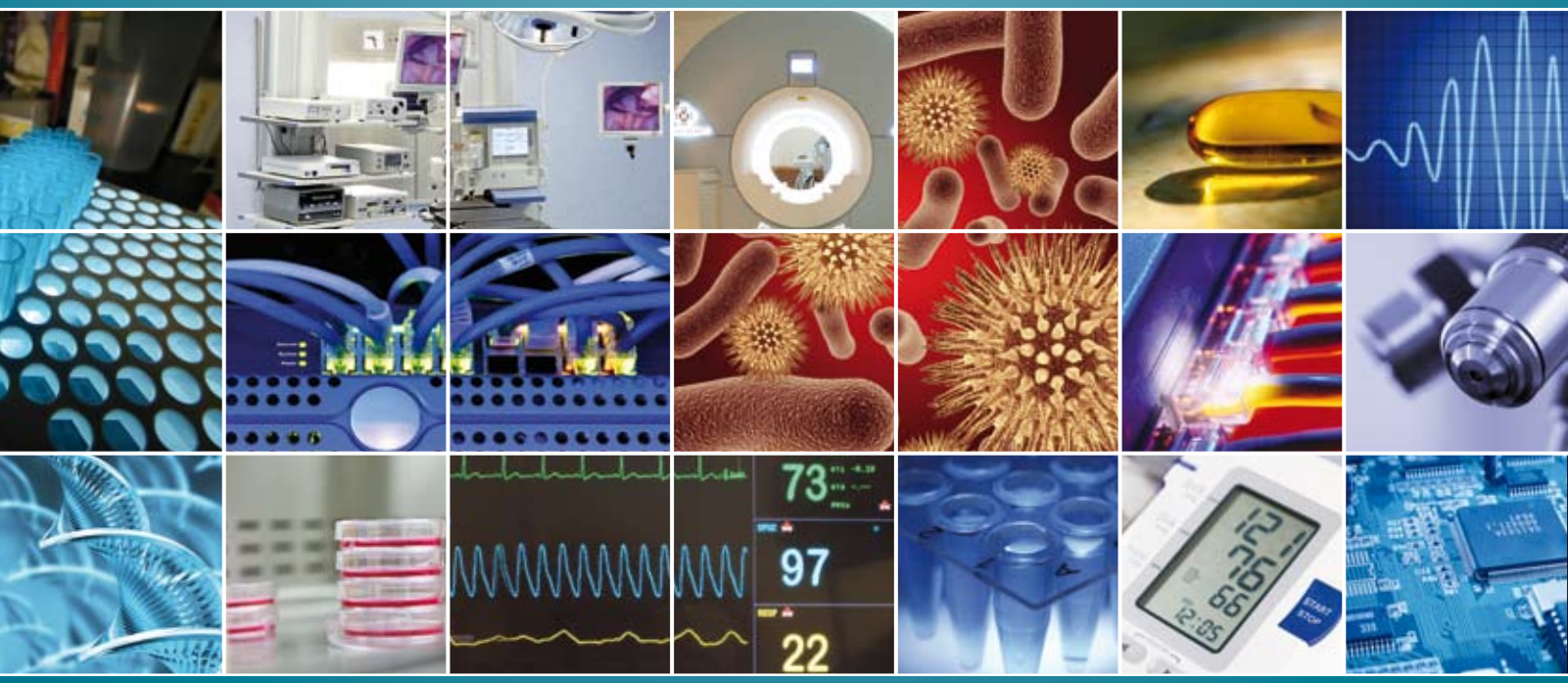
International Journal of Telemedicine and Applications

Pervasive Health Care Services and Technologies

Guest Editors: Laurence Yang, Sajid Hussain, Frédérique Laforest, and Christine Verdier



Pervasive Health Care Services and Technologies

International Journal of Telemedicine and Applications

Pervasive Health Care Services and Technologies

Guest Editors: Laurence Yang, Sajid Hussain, Frédérique Laforest, and Christine Verdier

Contents

Hindawi Publishing Corporation
International Journal of Telemedicine and Applications
Volume 2008, Article ID 946082, 2 pages
doi:10.1155/2008/946082

Editorial
Pervasive Health Care Services and Technologies

Sajid Hussain,[1] Laurence T. Yang,[2] Frédérique Laforest,[3] and Christine Verdier[4]

[1] *Jodrey School of Computer Science, Acadia University, Wolfville, NS, Canada B4P 2R6*
[2] *Department of Computer Science, St. Francis Xavier University, Antigonish, NS, Canada B2G 2W5*
[3] *CNRS UMR 5205, LIRIS, INSA de Lyon, Bâtiment Blaise Pascal, 7 avenue Jean Capelle, 69621 Villeurbanne Cedex, France*
[4] *CNRS UMR 5217, LIG-SIGMA, Joseph Fourier University, Bâtiment C, 220 route de la chimie, 38400 Saint-Martin-d'Hères, France*

Correspondence should be addressed to Sajid Hussain, sajid.hussain@ieee.org

Received 30 July 2008; Accepted 30 July 2008

Due to recent developments in pervasive and ubiquitous computing, health care systems can provide well-informed and high-quality patient care services. A care provider can receive a simplified, adaptive, and latest view of the medical data. Although the medical file is unique, the distributed record could be accessible from any place at any time by care providers. Pervasive health care information systems allow overcoming the problem of heterogeneity of technologies and services in this domain and the papers selected for this special issue represent a good panel for addressing this challenge. Of course, the selected topics and papers are not an exhaustive representation of the area of Pervasive Health Care Services and Technologies. Nonetheless, they represent the rich and many-faceted knowledge, that we have the pleasure of sharing with the readers. We would like to thank the authors for their excellent contributions and patience in assisting us. Finally, the fundamental work of all reviewers on these papers is also very warmly acknowledged.

This special issue contains ten papers, where three papers are related to cardiac and one paper covers the identification of sleep problems. Four papers are regarding the architecture and implementation of smart homes for tele-health. Finally, two papers address security and privacy issues.

In the first paper entitled "Feasibility study and design of a wearable system-on-a-chip pulse radar for contactless cardiopulmonary monitoring," Zito et al. present a system-on-a-chip radar sensor for next-generation wearable wireless interface for health care and safeguard. The system consists of a radar sensor for detecting the heart and breath rates and a low-power IEEE 802.15.4 ZigBee radio interface, which provides a wireless data link with remote data acquisition and control units.

In the second paper, "Real-time and secure wireless health monitoring," Dağtaş et al. present a framework for a wireless health monitoring system using wireless networks such as ZigBee. The proposed framework provides the detection of signals using a wireless body sensor network (BSN), low-power and reliable data transmission through ZigBee network nodes, secure transmission of medical data over BSN, efficient channel allocation for medical data transmission over wireless networks, and optimized analysis of data using an adaptive architecture that maximizes the utility of processing and computational capacity at each platform.

In the third paper, "Development of a novel contactless mechanocardiograph device," Tavakolian et al present a novel method of detecting mechanical movement of the heart, Mechanocardiography (MCG), with no connection to the subject's body. This measurement is based on radar technology and it has been proven through this research work that the acquired signal is highly correlated to the phonocardiograph signal and acceleration-based ballistocardiograph signal (BCG) recorded directly from the sternum.

In the fourth paper, "PATHOS: Pervasive at home sleep monitoring," Obermiller and Ahamed investigate home sleep monitoring. PATHOS focuses not only on analyzing patterns during the night, but also on collecting data about the subject lifestyle that is relevant and important to the diagnosis of his/her sleep. Using existing technology, PATHOS provides wireless, inexpensive, portable, and customized solution.

In the fifth paper, the research of Rammal et al entitled "An adaptive system for home monitoring using a multi-agent classification of patterns," proposes a software architecture to monitor elderly people at home. In order to monitor efficiently the aged people, they build dynamically patterns from sensors' data and define a multi-agent method of classification used at an individual level to evaluate a risk or prevent a medical failure.

In the sixth paper entitled "Location estimation in a smart home," Rahal et al. present the implementation of a location system of elderly people living alone at home. To do so, their system uses Bayesian filtering and a set of anonymous sensors. Their experiments have shown good accuracy and robustness.

In the seventh paper, Paganelli et al present "ERMHAN: A context-aware service platform to support continuous care networks for home-based assistance" . The main goal of this platform is to enhance appropriately information sharing between care providers in home care. They also present the architecture of their platform and preliminary experimental results.

In the eighth paper entitled "Building application-related patient identifiers: what solution for a european country?" Quantin et al. pose the problem of a global patient identifier. In France, like in other European Countries, the patient's national identifier is unauthorized and makes the linkage between patient's data difficult. The authors propose a method using a derived social number to be used as the national identifier for patients. The solution is based on the utilization of anonymity techniques.

In the ninth paper entitled "A Tamper-resistant and portable healthcare folder," Anciaux et al. present the idea of a new hardware portable device to give back to the patient the control over his/her medical data. The proposed architecture assesses a secure access to data hosted both by the device and by a traditional EHR server. They also present their experiment in the context of a medico-social network for the elderly.

Sajid Hussain
Laurence T. Yang
Frédérique Laforest
Christine Verdier

Hindawi Publishing Corporation
International Journal of Telemedicine and Applications
Volume 2008, Article ID 328597, 10 pages
doi:10.1155/2008/328597

Research Article

Feasibility Study and Design of a Wearable System-on-a-Chip Pulse Radar for Contactless Cardiopulmonary Monitoring

Domenico Zito,[1] Domenico Pepe,[1] Bruno Neri,[1] Fabio Zito,[2] Danilo De Rossi,[3] and Antonio Lanatà[3]

[1] *Radio-frequency and Microwave Integrated Circuits Laboratory (RFLab), Department of Information Engineering (DIIEIT), University of Pisa, Via Caruso 16, 56122 Pisa, Italy*
[2] *Dipartimento di Informatica, Matematica, Elettronica e Trasporti (DIMET), Università "Mediterranea" di Reggio Calabria, Via Graziella, Feo di Vito, 89060 Reggio Calabria, Italy*
[3] *Interdepartmental Research, Center E. Piaggio, University of Pisa, Via Diotisalvi 2, 56122 Pisa, Italy*

Correspondence should be addressed to Domenico Zito, d.zito@iet.unipi.it

Received 28 September 2007; Accepted 15 January 2008

Recommended by Sajid Hussain

A new system-on-a-chip radar sensor for next-generation wearable wireless interface applied to the human health care and safeguard is presented. The system overview is provided and the feasibility study of the radar sensor is presented. In detail, the overall system consists of a radar sensor for detecting the heart and breath rates and a low-power IEEE 802.15.4 ZigBee radio interface, which provides a wireless data link with remote data acquisition and control units. In particular, the pulse radar exploits 3.1–10.6 GHz ultra-wideband signals which allow a significant reduction of the transceiver complexity and then of its power consumption. The operating principle of the radar for the cardiopulmonary monitoring is highlighted and the results of the system analysis are reported. Moreover, the results obtained from the building-blocks design, the channel measurement, and the ultra-wideband antenna realization are reported.

1. INTRODUCTION

In February 2002, the Federal Communications Commission (FCC) gave the permission for the marketing and operation of a new class of products incorporating ultra-wideband (UWB) technology [1]. The FCC, through a modification of the 47 CRF Part 15 regulations [2], decided to allocate for the UWB systems an unlicensed band 7.5 GHz wide (for the first time, in a nonexclusive way), in the range of the radio-frequency spectrum 3.1–10.6 GHz. Moreover, the FCC allocates the unlicensed radio-frequency spectrum between 22 and 29 GHz, for UWB short-range vehicular radars.

Since UWB systems have been released to operate in regions of spectrum in which other services are already operating, the mask of the maximum power spectral density (PSD) allowed for UWB devices has been set to very low values (−41.3 dBm/MHz in the 3.1–10.6 GHz band).

UWB devices can be employed for several applications: ground penetrating radar (GPR), medical imaging, wall imaging, through-wall imaging, surveillance, and high data rate communication systems.

One of the most promising class of applications of the UWB systems consists of the medical imaging (field disturbance sensors designed to detect the location or the movement of objects within the body of a person or animal [2]). In particular, a UWB radar sensor can be employed to monitor the heart wall and chest movements, in order to detect in real time the heart and breath rates, respectively.

The idea at the base of the work presented herein, which has been presented recently [3], consists of the realization of a novel wearable wireless interface for the monitoring of the heart beat and breath rates. This work is a part of the ProeTEX Project (FP6-2004-IST-4-026987), a European integrated project aimed at developing a new generation of equipments for the market of emergency operators, like firefighters and civil protection rescuers.

Modern silicon technologies (e.g., transistors of the standard CMOS 90 nm technology of ST Microelectronics have

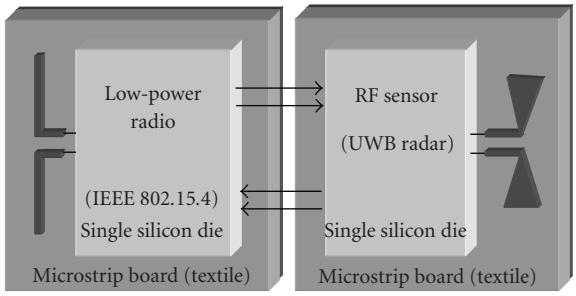


Figure 1: Wearable wireless interface for the heart monitoring: system idea.

cutoff frequencies higher than 150 GHz) allow us to realize miniaturized and ultra-small- and ultra-low-power wireless UWB sensors for WBAN (wireless body area network) applications. WBANs consist of sensor networks, in which a set of small sensors is placed around the human body or implanted in it, in order to monitor constantly the vital parameters and movements of the person under observation. The information collected by these sensors can be sent, by means of radio-frequency data link, toward remote data acquisition and signal processing units or even to a personal server, which can forward the data to the medical centers and hospitals by means of the Internet. In this way, the medical staff can investigate the manifesting of the heart diseases over all the daily activities of the subject under observation.

With respect to the preliminary study reported in [3], several details concerning the system analysis have been dealt with and reported in this paper. Moreover, this paper reports the advances and the present status of this research. In particular, the performance of the building blocks designed on silicon and the experimental characterization of the intra-body channel and the UWB antenna are reported. The paper is organized as follows. In Section 2, a system overview of the next-generation wearable wireless interface for the heart monitoring is presented. In Section 3, after a brief introduction on the techniques adopted to monitor the heart activity, the principle of operation of UWB radar sensor for the heart monitoring is highlighted. In Section 4, the feasibility study of the UWB radar sensor is discussed. In Section 5, a survey on the channel-loss measurements and the building blocks of the radar is reported, and future developments are discussed. Finally, in Section 6, the conclusions derived by this work are drawn.

2. WEARABLE WIRELESS INTERFACE FOR HEART MONITORING: SYSTEM OVERVIEW

The aim of the presented work consists of realizing a novel wearable wireless interface for the heart monitor.

The overall system idea is shown in Figure 1. It consists of a fully integrated UWB radar sensor and a low-power radio interface, where each section is realized on a single silicon die. The radar sensor and the low-data-rate wireless transceiver are implemented in a standard CMOS 90 nm technology by ST Microelectronics.

Figure 2: Prototype of the inner garment in which the wearable wireless interface for the detection of the heart and breath rates will be included. In detail, the sensor will be placed around the circled area.

By referring to the scheme of Figure 1, each antenna is realized on a microstrip substrate; however, in the most advanced realization, they can be realized directly by means of proper conductive layer tissues within clothes [4]. These interfaces will be inserted into an inner garment worn by emergency operators, which is shown in Figure 2.

The data acquired by the UWB radar sensor are transferred to a personal or remote unit by means of the low-power radio data link realized by a wireless transceiver based on the IEEE 802.15.4 (ZigBee) standard. By means of the low-power transceiver, the UWB radar sensor can be remotely programmed, increasing the flexibility of the system.

A future perspective is the realization of both the radar sensor and the low-power transceiver on the same silicon die (see Figure 3), in order to raise the level of miniaturization and to reduce further on the final costs of the system.

3. UWB RADAR SENSOR: PRINCIPLE OF OPERATION

A survey on the current techniques for the cardiac monitoring and the novel radar system proposed in this paper are reported hereinafter.

3.1. Radar sensors for monitoring the cardiac activity

The most widespread system for the monitoring of the cardiac activity is the electrocardiograph (ECG). The information provided by ECG is related to heart electrical activity. Pulse oximetry allows us to detect the cardiorespiratory activity, by measuring the saturation level of the oxygen in the blood. Other systems for the monitoring of the cardiac activity are based on ultrasounds (echocardiograph or echo Doppler). Ultrasound-based systems are generally cumbersome and they can be used only by specialized operators. Anyway, all the presented measurement techniques require the direct contact with the body in order to carry out the measurement.

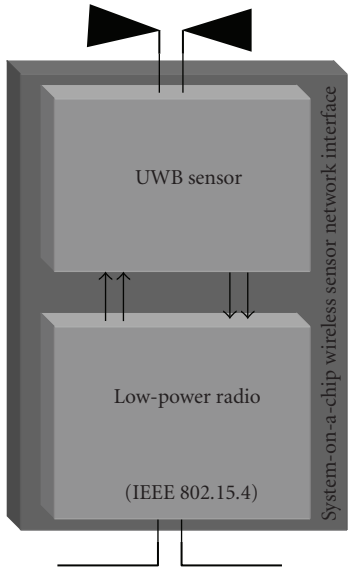


Figure 3: Future perspective: integration on a single silicon chip of the overall wearable wireless interface on a single silicon chip.

Unlike the traditional techniques (electrocardiograph, ecocardiograph, and pulsed oximetry), radar systems allow the monitoring of the heart activity in a noninvasive and contactless way for the patient [5]. Microwave Doppler radars have been used to detect the respiration rate since 1975 [6]. These first devices were bulky and expensive, but in recent times CMOS fully integrated versions of a radar for noncontact cardiopulmonary monitoring have been presented [7].

Doppler radars typically transmit a continuous wave signal and receive the echo reflected by the target. The frequency of the reflected signal varies from that of the transmitted one by an amount proportional to the relative velocity of the target with respect to the radar.

Another class of radar employed for the monitoring of vital parameters is based on pulse transmission. Pulse radars operate by sending short electromagnetic pulses and by receiving the echoes reflected by the target. The time delay between the transmission of the pulse and the reception of the echo is proportional to the distance from the target to the radar. Discrete prototypes of pulse radar for the detection of vital parameters are reported in literature [8, 9].

It is worth mentioning that radar sensors monitor the mechanical movement of the heart wall instead of the electrical activity of the heart (as the electrocardiograph), and then diseases of the heart that does not show anomalies in the electrical activity can be discovered as well. Moreover, the UWB pulses are not influenced by blankets or clothes [8].

From a circuit design point of view, UWB transceivers present a lower complexity with respect to traditional radio-frequency system, leading to a low-power consumption for a long life of the battery. In fact, with respect to the latter, UWB systems do not require a stable frequency reference, which typically requires a large area on silicon die and consumes a high amount of power.

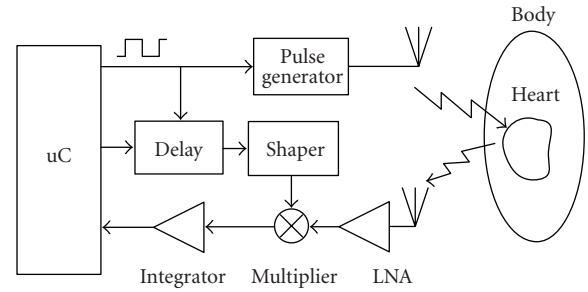


Figure 4: Block diagram of the proposed fully integrated UWB radar for the detection of heart and breath rates.

