

Self-Adaptive Data Collection and Fusion for Health Monitoring Based on Body Sensor Networks

Carol Habib, Abdallah Makhoul, Rony Darazi, Senior Member, IEEE, and Christian Salim

Abstract-In the past few years, wireless body sensor networks (WBSNs) emerged as a low-cost solution for healthcare applications. In WBSNs, biosensors collect periodically physiological measurement and send them to the coordinator where the data fusion process takes place. However, processing the huge amount of data captured by the limited lifetime biosensors and taking the right decisions when there is an emergency are major challenges in WBSNs. In this paper, we introduce a biosensor data management framework, starting from data collection to decision making. First, we propose an adaptive data collection approach on the biosensor node level. This approach uses an early warning score system to optimize data transmission and estimates in real time the sensing frequency. Second, we present a data fusion model on the coordinator level using a decision matrix and fuzzy set theory. To evaluate our approach, we conducted multiple series of simulations on real sensor data. The results show that our approach reduces the amount of collected data, while maintaining data integrity. In addition, we show the impact of sampling and filtering data on the accuracy of the taken decisions and compare our data fusion approach with a basic decision tree algorithm.

Index Terms—Adaptive sampling (AS), body sensor network (BSN), data fusion, early warning system (EWS), fuzzy theory, patient risk.

I. INTRODUCTION

OWADAYS, maintaining the quality of life and increasing the life expectancy are highly important. Therefore, monitoring patients constantly is becoming a requirement. Hence, distant patient monitoring is a solution providing constant surveillance of their vital signs, in order to control efficiently their health condition and to provide urgent treatment when an emergency occurs, such as an abnormal variation of the

Manuscript received September 6, 2015; revised December 7, 2015 and March 2, 2016; accepted May 11, 2016. Date of publication June 2, 2016; date of current version December 6, 2016. This work was supported in part by the Labex ACTION program under Contract ANR-11-LABX-01-01. Paper no. TII-15-1347.R2. (Corresponding author: A. Makhoul.)

- C. Habib, A. Makhoul, and C. Salim are with the Franche-Comt Electronics Mechanics Thermal Science and Optics—Sciences and Technologies Laboratory, Department of Computer Science and Complex Systems, University of Franche-Comté, Belfort 90000, France (e-mail: carol.habib@edu.univ-fcomte.fr; abdallah.makhoul@univ-fcomte.fr; christian.salim@edu.univ-fcomte.fr).
- R. Darazi is with the Telecommunications, Information and Computer Key Enabling Technologies Laboratory, Antonine University, Hadat-Baabda 40016, Lebanon (e-mail: rony.darazi@ua.edu.lb).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2016.2575800

heart rate. Wireless body sensor networks (WBSNs) are a subset of wireless sensor networks (WSNs) that allow continuous monitoring of health data for a given patient. WBSNs collect and analyze vital signs data by deploying different types of biomedical sensors (e.g., body temperature, heartbeat, blood pressure, ECG, EEG, etc.). WBSNs are generally used for inhome monitoring, or for surveilling elderly in nursing homes, avoiding unnecessary hospitalization, thus, reducing the general healthcare costs [1].

Patient monitoring involves periodic transmission of routine vital signs and alerting signals when vital signs cross a certain threshold. We assume a network of biosensors placed on or implanted in the body of patients. The biosensors send the sensed data to the coordinator, located on or near the body. The coordinator is assigned the aggregation, the fusion, and the forwarding of the collected data and its decisions to the sink node. The latter sends the received data to the healthcare center or any other destination [2]. Many challenges arise then in WBSNs, which are the energy consumption due to periodic transmission and the huge amount of heterogeneous raw data captured by the limited energy resources biosensors. Another challenge in WBSN is data fusion which enables combining information from several biosensor nodes to represent the global situation of a patient leading consequently to take a right decision. Our contribution in this paper is twofold.

First, vital signs can vary from critical values to normal values and vice versa, not to forget that the dynamics of the monitored conditions can slow down or speed up regarding the patient's situation. In this purpose and to reduce the energy consumption, the sampling rate in periodic data collection, such as WBSN, depends on how fast the condition varies and at what rate the characteristics need to be captured [3]–[5]. We are specifically interested in establishing an early warning system, where the biosensor nodes are capable of locally detecting emergencies and sending measurements to the coordinator only when a change in the criticality level of the vital sign is observed, thus improving power efficiency. Furthermore, adapting the sensing sampling rate is an effective method to reduce the energy consumption in WBSN. In our work, we propose a distributed adaptive sampling (AS) algorithm that is based on the sensed data variation while taking into account the patient criticality.

Our second objective in this paper is to provide data fusion for WBSN. Our main purpose is to obtain information of greater quality and make accurate decisions about the situation of the patient based on the collected data. Our proposed data fusion scheme uses Fuzzy set theory [6]. Thus, the coordinator

generates appropriate decisions according to the health status of the monitored patient by combining information from various biosensors. The raw data received during consecutive periods are aggregated using fuzzification procedures. Then, the decision having the closest feature values to the aggregated dataset is selected from a decision matrix.

The following of the paper is organized as follows. Section II presents the related work. In Section III, we describe the architecture of a WBSN and the behavior of the biosensor node while introducing an early warning system. The AS rate algorithm as well as the emergency detection mechanism are presented in Section IV. In Section V, we present the technique used for the fusion of several datasets and, therefore, the selection of the corresponding decision is described. Experimental results are given in Section VI. Section VII concludes the paper with some directions to a future work.

II. RELATED WORK

Various aspects and needs in WBSNs have been studied and discussed in the literature. Some of them treated routing issues and QoS such as in [7] and [1]. Others focused on analyzing and fusing the sensed data in order to produce useful information [8], [9]. Several solutions for supporting emergency messages in WBSNs have been proposed in the literature so far [10]-[13]. In [11], Phadat and Bhole proposed to locally classify the captured reading of the vital sign, based on a preset thresholds at each sensor. If the value of the vital sign is in the normal range, the corresponding packet will be classified as a normal packet and is put in a normal queue, otherwise it will be classified as a prioritized packet and is put in a precedence queue. A scheduler chooses first the packet in the priority queue and will put it in a transmission queue. In [10], Ganesan et al. proposed a system to help in finding the abnormalities of heartbeat rate and also medicine intake by the patient using the Bayesian algorithm. The aforementioned work highlights the need of personal caring given to the patient by the hospital, thus reducing unnecessary delay in providing treatment to a patient.

