sensing for real-time energy-efficient ecg compression on wireless body sensor nodes, Biomedical Engineering, IEEE Transactions on 58 (9) (2011) 2456–2466.

[33] S. Y. Lee, J. H. Hong, C. H. Hsieh, M. C. Liang, S. Y. C. Chien, K. H. Lin, Low-power wireless ecg acquisition and classification system for body sensor networks, IEEE Journal of Biomedical and Health Informatics 19 (1) (2015) 236–246. doi:10.1109/JBHI.2014.2310354.

[34] A. Tobola, O. Korpok, H. Leutheuser, B. Schmitz, C. Hofmann, M. Struck, C. Weigand, B. Eskofier, A. Heuberger, G. Fischer, System Design Impacts on Battery Runtime of Wearable Medical Sensors, in: D. Novák (Ed.), Proceedings of International Conference on Mobile and Information Technologies in Medicine and Health, 2014, pp. 1–4.

[35] A. Tobola, F. J. Streit, C. Espig, O. Korpok, C. Sauter, N. Lang, B. Schmitz, C. Hofmann, M. Struck, C. Weigand, et al., Sampling rate impact on energy consumption of biomedical signal processing systems, in: Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference on, IEEE, 2015, pp. 1–6.

[36] F. Rincón, J. Recas, N. Khaled, D. Atienza, Development and evaluation of multilead wavelet-based ecg delineation algorithms for embedded wireless sensor nodes, Information Technology in Biomedicine, IEEE Transactions on 15 (6) (2011) 854–863.

[37] J. Koomey, Growth in data center electricity use 2005 to 2010, Tech. rep., Analytics Press, Oakland, CA (2011).

[38] A. Venkatraman, Global census shows datacentre power demand grew 63% in 2012.

[39] J. K. Matt Stansberry, Uptime institute 2013 data center industry survey, Tech. rep., Uptime Institute (2013).

[40] M. Vallejo, J. Recas Piorno, J. L. Ayala Rodrigo, A link quality estimator for power-efficient communication over on-body channels, in: Embedded and Ubiquitous Computing (EUC), 2014 12th IEEE International Conference on, IEEE, 2014, pp. 250–257.

[41] M. Vallejo, J. Recas, J. L. Ayala, Proactive and reactive transmission power control for energy-efficient on-body communications, Sensors 15 (3) (2015) 5914–5934.

[42] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning), The MIT Press, 2005.

[43] MATLAB, version 8.5.0.197613 (R2015a), The MathWorks Inc., Natick, Massachusetts, 2015.

[44] J. W. Mason, D. J. Ramseth, D. O. Chanter, T. E. Moon, D. B. Goodman, B. Mendzelevski, Electrocardiographic reference ranges derived from 79,743 ambulatory subjects, Journal of electrocardiology 40 (3) (2007) 228–234.

[45] A. Lucero, Simulation of batch scheduling using real production-ready software tools, http://www.bsc.es/media/4856.pdf.

[46] EuroStat, Electricity price statisticcs, http://ec.europa.eu/eurostat/statistics-explained/index.php/Electricity_price_statistics.

[47] A. of Swiss electricity companies VSE-AES, Swiss electricity prices- development 1991-2015, http://www.strom.ch/de/metanavigation/download.html, accessed: 2016-06-01.

[48] INOGATE, A review of energy tariffs in inogate partner countries 2015, http://www.inogate.org/, accessed: 2016-06-01.

[49] J. M. P. Monteiro, Cefaleias: estudo epidemiológico e clínico de uma população urbana.

[50] K. R. Merikangas, A. E. Whitaker, H. Isler, J. Angst, The zurich study: Xxiii. epidemiology of headache syndromes in the zurich cohort study of young adults, European archives of psychiatry and clinical neuroscience 244 (3) (1994) 145–152.

[51] EuroStat, Population on 1 january by age and sex, http://ec.europa.eu/eurostat/en/web/products-datasets/-/DEMO_PJAN, accessed: 2016-06-01.