Moreover, the extremely low level of transmitted power density (lower than −41.3 dBm/MHz) of the UWB radar should reduce the risk of molecular ionization [10–15].

3.2. UWB (3.1–10.6 GHz) radar sensor for medical applications: principle of operation

The main block of the novel wearable wireless interface for human health care described herein is the UWB radar sensor. The block diagram of the proposed radar sensor for the detection of the heart and breath rates is shown in Figure 4. The radar exploits a correlation-based receiver topology followed by an integrator, which averages the received pulses in order to have an output signal containing the information on the heart and breath tones.

The operating principle of a cross-correlator radar is explained hereinafter. An electromagnetic pulse is transmitted toward the target. The echo received from the target is multiplied with a delayed replica of the pulse transmitted; the output signal of the multiplier is then integrated. Note that the output signal will reach its maximum in the case of perfect time alignment between the two signals at the input of the multiplier itself. In other terms, the cross-correlator has a frequency response equal to that of a matched filter. In particular, it can be demonstrated that the matched filter is the filter that allows to obtain the best signal-to-noise ratio at the output [16]. Moreover, this has been confirmed by preliminary system simulations (by means of the Ptolemy simulator within Agilent ADS2005A). In detail, the CAD system analysis has shown that this topology allows us to achieve the best performance in terms of output signal-to-noise ratio (SNR) and sensitivity to small variations of the position of the heart wall with respect to other topologies, like that in which the receiver is simply turned on by the command given by the delayed replica of the transmitted pulse [8, 9].

The principle of operation of the overall radar system shown in Figure 4 is explained hereinafter. A train of extremely short (about 200 picoseconds) Gaussian monocycle electromagnetic pulses is transmitted toward the heart. Since the heart muscle and the blood that flows inside have different characteristic impedance, a partial reflection of the energy associated with the radiated pulse occurs at the surface of separation of these two different media [8].

After a time delay approximately equal to the flight time of the pulse from the transmitter to the receiver (about two nanoseconds), a delayed replica of the transmitted pulse generated internally (by the delay and shaper blocks) is multiplied with the echo received. It is worth mentioning that, in practice, the shaper block can be replaced by an additional pulse generator. A pulse repetition frequency (PRF) greater than 1 MHz allows us to consider the heart almost "motionless" between two consecutive pulses. The amplitude of the signal at the output of the multiplier reaches its maximum when the received echo and the delayed replica of the transmitted pulse are perfectly time aligned. If the time delay is fixed, a fixed range gate is monitored by the radar, and the amplitude of the signal at the output of the multiplier is related to the position of the heart. The output voltage of the receiver front end is averaged by integrating over a large number of pulses. This operation allows us to increase considerably the signal-to-noise ratio at the output of the receiver, as explained in Section 4.2. Moreover, the amplitude of the continuous signal at the output of the integrator is related to the time-varying position of the moving object under observation, that is the heart wall in our case. Therefore, the output signal provided by the integrator includes the tones of the heart beat and breath frequencies.

4. FEASIBILITY STUDY OF THE UWB (3.1–10.6 GHz) RADAR SENSOR

A theoretical model of the intra-body channel in which the electromagnetic pulse propagates has been developed and a complete feasibility study of the UWB radar sensor has been carried out in order to demonstrate the effectiveness of such an approach. The specifications of the single blocks of the radar have been derived by taking into account the performance achievable by means of its implementation in a standard CMOS 90 nm technology by ST Microelectronics. System simulations have been performed by means of CAD tools in order to verify the feasibility of the proposed UWB sensor radar. In particular, CAD tool system analysis has been carried out by including accurate models at different levels of abstraction and nonideal effects (noise contribution and bandwidth limitation of the building blocks) associated with the technology process.

4.1. A simple theoretical model of the channel

A simple theoretical model of the channel has been developed in order to derive the system specifications of the fully integrated radar sensor, and then simulate the overall radar system.

A frequency-dependent channel-loss model (in the band 1–12 GHz) has been developed. The overall loss ($L(f)$) has been calculated by taking into consideration the contributions of the following: (i) path loss (PL(f)); (ii) attenuation in the tissues (Att(f)); and (iii) losses due to the reflections at the interface between different tissues (Rfl(f)). The properties of the body tissues have been extracted by the parametric model of the dielectric properties of the body tissues developed by C. Gabriel et al. at Brooks Air Force Base (USA)

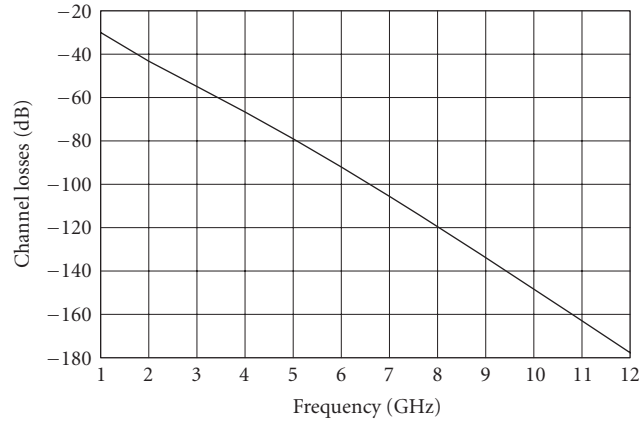


Figure 5: Channel loss versus frequency predicted by the near-field-based model.

[17, 18]. As for the thickness of the tissues layers, the model proposed in [5], based on the Visible Human Project and Gabriel's data book, has been considered. An antenna gain equal to 1.8 (2.5 dB) has been chosen to determine the contribution of the path loss, by referring to a UWB antenna realized, which has been proposed in [19]. Since we rely on applying the system-on-a-chip radar, and thus its antenna, very close to the skin in proximity of the chest, near field equations [20, 21] have been used in order to evaluate the path and reflection losses.

The detailed formulas employed to derive the frequency-dependent channel-loss model have been reported in [22]. The result is summarized herein (see Figure 5).

Simulation results show that the average power loss of the pulse in the 3.1–0.6 GHz band amounts to about 80 dB. Note that this result is in agreement with the channel measurements reported in [23].

The time of flight of the pulse has been estimated in about two nanoseconds (in accordance with the data reported in the literature [8]). The maximum time difference in the flight time of the pulses due to the heart displacement (1-2 cm) has been estimated in a few hundreds of picoseconds.

To validate this channel-loss model, a set of several antennas each operating in a different frequency subrange has been realized in order to perform channel-loss measurements over the frequency range of interest. The results are described in Section 5.

4.2. Theoretical system analysis and specifications

As for the transmitted electromagnetic pulse, it has to be very short in order to have a range resolution in the order of the centimeter, since the maximum displacement of the heart wall is of a few centimetres (1-2 cm). A duration time (τ) of the pulse of about 200 picoseconds makes possible the achievement of an accurate range resolution. Thus the maximum of the power spectral density results would be at 5 GHz (within the 3.1–10.6 GHz band), since the maximum of the

spectrum of a Ggaussian monocycle pulse is placed at the frequency $1/\tau$.

The Gaussian monocycle has been preferred to the Gaussian pulse because it has no dc component and its spectrum matches the FCC emission level mask. Furthermore, the Gaussian monocycle pulse can be implemented efficiently on silicon by means of differential transmitter [24]. The peak-to-peak voltage of the pulse has been chosen equal to 1.2 V, by considering that the radar will be realized in a standard CMOS 90 nm technology by ST Microelectronics, which is characterized by a power supply of 1.2 V (this is not a critical value since the theoretical maximum peak-to-peak voltage of the pulse generated by the differential transmitter amounts to the double of the supply voltage, without considering the voltage drop in the transistors).

As for the receiver, if $\mathrm{SNR}_{\mathrm{out}}$ is the advisable output signal-to-noise ratio, the minimum power (S_{min}) required at the input of the receiver amounts to

$$S_{\mathrm{min}} = k \times T \times B \times \mathrm{NF} \times \mathrm{SNR}_{\mathrm{out}}, \quad (1)$$

where k is the Boltzmann constant (1.38×10^{-23}), T is the equivalent temperature of the antenna, NF is the receiver noise figure, B the bandwidth of the receiver, and $\mathrm{SNR}_{\mathrm{out}}$ is the signal-to-noise ratio at the output of the receiver. If we consider an average channel loss equal to 80 dB in the 3.1–10.6 GHz band, a noise figure lower than −11.5 dB would be required in order to have an output SNR greater than 10 dB (for a single pulse).

This result is clearly not reachable in practice, so that an improvement is required. The output SNR of the radar receiver can be increased by integrating several pulses. The characteristic frequencies of the vital parameters under observation are within a few Hertz, then an integrator band of 100 Hz is wide enough to keep appropriately the heart and breath rates. The case of an integrator with a bandwidth of 1 KHz and a PRF in the range 1–10 MHz has been investigated. For a PRF equal to 10 MHz, 10 000 pulses are averaged. In particular, an integration over 10 000 pulses allows an improvement of the SNR ($\mathrm{SNR}_{\mathrm{imp}}$) at the output of the receiver of about 40 dB, as shown in the tables reported in [25]. This result is confirmed by the rough estimation given by the following equation:

$$\mathrm{SNR}_{\mathrm{imp}} \approx 10 \times \log\left(\frac{\mathrm{PRF}}{B_{\mathrm{int}}}\right) = 10 \times \log 10\,000 = 40\,\mathrm{dB}, \quad (2)$$

where B_{int} is the bandwidth of the integrator. The receiver front-end (LNA and multiplier) specification results are to be thus relaxed by the integration of a great number of pulses. In these conditions, the noise figure of the receiver front end has to be lower than

$$\mathrm{NF}_{\mathrm{max}} = -11.5 + 40 = 28.5\,\mathrm{dB}. \quad (3)$$

The specifications for the LNA are set in a power gain of 15 dB and a noise figure of 5 dB, whereas for the multiplier, a voltage gain of 0 dB and a noise figure of 10 dB. With these specifications, the noise figure of the overall receiver front end will result in largely lower than the maximum NF allowed for a proper detection. It is worth mentioning that the aforementioned specifications can be obtained by an implementation in a standard CMOS 90 nm technology.

Within this framework, a large integrator voltage gain is required. In fact, the dc component of the signal at the output of the multiplier, in the case of perfect time alignment between the signal received and the replica delayed of the transmitted pulse, has been estimated in a few hundreds of nanovolts. Thus, an integrator gain equal to 120 dB is required in order to have an amplitude voltage of a few hundreds of millivolts at the integrator output.

4.3. System analysis by means of CAD tool simulations

The overall radar system has been simulated by means of the Ptolemy simulator within the CAD tool ADS2005A by Agilent Technologies. Each block of the overall radar system has been implemented in the simulator by functional blocks which take into account their bandwidth limitations and noise contributions.

The channel model (of the human chest) has been implemented as a frequency-dependent S-parameter block, which has been included in the system analysis of the overall radar. This parameter description provides an equivalent behavior in terms of frequency response of the theoretical channel model shown in Section 4.1.

The simulation results show that the transmitted pulse is strongly distorted by the channel. To be noted that in narrow-band radar, the reflected wave has almost that same shape of the transmitted: it keeps the sinusoidal shape for signal down-conversion such as addition, subtraction, differentiation, and integration. In the case of UWB radar, the signal changes its shape during transmission, channel propagation, and reception and this impairs the signal-to-noise ratio.

An additional block realizing a periodically variable flight time has been included in order to emulate the cardiac movement. For the sake of clarity, this block emulates only two positions of the heart, providing a difference of 200 picoseconds in the arrival time of the pulse received for the two positions. The heartbeat period has been set in 20 milliseconds (this has been reduced with respect to the real heart movement in order to reduce the simulation time). However, this is not a limitation since the radar system reaches the steady state within ten milliseconds, thus the simulation results in terms of output signal and output SNR will not be influenced by this assumption.

Moreover, an additional white thermal noise source has been included in the channel in order to take into account the antenna noise at the input of the receiver. Both antennas, that of the transmitter and that of the receiver, have been included in the simulator by a 7.5 GHz band filter (centred at 6.85 GHz, with a slope of −20 dB/dec) each. The LNA block has a band of 3.1–10.6 GHz and an out-band slope of the frequency response equal to −20 dB/dec.

A two-step simulation has been carried out, since the time constants of the radio-frequency and the baseband parts of the radar are quite different. Actually, the radio-frequency part has to be simulated with a timestep of about one tenth of the pulse width (i.e., the duration time), whereas the

baseband part requires a simulation time of at least several tens of milliseconds. A timestep of 10 picoseconds has been adopted for the RF section. The integrator (with 120 dB of gain and 1 KHz of band) has been splitted in three low-pass filters with gain equal to 40 dB and band equal to 10 MHz, 100 KHz, and 1 KHz, respectively.

Simulation results are in agreement with those obtained by theoretical system analysis. In particular, the simulations show that the increase of the SNR from the output of the multiplier to the output of the integrator is of about 40 dB, as predicted by the theoretical system analysis.

Figure 6 shows the following: (i) the power spectral density (PSD) of the pulse sequence with duration time of 200 picoseconds, 1.2 Volt peak-to-peak amplitude, and PRF equal to 1 MHz, at the input of the antenna filter; (ii) the power spectral density (PSD) of the pulse sequence resulting and PRF equal to 1 MHz, at the output of the antenna filter; and (iii) the FCC mask for the UWB medical imaging applications. The Gaussian monocycle pulse at the input of the transmitter antenna filter and the pulse obtained at the output of the antenna filter are shown in Figure 7. The pulse at the input of the receiver (i.e., the output of the antenna filter of the receiver) is shown in Figure 8. Note the large amount of noise superimposed to the advisable signal. The output voltage of the integrator is shown in Figure 9. Note that the output signal reaches two different voltages for the two different positions of the heart.

It is worth mentioning that similar performance has been obtained by using sinusoidal pulses with the same duration and amplitude.

5. PRESENT STATUS AND FUTURE WORKS

Present and future works are addressed to the building blocks design in CMOS 90 nm technology by ST Microelectronics and their cointegration, UWB antenna design, channel model verification by means of measurements, system-on-a-chip prototyping, and experimental characterization.

A set of antennas resonating at the frequencies of interest has been realized in order to perform channel measurements. The results of these measurements are hereinafter reported.

Moreover, preliminary prototypes of UWB antenna have been realized.

The most critical radio-frequency blocks of the radar (low noise amplifier, pulse generator, and multiplier) have been designed and sent to the silicon foundry for the test chip prototyping.

A summary of the most relevant results achieved at the present status is reported.

5.1. Channel-loss measurements

Experimental characterizations of the intrabody channel have been reported in the literature [23]. It is important to remark that results obtained by our simple channel model have shown a wide agreement with those measured in [23]. In detail, this work reports that the mean loss in the UWB band between two antennas placed towards the chest on the same side is of about 80 dB.

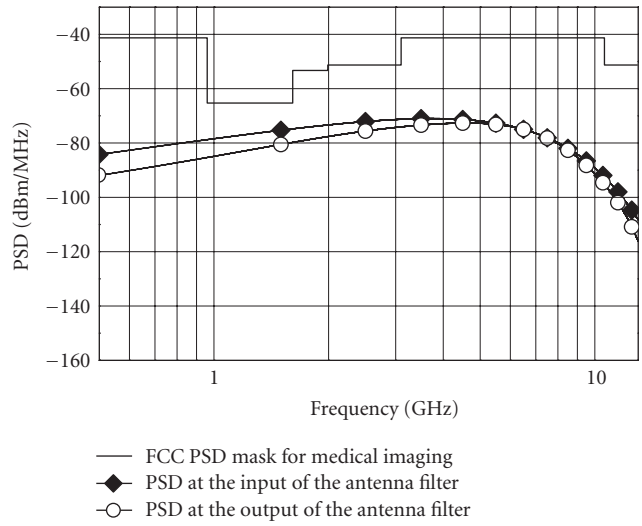


Figure 6: Power spectral density (PSD) of a pulse sequence with PRF equal to 1 MHz versus frequency.

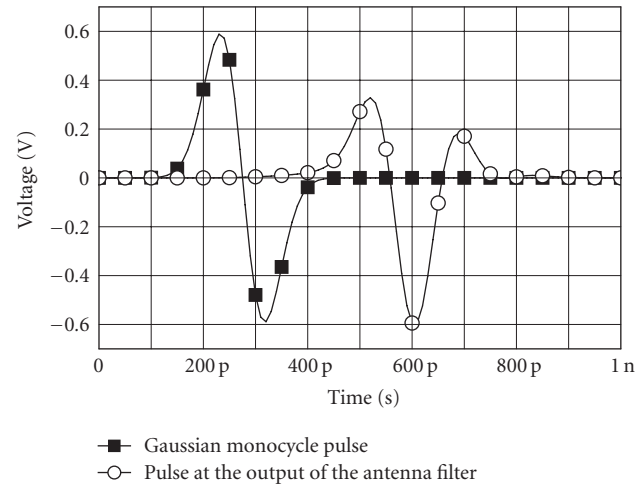


Figure 7: Gaussian monocycle pulse and pulse at the output of the antenna.

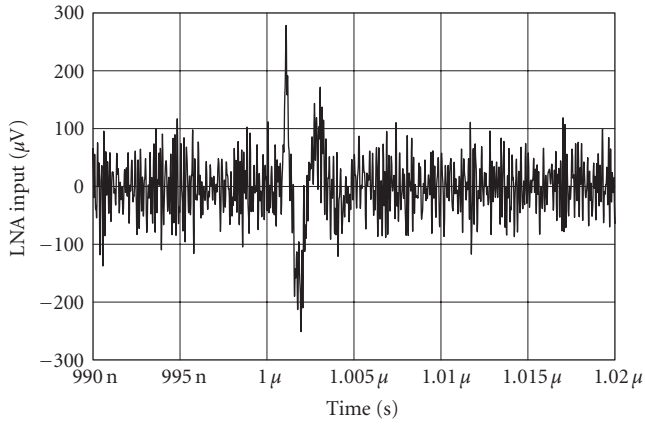


Figure 8: Voltage signal at the input of the LNA, including the noise contributions.

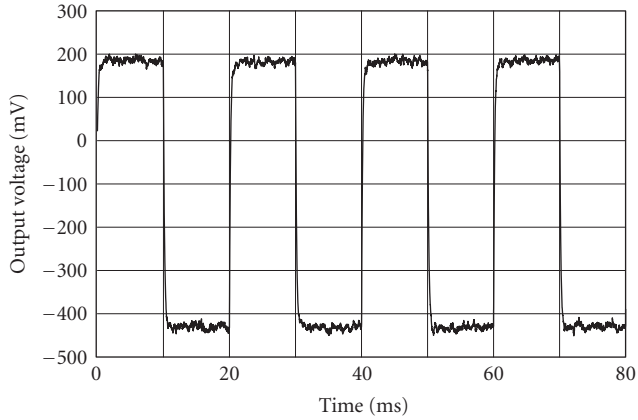


Figure 9: Voltage at the output of the integrator. The output signal has the same frequency of the movement imposed for the heart wall. A time-varying surface with a period of 20 milliseconds has been considered for the simulation (this period is short with respect to the real heart moving, in order to reduce the simulation time. This does not impair the analysis since the radar reaches widely the steady state within ten milliseconds).

Figure 10: Antenna prototype for the channel model verification.

Figure 11: Setup for the channel-loss measurement, setup between the front and the back of a human chest.