On the other hand, data fusion is used for multiple purposes depending on the aim of the proposed system. In fact, some systems are designed for physical activity recognition, others for monitoring the vital signs of patients using physiological parameters thresholds. While others are designed for aggregating and fusing heterogeneous contextual information to identify the context of the user and generate suggestions for him. In [14], [15], and [16], activity recognition models and systems are proposed. These work approaches use feature extraction, feature selection, and classifiers such as Bayesian networks [16], naive Bayes [16], decision trees (DTs)[15], [16], and K-nearest neighbor [14], in order to perform the information fusion. In [17], a scheme for human behavior recognition on WSNs is proposed. It transmits activities signals compressed by Hamming compressed sensing to the network server and performs behavior recognition by using a dimension reduction algorithm called rank preserving discriminant analysis and a nearest neighbor classifier. However, this approach has limitations since it does not take into consideration the energy consumption constraint in the WSN. In [18], an open-source programming framework called signal processing in node environment (SPINE) is presented and analyzed. This domain-specific framework is designed to support the rapid and flexible development of BSN applications. SPINE supports a variety of sensor and coordinator platforms allowing application designers to freely chose the hardware and software infrastructures. In [8], a new framework called C-SPINE for collaborative body sensor networks (CBSNs) is introduced. A three-layered architecture for multisensor data fusion in CBSNs is proposed. The system is based on a data fusion scheme to perform detection of handshakes between two individuals and capture of possible heart-rate-based emotion reactions due to the individuals meeting.

Unfortunately, the above related work assume that data acquisition and processing have an energy consumption that is negligible compared to the radio communication. Consequently, their research works aim only at minimizing radio transmission. On the other hand, in almost all the previous solutions there was no particular attention related to the optimization of raw data transmission and local emergency detection on the sensor node level. The studies just focused on detecting the emergency at the base station level where all the data are received from different sensors. For instance, some previous work [3], [5], [19] proposed an AS algorithm in order to reduce the sensors activity to periodic sensor networks. However, data transmission is still a significant issue and emergency detection is not being handled. Indeed, an early emergency detection along with energy saving and reduction of the huge amount of raw data captured by the sensors are the major challenges of WBSN. In [20], Elghers et al. proposed the LED algorithm which consists of detecting early emergencies while saving energy. In this approach, the authors suggest to send all critical values captured by the sensors without any sampling frequency adaptation. In this paper, our objective is to bring some modification to the LED* algorithm to further reduce the energy consumption and extend the lifetime of the network. Our aim is to study the sensor node's behavior when it detects a variation in a given vital sign where this variation could indicate an escalating or a deescalating state.

Furthermore, in all of the aforementioned related work, the data fusion process is used to identify either the emotional state of the user (stressed, relaxed, feelings intensity, etc.) or to identify the activity being performed using a set of recognized activities by the system. Therefore, the proposed solutions are very specific and application-oriented and rely on complex techniques. Moreover, the data fusion process is done on huge amounts of unfiltered raw data. Unfortunately, nowadays these methods are very complex, they need high and complex computations, and are not suitable for sensor nodes having limited resources. In our approach, we propose a data fusion technique using Fuzzy theory that allows the coordinator to generate decisions without the need of high processing capabilities.

III. BACKGROUND ON WBSN AND EMERGENCY DETECTION

In this section, we describe the architecture of a WBSN and the behavior of the biosensor node. We assume that a WBSN is deployed on a patient's body. The WBSN is composed of

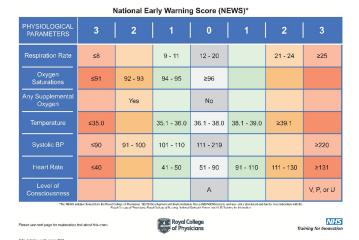


Fig. 1. Early warning system.

biosensors, each one of them sensing one vital sign, which we refer to as a feature. A biosensor node is defined as a traditional sensor node equipped with sensors that monitor vital signs. The role of each sensor in WBSN is to collect measurements in a periodic manner (with a given sampling rate) and send them to the coordinator to perform the data fusion and decision making. A coordinator can be a specific medical device, a mobile phone, a PDA, etc. It is also a gateway to other networks. We believe that the battery of this powerful device can be easily recharged or replaced, then it can be considered as a device unrestricted by power and computing resources. Therefore, it is possible to make a tradeoff between the energy of this device and the lifetime of the limited-energy body sensor nodes.

A. Early Warning Score System

An early warning score system (EWS) is a guide used by emergency medical services staff in hospitals to determine the degree of criticality of patient situation. An EWS is used as a systematic protocol to measure simple physiological parameters in all patients to allow early recognition of those presenting an acute illness or who are deteriorating [21]. For each vital sign, a normal healthy range is defined. Measured values outside of this range are allocated a score which is weighted and color-coded on the observation chart according to the magnitude of deviation from the normal range. The weighting reflects the severity of the physiological disturbance. Fig. 1 shows national EWS (NEWS) used in U.K. [22] that we have used in our experimental tests and examples. In the next section, we will show how we use this EWS in a local emergency detection algorithm.

B. Local Emergency Detection

In a traditional WBSN, each biosensor node collects data and sends them to the coordinator in a periodic manner. Thus, a huge amount of data is collected and sent every period to the coordinator. Therefore, we must find a model which reduces the amount of data while guaranteeing integrity and in the same time optimizes data transmission to reduce the energy consumption

Algorithm 1: Modified Local Emergency Detection Algorithm $Modified\ LED$.

```
Require:R_t (Instantaneous Sampling Rate).
   while Energy > 0 do
       for each period do
         takes first measurement r_0
4)
         sends first measurement r_0
         gets score S of r_0
         takes measurements r_i at R_t Rate
6)
         gets score S_i of measurement r_i
8)
         if S_i!=S then
           sends measurement r_i
10)
           S = S_i
         end if
12:
        end for
   end while
```

on nodes. The first intuition is to send the first captured measurement during a period as well as all the critical measurements to the coordinator as proposed in [20] and known as LED algorithm. Detection of abnormal situations is allowed by providing a local warning system on each node. Thus, the score of each captured data is calculated, which allow us to early and locally detect any emergency represented by a score different from zero. However, data transmission can be further optimized. Indeed, when an emergency is detected it is not always useful to send all the critical data. For instance, suppose a biosensor node capturing the respiration rate is running the LED algorithm. This latter will send huge amount of critical data if the respiration rate of the patient is abnormal for a long time. This case is very common in unstable and deteriorating health conditions where all the data sensed by the biosensor nodes are critical and redundant. Therefore, we propose to modify LED algorithm in order to further optimize data transmission and further reduce the energy consumption of the biosensor nodes and extend their lifetime.