[52] Y. Celik, G. Ekuklu, B. Tokuç, U. Utku, Migraine prevalence and some related factors in turkey, Headache: The Journal of Head and Face Pain 45 (1) (2005) 32–36.

[53] T. Steiner, A. Scher, W. Stewart, K. Kolodner, J. Liberman, R. Lipton, The prevalence and disability burden of adult migraine in england and their relationships to age, gender and ethnicity, Cephalalgia 23 (7) (2003) 519–527.

[54] V. Pfaffenrath, K. Fendrich, M. Vennemann, C. Meisinger, K.-H. Ladwig, S. Evers, A. Straube, W. Hoffmann, K. Berger, Regional variations in the prevalence of migraine and tension-type headache applying the new ihs criteria: the german dmkg headache study, Cephalalgia 29 (1) (2009) 48–57.

[55] M. Lantéri-Minet, D. Valade, G. Geraud, M. Chautard, C. Lucas, Migraine and probable migraine–results of framig 3, a french nationwide survey carried out according to the 2004 ihs classification, Cephalalgia 25 (12) (2005) 1146–1158.

[56] J. Matías-Guiu, J. Porta-Etessam, V. Mateos, S. Díaz-Insa, A. Lopez-Gil, C. Fernández, One-year prevalence of migraine in spain: a nationwide population-based survey, Cephalalgia 31 (4) (2011) 463–470.

[57] L. J. Launer, G. M. Terwindt, M. D. Ferrari, The prevalence and characteristics of migraine in a population-based cohort the gem study, Neurology 53 (3) (1999) 537–537.

[58] C. Dahlöf, M. Linde, One-year prevalence of migraine in sweden: a population-based study in adults, Cephalalgia 21 (6) (2001) 664–671.

[59] O. Sjaastad, L. Bakketeig, Migraine without aura: comparison with cervicogenic headache. vågå study of headache epidemiology, Acta Neurologica Scandinavica 117 (6) (2008) 377–383.

[60] M. B. Russell, N. Levi, J. Šaltytė-Benth, K. Fenger, Tension-type headache in adolescents and adults: a population based study of 33,764 twins, European journal of epidemiology 21 (2) (2006) 153–160.

[61] J. Bank, S. Marton, Hungarian migraine epidemiology, Headache: The Journal of Head and Face Pain 40 (2) (2000) 164–169.

[62] C. Lampl, A. Buzath, U. Baumhackl, D. Klingler, One-year prevalence of migraine in austria: a nation-wide survey, Cephalalgia 23 (4) (2003) 280–286.

[63] R. Zivadinov, K. Willheim, A. Jurjevic, D. Sepic-Grahovac, M. Bucuk, M. Zorzon, Prevalence of migraine in croatia: A population-based survey, Headache: The Journal of Head and Face Pain 41 (8) (2001) 805–812.

[64] Z. Katsarava, A. Dzagnidze, M. Kukava, E. Mirvelashvili, M. Djibuti, M. Janelidze, R. Jensen, L. Stovner, T. Steiner, et al., Primary headache disorders in the republic of georgia prevalence and risk factors, Neurology 73 (21) (2009) 1796–1803.

## Appendix A. Data Center Calculations and Sizing

The HPC and Cloud clusters of this paper can be either understood as either a centralized Data Centers that performs all the computation needed, but also as federation of Data Centers located in Europe. In order to compute the power consumption and performance of these two clusters, the following methodology has been followed:

1. Profiling and characterization of the GPML, training and validation stages in the Intel Xeon servers of the HPC cluster, obtaining the power consumption of each task for one instance of each task and when the server is fully utilized.
2. Profiling of GPML and prediction in the SandyBridge servers and the coordinators nodes, obtaining power and performance.
3. Consolidation analysis at maximum frequency for the virtualized cluster, to discover the maximum amount of instances that can be run in one server until utilization reached 100% without degrading performance. In this sense, we have found that the computational burden of GPML is of 260 ms, allowing up to 230 instances to run per core on the same VM without degrading performance. As for prediction, we can launch simultaneously 250 instances per VM, reaching a per-core utilization of 100% and not degrading performance.
4. Data Center sizing, *i.e.*, obtaining the amount of servers of each type for both the HPC and the Cloud cluster.
5. Generation of the incoming workloads, that consists on new models to be trained in the HPC cluster, as well as model re-training. The output of the HPC cluster, *i.e.*, the trained models, are the input to the Cloud cluster.