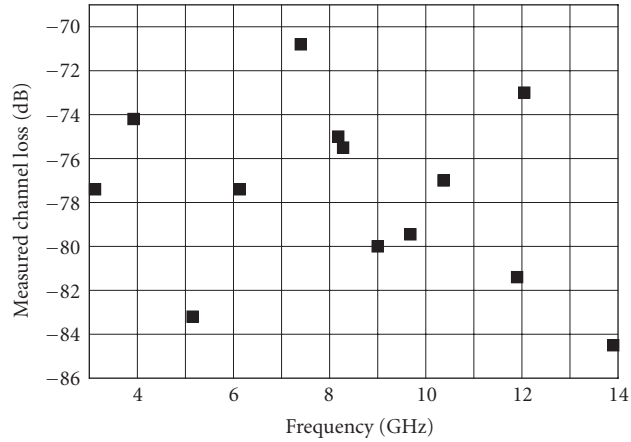


Figure 12: Measurement results of the intrabody channel loss.

In order to verify definitively through measurements the theoretical channel model we have developed, and to investigate some details concerning our application (e.g., the distance antenna-skin for a proper antenna operation), a set of antennas resonating at different frequencies in the spectrum range of interest has been realized (an example is shown in Figure 10). The antennas have been designed by means of Momentum, the EM simulator within Advanced Design System (ADS) by Agilent Technologies.

The measurements have been carried out using a vector network analyzer (VNA) placing two identical antennas, one in front of the chest of the subject under test (SUT) and the other on the back of the chest Although this measurement setup is not directly related with the antenna-heart-antenna path we modelled, this allows us to extract information concerning the attenuation in the tissues, avoiding any systematic error caused by proximity effects (i.e., direct coupling) of the two antennas placed on the same side. Each antenna irradiates towards the other antenna. The measurement setup is shown in Figure 11. The antennas are placed at one cm of distance from the skin (by means of plastic structure, which does not impair the electromagnetic behavior). Several tests have been carried out by varying the distance between the antenna and the skin surface, showing that the measurement results obtained do not significantly vary for distances of more than about five millimeters. If the distance antenna-skin decreases, then the radiation pattern of the antenna changes and the attenuation increases. This result is in agreement with the measurement reported by other researchers in [4].

The results of our measurements (and the resonating frequencies of the set of antennas realized) are reported in Figure 12. The mean value of $|S21|$ has resulted approximately equal to 78 dB. Note that the antennas we designed have a gain of about ten dB. Thus if we consider the typical path loss of a UWB antenna characterized by a gain of 2.5 dB (as considered in our channel model [22]), an additional loss of about 15 dB (10–2.5 dB for each antenna) has to be summed up. This means that the overall loss measured amounts approximately to 93 dB.

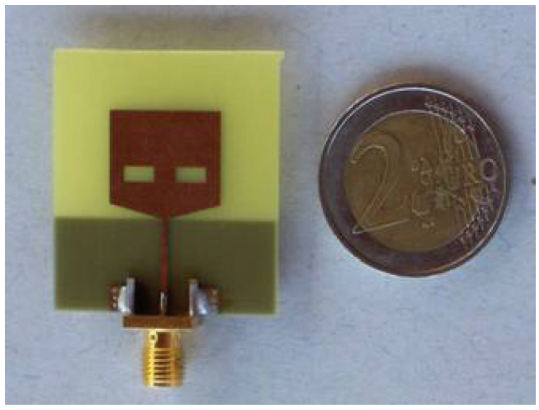


Figure 13: Prototype of UWB antenna realized.

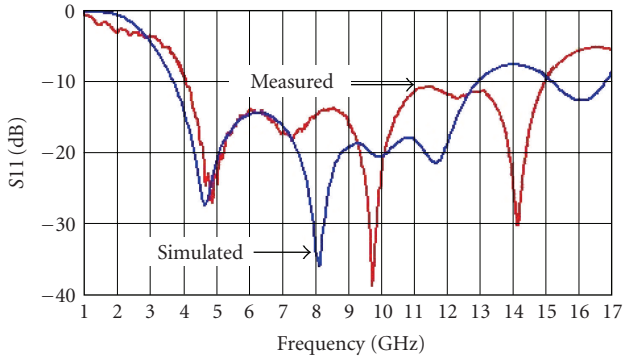


Figure 14: Simulated (blue) and measured (red) S11 parameters of the UWB antenna of Figure 13.

A direct comparison with the result (109 dB) reported in [23] cannot be carried out since, therein, the measurements have been carried out by means of antennas at contact with the skin. In fact, this causes an additional loss due to the change of the radiation pattern of the antenna itself (see above). Moreover, in [23], the directivity of the antennas used to carry out the test is not reported. Regardless of that, the two results are reasonably close to each other.

Anyway, the experimental results, in spite of being derived from a different setup, are in the order of those predicted by the model, especially in the lowest part of the 3.1–10.6 GHz frequency band. The increase of the attenuation at the upper part of the frequency spectrum predicted by the model is mitigated (in the measured results) by the increase of the antenna efficiency of the antennas realized.

Finally, this experimental investigation shows that the UWB radar, and especially its antenna, can be realized as wearable device close to the human body.

5.2. UWB antenna

Preliminary UWB antenna prototypes have been realized. Knight's helm shape antennas [26] have been realized in order to implement the wideband antenna. One of the antenna prototypes is shown in Figure 13.

The antenna has been realized on a microstrip substrate. As for the antenna, it is worth mentioning that the feasibility of textile UWB antennas has been demonstrated in [4].

The simulated and measured parameter S11 of this antenna is shown in Figure 14. It can be seen that the −10 dB band is between 4.081 GHz and 14.95 GHz. Note the large agreement between measurements and simulations up to the 7.5 GHz. However, the mismatch in the higher region of frequency (from 7.5 to 10.6 GHz) occurs and it is due to the approximations (2D and 1/2) of the geometry introduced by the EM simulator. The fractional bandwidth ($B\%$), can be defined as follows:

$$B\% = \frac{B}{f_c} \times 100, \qquad f_c = \frac{f_{\max} + f_{\min}}{2}. \tag{4}$$

Then note that the fractional bandwidth results are equal to 114.22%.

Present and future works on this task are addressed to realize a UWB antenna on a textile substrate with performance perfectly matched with the UWB band.

5.3. Building blocks in CMOS 90 nm by ST Microelectronics

A UWB 3.1–10.6 GHz low-noise amplifier (LNA) in CMOS 90 nm process by ST Microelectronics has been designed. The LNA consists of a common gate input stage and two subsequent common source gain stages. The common gate input stage allows the realization of a wideband input integrated matching to the source impedance of the antenna. This novel topology allows the achievement of a wideband input integrated matching by overcoming the issues of the classical wideband input matching technique implemented with passive low-pass filters, such as the parasitic effects and noise contributions introduced by the passive elements at the LNA input [27].

Schematic is reported therein [27]. With respect to the solution presented in [27], a three-stage topology has been adopted in order to reach the specifications in terms of gain.

If compared with other solutions with the same power consumption, the proposed LNA exhibits a very good performance tradeoff between gain, bandwidth, noise, and linearity. The design does not present any critical issues and exhibits an excellent design reliability.

Postlayout simulations have shown that the presented LNA provides a transducer gain (GT) of 19.34 dB at 5.4 GHz and a −3 dB band from 3.4 GHz to 8.3 GHz. Within the UWB band, the S11 parameter is equal to −22.33 dB at 3.1 GHz and −4.09 dB at 10.6 GHz. The noise figure (NF) is equal to 6.24 dB at 3.1 GHz and 5.13 dB at 10.6 GHz.

The overall LNA draws 41.71 mA on a 1.2 V power supply (biasing network included).

A novel fully integrated UWB pulse generator has been also designed in CMOS 90 nm process by ST Microelectronics [28]. Schematic is therein [22]. The pulse generator provides monocycle pulses with duration time close to 250 picoseconds and 1-V peak-to-peak amplitude. In detail, the

circuit provides a sinusoidal-like monocycle when activated by a negative edge of a trigger signal provided by a microcontroller. This activation can be delayed in the range 1–3 nanoseconds, by acting on a 5-bit programmable delay element, which provides a total set of 32 different delay times.

This new pulse generator provides a sinusoidal-like monocycle pulse with 1-V peak-to-peak and 250 picoseconds of duration time on a differential antenna with input impedance equal to 100 Ohm.

The overall pulse generator consumes 22 mW on a 1.2 V power supply, which represents an extremely low-power dissipation with respect to the other solutions presented in the literature.

The multiplier consists of a PMOS input matching stage, followed by a PMOS double balanced Gilbert cell terminated on a capacitive load, representing the input impedance of the following gain stage. Simulation results show a conversion gain G_C equal to 1.71.

To be noted that the performance of all the building blocks is in line or widely within the system requirements. For the sake of clarity, it has to be mentioned that the specifications derived for the receiver were more stringent than those strictly required for a proper operation.

Present work on this task is addressed to the design of the delay generator and integrator (note that feasibility has been widely demonstrated in several works presented in literature [29, 30]), and the cointegration of all the building blocks into a single silicon die.

6. CONCLUSIONS

The recent advances in silicon CMOS technology allow the realization of more and more miniaturized, low-cost, and low-power integrated system-on-a-chip sensors. These sensors can be employed in the wireless body area networks, for advanced and continuous monitoring of vital parameters.

In particular, the system overview of next-generation wearable wireless sensors for human health care and safeguard has been presented herein. Such a system is composed by a novel fully integrated ultra-wideband radar sensor for the detection of the heart and breath rates and a low-power radio interface (IEEE 802.15.4, ZigBee), which collects the data provided by the sensor and sends these data to a remote data acquisition unit or even in the Internet by means of a personal server. Thus the physiological data of a person under observation can be sent in real time to the hospital to be analyzed and then the doctors could act in time in case of anomalies in the vital parameters monitored.

A detailed feasibility study of the UWB radar on silicon technology (CMOS 90 nm) has been carried out, by means of both theoretical analysis and CAD tool simulations. The simulation results have shown a wide agreement with the theoretical model of the radar, demonstrating the feasibility of the proposed system-on-a-chip radar in a modern silicon technology.

Moreover, the main critical building blocks of the radar sensor have been designed and prototyped in CMOS 90 nm process by ST Microelectronics. Channel-loss model has been confirmed by means of several tests in laboratory. In particular, measurement results of the intrabody loss have shown results in quite agreement with the predicted ones especially in the lower portion of the 3.1–10.6 GHz band. Preliminary prototypes of UWB antennas have been realized and characterized to be implemented on a textile substrate.

Present and future works are addressed to the design of the noncritical building blocks and final system-on-a-chip prototyping, the cointegration with the textile antenna, and the medical characterization in clinical environment of this innovative UWB microsensor for contactless and noninvasive cardiopulmonary monitoring.

ACKNOWLEDGMENTS

The authors are grateful to the European Union for its financial support through the European Project ProeTex within the 6th FPQ (Project FP6-2004-IST-4-026987) and to Eng. Elisa Valentini for her help in the experimental characterization of the organic tissues and antenna.

REFERENCES

[1] "New public safety applications and broadband internet access among uses envisioned by fcc authorization of ultra-wideband technology," Federal Communications Commission, 2002, http://www.fcc.gov/Bureaus/Engineering_Technology/ews_Releases/2002/nret0203.html.

[2] "47 cfr part 15," Federal Communications Commission, 2002, http://www.fcc.gov/oet/info/rules.

[3] D. Zito, D. Pepe, B. Neri, D. De Rossi, and A. Lanatá, "Wearable system-on-a-chip pulse radar sensors for the health care: System overview," in *Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW '07)*, vol. 1, pp. 766–769, Niagara Falls, Canada, May 2007.

[4] M. Klemm and G. Troester, "Textile uwb antennas for wireless body area networks," *IEEE Transaction on Antennas and Propagation*, vol. 54, no. 11, pp. 3192–3197, 2006.

[5] E. M. Staderini, "UWB radars in medicine," *IEEE Aerospace and Electronic Systems Magazine*, vol. 17, no. 1, pp. 13–18, 2002.

[6] J. C. Lin, "Microwave sensing of physiological movement and volume change: a review," *Bioelectromagnetics*, vol. 13, no. 6, pp. 557–565, 1992.

[7] A. D. Droitcour, O. Boric-Lubecke, V. M. Lubecke, J. Lin, and G. T. A. Kovacs, "Range correlation and I/Q performance benefits in single-chip silicon doppler radars for noncontact cardiopulmonary monitoring," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 3, pp. 838–848, 2004.

[8] T. E. McEwan, "Body monitoring and imaging apparatus and method," U.S. Patent 5 573 012, November 1996.

[9] I. J. Immoreev and S. V. Samkov, "Ultra-wideband (uwb) radar for remote measuring of main parameters of patient's vital activity," *Radio Physics and Radio Astronomy*, vol. 7, no. 4, pp. 404–407, 2002.

[10] S. F. Cleary, F. Nickless, L. M. Liu, and R. Hoffman, "Studies of exposure of rabbits to electromagnetic pulsed fields," *Bioelectromagnetics*, vol. 1, no. 3, pp. 345–352, 1980.

[11] J. A. D'Andrea, B. L. Cobb, and J. O. De Lorge, "Lack of behavioral effects in the rhesus monkey: high peak microwave pulses at 1.3 GHz," *Bioelectromagnetics*, vol. 10, no. 1, pp. 65–76, 1989.

[12] T. J. Walters, P. A. Mason, C. J. Sherry, C. Steffen, and J. H. Merritt, "No detectable bioeffects following acute exposure to high peak power ultra-wide band electromagnetic radiation in rats," *Aviation Space and Environmental Medicine*, vol. 66, no. 6, pp. 562–567, 1995.

[13] C. J. Sherry, D. W. Blick, T. J. Walters, G. C. Brown, and M. R. Murphy, "Lack of behavioral effects in non-human primates after exposure to ultrawideband electromagnetic radiation in the microwave frequency range," *Radiation Research*, vol. 143, no. 1, pp. 93–97, 1995.

[14] J. H. Merritt, J. L. Kiel, and W. D. Hurt, "Considerations for human exposure standards for fast-rise-time high-peak-power electromagnetic pulses," *Aviation Space and Environmental Medicine*, vol. 66, no. 6, pp. 586–589, 1995.

[15] J. R. Jauchem, R. L. Seaman, H. M. Lehnert, et al., "Ultra-wideband electromagnetic pulses: lack of effects on heart rate and blood pressure during two-minute exposures of rats," *Bioelectromagnetics*, vol. 19, no. 5, pp. 330–333, 1998.

[16] E. Della Mese, "Appunti al corso di Teoria e Tecnica Radar," Ed. SEU Pisa.

[17] C. Gabriel and S. Gabriel, "Compilation of the dielectric properties of body tissues at rf and microwave frequencies," 2002, http://www.brooks.af.mil/AFRL/HED/hedr/reports/dielectric/home.html.

[18] D. Andreuccetti, R. Fossi, and C. Petrucci, "Calculation of the dielectric properties of body tissues in the frequency range 10 hz–100 ghz," 2002, http://niremf.ifac.cnr.it/tissprop/htmlclie/stfrtag.

[19] K. Kiminami, A. Hirata, and T. Shiozawa, "Double-sided printed bowtie antenna for uwb communications," *IEEE Antennas and Wireless Propagation Letters*, vol. 3, no. 1, pp. 152–153, 2004.

[20] H. Schantz, "Near field channel mode," IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs), 2006.

[21] A. Monorchio, "Appunti di compatibilità elettromagnetica," http://www2.ing.unipi.it/homepages/agostino.monorchio/emc.html.

[22] D. Zito, D. Pepe, B. Neri, and D. De Rossi, "Feasibility study of a low-cost system-on-a-chip UWB pulse radar on silicon for the heart monitoring ," in *Proceedings of the 3rd International Waveform Diversity & Design Conference (WDD '07)*, pp. 32–36, Pisa, Italy, June 2007.

[23] T. Zasowski, F. Althaus, M. Stager, A. Wittneben, and G. Troster, "UWB for noninvasive wireless body area networks: channel measurements and results," in *Proceedings of IEEE Conference on Ultra Wideband Systems and Technologies (UWBST '03)*, pp. 285–289, Reston, Va, USA, November 2003.

[24] K. Marsden, H.-J. Lee, D. S. Ha, and H.-S. Lee, "Low power cmos reprogrammable pulse generator for uwb systems," in *Proceedings of IEEE Conference on Ultra Wideband Systems and Technologies (UWBST '03)*, pp. 443–447, Reston, Va, USA, November 2003.

[25] M. I. Skolnik, *Radar Handbook*, McGraw Hill, New York, NY, USA, 1970.

[26] F. Barale and D. Zito, "UWB 3.1–10.6 GHz CMOS LNA," in *Proceedings of the 3rd Conference on Ph.D. Research in Microelectronics and Electronics (PRIME '07)*, pp. 45–48, Bordeaux, France, July 2007.

[27] F. Zito, D. Zito, and D. Pepe, "UWB 3.1–10.6 GHz CMOS transmitter for system-on-a-chip nano-power pulse radars," in *Proceedings of the 3rd Conference on Ph.D. Research in Microelectronics and Electronics (PRIME '07)*, pp. 189–192, Bordeaux, France, July 2007.

[28] Z. N. Low, J. H. Cheong, and C. L. Law, "Low-cost PCB antenna for UWB applications," *IEEE Antennas and Wireless Propagation Letters*, vol. 4, no. 1, pp. 237–239, 2005.

[29] M. Maymandi-Nejad and M. Sachdev, "A monotonie digitally controlled delay element," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 11, pp. 2212–2219, 2005.

[30] R. Rieger, A. Demosthenous, and J. Taylor, "A 230-nW 10-s time constant CMOS integrator for an adaptive nerve signal amplifier," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 11, pp. 1968–1975, 2004.

Hindawi Publishing Corporation
International Journal of Telemedicine and Applications
Volume 2008, Article ID 135808, 10 pages
doi:10.1155/2008/135808

Research Article

Real-Time and Secure Wireless Health Monitoring

S. Dağtaş,[1] G. Pekhteryev,[2] Z. Şahinoğlu,[3] H. Çam,[4] and N. Challa[4]

[1] *Department of Information Science, University Arkansas, Little Rock, AR 72204-1099, USA*
[2] *Teknoset Research, TUBITAK MAM, Gebze 41730, Turkey*
[3] *Mitsubishi Electric Research Labs, 201 Broadway Avenue, Cambridge, MA 02139, USA*
[4] *Computer Science and Engineering Department, School of Engineering Arizona State University Tempe, Arizona State University, Tempe, Arizona 85287, USA*

Correspondence should be addressed to S. Dağtaş, sxdagtas@ualr.educ

Received 16 October 2007; Accepted 27 February 2008

Recommended by Sajid Hussain

We present a framework for a wireless health monitoring system using wireless networks such as ZigBee. Vital signals are collected and processed using a 3-tiered architecture. The first stage is the mobile device carried on the body that runs a number of wired and wireless probes. This device is also designed to perform some basic processing such as the heart rate and fatal failure detection. At the second stage, further processing is performed by a local server using the raw data transmitted by the mobile device continuously. The raw data is also stored at this server. The processed data as well as the analysis results are then transmitted to the service provider center for diagnostic reviews as well as storage. The main advantages of the proposed framework are (1) the ability to detect signals wirelessly within a body sensor network (BSN), (2) low-power and reliable data transmission through ZigBee network nodes, (3) secure transmission of medical data over BSN, (4) efficient channel allocation for medical data transmission over wireless networks, and (5) optimized analysis of data using an adaptive architecture that maximizes the utility of processing and computational capacity at each platform.