Suppose $s = \{v_0, \dots, v_n\}$ is a series of sensed data at a R_t rate during a period p belonging to a given feature and $s_{\text{scores}} = \{\text{score}(v_0), \dots, \text{score}(v_n)\}\$ is the series of their corresponding scores computed using an EWS. The biosensor sends a sensed data v_i only if its score $score(v_i)$ is different from the score of the previous sent data in the same period (cf., Algorithm 1). Therefore, the transmission is optimized by eliminating the transmission of consecutive sensed data having the same score while maintaining data integrity by sending data each time a new score is detected. For example, suppose $s = \{v_0, v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ is a series of eight consecutive measurements for a given feature, $s_{\text{scores}} = \{1, 1, 0, 2, 2, 2, 2, 0\}$ is the series of the corresponding consecutive scores. By using Modified LED^* , only the following series $\{v_0, v_2, v_3, v_7\}$ is sent to the coordinator. This latter is able to regenerate the original data series since it considers the last value received at time t as the current one, while it has not received any new measurement at time t+1 from the biosensor during the same period.

The *Modified LED* Algorithm optimizes data transmission over the network without considering the energy consumed by data sensing. In the next section, we show how we can adapt the sampling frequency in order to save more node's energy.

IV. AS AND EMERGENCY DETECTION

Several approaches for energy saving in WSN are proposed in the literature. However, the majority of these works consider that data sensing and processing have energy consumption that is negligible compared to data transmission. Consequently, these approaches try to minimize the network's communications. However, this assumption is not always correct especially when the sensors collect data periodically, thus, a huge amount of sensing data is collected. Moreover, medical applications require specific sensors whose power consumption cannot be neglected [23]. For instance, popular radio equipment used in sensor nodes "CC1000" produced by "Texas Instruments" consumes 42 mW (at 0 dBm) for transmission and 29 mW for reception. On the other hand, an accelerometer "iMEMS" by "ADI" consumes 30 mW. Therefore, if we consider that the data acquisition phase is longer than the transmission phase, we can conclude that some sensors may consume more energy than radio communications. As such, the Modified LED algorithm which aims to minimize the network communications need to be complemented with an efficient energy management of the sensors.

A. Adaptive Sampling

In this section, we suggest an AS algorithm that adapts the sampling rates of the sensors to the vital sign dynamic evolution. Therefore, we propose a distributed AS algorithm based on the sensed data variation as explained in [4]. The idea here is to apply the ANOVA model with Fisher test in order to verify during a specific period if there is high variation in the captured measurements. In the affirmative case, the sampling rate must be at its maximum otherwise the sampling rate is adapted according to the variation and to the patient situation/risk. Our goal is to minimize the sensing activity and to reduce the amount of raw data sent to the coordinator. The algorithm operates in rounds where each round is equal to J periods. We define (ST) as the total variation, (SR) the within period variation, (SF) the between period variation, n the number biosensor nodes, J the number of periods, and N as the total number of measurements. Let us consider: ST = SR + SF. Using the Fisher test we consider:

$$F = \frac{\mathrm{SF}/(J-1)}{\mathrm{SR}/(N-J)}.$$
 (1)

Let $F_t = F_{1-\alpha}(J-1,N-J)$ with risk α . The decision is based on F and F_t .

Three situations are possible:

- 1) $F > F_t \Rightarrow$ the variance between periods are significant and the sampling rate is balanced to the maximum sampling rate.
- 2) $F \le F_t \Rightarrow$ the sampling rate is adapted depending on the Fisher test F. It can be increased in order not to miss

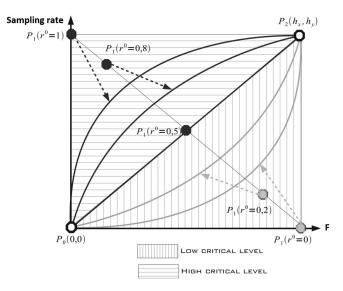


Fig. 2. Sampling rate adaptation using the behavior curve functions.

important measurements or decreased when the variance is less than the threshold.

3) If N < J, the sampling rate is balanced to the maximum sampling rate.

To fully exploit the sensor node capabilities, we propose to adapt the sampling rate of a node based on the result of the Fisher test *F* as shown in Fig. 2. Based on the results and residual of the variance test described above, the idea is that when a node notices high variance differences, it increases its sampling rate in order to prevent missing important measurements and decreases its sampling rate when the variance is less than the threshold. However, it is desirable to be able to adjust the sampling rate according to each patient situation.

We define the patient situation as a patient risk level and we express this level by a quantitative variable r^0 which can take values between 0 and 1 representing the low and the high risk level, respectively. Two risk levels are defined:

- 1) Level 1 "Patient with low risk," $0 \le r^0 < 0.5$: In some cases, patients should be monitored with low level of criticality. These patients include elderly people in caring houses, who are in good shape but need to be monitored occasionally. In other words, the biosensor nodes will preserve energy by sampling slowly.
- 2) Level 2 "Patient with high risk," $0.5 \le r^0 \le 1$: In other cases, patients should be monitored with a high level of criticality. These patients can be acute ill or at home after a surgical intervention.

On the one hand, we can see in Fig. 2 that the sampling rate of patients with a low risk level is represented by a hyperbolic BV function. Thus, the majority of the sensors will preserve their energy by slowing down the sampling process if there is no need to consider the maximum sampling rate. On the other hand, the sampling rate of patients with high risk level is increased. This characteristic is represented by a parabolic BV function. Thus, the majority of nodes capture measurements at a high sampling rate because of the patient situation. To find the exact sampling

rate according to the results of Fisher test and the risk r^0 , we use a behavior function BV which is is expressed by a Bezier curve that passes through three points as shown in Fig. 2. The three points are (0,0), $(F_t, MaximumSamplingRate)$, and r^0 .