Regarding Data Center sizing, we need to compute models for $P = 1,393,649$ patients (2% of the migraine sufferers in Europe). To this end, the HPC cluster is composed of with $S = 2275$ Intel Xeon servers, with 4 cores each, so that the total number of cores $C = 9100$ is able to tackle modeling. Models need to be re-trained each $R_{re-train} = 30$ days per patient.

As for the Cloud cluster, a similar computation needs to be performed. Again, we design Data Center utilization for an 80%. In this sense, considering that we run one VM per core. There is only one final consideration that must be taken into account, and is that even though all models are predicted, the data preprocessing is not run every time, only when data are lost. As we have experimentally calculated the probability loss of each sensor separately, to compute the workload we need to obtain the number of GPML instances that need to be run, on average, for each model. Once the probability is obtained, we can use the same formula than for the HPC cluster.

## Appendix B. Migraine savings

The migraine affects approximately 15% population in Europe. This value ranges from 8.8% in Portugal [49] to 24.6% in Switzerland [50]. Each migraine patient leads to costs of €1222 per year in Europe, and according to the study in [5], Linde *et al.* distribute these costs in the way shown in Table B.12. Reducing indirect costs would lead to a huge amount of savings for the National Health Services and private health companies.

The migraine is a stable chronic disease whose affection rates in population vary very slow in time; thus, let's suppose the percentages of migraineurs in Table B.13 are still real (despite some of them are studies from more than 20 years ago). Table B.13 shows the number of migraine sufferers in different European countries. We work with the European population in 2014 [51] (excluding Andorra, Bosnia, Kosovo, Monaco, European area of Russia, San Marino and Armenia).

Second-to-last column in Table B.13 show the amount of patients in the migraine prediction study. We have designed a Data Center able to manage this population (see Section 3.3). This population is 2% of the migraine population.

The N4SID state-space migraine prediction models used in this paper were generated in a previous work [9]. In that paper a methodology was proposed in order to change between prediction models according to the availability of sensors. The probability of failure for each sensor is shown in Table 2. With these probabilities, and the prediction results in [9], the percentage of avoided migraines has been calculated. This value (76.0%) appears as an average result in Table B.14 and indicates the saving costs per patient per year.

Table B.14 shows the results for two patients. Availability (%) means the probability of usage of each kind of model depending on the combination of features (sensors). The availability has been calculated with probabilities of failure in Table 2. Eq. B.1 shows the most general expression to calculate the probability of usage of a kind of model $M_k$, $k = 1, 2, \ldots, 5$.

$$P_U(M_k) = P_A(f_1) * P_A(f_i) * \ldots * P_A(f_N) = (1 - P_F(f_1)) * (1 - P_F(f_i)) * \ldots * (1 - P_F(f_N)) \quad \text{(B.1)}$$

where:

$P_U(M_k)$ is the probability of usage or availability of model $M_k$. $P_A(f_i)$ is the probability of availability of sensor $f_i$, and $P_F(f_i)$ is the probability of failure of sensor $f_i$ in Table 2, $i = 1, 2, \ldots, 4$.

The *Alarms* column in Table B.14 represents the True Positive Rate (TPR) of each model (results in [9]). These models do not report any False Positive (FP) event, thus, TPR equals the percentage of real migraines detected by the system. The more accurate a model ism the more time it is used. Time of usage $T_U(M_{k|k>1})$ (%) of model $M_{k|k>1}$ in Table B.14 is calculated as shown in Eq. B.2:

$$T_U(M_{k|k>1}) = (1 - T_U(M_1) - \ldots - T_U(M_{k-1})) * P_U(M_k) \quad \text{(B.2)}$$

and:

$$T_U(M_1) = P_U(M_1) \quad \text{(B.3)}$$

As an example, for both of the patients, the models of features combinations TEMP-HR-SpO2 only detect 0.3% of all the migraines of each patient. Models detect 63.1% of migraines of Patient A and 88.9% of migraines of Patient B when hierarchical change of models is applied taking into account the probability of failure of each sensor. On average we can extrapolate that this system detects 76.0% of migraines without false alarms.