1. INTRODUCTION

As numerous wireless personal area networking (WPAN) technologies emerge, the interest for applications such as health monitoring, smart homes, and industrial control has grown significantly. ZigBee is the first industrial standard WPAN technology [1] that provides short-range, low-power, and low-data-rate communication, and supports mesh networking and multihopping. While many smart home application areas such as lighting, security, and climate control have been suggested using the ZigBee standard, health-care applications have not received much attention despite their importance and high-value added. Here, we present a wireless communication system for real-time health monitoring with secure transmission capability.

Early clinical trials conducted in the mid-nineties by the National Institute of Health in the Mobile Telemedicine project [2] has indicated that the transmission of critical patient data during emergencies can make significant difference in patient outcomes. This result has led to a proliferation of health-care projects, including CodeBlue [3], PPMIM [4], CustoMed [5], MobiCare [6], LiveNet [7], PCN [8], UbiMon [9], MobiHealth [10], AMON [11], and PadNET [12]. Various types of wearable health monitoring sensor devices have been developed and integrated into patients' clothing [13–16], an armband [17], or wristband [18].

Most of the existing mobile patient monitoring projects such as PPMIM [4], MobiCare [6], MobiHealth [10], and AMON [11] employ cellular networks (e.g., GSM, GPRS, or UMTS) to transmit vital signs from BSN to health centers. For instance, in PPMIM project, a remote medical monitoring three-tier architecture with a GSM/GPRS peer-to-peer channel is presented, and the concept of multiresolution is introduced to identify useful information and to reduce communication costs. In MobiCare, a body sensor network (or MobiCare client) and health-care servers employ short-range Bluetooth between BSN and BSN Manager, and GPRS/UMTS cellular networks between BSN Manager and health-care providers. The UbiMon (Ubiquitous Monitoring Environment for Wearable and Implantable Sensors) Project aims to provide a continuous and unobtrusive monitoring system for patients in order to capture transient but life threatening events [19]. Among major projects in the area,

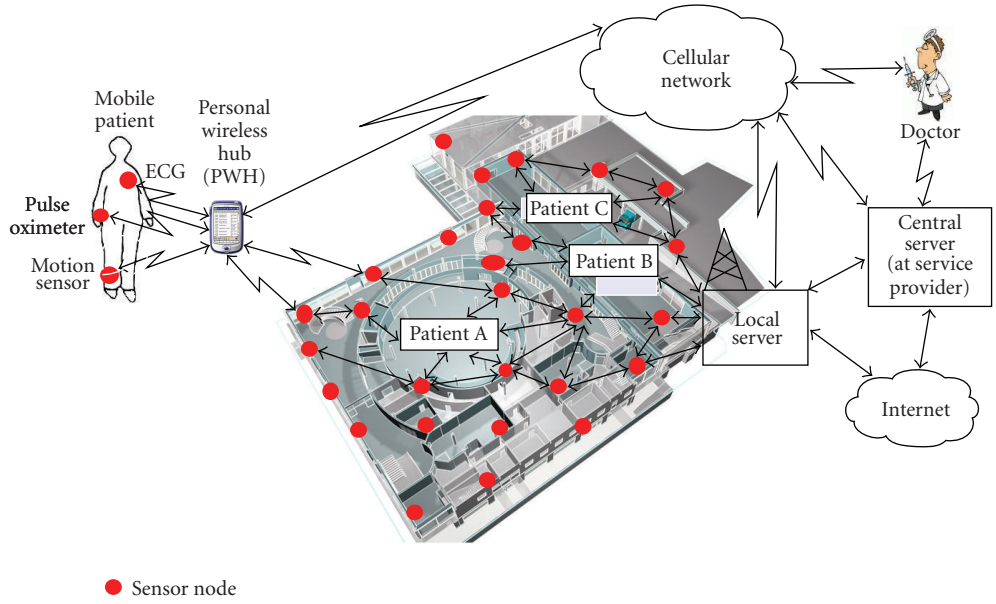


Figure 1: Wireless health monitoring in a smart environment.

CodeBlue [20] is the only existing project that employs wireless sensor networks in emergency medical care, hospitals, and disaster areas as an emergency message delivery system. With MICA motes [21], CodeBlue uses pulse oximetry and electrocardiogram (ECG) sensors to monitor and record blood oxygen and cardiac information from a large number of patients. However, most of the existing systems lack two key features: (1) wireless communication technology that conforms to standards, (2) integration with wireless sensor network platforms such as smart home systems, and (3) secure transmission capability that addresses the resource constraints optimally.

Our ZigBee-based architecture is based on the premise that the secure wireless communications combined with the widespread infrastructure provided by applications such as smart homes will be key to the effective use of future medical monitoring systems. This is due to the fact that practicality of the sensing, transmission, and processing steps is often the major obstacle against common use of such devices. Therefore, we believe that medical monitoring based on the emerging smart home wireless technology, ZigBee, has a great potential.

In addition, optimized processing of the collected data plays a key role. With optimization, we refer to the best use of computation and storage capacity at each of three different stages, namely, the mobile device, home server, and the central server. For example, the mobile device can play an important role in alerting the user in case of emergencies and therefore should be used for detecting the most urgent situations, especially in the absence of the wireless link. The home server typically has greater capacity and thus can perform much more complex tasks.

Another objective of this paper is to provide a secure and efficient scheme to meet the quality of service (QoS) requirements of medical and context data while they are transmitted from body sensors of mobile patients to health-care centers over body sensor networks and wireless networks of various types. In this regard, this paper also addresses two essential issues: (i) secure data transmission from body sensors to the mobile device over a body sensor network and (ii) efficient channel allocation and data security for transmission of medical and context-aware data to health-care center.

Earlier, we have proposed a wireless health monitoring system [22] that presents a three-tier architecture integrated to smart homes. Here, we enhance that architecture with (1) ZigBee profile descriptions for standards-based wireless communication. (2) A secure transmission mechanism that optimizes the use of system resources in medical monitoring applications. In addition, we provide the particulars of the system that we have developed that will be implemented in a pilot market. The next section presents a brief overview of the ZigBee technology. Then, we provide an introduction to ECG systems, which is the primary data we collect and process at the first stage of our development. The section that follows provides some specific aspects of our approach followed by some concluding remarks and discussion of our ongoing work.

2. MOBILE ECG SENSING, TRANSMISSION, AND ANALYSIS

Our system consists of several modules, each associated with a certain function as shown in Figure 1. Data collection module collects vital data, particularly ECG signal and provides local storage and transmission functionalities. The local server receives the data and preliminary analysis results and communicates with the central server as required. The central server follows the guidelines set by a particular service provider to initiate a sequence of response actions, including contacting the health professional.

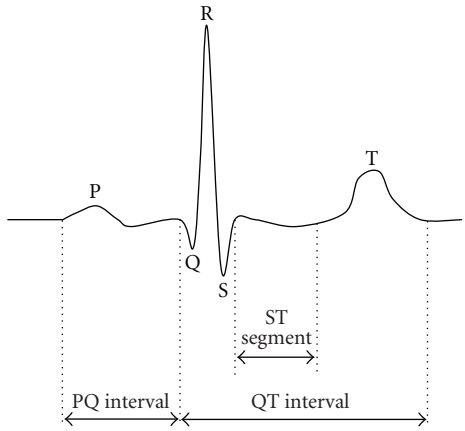


Figure 2: Illustration of the key ECG features.

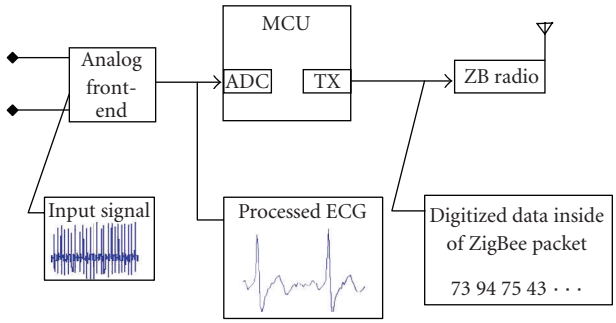


Figure 3: Block diagram of the ECG measurement platform.

Our system first measures the raw ECG signal from up to three electrodes and locally analyzes heart rate variability. If an arrhythmia risk is detected, an alert is transferred to the home server over the ZigBee network controller. At the same time, the raw ECG is measured and transmitted continuously to the home server, and then the home server analyzes the ECG records. If any anomaly is detected, patient's doctor is contacted. The ZigBee protocol does not have any transport layer functionality. Continuous transmission of the ECG data requires support for segmentation and reassembly, which is not offered by the current version of the ZigBee standard. We have implemented these functionalities at the application layer.

2.1. ECG basics

The electrocardiogram is primarily a tool for evaluating the electrical events within the heart. The action potentials of cardiac muscle cells can be viewed as batteries that cause charge to move through the body fluids. These moving charges can be detected by recording electrodes at the skin surface. Figure 2 illustrates the typical lead II ECG where the active electrodes are placed on the right arm and left leg.

The first deflection, the P wave, corresponds to a current flow during atrial depolarization. Normal P waves have various shapes, from flat to sharply-peaked with amplitudes ranging from 0 to 0.3 mv. The PQ interval, extending from the beginning of the P wave to the first component of the QRS complex, corresponds to the depolarization of the atria, AV node, AV bundle and its branches, and the Purkinje system. The second deflection, the QRS complex, is due to ventricular depolarization. The final deflection, the T wave, is the result of ventricular repolarization. Atrial repolarization is usually not evident on the ECG, because it occurs at the same time as the QRS complex.

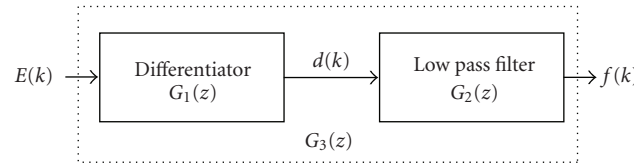


Figure 4: Filters for ECG signal conditioning.

2.2. Sensing and analysis of ECG data

We have built a device for sensing and filtering ECG signals. Key steps consist of low-noise amplification, quantization, digital filters, and feature detection algorithms. The processed digital data is then sent to the home server over the ZigBee network.

Typical ECG signal level on the human body surface is around 2 mV. The AD converter used in our setup accepts voltages from 0 to 3 V. Therefore, we first add 1.5 V offset to center the ECG waveform prior to amplification. The amplified signal is then quantized to 8 bits by the ADC within the M16C microcontroller. The discrete waveform is passed through a differentiator and lowpass filter as shown in Figure 4, where $E(k)$ represents the quantized ECG signal. The sampling rate in our implementation is 320 Hz. The filter transfer functions are as follows:

$$G_1(z) = 1 - z^{-1}. \tag{1}$$

$G_1(z)$ is a differentiator filter; and is used to obtain slope of the QRS complex. $G_2(z)$ is a lowpass filter to avoid residual noise and intrinsic differentiation noise. Overall filter response maximizes the energy of the QRS complex and improves detection of R wave peaks.

Detection of QRS peaks has a great value in diagnosis of many medical anomalies. For detection of QRS peaks, numerous adaptive threshold detection techniques are used in the literature. Normalized amplitude distribution function of a smoothed ECG signal is used in [23] to detect whether a QRS complex is present within an observation interval. The slope of the distribution function becomes very sharp when the QRS complex does not exist inside the interval. Threshold value is set to be proportional to the R-R interval. In [24], two thresholds are set and the number of crossings are used to determine the QRS complex. The threshold values are determined adaptively from the signal amplitude. The higher threshold γ_h is set to be $\gamma_H = \max(0.5P_{\mathrm{QRS}}, \mathrm{gamma}_L)$, and the lower threshold γ_L is set to be $\gamma_L = 0.75\gamma_L + 0.25P_{\mathrm{QRS}}$, where P_{QRS} indicates the peak of the latest QRS complex. These two algorithms are not very effective in presence of strong baseline drift.

In our work, a more robust adaptive threshold setting similar to that in [25] is used to detect QRS peaks. Note that due to severe baseline drifting and movement of patients, an ECG signal waveform may vary drastically from a heart beat to the next. With an adaptive threshold, probability of missing a QRS peak can be decreased.

In our platform, the first five seconds of the absolute value of the lowpass filtered digitized ECG data, $f(n)$, is searched for its highest peak. Let us denote the magnitude of this peak as $p[0]$. Then, the threshold τ is initialized to $\tau[0] = \alpha p[0]$, where $\alpha < 1$. In our implementation, we set $\alpha = 0.65$.

Let $p[i]$ denote the first local peak of $f(n)$ after a threshold crossing. After determining the slopes on both sides of $p[i]$, the zero crossing between $p[i]$ and the peak of the highest slope is chosen as the ith R wave peak location. The next threshold is set as follows:

$$\tau[i] = \alpha\tau[i-1] + (1-\alpha)p[i-1]. \quad (2)$$

If an R-to-R interval is measured as β times longer than the previous interval, where $\beta > 1$, a search is repeated only within that section of the ECG with a lower threshold to detect a possibly missed heart beat. We set β to 1.5 as a result of empirical observations.

The inverse of the interval between two consecutive R wave peaks gives the instantaneous heart rate. Their sequence shows how heart rate varies, which is another important medical data.

In Figure 5, a raw ECG trace and the output of the implemented R peak detector is shown. It is observed to be robust against baseline drifting caused by patient movements.

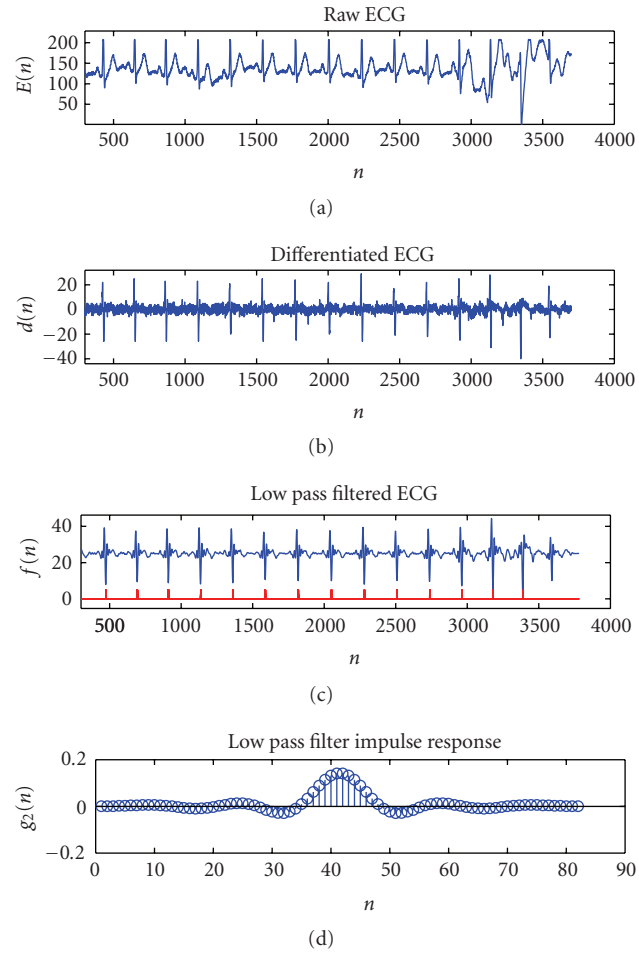


Figure 5: (a) Raw measured ECG signal, (b) ECG signal $d(n)$ after the differentiator, (c) ECG signal $f(n)$ after lowpass filtering. The detected R peaks are overlaid on the plot, and (d) impulse response of the lowpass filter, $g_2(n)$.

2.3. Secure transmission of vital sign data

Body sensors are used to sense vital sign data such as ECG for performing real-time health monitoring of mobile patients. Due to the transmission of sensitive medical data, it is imperative to build a strong security mechanism in order to protect the information transmitted as it may be susceptible to both active and passive attacks. Key distribution is central to any security mechanism based on cryptographic techniques. Whenever it is not possible to meet the power and computational requirements of asymmetric security techniques in small devices such as body sensors, symmetric key cryptography is employed to establish a secret session key for safeguarding data against various security attacks. Key management is fundamental to provide in body sensor networks (BSNs) because it provides and manages cryptographic keys to enable security services such as authentication, confidentiality, and integrity. Next, we will first introduce an attacker model and then present a secure key establishment and authentication algorithm for transmitting medical data from body sensors like ECG sensors to a handheld device of mobile patient, called personal wireless hub (PWH).

Our attacker model holds the following assumptions and properties. Prior to attaching body sensors to a mobile patient, we assume that they are not compromised and are embedded with a common global key, K_{CG}, at a secure place like home. K_{CG} is initially used to set up a session key at a secure place and then deleted. A compromised node of a BSN of a patient may eavesdrop on data being transmitted in its own BSN in order to break the session key. A compromised body sensor node of another close-by patient or intruder may eavesdrop on the medical data being transmitted in BSN of a neighboring patient. A compromised node from another BSN may try to inject false data into the BSN in order to force the PWH to establish a compromised session key. A compromised node may engage in replay attacks. A replay attack is a form of network attack in which a valid data transmission is maliciously or fraudulently repeated or delayed. This is carried out by a compromised node who intercepts the data and retransmits it at some later stage. This attack can be carried out by a compromised node which can be either internal or external to the BSN.

Our objective is now to establish two symmetric keys, namely $K_{\text{PWH-BSN}}$ and $K_{\text{BSN-BSN}}$ using an algorithm of three phases. The symmetric key $K_{\text{PWH-BSN}}$ is used to encrypt

data between BSN sensors and PWH, while the symmetric key $K_{\text{BSN-BSN}}$ is employed among BSN sensors. The first phase is required to be implemented in a more secure place like home, though the second and third phases can be implemented at any place. In the *first phase*, PWH first generates an initialization key K_{PWH} using its random number generator, XORs the key K_{PWH} with the timing information of the last ECG signal sent by the BSN, and then sends the XORed result to the BSN sensors. At the end of the first phase, BSN sensors first recover the initialization key K_{PWH} using the same last ECG signal sent to PWH, and then agree on a nonce N_{BSN} among themselves. In the *second phase*, all BSN sensors compute the symmetric key $K_{\text{PWH-BSN}} = K_{\text{PWH}} \oplus N_{\text{BSN}}$, encrypt N_{BSN} with K_{PWH}, and send the encrypted message $E_{K_{\text{PWH}}}(N_{\text{BSN}})$ to PWH. At the end of the second phase, PWH first recovers N_{BSN} by decrypting $E_{K_{\text{PWH}}}(N_{\text{BSN}})$ with the initialization key K_{PWH} and then recovers the symmetric key $K_{\text{PWH-BSN}}$. In the *third phase*, PWH generates a nonce N_{PWH}, computes the symmetric key $K_{\text{BSN-BSN}} = K_{\text{PWH}} \oplus N_{\text{PWH}}$, encrypts N_{PWH} with K_{PWH}, and sends the encrypted message $E_{K_{\text{PWH}}}(N_{\text{PWH}})$ to BSN sensors. At the end of the third phase, BSN sensors first recover N_{PWH} by decrypting $E_{K_{\text{PWH}}}(N_{\text{PWH}})$ with the initialization key K_{PWH} and then recover the symmetric key $K_{\text{BSN-BSN}}$. The established symmetric key can be used to encrypt data in those techniques supporting authentication, data integrity, and confidentiality. For instance, PWH can authenticate a BSN sensor node by comparing its aggregated timing information of heart beats with the heart beats' timing information sent by the BSN sensor node.