Algorithm 2: Modified Local Emergency Detection with Adaptive Sampling Algorithm $Modified\ LED^*$.

```
Require: m (1 round = m periods), R_{\text{max}} (maximum
    sampling rate).
Ensure: R_t (instantaneous sampling speed).
    R_t \leftarrow R_{\text{max}}
    while Energy > 0 do
       for each round do
           for each period do
              Run Modified LED (Emergency Detection)
 6)
           end for
           compute SR, SF and F.
           if N < m then
 9)
              R_t \leftarrow R_{\text{max}}
           else
              find F_t
 12)
              if F < F_t then
                R_t \leftarrow BV(F, F_t, r^0, R_{\text{max}}) (BV behavior
                  R_t \leftarrow R_{\max}
 15)
           end if
 18)
        end for
    end while
```

B. Adaptive LED Algorithm

Using the behavior function (quadratic Bezier curve) and according to the Fisher test results for the ANOVA model, $Modified\ LED^*$ is proposed to adapt the sampling rate according to the variances within the captured measurements during one period and between J periods. $Modified\ LED^*$ takes into consideration the criticality of each measurement, by verifying the score condition. Thus, the $Modified\ LED$ algorithm is optimized in order to allow biosensor nodes to adapt their sampling rate according to the comparison between the means and variances of the measurements taken in different periods (cf., Algorithm 2). It is a recurrent situation in WBSN that sensed data is the same during several periods. In this case, the sampling rate is balanced to the minimum sampling rate during the next period.

After presenting the AS rate algorithm in the next section, we present how the coordinator can fuse the received data and take the right decisions.

V. DATA FUSION SCHEME

A. Data Fusion and Decision Making

In our approach, we assume that one sensor is activated to monitor one of the features of interest. Assuming there are m

features to be monitored, there will be m sensors with the ith sensor observing feature F_i . The m readings are to be aggregated by the data fusion processor to reach a decision concerning the occurrence of an event of interest (e.g., an infraction or an emergency assessment).

We assume each coordinator fusion processor is provided with a local decision matrix D defined as:

$$D = [D_1, D_2, ..., D_n]$$

where D_k is a vector of score values corresponding to feature values ($[f_{1,k}, f_{2,k}, ..., f_{m,k}]$) and supporting decision d_k .

As an example, let us consider an application where a sensor network is used for the detection if any supplemental oxygen must be given to the patient or not. The set of monitored features may consist of the following:

 F_1 : Respiration rate.

 F_2 : Oxygen saturation.

 F_3 : Heart rate.

The decisions are:

 d_1 : supplemental oxygen is needed.

 d_2 : no supplemental oxygen is needed.

A decision matrix can be defined based on any EWS. We chose the NEWS [22] as an example [the elements of the matrix corresponds to the score (NEWS) see Fig. 1]:

$$\begin{pmatrix} d_1 & d_2 \\ -- & -- \\ f_{11} & f_{12} \\ f_{21} & f_{22} \\ f_{31} & f_{32} \end{pmatrix} = \begin{pmatrix} d_1 & d_2 \\ -- & -- \\ 1 & 0 \\ 2 & 0 \\ 1 & 1 \end{pmatrix}$$

where f_{ij} is the score of decision d_j corresponding to feature F_i .

The above example decision matrix indicates that if feature F1 (respiration rate) is between 9 and 11, feature F2 (oxygen saturations) is between 92 and 95, and F3 (heart rate) is greater than 91 and less than 110 then decision d_1 is taken (i.e., supplemental oxygen is needed); and if F1 (respiration rate) is between 12 and 20, feature F2 (oxygen saturations) is greater or equal to 96, and F3 (heart rate) is greater than 91 and less than 110 then decision d_2 is taken (i.e., no supplemental oxygen is needed).

Given an actual sampling vector S collected by m sensors during one period and composed of m scores corresponding to m readings/features is represented as follows:

$$S = [s_1, s_2, ..., s_m].$$

The coordinator node takes decision d_i so that

$$\sum_{i=1}^{m} (f_{ij} - s_i)^2 \le \sum_{i=1}^{m} (f_{ik} - s_i)^2$$
 (2)

for all decisions d_k and for all k going from 1 to n.

Note that $\sum_{i=1}^{m} (f_{ik} - s_i)^2$ is a measure of how close the actual data collected by the sensors is to the feature values expected to support decision d_k . Then, we define the strength of a decision d_k , Str_k , as the inverse of the Cartesian distance as follows:

$$Str_k = \frac{1}{\sum_{i=1}^{m} (f_{ik} - s_i)^2}.$$
 (3)

The smaller the distance $\sum_{i=1}^{m} (f_{ik} - s_i)^2$ of the data reading scores S to the feature values expected to supported decision d_k , the stronger the decision.

These decisions can be taken at one period, where only one set of reading is sent to the coordinator. Otherwise, in WBSN the decision is taken based on several reading values corresponding to several number of periods. In the next section, we describe how we use fuzzy sets in order to allow the coordinator to make a decision based on several data readings.

B. Several Datasets Fusion

We start this section by a brief overview of fuzzy sets.

1) Overview on Fuzzy Sets: "Fuzzy sets are sets whose elements have degrees of membership" [6]. A fuzzy set is composed of a set U and a membership function $M: U \rightarrow [0,1]$. The membership function is a generalization of the characteristic function of an ordinary set.

For each $x \in U$, the value M(x) is called the grade of membership of x in (U, M). It denotes the degree to which the element x is a member of fuzzy set U.

For an ordinary set $U = \{x_1, \ldots, x_n\}$, we have

$$M(x) = \begin{cases} 1, & \text{if } x \in U \\ 0, & \text{otherwise.} \end{cases}$$

Then, x is called not included in the fuzzy set of U if M(x)=0, x is called fully included if M(x)=1, and x is called a fuzzy member if 0 < M(x) < 1. For a fuzzy set, $0 \le M(x) \le 1$. For notational convenience, we do not distinguish between the membership function and the fuzzy set itself. Therefore, the membership function M is the fuzzy set of U. When the set $U=\{x_1,\ldots,x_n\}$ is finite, we represent fuzzy set M as $M=\{M(x_1)/x_1,\ldots,M(x_n)/x_n\}$, where M(x) denotes the degree to which x belongs to M or the confidence of the belief that x belongs to M.

2) Decision Making After Several Readings: As in the one period readings case, the coordinator uses a decision matrix: $D = [D_1, D_2, ..., D_n]$, where D_l is a score vector for features $[f_{1l}, f_{2l}, ..., f_{ml}]$ of the ideal score values supporting decision d_l . Ideally, if the score values computed by the coordinator (based on its sensor readings) are exactly equal to vector D_l , then the coordinator (i.e., the data fusion processor) should take the decision d_l . However, in the case of several readings sets the

computed score values are not the same and will not all match exactly any of the vectors of D. The objective of our decision-making process is to select the decision that matches best the computed scores values.