76.0% are the savings applied for each row in Table B.13. Total savings of migraine costs over the migraineurs to whom one Data Center can provide its prediction benefits leads to € 1271.8 million.

Table B.12: Average direct and indirect costs of migraine disease in Europe per patient per year [5].

| Indirect (93%) | | Direct* (7%) | | | | |
|---|---|---|---|---|---|---|
| Productivity (€) (2/3) | Absenteeism (€) (1/3) | Outpatient care (€) | Diagnostic investigations (€) | Hospitalizations (€) | Acute medications (€) | Prophylactics (€) |
| 758 | 379 | 30 | 19 | 16 | 16 | 5 |

*Acute medications and prophylactics costs are not reduced by predictions of migraines. Patients will continue with the medical treatment.

Table B.13: Migraine sufferers in Europe for the target population of 2% of migraineurs for this research study. Total savings in direct and indirect costs due to migraine prediction with 76% of average rate of prediction success.

| Country | Population [51] | Age range | Migraineurs (%) [Reference] | Total migraineurs | Target population | Savings (M€) |
|---|---|---|---|---|---|---|
| Turkey | 76,667,864 | ≥14 | 19.9 [52] | 15,256,905 | 305,138 | 278.5 |
| UK | 64,308,261 | 16-65 | 14.3 [53] | 9,196,081 | 184,150 | 168.0 |
| Germany | 80,767,463 | ≥20 | 11.4 [54] | 9,207,491 | 183,922 | 167.8 |
| Italy | 60,782,668 | 43±13 | 15.0 [4] | 9,117,400 | 182,348 | 166.4 |
| France | 65,835,579 | ≥18 | 11.2 [55] | 7,373,585 | 147,472 | 134.6 |
| Spain | 46,512,199 | 18-65 | 12.6 [56] | 5,860,537 | 117,211 | 107.0 |
| Netherlands | 16,829,289 | 20-65 | 23.2 [57] | 3,904,395 | 78,088 | 71.3 |
| Switzerland | 8,139,631 | 29-30 | 24.6 [50] | 2,002,349 | 40,047 | 36.5 |
| Sweden | 9,644,864 | 18-74 | 13.2 [58] | 1,273,122 | 25,462 | 23.2 |
| Norway | 5,107,970 | 18-65 | 23.0 [59] | 1,174,833 | 23,497 | 21.4 |
| Denmark | 5,627,235 | 12–41 | 19.1 [60] | 1,074,802 | 21,496 | 19.6 |
| Hungary | 9,877,365 | 15-80 | 9.6 [61] | 948,227 | 18,965 | 17.3 |
| Portugal | 10,427,301 | - | 8.8 [49] | 917,602 | 18,352 | 16.7 |
| Austria | 8,506,889 | ≥15 | 10.2 [62] | 867,703 | 17,354 | 15.8 |
| Croatia | 4,246,809 | 15-65 | 19.0 [63] | 806,894 | 16,138 | 14.7 |
| Georgia | 4,490,498 | ≥16 | 15.6 [64] | 700,518 | 14,010 | 12.8 |
| Total | 477,771,885 | | 15.7±5.2 | 69,682,444 | 1,393,649 | **1271.8** |

Table B.14: Availability of models depending on features and sensors' status. Total of detected events on average to compute accuracy of the prediction system.