Once a session key is securely established between the body sensors and the PWH, the body sensors can use this session key for transmitting the data securely to the PWH. The established session key is known globally by all the body sensors and the PWH. Each body sensor also establishes a pairwise key with the PWH which is only known to the corresponding sensor and the PWH. To provide with added security, each body sensor provides with data confidentiality by encrypting the data with the session key and uses the pairwise key for data authentication. We can provide data confidentiality by encrypting the data with the session key established. To provide data integrity, a keyed-hash message authentication codes (HMACS) can be used along with a key as an input. To make it more secure from intruders, the key used for data confidentiality should be different from the key used to establish data integrity (to calculate the HMAC). The body sensors use the session key to encrypt the data to provide data confidentiality and use the pairwise key to calculate the HMAC to provide data integrity. Using this scheme, an intruder will need to have the knowledge of both the keys in order to spoof the PWH.

2.4. ZigBee for wireless sensing and transmission of medical data

Many medical applications will benefit from standards-based wireless technologies that are reliable, secure, and run on low power. Established standards for wireless applications, such as Bluetooth and IEEE 802.11, allow high transmission rates, but poses disadvantages such as high power consumption, application complexity, and cost. ZigBee networks on the other hand, are primarily intended for low duty cycle sensors, those active for less than 1% of the time for improved power consumption. For instance, an off-line node can connect to a network in about 30 milliseconds. Waking up a sleeping node takes about 15 milliseconds, as does accessing a channel and transmitting data. In addition, with their support of mesh networking and rapidly increasing popularity in wireless sensor network environments and smart homes make this networking technology a strong candidate for health applications as well.

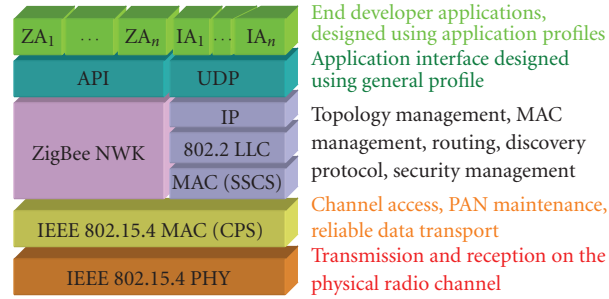


Figure 6: Illustration of ZigBee stack.

ZigBee is best described by referring to the 7-layer OSI model for layered communication systems Figure 6. The network name comes from the zigzagging path a bee (a data packet) follows to get from flower to flower (or node to node) [26]. The alliance specifies the bottom three layers (physical, data link, and network), as well as an Application Programming Interface (API) that allows end developers the ability to design custom applications that use the services provided by the lower layers. It should be noted that the ZigBee Alliance chose to use an existing data link and physical layer specifications. These are the recently published IEEE 802.15.4 standards for low-rate personal area networks. Complete descriptions of the protocols used in ZigBee can be found in [1, 27].

2.5. ZigBee network configuration: a medical profile proposal

Profiles in ZigBee networks provide a common protocol for communication within the network for a particular application to form an industry standard. So far, ZigBee Alliance has not issued an approved profile for use in health-care applications. In this section, we propose a proprietary profile description we have developed as part of this project that may also form a basis for future standardization efforts.

ZigBee network is configured in such a way that it uses one PAN per unit being monitored, that is, apartment, hospital section. Every device is configured as a ZigBee end-device. Several devices may coexist and report data simultaneously. In order to completely cover the monitored area, several additional ZigBee routers may be required. At the initial configuration, every router device discovers the route to the controller (PAN coordinator in our case). Every

router then can respond to the route discovery request from the ECG device and does not need to rediscover the entire route.

The application software running on ZigBee devices is responsible for the creation of proper payload that carries corresponding commands, responses, and data. Once the payload is created, it is passed to the ZigBee APS layer for the transmission over the air using API provided by ZigBee stack manufacturer. Application endpoint has one incoming and two outgoing. Incoming cluster is used for command and control messages, one of the outgoing clusters is used to send response to control messages and the other to send raw ECG data. Node can receive command messages such as "start", "stop" to control transmission, "setFQ" to set sample frequency, and others may be defined in the future. At 320 Hz, sample rate device produces 4 data packet per second. Depending on the hardware configuration (e.g., available RAM) some amount of data can be stored locally in case of temporary network failures.

ZigBee device endpoint consists of 2 incoming and 1 outgoing clusters. Outgoing cluster is for command and control interface. Incoming clusters receive command responses and ECG data. Upon receiving the data, ZigBee coordinator passes it to the server for further processing and analysis.

The application software running on ZigBee devices is responsible for creation of proper payload that carries corresponding commands, responses, and data. Once the payload is created, it is passed to the ZigBee APS layer for the transmission over the air using API provided by ZigBee stack manufacturer.

ZigBee Device Profile defines a set of commands and responses for a particular application. These are contained in *clusters* with the cluster identifiers enumerated for each command and response. Each ZigBee Device Profile message is then defined as a cluster. A cluster is a related collection of commands and attributes, which together define an interface to a specific functionality. Typically, the entity that stores the attributes of a cluster is referred to as the *server* of that cluster and an entity that affects or manipulates those attributes is referred to as the client of that cluster. Commands that allow devices to manipulate attributes, for example, the read or write attribute commands, are sent from a client device and received by the server device. Any response to those commands are sent from the server device and received by the client device. Conversely, the command that facilitates dynamic attribute reporting, that is, the report attribute command is sent from the server device (as typically this is where the attribute data itself is stored) and sent to the client device that has been bound to the server device. The clusters supported by an application object within an application profile are identified through a simple descriptor (see [1]), specified on each active endpoint of a device. In the simple descriptor, the application input cluster list contains the list of server clusters supported on the device, and the application output cluster list contains the list of client clusters supported on the device.

In our system, the server is the source node, that is, the mobile medical data collection device and client is the data collector node, that is, the home server. Our profile defines attributes and commands to configure ECG acquisition process. The server can receive the following commands: *SetSamplingFreq* to set sampling frequency of the analog signal at the ECG circuit, *Start* to start reporting ECG data, and *Stop* to stop reporting ECG data.

Table 1: ZigBee ECG profile.

Type	ZigBee headers	Timestamp	Freq.	ECG Data
Length	Variable	4 bytes	1 byte	80 bytes
Example		00 00 00 00	00	00 00 00 · · · 00

The server uses a reporting mechanism to send raw ECG data to the client for further processing. The format of the ECG data packet is summarized in Table 1. In the profile, Timestamp refers to the consecutive number that increments with every packet sent and allows to reconstruct ECG trace at the server in case packets come unordered or if some packets are missing; Freq is the sample frequency index, which is by default 320 Hz, and can be changed from the controller; ECG data is 80 bytes of digitized ECG signal, where each byte is a sample of the analog signal generated by the acquisition circuit.

3. EFFICIENT CHANNEL ALLOCATION FOR PATIENT DATA TRANSMISSION OVER WIRELESS NETWORKS

Future generation wireless networks will experience huge demands from mobile telemedicine applications. Mobile telemedicine will allow patients to do their daily activities while they are being monitored continuously anytime, anywhere. Typical telemedicine applications include transmission of ECG signals from the patient to the doctor, voice conversation between the doctor and the personnel in the emergency vehicle, transmission of X-rays, live video and medical images from the emergency vehicle, or the patient to the doctor at the health-care center. These applications require communication between a mobile patient and a health-care center or central server. But, these telemedicine applications will have to deal with the characteristics of wireless networks such as low bandwidth, channel fluctuations, and coverage changes. Further, a single network alone would not be able to meet the bandwidth requirements of applications at all locations. Fortunately, future mobile devices are expected to support multiple wireless interfaces, so that they can communicate through more than one wireless network at the same time. This allows the mobile applications to take advantage of various wireless networks, depending on their availability, channel conditions, coverage, and bandwidth. In this section, we present an efficient channel allocation algorithm to enable a PWH to allocate channels efficiently from wireless networks for mobile telemedicine applications.

Environmental factors can degrade the precision of ECG measurements. In addition, the human body can respond differently in cold temperatures and/or high altitudes, which may, for instance, prevent a finger pulse oximeter from taking an accurate measurement because blood flow to the fingers may get restricted. Therefore, not only ECG measurements,

but also context information such as the time, location, environment and weather information associated with the ECG measurements should be sent from the mobile device to the server for analysis. Another important characteristic of ECG data is the difference in the periodicity and sporadic nature of the data. When the patient is under good conditions, the ECG data of the patient is typically sent to the health-care center periodically to monitor the condition of the patient. In addition to the periodic data, when the patient's condition deteriorates, sporadic emergency data needs to be sent to the health-care center, which may impact the periodicity and thus the wireless bandwidth requirement for the ECG data.

We assume that a patient exists in one of three states at a given moment with respect to a particular vital sign. At any given time, a patient's health status sign would be *GOOD*, *FAIR*, or *CRITICAL*. A patient's health status is treated as *GOOD* if the vital signs are stable and within normal limits. A patient's health status is treated as *FAIR* if the vital signs show slight instability and the patient may be uncomfortable. A patient whose vital sign data are unstable and not within normal limits, the patient's status is treated as *CRITICAL*. A simple way to predict a patient's status is through the use of thresholds. For example, we can measure the number of times that a vital sign value from a body sensor exceeds the acceptable value specified by the doctor. If out of x consecutive vital sign values, y of those vital sign values exceed the acceptable value, then the patient's status can be concluded as moving from *GOOD* to *FAIR* or *CRITICAL*. The exact way of determination of patient's status is specified by the medical doctor based on the patient's health history and the particular vital sign being measured. Determination of patient's status should also take into account the current activity level of the patient. That is, when the patient is exercising or doing some other highly intense activity, a different threshold should be used in determining the status of the patient from the threshold used under normal activity conditions. The current activity level of the patient can be determined from observing the outputs of body sensors that record the patient's speed, elevation, angle, room temperature, and so forth.

Wireless bandwidth can be reserved for periodic ECG data since we know the amount and time of occurrence of the data. In contrast to this, reserving bandwidth for emergency ECG data is not efficient in terms of resource usage. At the same time, emergency ECG data should be delivered on time without too much delay. To solve this problem, we propose a scheme, where the periodic ECG data is differentiated and if the differences in the periodic data exceeds a threshold, the PWH will start reserving bandwidth on the wireless networks in order to ensure the availability of the wireless bandwidth for any possible emergency ECG data. This reduces the wastage of bandwidth resources by not reserving bandwidth for emergency ECG data all the time. At the same time, by predicting the occurrence of an emergency situation and reserving resources beforehand based on the prediction, it also improves the probability of bandwidth availability under emergency situations.

We now present the dynamic channel allocation algorithm that is invoked whenever bandwidth is not available to satisfy an ECG data call request for emergency ECG data in a cellular network. The dynamic channel allocation algorithm presented here dynamically requests and migrates bandwidth from neighboring base stations. The base stations are responsible for bandwidth allocation. The load on all base stations is not going to be uniform. Also, every possible attempt should be made to not drop an ECG data call. Therefore, a dynamic and adaptive channel allocation scheme is essential, where channels are shared dynamically by various base stations of the same network access technology by adapting to the requirements of the clients.

The adaptive dynamic channel allocation (ADCA) algorithm presented here does the channel allocation to an ECG or a non-ECG call. A base station uses a different frequency, time slot, or code for each connection with a client. We also assume that each BS knows its neighboring BSs, that is, the network is already established and it remains fixed. The base stations do not move, however, the wireless clients can move from the coverage area of one base station to another. It is possible that some of base stations may become more loaded than the others. In such a situation, some channels have to be transferred from one base station to another.

The basic steps of the ADCA algorithm are as follows. On line 2, each base station computes and sends its call blocking probability for ECG and non-ECG calls to all of its neighbors. Based on the knowledge of its own call blocking probabilities and of that of its neighbors, on line 3, each base station determines whether a request can be made to borrow channels from any of the neighboring nodes. Based on the determination, neighboring channels are requested. If this base station receives a channel borrowing request from the neighboring node, it first checks if the number of free channels under the base station is greater than the threshold of free channels. If so, an appropriate free channel is moved from the base station to the requesting neighbor on line 5. On lines 6 to 9, a channel is allocated to an ECG call. An ECG call is assigned a channel as long as there are free channels available. On lines 10 to 13, a channel is allocated to a non-ECG call. A non-ECG call is allocated a channel only if the number of free channels is greater than the threshold of guard channels, TGC.

The ADCA algorithm makes use of two thresholds while assigning channels to an ECG data or a non-ECG data call. The objective is to ensure low call blocking probability for ECG data calls. Every base station maintains two thresholds, TFC and TGC, where TFC is the threshold of free channels and TGC is the threshold of guard channels for ECG data calls (TFC $>$ TGC). Every base station periodically sends the call blocking probabilities of ECG and non-ECG calls to its neighbors. A non-ECG call is assigned a channel only when the number of free channels under the base station is greater than TGC, that is, TGC number of channels are always reserved for ECG data calls. In situations when an ECG data call cannot be allocated a channel even from the set of guard channels, the algorithm attempts to transfer a channel from a neighboring base station. A base station is allowed to transfer a channel only when the number of free channels under that base station is greater than TFC. Since TFC is greater than TGC, this transfer of channels does not

Input: A wireless network has N channels and M base stations. Initially, each BS is assigned an equal number of channels.
Output: Based on the blocking probabilities of ECG and non-ECG calls, channels are dynamically assigned among BSs. Channels are allocated to reduce the blocking and dropping probability of ECG data calls.
begin
1: *for* each base station, BS_i *do*
2: base station BS_i first computes the blocking probabilities of ECG and non-ECG calls and sends this information along with the list of occupied channels to the neighboring base stations.
3: Using the information of blocking probabilities in the local and the neighboring base stations, base station BS_i decides whether a request should be made to move an appropriate free channel from neighboring base stations, and then implements its decision.
4: *if* (number of free channels under base station, BS_i > TFC) and (a neighboring base station requests a free channel) *then*
5: An appropriate free channel is moved from base station BS_i to the neighboring base station requesting a free channel.
6: *else if* (an ECG-data call arrives) *then*
7: *if* (a free channel is available) *then*
8: Assign a free channel to the ECG-data call.
9: *end if*
10: *else if* (a non-ECG-data call arrives) *then*
11: *if* (number of free channels under base station$BS_i \geq$ TGC) *then*
12: Assign a free channel to the new call.
13: *end if*
14: *end if*
15: *end for*
end

Algorithm 1: Algorithm ADCA.

affect the call blocking probability of ECG data calls under the base station that transfers the channels.

4. A SMART HOME SYSTEM WITH INTEGRATED MEDICAL MONITORING

In the previous sections, we have presented techniques for the capture, analysis, and secure transmission of ECG data for real-time monitoring of persons/patients in their homes. We envision that such techniques can work best in practice when they are integrated to existing platforms in homes such as wireless smart home systems. We have devised a three-stage approach where the vital signals are processed at the mobile device for immediate, life-threatening situations as described in the previous section. Here, we briefly present the processing at the next two levels: processing at the (local) home server and processing at the central service provider.

For improved performance, we perform the basic processing of ECG signals at the mobile device. This part entails the measurement pulse rate through QRS peak detection algorithm discussed earlier. This is done to facilitate a reliable warning in case of a network failure. Next, raw data basic analysis results are transmitted to the home server and central server as described in the next two sections.

4.1. Medical data processing at the home server

Digitized ECG data is continuously transmitted to the home server via a ZigBee network. Additionally, results of the analysis at the mobile device are sent to the server and stored here for future reference. The goal is to provide a repository for more detailed analysis of the data by medical professionals or detection algorithms. In addition, the stored data is processed for more detailed and accurate analysis of ECG signals for detections such as Q-T interval and T wave detection.

The main responsibilities of the home server are: (1) coordinate the ZigBee in-home wireless network, (2) store incoming data, (3) conduct accurate and detailed analysis of the data, and (4) communicate with the central service

provider for transmission of the data and notifications for detected anomalies. We are currently developing a Linux-based architecture housed on a PC-platform that allows remote access through a web-based interface. The server also is connected to the ZigBee network through a coordinator module connected via USB or serial port. Routers on the smart home network will be continuously powered and distributed throughout the home, possibly one for each room. An existing ZigBee network normally used for lighting or security can be used for this purpose as well. The data repository is made available for future use by service provider as well as the user and is backed up against losses through a data warehousing service.

4.2. Medical data processing at the central server

Continuous recording and analysis of ECG data provides an excellent basis for automated detection as well as professional diagnosis of many cardiological symptoms. According to our model, the last and the third piece is the central data processing center where servers as well as medical personnel can provide a variety of services such as storage, early diagnosis, and in-home care. The home server transmits periodic reports and makes stored data available to the central server.

A key point for the central server processing is the optimization of the use of resources at the home server resources and the central server. As the number of users increase, the central server can allocate only a limited amount of computational capacity to each user. Therefore, data analysis is performed at the home server as much as possible.

The central server also functions as an entry point for the professional staff to monitor the data and reports generated by the home server. In addition, the alerts initiated by the mobile device are transmitted to the central server through the home server. The central server also keeps records of all transactions through an event management system.

5. CONCLUSION

We have presented a real-time and secure architecture for health monitoring in smart homes using ZigBee technology. Our research has outlined many specific issues that relate to a collection of emerging technologies, namely, wireless communication, secure transmission, and processing of medical data within the context of "smart environments". In particular, we believe that our work has made the following contributions:

(i) description of a ZigBee profile for transmission of ECG data over a wireless sensor network,

(ii) a security model for secure transmission of data over wireless sensor network,

(iii) an efficient mechanism for channel allocation in wireless networks to transmit medical data, and

(iv) a three-tier architecture for optimized analysis of data using an adaptive mechanism that maximizes the utility of processing and computational capacity at each of three stages.

Our project continues towards completion of the implementation of local and central server components for an end-to-end service. In addition, we are expanding the collection of information to other modalities such as oxygen, temperature, and glucose level measurements.

REFERENCES

[1] ZigBee Alliance Document 02130, Network Layer Specification, July 2004.

[2] Final Report, "Mobile Telemedicine Testbed for National Information Infrastructure," National Institute of Health, Project N0-1-LM-6-3541, http://collab.nlm.nih.gov/tutorialspublicationsandmaterials/telesymposiumcdbdmfinal.pdf/, August 1998.

[3] D. Malan, T. Fulford-Jones, M. Welsh, and S. Moulton, "Codeblue: an ad-hoc sensor network infrastructure for emergency medical care," in *Proceedings of the 1st International Workshop on Wearable and Implantable Body Sensor Networks (BSN '04)*, London UK, April 2004.

[4] D. Jea and M. B. Srivastava, "A remote medical monitoring and interaction system," in *Proceedings of the 4th International Conference on Mobile Systems, Applications, and Services (MobiSys '06)*, Uppsala, Sweden, June 2006.

[5] R. Jafari, F. Dabiri, P. Brisk, and M. Sarrafzadeh, "CustoMed: a power optimized customizable and mobile medical monitoring and analysis system," in *Proceedings of ACM HCI Challenges in Health Assessment Workshop in Conjunction with Proceedings of the Conference on Human Factors in Computing Systems (CHI '05)*, Portland, Ore, USA, April 2005.

[6] R. Chakravorty, "A programmable service architecture for mobile medical care," in *Proceedings of the 4th Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom '06)*, pp. 532–536, Pisa, Italy, March 2006.

[7] A. Y. Benbasat and J. A. Paradiso, "A compact modular wireless sensor platform," in *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks (IPSN '05)*, pp. 410–415, Los Angeles, Calif, USA, April 2005.

[8] Patient-Centric Network, http://nms.lcs.mit.edu/projects/pcn/.

[9] J. W. P. Ng, B. P. L. Lo, O. Wells, et al., "Ubiquitous monitoring environment for wearable and implantable sensors (UbiMon)," in *Proceedings of the 6th International Conference on Ubiquitous Computing (UBICOMP '04)*, Nottingham, UK, September 2004.

[10] European MobiHealth Project, http://www.mobihealth.org/.