After p periods, each sensor took p readings corresponding to p scores of each feature F. This forms a set of scores $S(F) = \{s_1, s_2, ..., s_p\}$ of the feature F. Then, we define the frequency $\operatorname{Freq}(s)$ of the score s as the number of the subsequent occurrence of the same score in the same set S(F). For instance, if after six readings of the feature F (respiration rate) we obtain $S(F) = \{1, 1, 0, 1, 0, 2\}$, then we have $\operatorname{Freq}(0) = 2$, $\operatorname{Freq}(1) = 3$, and $\operatorname{Freq}(2) = 1$.

Let $S = [\hat{s}_1, \hat{s}_2, ..., \hat{s}_m]$ be a set of fuzzy membership functions computed by the coordinator based on its sensor readings, where \hat{s}_i is the membership function for feature F_i . Note that each \hat{s}_i is a fuzzy membership function computed using the procedure described in the previous section. Using the notation of the previous section:

$$\hat{s}_i = \{M(s_{i1})/s_{i1}, \dots, M(s_{ip})/s_{ip}\}.$$

In this approach, we define the membership function as:

$$M(s_{ik}) = \frac{\operatorname{Freq}(s_{ik})}{\sum_{j=1}^{p} \operatorname{Freq}(s_{ij})}$$
(4)

and the strength of decision d_l as:

$$Str_{l} = \min_{i=1}^{m} (\max_{k=1}^{q} (M(s_{ik})e^{-(s_{ik}-f_{il})^{2}}))$$
 (5)

where $q = |\hat{s}|$ of feature m.

Noting that the min function produces the weakest link among a set of series links and the max function generates the strongest link among a set of parallel links, we use the max function to weed out readings that either have low confidence level or large distance to the ideal value for each feature, then we use the min function to represent the strength of a decision with the strength of its weakest feature reading. Other aggregate functions such as the mean can also be used instead of the min and the max functions.

As an example, let us assume the decision matrix below with two features F_1 and F_2 and decisions d_1 , d_2 , and d_3 :

$$\begin{pmatrix} d_1 & d_2 & d_3 \\ -- & -- & -- \\ 1 & 3 & 0 \\ 2 & 0 & 1 \end{pmatrix}.$$

Assume that the coordinator got the following feature membership values based on its sensor readings and using the fuzzification procedure of the previous section:

$$\hat{s}_1 = \{0.8/1, 0.2/3\}$$

$$\hat{s}_2 = \{0.6/0, 0.4/2\}.$$

Intuitively, the readings can support decisions d_1 , d_2 , and d_3 since d_1 and d_2 's values match the readings values but with different levels of confidence and d_3 's values are close to the reading values. However, it is clear to see that d_1 should be the strongest since its values are the closest to the readings with

the highest confidence levels. Applying (5), we get the strength of each decision:

$$\begin{split} Str_1 &= & \min[\max(0.8e^{-(1-1)^2}, 0.2e^{-(3-1)^2}), \\ & & \max(0.6e^{-(0-2)^2}, 0.4e^{-(2-2)^2})] \\ Str_2 &= & \min[\max(0.8e^{-(1-3)^2}, 0.2e^{-(3-3)^2}), \\ & & \max(0.6e^{-(0-0)^2}, 0.4e^{-(2-0)^2})] \\ Str_3 &= & \min[\max(0.8e^{-(1-0)^2}, 0.2e^{-(3-0)^2}), \\ & & \max(0.6e^{-(0-1)^2}, 0.4e^{-(2-1)^2})]. \end{split}$$

Then, we find: $Str_1 = 0.4$, $Str_2 = 0.2$, and $Str_3 = 0.22$. Based on the above results, decision d_1 is the strongest, which intuitively is what we expected, while decisions d_2 and d_3 are weaker with d_2 slightly weaker than d_3 .

3) Decision-Making Algorithm: In this section, we describe the proposed decision-making algorithm. The data fusion mentioned algorithm is implemented at the coordinator level. We assume that we have a WBSN composed of m biosensors and one coordinator. The role of each sensor in a WBSN is to collect measurements in a periodic manner (with a given sampling rate). The collected data are sent to the coordinator where the data fusion process takes place. We assume that the coordinator has a monitoring period P such as $P = k \times p$ with $k \in \mathbb{N}^*$. Based on the *Modified LED** algorithm presented above, the biosensors send only the first sensed measurement at the beginning of each period p as well as the measurements sensed during p which indicate a change in the criticality level of the captured vital signs, if any exist. The coordinator should take the appropriate decisions depending on the received data (cf., Algorithm 3).

Algorithm 3: Decision-making and monitoring algorithm DMM

Require: P (monitoring period (round) = k * p), k (number of sensor periods p) while Energy > 0 do for each round do

Read received measurements from each sensor for each period \boldsymbol{p}

4) Compute *Score* of each measure and increment its occurrence.

if Score! = 0 then

Send query to the other sensors asking for measurements at current time

8) Update fuzzy sets for each sensor Take Decision using fuzzy procedure

end if

Take decision at the end of monitoring period (round)

end for

12) end while

4) Scenario and Illustrative Example: At each monitoring period (round), the coordinator receives at least k mea-

surements according to $Modified\ LED^*$ from each biosensor. When reading a new measurement, the coordinator identifies the sending biosensor, computes the score using an EWS, and increments the occurrence of the computed score for the corresponding feature. Then, it checks whether the score is different from the normal score zero. In the affirmative case, it queries all the other biosensors to get their sensing data at the same time of criticality. This allows the coordinator to make sure that the patient's health is at risk. The coordinator updates the frequencies of the scores for each feature (biosensor) and then updates the fuzzy sets in order to take a decision by applying the data fusion process explained in Section V-B. The coordinator takes a decision based on the collected measurements sent in the previous periods of the current round and the ones sent at the critical detection moment. At the end of each monitoring period P, the coordinator takes a decision called global decision based on the fuzzy sets of all the monitored features during this round. It is called a global decision because it is not triggered by an emergency but is required for monitoring purposes and is based on the overall health condition of the patient during one round (cf., Algorithm 3).

In the next section, we present our simulation results.

VI. EXPERIMENTAL RESULTS

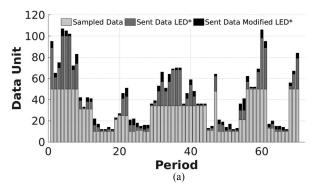
To verify our suggested approach, we conducted multiple series of simulations using a custom Java-based simulator. The objective of these simulations is first to confirm that our adaptive data collection and detection technique can successfully detect locally any emergency while taking into consideration desirable energy conservation objectives. Second, we show that our data fusion technique can cope with our adaptive data collection and detection technique. Therefore, in our simulations, we used real medical readings collected from the online MIMIC Database [24].