| | Model ($M_k$) | Availability (%) | Alarms = TPR (%) | Time of usage (%) | Detected events (%) |
|---|---|---|---|---|---|
| Patient A | EDA-HR-SpO2 | 61.9 | 67.0 | 61.9 | 41.4 |
| | TEMP-EDA-SpO2 | 61.9 | 60.0 | 23.6 | 14.2 |
| | TEMP-EDA-HR-SpO2 | 58.0 | 53.0 | 8.4 | 4.5 |
| | TEMP-EDA-HR | 82.5 | 53.0 | 5.0 | 2.7 |
| | TEMP-HR-SpO2 | 61.9 | 47.0 | 0.7 | 0.3 |
| | Total | | | | 63.1 |
| Patient B | TEMP-EDA-SpO2 | 61.9 | 90.0 | 61.9 | 55.7 |
| | TEMP-EDA-HR | 82.5 | 90.0 | 31.5 | 28.3 |
| | TEMP-EDA-HR-SpO2 | 58.0 | 90.0 | 3.9 | 3.5 |
| | EDA-HR-SpO2 | 61.9 | 60.0 | 1.7 | 1.0 |
| | TEMP-HR-SpO2 | 61.9 | 50.0 | 0.7 | 0.3 |
| | Total | | | | 88.9 |
| | Average | | | | 76.0 |

# Power Transmission and Workload Balancing Policies in eHealth Mobile Cloud Computing Scenarios Biographies

Josué Pagán Marina Zapater José L. Ayala

December 13, 2016

Josué is a Ph.D. student in Complutense University of Madrid (UCM) and works as a research assistant in the Technical School of Telecommunication Engineering in the Technical University of Madrid (UPM). His work focuses on develop robust methodologies for information acquisition in biophysical and critical scenarios. He has worked developing models for prompt prediction and classification of neurological diseases. On summer 2016 he did a 12-week research stay at the Embedded Pervasive Systems Lab at Washington State University under the supervision of Prof. Hassan Ghasemzadeh. Previously, on fall 2015 he did a 16-week research stay at the Pattern Recognition Lab. at Friedrich Alexander University under the supervision of Prof. Björn Eskofier. He achieved his MsC at Universidad Politécnica de Madrid in september 2013 with Honor Mention. He also achieved a Bachelor in Telecommunication Engineering by the Universidad Pública de Navarra in 2010.

Marina Zapater-Sancho is Visiting Professor in the Department of Computer Architecture and Automation at Complutense University of Madrid (UCM), and post-doc researcher at the ESL lab at École Polytechnique Fédérale de Lausanne (EPFL). She received her PhD degree on 2015 from Universidad Politécnica de Madrid, and a Master in Telecommunication Engineering and in Electronic Engineering from the Universitat Politècnica de Catalunya, Spain, both in 2010. She was part-time lecturer during the academic year 2014-2015. She visited the PeacLab group in Boston University during 2014 and 2012 under the supervision of Prof. Ayse Coskun. On 2011, she was awarded a predoctoral fellowship by the Program for Attracting Talent (PICATA) of the Campus of International Excellence of Moncloa. Her research focuses on thermal optimization of complex heterogeneous systems, proactive and reactive thermal-aware optimization of datacenters.

Jose L. Ayala is currently an Associate Professor in the Department of Computer Architecture and Automation at the Complutense University of Madrid. He received his MSc and PhD degrees in Telecommunication Engineering from the Technical University of Madrid, Spain, in 2001 and 2005, respectively. He is member of the HiPEAC European Network of Excellence, IEEE, ACM, IFIP 10.5 and the Council of Electronic Design Automation. He has organized several international events as General Chair and Program Chair, such as VLSI-SoC, GLSVLSI and PATMOS. He has served as TPC member of many conferences, including DATE, DAC, VLSI-SoC, ICCAD, GLSVLSI, etc. He has leaded and participated in a large number of international research projects and bilateral projects with industry, in the fields of power and energy optimization of embedded systems, and non-invasive health monitoring. His current research interests focus on thermal- and energy-aware design and management of processor-based systems, design of embeded processors, thermal estimation, 3D integration, health monitoring and wireless sensor networks.

**Contact Author: Josué Pagán (jpagan@ucm.es)**

josue.png

Click here to download high resolution image

**marina.jpg**

**Click here to download high resolution image**

jayala.jpg

Click here to download high resolution image