[11] U. Anliker, J. A. Ward, P. Lukowicz, et al., "AMON: a wearable multiparameter medical monitoring and alert system," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 4, pp. 415–427, 2004.

[12] H. Junker, M. Stäger, G. Tröster, D. Blttler, and O. Salama, "Wireless networks in context aware wearable systems," in *Proceedings of the 1st European Workshop on Wireless Sensor Networks (EWSN '04)*, pp. 37–40, Berlin, Germany, January 2004.

[13] S. Park and S. Jayaraman, "Enhancing the quality of life through wearable technology," *IEEE Engineering in Medicine and Biology Magazine*, vol. 22, no. 3, pp. 41–48, 2003.

[14] R. Paradiso, "Wearable health care system for vital signs monitoring," in *Proceedings of the 4th IEEE International Conference on Information Technology Applications in Biomedicine (ITAB '03)*, pp. 283–286, Birmingham, UK, April 2003.

[15] E. Jovanov, A. Milenkovic, C. Otto, and P. C. de Groen, "A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 2, article 6, pp. 1–10, 2005.

[16] K. Van Laerhoven, B. P. L. Lo, J. W. P. Ng, et al., "Medical healthcare monitoring with wearable and implantable sensors," in *Proceedings of the 3rd International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications (UbiHealth '04)*, Nottingham, UK, September 2004.

[17] BodyMedia, http://www.bodymedia.com/.

[18] IBM Linux Wrist Watch, http://www.research.ibm.com WearableComputing/.

[19] B. P. L. Lo and G.-Z. Yang, "Implementations of body sensor networks," in *Proceedings of the 2nd International Workshop on Body Sensor Networks (BSN '05)*, London, UK, April 2005.

[20] CodeBlue project, http://www.eecs.harvard.edu/ mdw/proj /codeblue/.

[21] J. Polastre, R. Szewczyk, C. Sharp, and D. Culler, "The mote revolution: low power wireless sensor network devices," in *Proceedings of the 16th Symposium on High Performance Chips (Hot Chips '04)*, p. 10, Los Alamitos, Calif, USA, August 2004.

[22] S. Dağtaş, G. Pekhteryev, and Z. Şahinoğlu, "Multi-stage real time health monitoring via ZigBee in smart homes," in *Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW '07)*, vol. 1, pp. 782–786, Niagara Falls, Canada, May 2007.

[23] K. Akazawa, K. Motoda, A. Sasamori, T. Ishizawa, and E. Harasawa, "Adaptive threshold QRS detection algorithm for ambulatory ECG," in *Proceedings of the 18th Annual Conference on Computers in Cardiology*, pp. 445–448, Venice, Italy, September 1991.

[24] J. C. T. B. Moraes, M. M. Freitas, F. N. Vilani, and E. V. Costa, "A QRS complex detection algorithm using electrocardiogram leads," in *Proceedings of the 29th Annual Conference on Computers in Cardiology*, pp. 205–208, Memphis, Tenn, USA, September 2002.

[25] Z. Şahinoğlu, "Analysis of multi-lead QT dispersion by means of an algorithm implemented on labview," M.S. thesis, New Jersey Institute of Technology, Newark, NJ, USA, January 1998.

[26] ZigBee Networks Open the Door to More Wireless Medical Devices, Medical Design, April 2005.

[27] IEEE Standard for Part 15.4, Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications for low rate Wireless Personal Area Networks (WPANs), 2003.

Hindawi Publishing Corporation
International Journal of Telemedicine and Applications
Volume 2008, Article ID 436870, 5 pages
doi:10.1155/2008/436870

Research Article

Development of a Novel Contactless Mechanocardiograph Device

Kouhyar Tavakolian, Faranak M. Zadeh, Yindar Chuo, Ali Vaseghi, and Bozena Kaminska

Computational and Integrative Bio-Engineering Research laboratory (CIBER), School of Engineering Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada V5A 1S6

Correspondence should be addressed to Kouhyar Tavakolian, kouhyart@sfu.ca

Received 13 October 2007; Revised 15 January 2008; Accepted 3 March 2008

Recommended by Sajid Hussain

A novel method of detecting mechanical movement of the heart, Mechanocardiography (MCG), with no connection to the subject's body is presented. This measurement is based on radar technology and it has been proven through this research work that the acquired signal is highly correlated to the phonocardiograph signal and acceleration-based ballistocardiograph signal (BCG) recorded directly from the sternum. The heart rate and respiration rate have been extracted from the acquired signal as two possible physiological monitoring applications of the radar-based MCG device.

1. INTRODUCTION

The ballistocardiograph (BCG) is a vital signal in the 1–20 Hz frequency range which is caused by the movements of the heart and blood. The BCG can be recorded from the surface of human body by noninvasive means. In the early 1930s, Isaac Starr recognized that the BCG signals closely reflect the strength of myocardial contraction and classified them into four groups to distinguish normal and abnormal heart performances [1, 2]. As a result of his valuable research, clinicians and medical experts for almost four decades studied the effects of different heart malfunctions by means of the BCG and proved that these malfunctions can be related to typical patterns in BCG signals [3, 4]. The ideal BCG waveform consists of seven waveform peaks labeled H through N as defined by the American Heart Association, and this annotation can be seen in Figure 1.

In addition to a number of clinical studies that have been performed with the BCG, specialized BCG instruments, including beds and chairs, have been developed by different research groups [5–7]. Due to the unrefined nature of the previous BCG signal acquisition technologies, the lack of interpretation algorithms and the lack of practical devices, the current health care systems do not use BCG for clinical purposes. New microsensor and digital wireless technologies provide more accurate BCG signal acquisition and processing, therefore, opening new possibilities to use BCG, and in general mechanical signals of the heart, clinically as a new additional tool in diagnosis and identification of heart malfunctions.

Based on these observations, we are improving the BCG signal acquisition system that reflects the mechanical functioning of the heart. In our first study [8], we used a high-precision accelerometer, which is factory calibrated, weighs 54 g, and is attached to the sternum, the same as used in reference [9]. In our second study, we improved the previous sensor by developing a small mechanically flexible microsensor that could be easily attached to the sternum. It has been proven that all the aspects of a classical BCG signal, (H, I, J, K) waves similar to that measured through a high-precision accelerometer [9], can be identified in the novel microsensor recordings [10].

The current research, reported in this paper, is aiming to go one step further by applying contactless measurement of the mechanical movement of the heart, using a radar-based MCG device operating in the range of 2.45 GHz [11]. The extracted signal is considered a mechanocardiograph (MCG) reflecting the mechanical functioning of the heart and is highly correlated to the BCG signal acquired in previous studies [12].

This paper is organized as follows: in the second section of the paper our signal acquisition methodology is explained

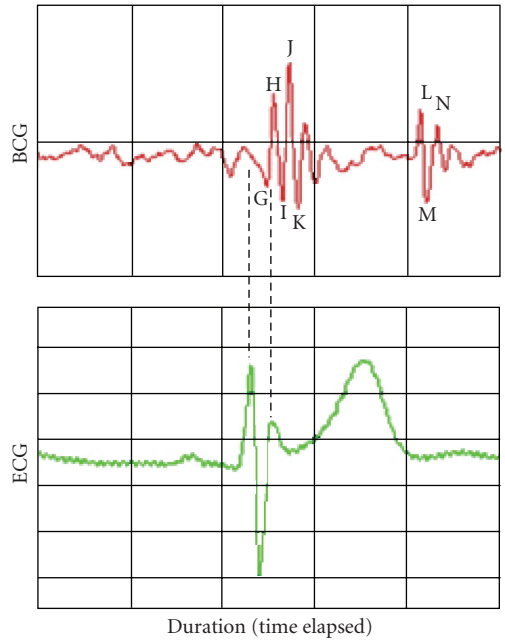


Figure 1: BCG and ECG signals of a single heart-cycle as recorded by the microsensor with the annotation referenced to ECG.

and then the comparison of the acquired MCG signal to our previous BCG recordings is presented. In section three, the electronic circuit board design developed for the radar-based MCG device is explained. In section four the results of breathing rate and heart rate measurements using the new technology are presented. The comparison results of the phonocardiograph and the acquired signal are also presented clearly showing the correspondence of our MCG signal to cardiac events.

2. MCG SIGNAL ACQUISITION

The block diagram in Figure 2 shows the general methodology followed in our research for processing of the radar-based MCG signal. The MCG signal was acquired by a microprocessor controlled, radar-based MCG device placed at distance of 10 cm away from the subject's chest. The data acquisition, for the results presented in this paper, involved the measurement of ECG and respiration signals too. These signals were acquired by a Biopac biological data acquisition system [13]. The design principle of radar-based MCG device is presented in section three of the paper.

While monitoring the heart from the outside of the body contactlessly from the sternal location, the signal passes through only a few layers of different tissues between the sternum and the heart which can be seen in Figure 3. The tissue layers between the sensor and heart muscle include: skin, sternum, lung and pleural tissue, pericardium and pericardial space. From the sternum position, these tissue layers are thinner compared to the other positions. Therefore, the best position to record the heart's MCG signal is from the sternum (normally on the first third of the sternum length, as this is the closest path to get to the heart muscle).

RF signal with the carrier frequency of 2.45 GHz was transmitted toward the subject's chest, and the reflected signal was band-passed filtered between 0.5 Hz to 25 Hz. The filtered signal was differentiated and then band-pass filtered again between 4 Hz to 20 Hz. The comparison of the processed MCG signal to the BCG signal recorded simultaneously from the sternum can be seen in Figure 4 together with the synchronized ECG signal.

It can be observed that there is a close correlation between the signal acquired from the radar-based MCG device and the signal acquired from the accelerometer-based BCG device attached to the sternum. The systolic and diastolic phases of cardiac cycle are shown to identify the correlations of these mechanical signals to the heart functioning.

Phonocardiograph signal reflects the heart sounds that can be hared using stethoscope. Heart sound S1 corresponds to systolic phase of the heart cycle, and S2 corresponds to the diastolic phase of it. For the comparison purpose, the MCG signal was acquired synchronously with phonocardiograph signal and it was observed that S1 and S2 sounds of the phonocardiograph signal corresponded to the similar complexes on the MCG signal. The clear correlation of this new signal to the heart cycle's events can be observed in Figure 6.

3. CONTACTLESS MCG DEVICE

As stated before, in our experiments heart's MCG signal was measured by a MCG device operating in the range of 2.45 GHz. At this electromagnetic frequency, the surface of the body is highly reflective to incident electromagnetic field. Biological tissue is very lossy at this frequency, and there is minimal penetration of radiated electromagnetic energy into the body. Therefore, a return signal from a radiated electromagnetic field incident on the body will primarily contain information associated with movement events.

The effects of radio-frequency fields on human health have been monitored and studied by the scientists around the world for about 60 years now. The 2.45 GHz frequency is considered to be safe for human health. This is the same frequency range as emitted from a cell phone [14]. The specific absorption rate, SAR, is 0.42 w/kg which is below the limits as recommended by international commission on nonionizing radiation protection [15].

The principal design of the radar-based MCG device is shown in Figure 5. The antenna mounted in the device is HFMD24 by Siemens and contains a transmitter and a receiver in the same housing. The operating frequency is 2.45 GHz. The transmitter transmits continuous wave radio frequency energy towards subject's body. The returned signal is frequency shifted due to the Doppler effect, which is the apparent change in frequency and wavelength of a wave that is perceived by an observer moving relative to the source of the waves. This effect permits the measurement of slight body movements from which physical heart movement signal can be obtained by the MCG receiver.

Figure 2: Radar-based MCG signal processing methodology.

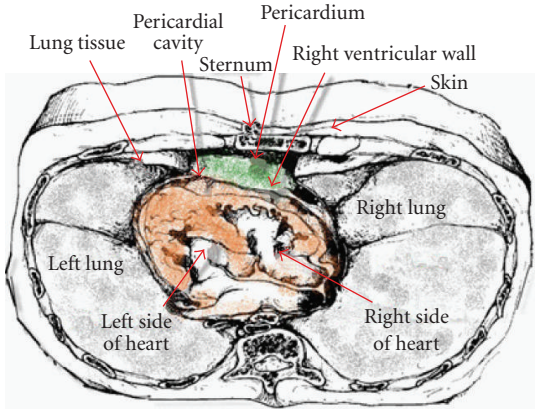


Figure 3: Positioning of different layers of tissues that the radar signal will go through. The layers between sternum and right ventricle include pericardium, pericardial cavity, and thin layers of pleural tissues.

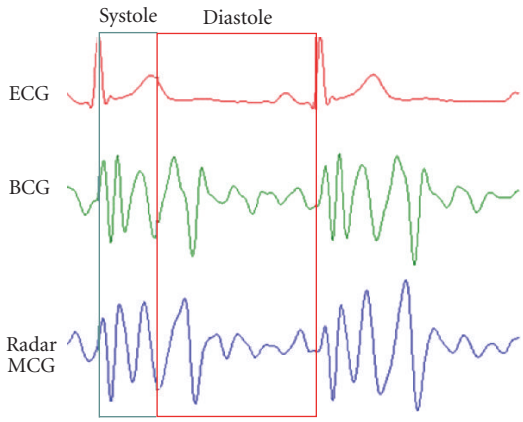


Figure 4: Two cycles of the filtered radar-based MCG signal (bottom), the filtered signal recorded directly from the sternum using a high-precision accelerometer (middle), and the lead I of ECG signal (top).

The output signal from the receiver is filtered and amplified and is sent to the microprocessor for further processing. The cutoff frequencies for the band-pass filter shown in the schematic are 1 Hz and 100 Hz, and the gain of the amplifier is around 800. After filtering and amplifying, the MCG signal is sent to the A/D unit and then to the ATMEL CPU for further processing. The CPU is connected to a thin-film transistor (TFT) display via its serial peripheral interface (SPI) port. The MCG device operates with two AA batteries (2.45 V).

Considering that the MCG device has its own CPU and monitor, it can be used stand alone to acquire and process the MCG signal. There is also another option of sending the data to a personal computer for more advanced processing of the data using Matlab. To have this option on our device, the digitized MCG signal is transformed to packets and sent through UART to the USB and finally to the host personal computer for possible further processing.

4. RESULTS AND DISCUSSION

The data from subjects were acquired at Burnaby General Hospital (British Columbia, Canada) and in all the measurements presented in this paper the distance from the sensor to the chest was set to 10 cm. The processed MCG signal has been superimposed on the phonocardiograph signal, and it has been noticed that S1 and S2 sounds of the phonocardiograph signal corresponded to the similar complexes on the radar MCG signal. The correlation of the MCG signal to the heart cycle's events is observed, and the results of this superimposition can be seen in Figure 6.

For the heart rate measurement our experimental setup included the acquisition of the MCG signal and two leads of ECG as a reference. For respiration rate measurement, the setup included the acquisition of MCG signal together with the respiration signal as the reference. Eight subjects took part in the respiration measurement tests, and six of these subjects took part in the heart rate measurement tests too. Breathing rate measurement experiments were 60 seconds long while heart rate measurement experiments were 15 seconds long.

For detection of breathing rate, the radar signal was low pass filtered under 0.4 Hz, and the peaks were counted and compared to the results acquired from a strain gauge transducer that measures the changes in thoracic circumference, using a belt which is fastened to the subject's thorax. As can be seen in Table 1, the accuracy of respiration rate measurement was 91.35 percent over all eight subjects. The heart rate was measured using radar-based MCG device and was compared to the heart rate calculated from the simultaneous ECG signal for six subjects. The average of the heart rate accuracy on these subjects was calculated to be 92.9 percent.

As an example, for subject six we had 14 cycles of respiration from which 13 of them where correctly identified. Thus, for this subject, the 92 percent accuracy for respiration rate detection was achieved. For the same subject and out of 19 heart beats, 17 of them were correctly identified. Thus, for heart beat detection for this subject, the accuracy of 89.4 percents was achieved. Using our device, the data can be seen real time and can be processed in near real time. The reason for quick processing time is that the algorithm includes filtering of the data and then peak searching to find

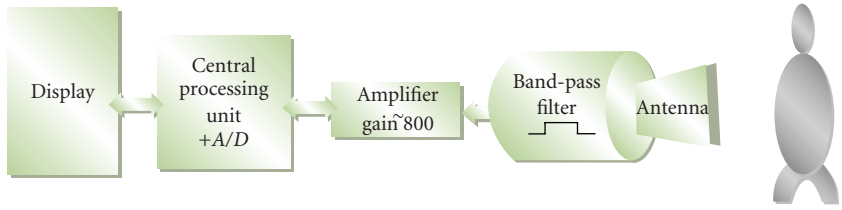


Figure 5: Block diagram of the principle design of the contactless MCG device.

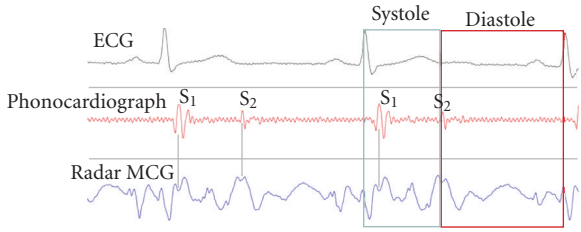


Figure 6: Two cycles of synchronous radar MCG, phonocardiograph, and ECG signal showing the correlation of cardiac cycle events to radar MCG signal. Systolic and diastolic complexes can be identified in the radar MCG signal corresponding to S1 and S2 of heart's sounds.

Table 1: The heart rate and respiration rate measurements using the MCG device for eight subjects. The numbers represent the percentage of correctly detected heart beats or breathing cycles to the total number of heart beats or breathing cycles.

Subjects	Heart rate	Respiration rate
1	93	90
2	—	94.2
3	84	90
4	91	90
5	100	87.1
6	89.4	92
7	100	100
8	—	87.5
Averages	**92.9**	**91.35**

the corresponding peaks to the inspiration and heart beat events.

The sources of interference for contactless heart motion signal acquisition can be categorized into two main categories. First, foreign electromagnetic radiation that may disrupt the desired signal on its way back to the antenna, and second, motion interference of nearby subjects. As discussed previously, the radiation frequency of the contactless MCG device is centered about 2.4 GHz. This frequency falls within the Industrial Scientific and Medical (ISM) bands. The ISM bands are recognized internationally, where it has been traditionally reserved for industrial, scientific, and medical purposes excluding communications [16].

The 2.4 GHz ISM band in many urban settings is in fact quite saturated with devices. Examples include the very common wireless local area network (WLAN) devices, Bluetooth devices, microwaves ovens, and household cordless phones. These devices can exhibit relatively high signal broadcast power levels during full power transmission. In the North American/US market, the safe power level is governed by the FCC [16]. As a result, it is possible that other 2.4 GHz, ISM devices operating in the vicinity of measurement could pose undesirable electromagnetic signal interference, particularly when these devices are in-motion. However, most cellular phone devices operate in the Global System for Mobile Communications (GSM) Bands, and thus will not pose electromagnetic interference concerns, in the sense of disturbing the Doppler reading, for the contactless MCG device proposed.

Motion interference is another possible source of signal disruption when operating the device in a high-density, busy setting. The antenna patch on the contactless MCG device is a printed patch antenna and at ranges greater than what is specified operational for this application (i.e., more than 10 cm), its side-lobes begin to pickup motion in its path. Also the main-lobe of the antenna begins to exhibit relatively wide angle. In attempts to acquire useful heart motion signals at longer ranges, more than 10 cm, the motion interference from the subject's general body movement or nearby movements that are not subject-related will both disrupt the heart motion signal acquisition greatly.