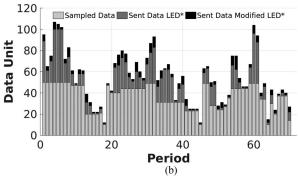
We have run the different algorithms during 70 periods (approximately 2 h) using a Fisher risk $\alpha=0.05$. In this paper, we are interested in two fields of biosensor measurements: the respiration rate and the body temperature. We have taken into consideration two different situations for a patient, low and high risks, respectively. We have evaluated the performance of the approach using the following parameters: 1) the time t (periods), 2) m the number of periods per round, and 3) the patient criticality level (r^0). We use four metrics in our simulations:

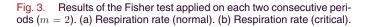
- 1) the instantaneous sampling speed after each round which reflects the amount of data reduction,
- 2) the energy dissipation,
- 3) the data integrity,
- 4) the decision making.

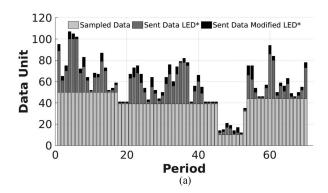
A. Instantaneous Sampling Rate and Data Reduction

The main goal of this section is to show how our algorithm is able to reduce and adapt its sampling rate according to the patient risk level. We consider two situations, a patient with low risk level ($r^0 = 0.4$) and a patient with high risk level ($r^0 = 0.9$). In the following, we will use normal and critical situations to identify these two situations, respectively.









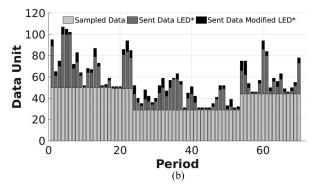


Fig. 4. Results of the Fisher test applied on each three consecutive periods (m=3). (a) Respiration rate (normal). (b) Respiration rate (critical).

In Figs. 3 and 4, we show the number of sampled data in each period. We have fixed the maximum sampling rate to 50 measurement per period and the minimum sampling rate to 10 measurement per period. Then, we compare the quantity of sent data between LED^* (LED coupled with the AS) and Modified LED^* (Modified LED coupled with AS).

First, it is clear to see that our approach adapts the sampling frequency according to the patient's situation. We compare the results obtained for a normal patient $(r^0=0.4)$ and a critical patient $(r^0=0.9)$. When comparing Fig. 3(a) and (b), we can see that the sampling rate in the case of a critical patient presents higher values over the periods where the sampling rate is adapted using similar periods (same Fisher test result). For example, at period 9, the sampling rate is decreased to 47 in the case of a critical patient, however, it is decreased to 31 when the patient is normal. In fact, when the situation of the patient is critical, it is a necessity to monitor the physiological parameters with a higher sampling rate, in order to keep tracking any changes which might have effects on the patient's health.

Second, another parameter we took into consideration in our simulations is the number of m periods per round. This parameter indicates to the biosensor nodes after how many consecutive periods they must apply the ANOVA model to find the instantaneous sampling rate. We compare the results of our approach while assigning the values 2 and 3 to m. These comparisons are made in Figs. 3 and 4. We can observe that the sampling rate varies much more for low values of m (2 in our case) than high values. This means that when m increases (m = 3), the variation between the sensed measurement increases also. The

sampling rate becomes constant near the maximum sampling rate especially when there is high variations in the monitored feature. For example, if we consider the respiration rate (critical case): when m=2 (round $=2\times$ period), the variation between the measurement remains important but without high variations contrary to when m=3 (round $=3\times$ period). Therefore, when m=2, the sampling rate varies more precisely with the monitoring needs of the biosensor. This is due to having a standard deviation between the measurement lower than the one when m=3.

Third, we compare the quantity of sent data in each period when adopting LED^* and $Modified LED^*$. We can observe that both algorithms minimize the amount of data transmitted to the coordinator (not all the sampled data are sent). In the case of the temperature, LED^* and $Modified LED^*$ have the same performance since the vital sign presents stable normal score measurements over the 70 periods. In both the algorithms, only the first sensed data in a period is sent. However, in the respiration rate case, $Modified\ LED^*$ algorithm outperforms LED^* and allows data reduction 50% more than LED^* and from the sampled data. The reason behind is that the respiration rate of this patient is abnormal and presents critical scores for the majority of the periods. LED^* sends all the critical sensed data during a period and, therefore, it is not reducing the transmitted data compared to the sensing data in this case. However, Modified LED^* sends only the measurements indicating changes in the respiration rate state and, therefore, reduces redundancy and optimizes the transmission. Data integrity is studied next by

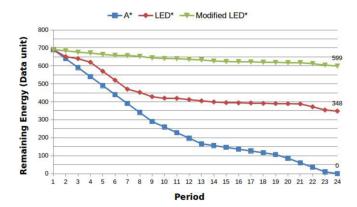


Fig. 5. Comparison of residual energy on the respiration rate node.

showing the impact of applying AS for collecting the data on the sensor node level (see Section VI-C).

B. Energy Consumption

Fig. 5 illustrates the energy consumption on the node responsible of capturing the respiration rate. We assume that the node has an energy level arbitrarily fixed to 700 units. Each captured and sent measurement consumes 0.3 and 1 unit, respectively. The values correspond to a normal patient over 24 periods (40 min). We have compared our algorithm $Modified LED^*$ with A^* and LED^* . All of the three algorithms adapt the sampling rate of the node to the respiration rate dynamic evolution (Fisher test m=2 and $\alpha=0.05$). However, in A^* all the sensed data are sent, in LED^* the sent data are determined by LED (cf., Algorithm 1), and in $Modified LED^*$ the sent data are determined by Modified LED (cf., Algorithm 2 and 3). As shown in Fig. 5, the $Modified LED^*$ algorithm consumes less energy than LED^* and A^* algorithms since the transmission is optimized. We can see that our algorithm saves energy up to three to four times more than LED^* and and up to seven times more than A^* .

C. Data Integrity

In this section, we examine the effect of AS on data integrity. We have run the AS algorithm (m=2, see Section IV) for 70 consecutive periods (approximately 2 h). We have fixed the maximum sampling rate to 50 measurements/period and the minimum sampling rate to 10 measurements/period. Then, we have compared the sensed data collected at each period to the sensed data collected when we do not apply AS on the node. This is done by comparing the distribution of scores (NEWS). Originally, the sensed data = 100 measurements/period (no AS case).