In efforts to evaluate the reliability of this device considering the two major sources of interference, in addition to the measurements at Burnaby General Hospital, as explained previously, measurements on 200-plus participants during the Wired NextFest convention were obtained [17]. The environment of the convention hall naturally provided many possible sources of both electromagnetic interference and motion interference. In fact, the convention hosted more than 150 stations, many equipped with audio-video setup, wireless network stations, and up to thousands of visitors at any given moment, likely each carrying some sort of personal mobile device.

The heart motion measurements were compared simultaneously with the corresponding heart beat of each individual participant as assessed via wrist pulse by research staff members. It was observed that when the subject held still, and without other people within 50 cm of the device radius, approximately only one out of ten subjects during 30 second recording session would exhibit grossly delayed or out-of-sync heart rates between the wrist pulse and device reading.

Although it is unknown how many wireless broadcasting devices, using the 2.4 GHz band, were present during the convention, nor the associated power levels, and hence the random interference power density in the foot-ball sized convention hall, the quasi field-test of the device provided assurance that even in the busiest everyday environments, filled with urban electronics and random motion, when the subject recordings are controlled properly, the reliability of the heart rate detection is seemingly good.

Further testing on the effects of both electromagnetic and motion interference is necessary to fully assess the performance of the proposed contactless MCG device. Though, it suffices to say that the observations through the combined evaluations and field-trials give high-hopes on the performance and for the eventual perfection of this device.

5. CONCLUSION

In this research work, the MCG signal was extracted from the radar-based MCG device and by using two other heart related signals, BCG and phonocardiograph, it has been proven that the extracted MCG signal corresponds to the mechanical functioning of the heart.

Two physiological signals, heart rate and respiration rate, were measured using the MCG signal and a noticeable accuracy of 91.35% for respiration rate, and 92.9% for heart rate were achieved. It should be noticed that these results are acquired from the first generation of our device, and the accuracy of the detection can be improved by focused targeting of the MCG device on the sternum and further hardware improvements. Another important factor, which is an asset of our device, is the fact that all the electronics and the radar sensor can be fabricated for less than $200 USD which makes our device economical for general heart monitoring purposes.

Considering the short amount of processing and the low cost of the contactless measurement of heart rate and breathing rate as we presented, our device can have numerous applications in the health care system such as, prevention of sudden infant death syndrome (SIDS), sleep apnea, and other areas in which contactless monitoring is desired.

In this research work, we have proposed a new methodology and a new device based on radar technology. We proved the correspondence of our acquired MCG signal to sternal BCG and phonocardiograph signals. Further improvement of the current device can provide us with more accurate MCG signal corresponding to mechanical performance of the heart and ultimately development of an additional contactless heart diagnostic tool.

ACKNOWLEDGMENTS

The authors would like to thank Burnaby General Hospital for providing them with subjects and the research environment and also Brandon Ngai, Brent Carmichael, and Tim Snider for their support during signal acquisition experiments.

REFERENCES

[1] I. Starr and E. C. Wood, "Twenty-year studies with the ballistocardiograph. The relation between the amplitude of the first record of "Health" adults and eventual mortality and morbidity form heart disease," *Circulation*, vol. 23, no. 5, pp. 714–732, 1961.

[2] I. Starr, "Prognostic value of ballistocardiograph," *Journal of American Medical Association*, vol. 187, no. 7, pp. 511–517, 1964.

[3] J. L. Nickerson, G. H. Humphreys, R. A. Deterling, and T. C. Fleming, "Diagnosis of coarctation of the aorta with the aid of the low frequency, critically damped ballistocardiograph," *Circulation*, vol. 1, no. 4, pp. 1032–1036, 1950.

[4] B. Phibbs, C. R. Lowe, and R. W. Holmes, "The ultra low frequency force ballistocardiograph in acute cardiomyopathy," *Circulation*, vol. 36, no. 1, pp. 92–100, 1967.

[5] B. H. Jensen, B. H. Larsen, and K. Shankar, "Monitoring of the ballistocardiogram with the static charge sensitive bed," *IEEE Transaction on Biomedical Engineering*, vol. 38, no. 8, pp. 748–751, 1991.

[6] S. Junnila, A. Akhbardeh, L. C. Barna, I. Defee, and A. Värri, "A wireless ballistocardiographic chair," in *Proceedings of the 28th IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBS '06)*, pp. 5932–5935, New York, NY, USA, August-September 2006.

[7] W. Dock, H. Mandelbaum, and R. Mandelbaum, *Ballistocardiography: The Application of the Direct Ballistocardiograph to Clinical Medicine*, Mosby, St. Louis, Mo, USA, 1953.

[8] B. Kaminska, K. Tavakolian, B. Ngai, Y. Chuo, A. Vaseghi, and B. Carmichael, "Prevention and risk reduction in heart health," in *Proceedings of the Nanotechnology Conference and Trade Show (Nanotech '07)*, vol. 2, pp. 531–534, Santa Clara, Calif, USA, May 2007.

[9] W. P. S. McKay, P. H. Gregson, B. W. S. McKay, and J. Militzer, "Sternal acceleration ballistocardiography and arterial pressure wave analysis to determine stroke volume," *Clinical and Investigative Medicine*, vol. 22, no. 1, pp. 4–14, 1999.

[10] Y. Chuo and B. Kaminska, "Multiparameter single locus integrated multilayer polymer microsensor system," in *Proceedings of the International Conference on Biomedical Electronics and Devices (BIODEVICES '08)*, Funchal, Portugal, January 2008.

[11] F. M. Zadeh, F. Taghibakhsh, and B. Kaminska, "Contactless heart monitoring," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE '07)*, pp. 583–585, Vancouver, BC, Canada, April 2007.

[12] K. Mohri, T. Jinnouchi, and K. Kawano, "Accurate mechanocardiogram sensors using amorphous star-shaped core multivibrator combined with a magnet," *IEEE Transactions on Magnetics*, vol. 23, no. 5, pp. 2212–2214, 1987.

[13] "Biopac biological signal acquisition systems," http:// www.biopac.com/.

[14] September 2007, http:// www.scientificamerican.com/.

[15] December 2007, http://www.icnirp.de/.

[16] January 2008, http://www.fcc.gov/.

[17] December 2007, http://www.wirednextfest.com/.

Hindawi Publishing Corporation
International Journal of Telemedicine and Applications
Volume 2008, Article ID 290431, 9 pages
doi:10.1155/2008/290431

Research Article

PATHOS: Pervasive at Home Sleep Monitoring

Ian Obermiller and Sheikh I. Ahamed

Mathematics, Statistics and Computer Science Department, Marquette University, Milwaukee, WI 53201-1881, USA

Correspondence should be addressed to Sheikh I. Ahamed, sheikh.ahamed@mu.edu

Received 16 October 2007; Accepted 15 January 2008

Recommended by Sajid Hussain

Sleeping disorders affect a large percentage of the population, and many of them go undiagnosed each year because the method of diagnosis is to stay overnight at a sleep center. Because pervasive technologies have become so prevalent and affordable, sleep monitoring is no longer confined to a permanent installation, and can therefore be brought directly into the user home. We present a unique solution to the problem of home sleep monitoring that has the possibility to take the place of and expand on the data from a sleep center. PATHOS focuses not only on analyzing patterns during the night, but also on collecting data about the subject lifestyle that is relevant and important to the diagnosis of his/her sleep. PATHOS means "evoking emotion." Here, we mean Pathos will help us to keep healthy: both mentally and physically. Our solution uses existing technology to keep down cost and is completely wireless in order to provide portability and be easily to customize. The daytime collection also utilizes existing technology and offers a wide range of input methods to suit any type of person. We also include an in-depth look at the hardware we used to implement and the software providing user interaction. Our system is not only a viable alternative to a sleep center, it also provides functions that a static, short-term solution cannot provide, allowing for a more accurate diagnosis and treatment.

1. INTRODUCTION

Sleep problems affect more than 70 million people in the United States alone [1]. Yet, the public as a whole is largely unaware of the causes and consequences of the disorder that one in four people have. Additionally, the majority of sleep disorders go undiagnosed and untreated. Sleep disorders range from snoring to insomnia, and can be as serious as obstructive sleep apnea, when a person stops breathing multiple times during the night. Because sleep problems cost Americans over $100 billion a year in lost productivity and medical expenses, it is imperative that more people are diagnosed and treated for such a common illness.

Sleeping disorders are usually diagnosed only when a person goes to a sleep center to have diagnostics taken over night. During such a study, known as a polysomnogram, a patient goes to a sleep center, spends around 45 minutes getting hooked up to a machine via electrodes, sensory belts, a microphone, and various other sensors. After the process, which is painless and made to be as nonintrusive as possible, the patient goes to sleep and the numbers are gathered. There is no doubt that this method of sleep monitoring works, or that it is the main mode of diagnosing many important sleeping disorders. However, sleep centers have several drawbacks, including cost, time, and environment.

In order for a truly universal system for monitoring sleep and diagnosing disorders to become a reality, it must be moved from the center to the home. Although this is not a new idea, the technology has only recently caught up with the proposal. There is a continual research to improve the sensors used in polysomnography. Researchers at the University of Aizu [2], for example, have created an under-pillow sensor to measure respiratory rhythm and pulse during sleep. A more complete system, with more sensors, has been designed at the University of Cairo [3] and is presented with a detailed hardware design. The University of Washington [4] has researched using inexpensive multimodal sensors to detect sleep conditions in a completely noninvasive manner.

The majority of previous research in the area of polysomnography technology has been directed toward the development hardware designs that would be more practical in a sleep center than in a home. The aim of the project at the University of Cairo [3], however, was a home monitoring solution, but its purpose was mainly to decide whether or

not a person should go through with a full evaluation at a dedicated sleep center.

Our approach to the field of polysomnography combines several aspects of the previous research and seeks to provide an all-encompassing solution that negates the need for a person to ever spend the night at a costly sleep center. The three main goals of PATHOS are to be (1) simple and inexpensive, using existing technology, (2) performed in a natural environment, namely, the subject's home, and (3) a continuous solution, not limited to gathering data during the night. For night monitoring, we propose using a mesh network of simple sensors that wirelessly transmit data to a subject's mobile device or pc. The sensors are inexpensive and require interpretation which can be performed by small motes which form the nodes of the network. Additionally, data regarding factors that affect sleep is collected throughout the day using the subject's cell phone. Because the solution is contained entirely at the person's home, the steep cost and unfamiliar location of a sleep center can be bypassed. Furthermore, data can be collected over a longer period of a time, producing a more accurate diagnosis.

Our paper is laid out as follows: Section 2 describes the motivation for our choice of design; Section 3 details the characteristics of our approach; Section 4 explains our approach; Section 5 focuses on our implementation; Section 6 contains information collected from a user survey and evaluation of our proposal; Section 7 discusses in greater depth related research; and finally, Section 8 presents the conclusion or our plan and the direction of our future work.

2. MOTIVATION

Consider the following scenarios.

2.1. Scenario one

Arin is an older woman who has never suspected that she had a sleeping disorder. Because an overnight sleep center is unpractical and far too expensive, she is never diagnosed for a disorder even though she is frequently tired during the day and never feels fully rested.

2.2. Scenario two

Grahame is a middle aged man who suspects he may have a sleeping disorder. However, he is uncomfortable staying overnight in an unfamiliar place and does not have the time due to his busy work week, so he refuses to go to a sleeping center for a diagnosis. Years later, his wife convinces him to go, and he finds that he has sleep apnea, a dangerous disorder that could have been treated long ago and has resulted in many days and nights of needless symptoms.

2.3. Scenario three

Finally, consider Joye, a young woman who also suspects she may have a sleeping disorder. Because there are no other options, she goes to a sleep center for diagnosis three nights in a row. During the night, she is more restless than usual. The next day, she is asked a series of questions about her lifestyle, and the doctors conclude that she has a minor sleep disorder. In reality however, she has a more major sleep disorder, and her lifestyle is making it progressively worse. The short amount of time spent monitoring her sleep and the survey that reflected her opinions about herself more than how she actually acts misrepresented her situation and caused a misdiagnosis.

2.4. Interpretation

Our motivation for PATHOS stems from the previous scenarios. First, a sleep center is not the most practical way for a person to be diagnosed. They are costly due to sophisticated equipment and the trained professionals who must operate the equipment. Second, not everyone has time to stay overnight somewhere multiple times a week. Third, it may be very uncomfortable for someone to sleep in a foreign place, which may lead them to sleep differently than usual because of nerves. And, if the readings at a sleep center are sufficiently different than they would otherwise be at home, a misdiagnosis could occur.

3. CHARACTERISTICS

The characteristics that make our approach to polysomnography novel and unique are the following.

3.1. Cost effective

Polysomnography is typically a very costly process, especially when done at a sleep center. Even portable systems available today are not inexpensive enough for the average person to be compelled to verify whether or not she has a sleeping disorder, especially if it is a minor one. Our system uses inexpensive, multimodal sensors and takes advantage of the existing technology a user would already have in her home, such as computer, cell phone, or PDA. In addition, the system can handle multiple users, if, for instance, a couple wanted to monitor both of their sleeping habits but only wanted to have one unit. They could trade-off each night, indicating who is wearing the sensors, and send in separate data during the day.

3.2. Accurate

In order to be comparable to the diagnosis a sleep center would give, a portable solution must keep accuracy as an important goal. To balance cost and accuracy, the chosen sensors are in some cases multifunctional (the pulse oximeter) and in all cases deliver the best results for the price. Also, because our approach can be used as a long-term monitoring solution, extreme sensor readings will average out over the longer period of time, producing a more accurate assessment.

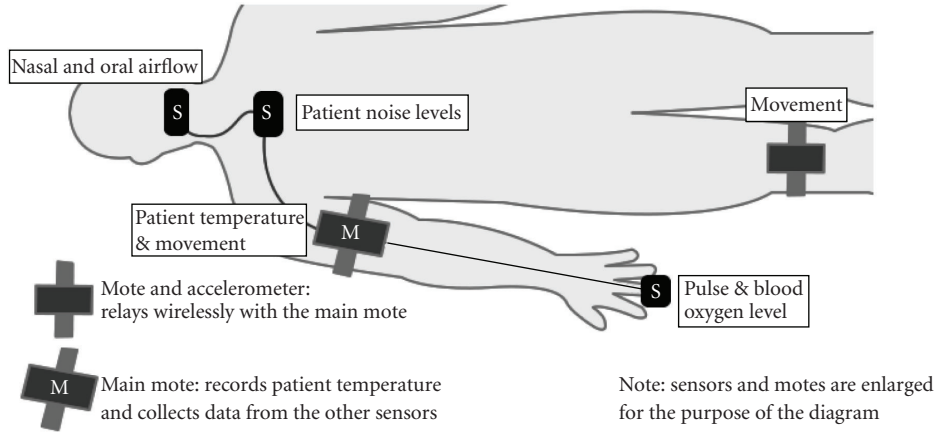


Figure 1: Diagram of the sensor network used for nighttime monitoring.

3.3. Reliable

Because we have designed our sleep monitor with long-term monitoring as an objective, the system itself must be reliable. It should be able to be used consistently for weeks or even months at a time without degrading in accuracy or requiring any maintenance on the part of the user. Although it is impossible to prepare for all situations that may occur, the system is easily customizable for any age of the user, allowing it to respond to unique needs more effectively than a "one size fits all" solution could.

3.4. Privacy aware

Because there is a system that monitors sleeping habits and collects lifestyle data, it is natural that users would want to control the use of their confidential information. Our solution allows the user to customize how their information is shared and who can have access to it. By default, only the doctor who will diagnose the user will have access to the information, and data on the system itself will be restricted to the specific user who collected it.

3.5. Simple yet meaningful GUI

One of the major target age groups of an easy-to-use, portable sleep monitoring system is the elderly. As time progresses, the older age groups will become more and more familiar with computers and mobile devices, but to accommodate everyone now and in the future, the user interface of our application is very simple. Text is easy to read, and most importantly a new user should have little problems utilizing the full potential of the system. For the completely technophobic, it can be nearly autonomous, requiring the user to simply wear the sensors and then visit the doctor when collection is complete. However, it will also cater to the other end of the spectrum by being very customizable and revealing more options to more experienced users.

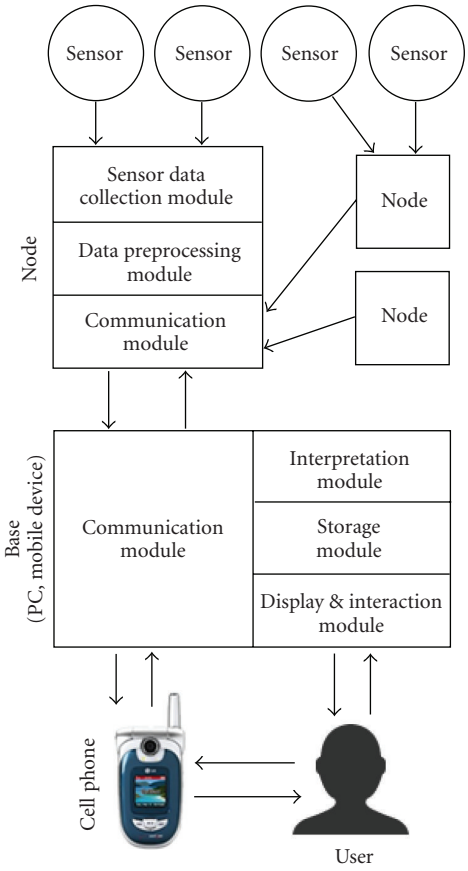


Figure 2: Diagram of the software architecture.

3.6. Easy and noninvasive to setup and use

A user will be expected to use the system every night in order to get an accurate assessment of her sleeping patterns. For this reason, the device must be easy to setup each night and must not interfere with a good night's rest. Pervasive technologies make this especially easy, but present some challenges as well. The communication between devices is wireless, so the user does not need to worry about many

wires being tangled up during sleep. However, the number of wireless motes needs to be balanced as well, because the more parts a system contains the less easy it is to setup and use. Because of the expandability of the system, it is easy to incorporate many wireless nodes or stick to fewer nodes with more sensors attached to each one. Figure 1 shows one such configuration.

In addition to the nighttime sensor network, the daytime data collection must be quick and simple as well. By allowing the user multiple ways to enter information, she can choose the most convenient way and is therefore more likely to report throughout the day.

3.7. Portable

Portability is the key aspect that draws users to our system. Besides the advantages of allowing a user to sleep in her own bed each night and not be bothered by traveling to and from a sleep center; the system is also readily portable to be taken to a different sleeping location. For instance, when the user goes on vacation or temporarily is staying at another place, the system can be brought along; data collection does not have to stop.

3.8. Customizable

For a system to be truly universal, it must be easy to customize for different users. In our system, this is accomplished in two ways. First, as explained in Section 3.6, our wireless sensor network allows nodes to be easily added and removed. Sensors that are not applicable to a certain users diagnosis can be left out, saving time and money. Further, if a doctor believes a patient needs more advanced monitoring, additional sensors can be added to the system to provide a more detailed report. Second, the main aspects of the system can be used in several different ways to achieve the same end result. See Section 6B for an example of multiple input methods.

Such a customizable system would lead to separate base configuration for different user groups, for example, divided by age.

On the one hand, an elderly person may be technophobic, and would want a system that requires very little interaction on his part. His system could include a passive infrared video monitor that could be setup once and then used every night. Also, instead of entering data throughout the day, he could enter a simple phone survey once each night. All the data collected would be automatically sent to the doctor and analyzed without any interaction.