In Table I, we can clearly see that the AS does not influence considerably on the distribution of scores and, therefore, on the integrity of data and information required for the decision making. Since the temperature of the patient is normal over the 70 periods, when adapting the sampling rate we do not lose information (average difference in distribution = 0.02%) and we reduce data to 75.8% when $r^0 = 0.9$ and to 85.7% when

TABLE I

DATA REDUCTION AND AVERAGE OF DIFFERENCE BETWEEN THE

DISTRIBUTIONS OF SCORES IN THE AS CASE AND THE NO ADAPTATION OF

SAMPLING CASE DURING ONE PERIOD

Patient risk Average of difference between scores of adaptive sampled data and non sampled	Respiration rate		Temperature	
	0.4 4.5%	0.9 2.9%	$0.4 \\ 0.02\%$	0.9 0.02%
data during one period Data reduction	71.8%	63.5%	85.7%	75.8%

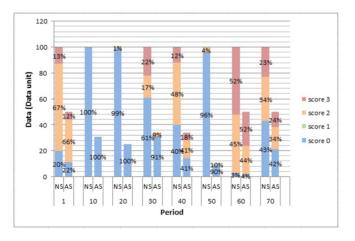


Fig. 6. Comparison between the quantity of sensed data and the distribution of scores when using AS and when no AS is applied (NS) on the respiration rate node.

 $r^0=0.4$. However, the respiration rate of the patient is unstable and presents many variations and critical scores. Therefore, it is very important to make sure that when applying AS, we do not lose important measurements and change the distribution of scores during periods. Our results show (see Table I) that the average difference between the distribution of scores obtained when we apply AS and when we do not is only 4.5% during one period, while the data reduction is about 71.8%.

Fig. 6 shows the quantity of sensed data and the distribution of scores when using AS and when not using AS (NS) on the respiration rate node ($r^0=0.4,\ m=2$). The results of 8 periods chosen from the 70 periods show that when using the AS algorithm we reduce the quantity of sensed data by 64.5% while maintaining 88% of the time very close distributions to the original ones (NS). For example in period 1, AS reduces the sensed data to 50% compared to NS while maintaining approximately the same distributions of scores. Confirming that AS has no influence on the decisions taken by the coordinator.

D. Decision Making

In this next section, we compare the decisions taken by the coordinator when the biosensor nodes run on the one hand $Modified\ LED$ and on the other hand $Modified\ LED^*$ ($Modified\ LED$ coupled with AS). Then, we compare the performance of our data fusion model with the basic DT algorithm.

Decision Accuracy: The Modified LED algorithm (cf., Algorithm 1) preserves data integrity regardless of the sequence of sensed data, since it sends data to the coordinator each time a new score is detected. Moreover, the coordinator considers the last measurement received from a biosensor x valid while it has not received any new measurement during one round. Therefore, the original sequence is regenerated at the coordinator level. We have run the proposed DMM algorithm for 70 periods (round = period) and we have compared the decisions taken (immediate and global) during each round when the biosensors run the Modified LED to the ones taken when the biosensors run the $Modified\ LED^*\ (m=2,\ r^0=0.4)$. The monitored features are the following: temperature, oxygen saturation, heart rate, respiration rate, and systolic blood pressure. Therefore, the WBSN is composed of five bisosensors. The decision matrix used in our simulations is composed of six decisions. The results show that about 99% of the decisions taken on the original data ($Modified\ LED$) are the same as the ones taken on the sampled data ($Modified LED^*$). Therefore, our data fusion model based on fuzzy procedures is tolerant to the changes in the distribution of scores due to our AS algorithm (see Fig. 6 and Table I).

Comparison to a basic DT approach: We have constructed a basic DT based on the decision matrix used in our simulation (shown above). We have run the DMM algorithm but instead of applying our fuzzy data fusion approach, we have used the DT algorithm to take decisions. Therefore, a decision is taken only and only if the current scores of the five features correspond exactly to one of the decisions supported by the DT. The simulation is run over 70 consecutive periods. The results show that only 10% of the decision-making processes were successful where the current scores of the five features corresponded exactly to one of the decisions supported by the DT. On the contrary, our data fusion approach takes into account the fuzzy sets (confidence level) of each feature and chooses the decision that has the feature values the closest to the readings described by fuzzy sets.

VII. CONCLUSION

In this paper, we have proposed a new framework for the data management and processing in WBSNs. We have conducted a series of simulations on real medical data recordings to show the effectiveness of our algorithms and approaches. The results show that our approach reduces considerably the sensed and the transmitted data and the energy consumption while maintaining data integrity and decision accuracy. Monitoring the health condition of employees in factories or in chemical laboratories where accidents might occur could be interesting for the industry field. Vital signs are highly correlated with the surrounding environment. This solution alerts the employees when their vital signs become abnormal. Therefore, allowing the employers to automatically detect accidents and ensure safe work conditions that suits the health condition of the employees. In addition, this solution can monitor the stress level of employees allowing the employers to ensure better working conditions. For future work, we intend to test our proposed scheme in a reallife WBSN application and to propose a method for the fusion and the aggregation of heterogeneous data in a context-aware WBSN.

REFERENCES

- J. I. Bangash, A. H. Abdullah, M. H. Anisi, and A. W. Khan, "A survey of routing protocols in wireless body sensor networks," *Sensors*, vol. 14, no. 1, pp. 1322–1357, 2014.
- [2] N. Bradai, L. C. Fourati, and L. Kamoun, "WBAN data scheduling and aggregation under WBAN/WLAN healthcare network," Ad Hoc Netw., vol. 25, Part A, pp. 251–262, 2015.
- [3] A. Makhoul, D. Laiymani, H. Hareb, and J. Bahi, "An adaptive scheme for data collection and aggregation in periodic sensor networks," *Int. J. Sensor Netw.*, vol. 18, no. 1/2, pp. 62–74, 2015.
- [4] D. Laiymani and A. Makhoul, "Adaptive data collection approach for periodic sensor networks," in *Proc. 9th Int. Wireless Commun. Mobile Comput. Conf.*, Jul. 2013, pp. 1448–1453.
- [5] A. Makhoul, H. Harb, and D. Laiymani, "Residual energy-based adaptive data collection approach for periodic sensor networks," *Ad Hoc Netw.*, vol. 35, pp. 149–160, 2015.
- [6] B. Davvaz and I. Cristea, Fuzzy Algebraic Hyperstructures: An Introduction. Berlin, Germany: Springer, 2015.
- [7] J. Bangash, A. Abdullah, A. Khan, M. A. Razzaque, and R. Yusof, "Critical data routing (CDR) for intra wireless body sensor networks," *TELKOM-NIKA*, vol. 13, no. 1, pp. 181–192, 2015.
- [8] G. Fortino, S. Galzarano, R. Gravina, and W. Li, "A framework for collaborative computing and multi-sensor data fusion in body sensor networks," *Inf. Fusion*, vol. 22, pp. 50–70, 2015.
- [9] L. Chen, C. Nugent, and H. Wang, "A knowledge-driven approach to activity recognition in smart homes," *Knowl. Data Eng.*, vol. 24, no. 6, pp. 961–974, 2011.
- [10] M. Ganesan, A. P. Kumar, S. K. Krishnan, E. Lalitha, B. Manjula, and J. Amudhavel, "A novel based algorithm for the prediction of abnormal heart rate using Bayesian algorithm in the wireless sensor network," in *Proc. Int. Conf. Adv. Res. Comput. Sci. Eng. Technol.*, 2015, pp. 53:1–53:5.
- [11] D. Phadat and A. Bhole, "Sensor network for patient monitoring," Int. J. Res. Advent Technol., vol. 2, no. 1, Jan. 2014. [Online]. Available: http://www.ijrat.org
- [12] E. Seto et al., "Opportunistic strategies for lightweight signal processing for body sensor networks," in Proc. 3rd Int. Conf. Pervasive Technol. Related Assistive Environ., 2010, pp. 56:1–56:6.
- [13] J.-Y. Chang and P.-H. Ju, "An energy-saving routing architecture with a uniform clustering algorithm for wireless body sensor networks," *Future Gener. Comput. Syst.*, vol. 35, pp. 128–140, 2014.
- [14] L. Atallah, B. Lo, R. King, and G. Z. Yang, "Sensor placement for activity detection using wearable accelerometers," in *Proc. Int. Conf. Body Sensor Netw.*, 2010, pp. 24–29.
- [15] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," ACM SIGKDD Explorations Newslett., vol. 12, no. 2, pp. 74–82, 2010.
- [16] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "Fusion of smartphone motion sensors for physical activity recognition," *Sensors*, vol. 14, pp. 10146–10476, 2014.
- [17] D. Tao, L. Jin, Y. Wang, and X. Li, "Rank preserving discriminant analysis for human behavior recognition on wireless sensor networks," *IEEE Trans. Ind. Informat.*, vol. 10, no. 1, pp. 813–823, Feb. 2014.
- [18] G. Fortino, R. Giannantonio, R. Gravina, P. Kuryloski, and R. Jafari, "Enabling effective programming and flexible management of efficient body sensor network applications," *IEEE Trans. Human-Mach. Syst.*, vol. 43, no. 1, pp. 115–133, Jan. 2013.
- [19] C. Alippi, G. Anastasi, M. Di Francesco, and M. Roveri, "An adaptive sampling algorithm for effective energy management in wireless sensor networks with energy-hungry sensors," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 2, pp. 335–344, Feb. 2010.
- [20] S. Elghers, A. Makhoul, and D. Laiymani, "Local emergency detection approach for saving energy in wireless body sensor networks," in *Proc.* IEEE 10th Int. Conf. Wireless Mobile Comput., Netw. Commun., Oct. 2014, pp. 585–591.
- [21] G. K. Ragesh and K. Baskaran, "A survey on futuristic health care system: WBAN," *Procedia Eng.*, vol. 30, pp. 889–896, 2012.
- [22] National Early Warning Score (NEWS), Royal College of Physicians, London, U.K., May 2015. [Online]. Available: http://www.rcplondon. ac.uk/resources/national-early-warning-score-news

- [23] V. Raghunathan, S. Ganeriwal, and M. Srivastava, "Emerging techniques for long lived wireless sensor networks," *IEEE Commun. Mag.*, vol. 44, no. 4, pp. 108–114, Apr. 2006.
- [24] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," Circulation, vol. 101, no. 23, pp. e215–e220, Jun. 2000.



Carol Habib received the M.S. degree in computer science and telecommunication engineering from Antonine University, Hadat-Baabda, Lebanon, and the M.S. degree in mobile and distributed computing from the University of Franche-Comté (UFC), Belfort, France, both in 2015. She is currently working toward the Ph.D. degree at the University of Franche-Comté.

Her research interests include wireless sensor networks, multisensor data fusion, and data aggregation.



Abdallah Makhoul received the M.S. degree in computer science from INSA Lyon, Lyon, France, in 2005, and the Ph.D. degree in the problems of localization, coverage and data fusion in wireless sensor networks from the University of Franche-Comté, Belfort, France, in 2008.

Since 2009, he has been an Associate Professor with the University of Franche-Comté. His research interests include Internet of Things, structural health monitoring, and real-time is-

sues in wireless sensor networks.

Dr. Makhoul has been a TPC Member of several networking conferences and a Reviewer for several international journals.



Rony Darazi (SM'16) received the M.S. degree in computer science and telecommunication engineering from Antonine University (UA), Baabda, Lebanon, and the Ph.D. degree in engineering sciences from the Université catholique de Louvain (UCL), LouvainlaNeuve, Belgium, in 2005 and 2011, respectively.

His Ph.D. was entitled "Towards a combining scheme for compression and watermarking for 3D stereo images." He is currently an Associate Professor at UA. He was a Researcher in the

ICTEAM Institute at UCL from 2006 until 2012, and is a Member of the TICKET Lab at UA since 2010. His research interests include information security and digital watermarking, digital 2D and 3D image processing, sensor networks, and eHealth.

Dr. Darazi cochaired the International Conference on Applied Research in Computer Science & Engineering, sponsored by IEEE in 2015, and has been actively involved as a Reviewer of the Signal, Image, and Video Processing Journal (Springer), IEEE TRANSACTIONS ON INFORMATION FORENSICS & SECURITY and the International Conference on Image Processing. In 2009, he received the Best Paper Award, second prize by the Digital Watermarking Alliance, and the IS&T/SPIE International Conference on Media Forensics and Security XII.



aggregation.

Christian Salim received the M.S. degree in computer science and telecommunication engineering from Antonine University, Hadat-Baabda, Lebanon, and the M.S. degree in distributed and mobile computing from the University of Franche-Comté (UFC), Belfort, France, both in 2015. He is currently working toward the Ph.D. degree at UFC, codirected by the Antonine University

His research interests include wireless sensor networks, multisensor data fusion, and data