A middle aged person, on the other hand, would perhaps want a great deal of control over his system. He would choose the most accurate sensors that attach directly to his body, such as the accelerometer and thermistors for airflow, even though for other people they may be more difficult to setup. Although he does not use texting for cell phone data entrance, he periodically takes a quick survey from his work or home computer to enter daytime information. Every morning when he wakes up he checks out his sleeping and daytime scores, and looks to see how he can improve his sleep by changing his habits.

There is also the case of a younger person, who does not mind entering her data throughout the day via simple text messaging. At night however, she prefers a similar setup to the elderly person: easy to setup and quick to activate. Also like the elderly person, she prefers to have her doctor look at the data and thus has it sent directly to him.

The preceding customizations are only a few ways by which the system could be configured. It is adaptable to any circumstances, and can be updated with new sensors just as easily as old ones can be removed.

4. OUR APPROACH

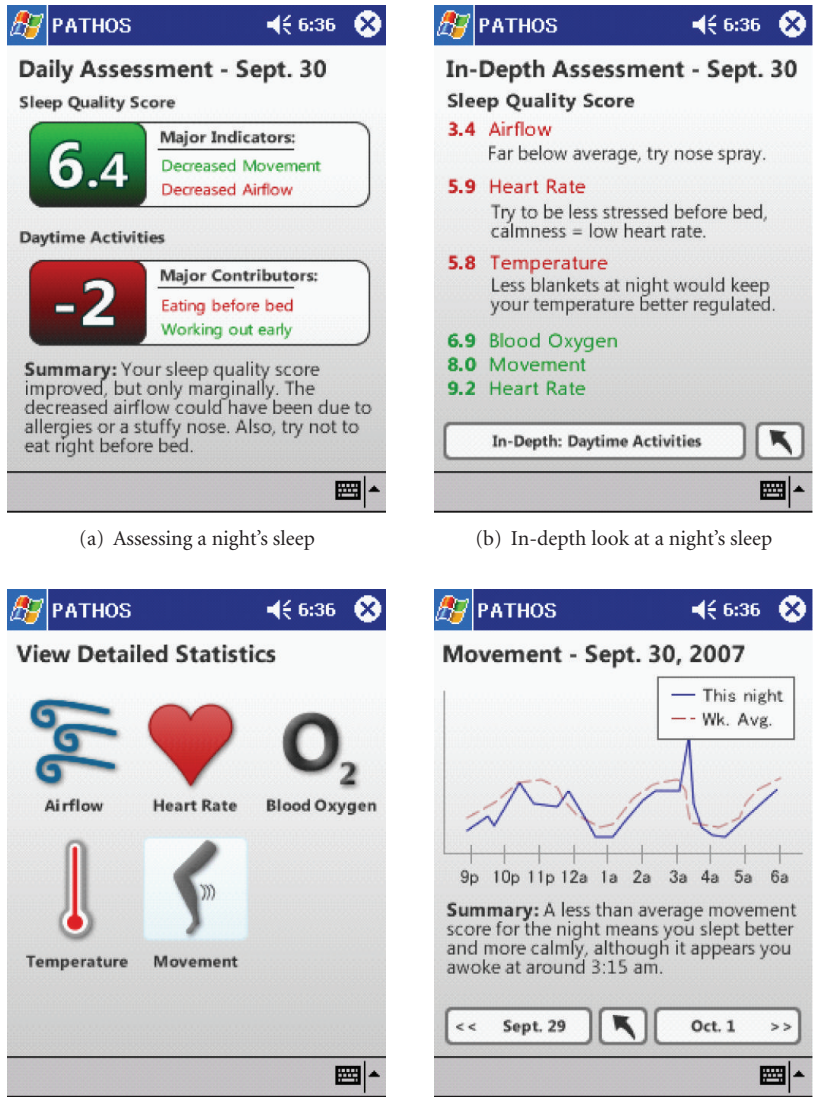
We have focused on making PATHOS low cost, suitable for home use, and most importantly a continuous, long-term system. Our approach can be split into three major sections: nighttime data collection, daytime data collection, and analysis.

4.1. Nighttime data collection

In order to make a sensor-based system that is user friendly and not too invasive during sleep, we will use a network of sensors and motes that communicate wirelessly using the Zigbee protocol. Figure 1 details a possible setup for the patient during nighttime monitoring. Three of the sensors are connected via wires to the main mote located on the person arm. The nasal and oral airflow sensors are thermistors that measure the patient's breathing, a characteristic especially important when diagnosing sleep apnea. There is also a microphone near the base of the subject's neck to record sound such as snoring. The pulse and blood oxygen sensor uses a simple finger clip-on, and is then connected to the main mote by a wire running up the arm. Because of their proximity to the main mote, it would have been more cumbersome for the sensors to communicate wirelessly via their own motes. The main mote itself contains an accelerometer to measure arm movement and a thermometer to measure the subject's temperature. The leg movement sensor (accelerometer) is connected to its own mote and transmits data to the main mote wirelessly. Additional sensors could be added wirelessly to the network very easily. For example, a simple passive infrared camera could be setup to record nighttime sleep movement patterns with more detail than the accelerometers could.

The setup that we have depicted is one of many. The system can be adapted to include more sensors in different ways, depending on the age group of the user, and what and how they want their sleep to be monitored.

Not pictured in the diagram is the receiving mote connected to the person's personal computer. If the pc is in the same room, the mote connected to it could be used to record the ambient pressure and humidity of the sleeping environment. If not, another mote could easily be added to the network and placed on a bedside desk. The main mote on the patients arm cannot be used to measure ambient



(a) Assessing a night's sleep

(b) In-depth look at a night's sleep

(c) Iconic, easy to use menu

(d) Detailed graph and summary

Figure 3: Some screenshots of our implementation.

temperature and humidity because it may be covered by blankets or moved frequently during sleep.

4.2. Daytime data collection

The major aspect that differentiates our approach from the others is that we seek to continuously monitor the person's lifestyle and habits that factor into the condition of her sleep. This allows us to make a more accurate assessment of factors that influence sleep and weigh the data collected during the night accordingly. Since many sleep assessments rely on survey data to collect habits and lifestyle information from patients, most of which are less than accurate, an easy to use system that gathers information as it happens could be a valuable resource for doctors diagnosing a sleep disorder. Some of the factors that will be monitored throughout the day are eating and drinking patterns (caffeine, alcohol, heavy/light meals, snacking, drinking before bed), exercise patterns (frequency, time of day, difficulty), and lifestyle (smoking, relaxing or stimulating activities before bedtime, sleeping schedule, amount of fatigue, and stress level). This data could be entered in a variety of ways, but must be convenient for the user or it will not get entered at all. Therefore, the system will rely on everyday technology that the user already has access to. For instance, she could enter information through a text message or voice automated prompt on a cell phone. Or she could go to a website and fill out a quick survey form. If none of the above is a viable option, the user could keep track in a paper log and enter the data manually later in the day.

4.3. Data analysis

All of the data collected throughout the day and night is relayed to a central location, which could be on the user's

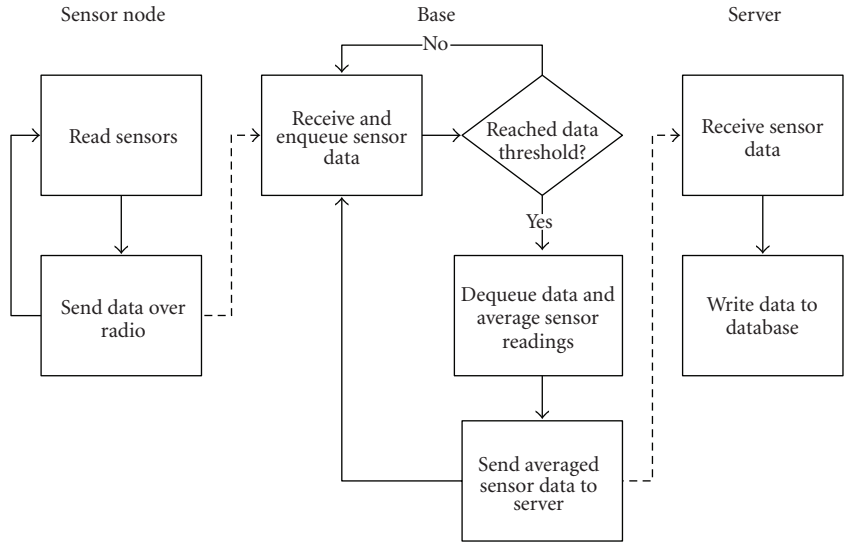


Figure 4: Flow diagram of the implementation.

pc or at a central database accessed through the internet. The central database would allow access to the patient and valid doctor's only to insure user privacy. Because data is stored on a computer or server, there is plenty of hard drive space for many weeks worth of data. Thus instead of data from only three nights at a sleep center, a user could get a diagnosis from her last month of sleeping habits. This long-term solution is also viable because the technology is noninvasive and would be practical for a patient to use every night for a month.

Additionally, the system itself can give recommendations and customized reports based on the data it collects. This is discussed in further detail in architecture—software design Section 5.2.

5. ARCHITECTURE

5.1. Hardware

The hardware we have chosen for PATHOS includes the following.

(i) Tmote Sky [5]: every sensor node consists of a Tmote Sky with either embedded sensors or sensors attached through analog output. The motes provide the Zigbee wireless connection and are battery powered, allowing complete freedom of placement.

(ii) BCI Micro Power Oximeter Board [6]: an extension for the Tmote Sky that allows a finger pulse oximeter to be connected directly to the mote. This sensor checks blood oxygen level and also records pulse.

(iii) EasySen SBT80—Multi Modality sensor board [7]: another extension board for the Tmote Sky, this board provides additional sensors to the mote, the most important of which are two dual-axis accelerometers which, when combined, can measure motion in three dimensions.

(iv) Handheld device or personal computer: forming the base of the network and is responsible for managing the collected data. The program will be implemented so that it is deployable on the user's existing hardware, cutting down costs and the amount of devices present.

(v) Cellular phone: used for collecting data during the day. The daytime collection software can be implemented as either a program that runs on the cell phone or a program running server-side that collects information through text messages or an automated phone system.

5.2. Software design

Our software design consists of two separate programs: one running on the nodes and one running on the base station. Figure 2 provides an overview of the programs and how they interact with each other and the user.

The program running on the motes is responsible for collecting data from the sensors and processing the data before sending it to the base. The sensors are either built into the nodes (temperature, humidity, movement), or are provided via a wired external sensor (pulse oximeter, thermistors, acoustic). The nodes run the TinyOS operating system and as a result are quite capable of preprocessing the data and transforming the analog voltage input to a format that can be interpreted easily by the base. As soon as the data is processed, it is sent wirelessly to the base. Nodes also have the ability to relay signals from another node that is too far away to connect to the base. Because the Tmote Sky's ZigBee wireless has a range of over 50 meters (indoor), a node will typically never be out of reach of the base, unless the wireless is scaled back in order to ensure a more secure connection.

The software running on the base is more important from the user's point of view because it is the only aspect of the system that he interacts with regularly. The most important part of the base software is the communication module, which is responsible for gathering the data from

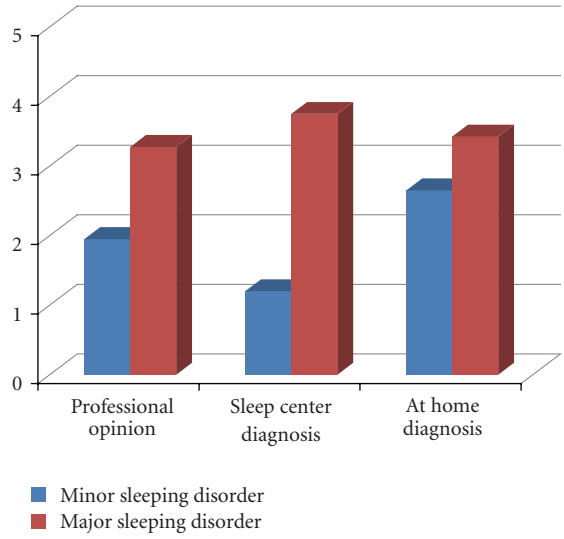


Figure 5: Graph of user response.

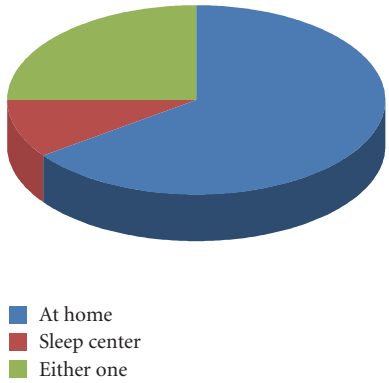


Figure 6: Chart of preference for home monitoring or sleep center.

multiple incoming motes, receiving data throughout the day from a cell phone or internet source, and sending collected data to the doctor for review. Because very few computers or handheld devices implement the ZigBee wireless protocol, a node must be plugged into the handheld device or computer via a USB port. That node then relays all of the information from the incoming nodes by UART to the base.

The interpretation module is responsible for identifying outlying data, for example if a sensor falls off during the night, or becomes disconnected the data will not be included in calculations, but it will still be stored for future reference. The interpretation module is also important because it computes and stores data results as it is collected. For instance, every day the module will calculate averages, maxes, and minimums, and store this data so that it does not have to be calculated again when the user wants to run longer, more detailed reports.

A separate storage module is necessary because it must decide where exactly to store the collected information. Because the system is collecting data all night for up to many months at a time, the amount of space needed can easily

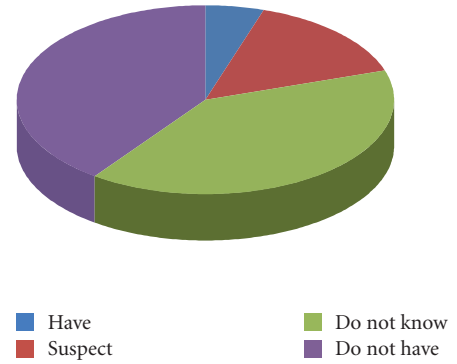


Figure 7: Chart of the participant's knowledge of their own conditions.

exceed the memory on a small handheld device. Normal computer should have little problem storing the data, however in both cases the storage module will utilize data compression techniques to remove statistical redundancy and efficiently store data in the smallest possible space.

The display and interaction module is visible in the screenshots in Figure 3. Every morning when a user awakes she has the option to look at a daily assessment of the last night's sleep. Figure 3(a) shows this function, where the program assigns a sleep quality score based on the collected data. Screenshot 3(a) also shows how the daytime activities affected sleep and a summary of the entire page. The results of the screenshot are simulated to develop such a score would require extensive processing of the data which we have not developed yet. However, it is possible for a doctor to receive a week's worth of data and then send the patient the information listed in screenshots 3(a) and 3(b). The doctor would make the observations using his own judgment, and when the data is sent to the base it informs the user that a new assessment has been delivered. This would effectively allow a doctor to make a remote diagnosis, and if the condition is severe enough the diagnosis could be done in person as well.

Screenshots 3(c) and 3(d) are fully implemented, as they rely on collected data. The summary section in 3(d) is implemented using a predefined set of threshold values that take into account the users long-term and short-term average scores. Screenshot 3(c) displays the program's simple, iconic, point-and-click interface (or one style tap, in this case).

Also, in screenshot A the large numbers allow a user to quickly identify how well they slept according to the sensors and what factors during the day influenced their sleep at night. The numbers themselves are clickable, leading to the in-depth assessment found in screenshot 3(b).

6. IMPLEMENTATION

We have partially implemented the aforementioned hardware and software architecture of PATHOS. The flow of data can be seen in Figure 4. The sensor node collects data and sends it over the radio; the base receives and processes the data for sending to the server, which in turn stores it in a database.

All of the sensors we have implemented are accessed via the Tmotes external ADC ports. They return a voltage that can be converted into the specified units for the type of data the sensor collects. This conversion is performed in the base station in order to keep the mote software simple, so that the formula's can be modified as necessary.

The base implements a queue for the sensor data it receives in order to process them efficiently and without worry of losing data. After the threshold limit for the queue has been reached, the base program dequeues the data, averages each reading, and sends it to the server. Since the sensor nodes send in data twice a second, an optimal threshold limit is 120 data packets, or one minute's worth of data. One minute is enough time to average out any sensor hiccups and also a small enough time span in which graphs and analysis can have fine grain data control. Larger limits can be used as well, and we will have to do more research to find the optimal limits for different sensor types.

The server side programming is currently implemented with a PHP file. The base posts the sensor data to the website, and the server then stores the data in a flat file xml system. When the site is accessed by the user in a browser, the sensor data collected over the last hour is displayed. The completed system will have authentication by the base and user account for each individual user. They will be able to login to the website and view statistics for the last day, month, or even year. The website will be the user's primary means for accessing analysis of the data and doctor's recommendations, and the data will also be sent to the base for easier display.

7. EVALUATION

To help us evaluate the idea of a home-based sleep monitoring system and continuous information gathering, we took a survey of 20 participants. Our survey confirmed the statement that a home-based system is a better alternative for many people than an overnight sleep center. Figure 5 details the results of our survey, where the subjects were asked if how likely they were to seek professional opinion, if they suspected they had a minor or major sleeping disorder, and then how likely they would go to a sleep center or stay at home for a diagnosis. The results show that the participants were 10% more likely to receive home monitoring than go to a sleep center if they suspected they had a major sleeping disorder. The difference increased dramatically when a minor sleeping disorder was concerned: home monitoring was favored 29% more than a sleep center. Our survey also showed that the participants were twice as likely to seek professional opinion if they suspected they had a major sleeping disorder rather than a minor sleeping disorder.

Finally, in a question where participants were asked outright which they would prefer, only 10% chose the sleeping center, versus 75% who would prefer to have the monitoring done at home. Figure 6 shows the distribution of participants who would prefer at home diagnosis or a sleep center. The majority of the participants would prefer to be monitored at home, which could be due to several reasons, including increased comfort and convenience. This survey data confirms that there is a large need for inexpensive, easy to use home sleep monitoring.

Also, over 70% of the participants said they would be willing to enter data throughout the course of the day if it were convenient and helped the diagnosis of a sleep disorder. Although this figure is favorable, it could be higher, which means a prototype system must be very convenient and easy to use. Another possibility is that the user is asked a short series of questions once or twice per day, which would cut down on the number of times a user has to remember to interact with the system.

Finally, the participants were asked whether they have, suspect they have, are unsure, or do indeed have a sleeping disorder, the results of which are shown in Figure 7. Even from this small sample of participants, it is clear that there is a need for a simple, at-home sleep monitoring solution because around 40% of participants were unsure whether or not they have a sleeping disorder. If disorders were easier to diagnose, that number would be much lower.

8. RELATED WORKS

Many projects are underway that focus on general health monitoring. A long term monitoring system known as Terva [8] has been implemented to collect critical health data such as blood pressure, temperature, sleeping conditions, and weight. The problem with Terva is that although it is self contained, it is housed in a casing about the size of a suitcase, which seriously dampers mobility. As a result, Terva is only practical inside the home. IST VIVAGO is a system used to remotely monitor activity and generate alarms based on received data [9]. In contrast with Terva, our system is small and completely wireless, allowing it to easily adapt to new situations.

Another system, wireless wellness monitor (WWM), is builtspecifically to manage obesity [10]. The system has measuring devices, mobile terminals (handheld devices), and a base station home server with a database. It uses Bluetooth and Jini network technology and everything is connected through the internet. The MobiHealth project [11] is similar to WWM as it monitors a person's health data using small medical sensors which transmit the data via a powerful and inexpensive wireless system. A combination of these sensors creates a body area network (BAN), and the project utilizes cell phone networks to transmit a signal on the fly from anywhere the network reaches.

Students at Duke University [5], as part of their DELTA Smart House design, described a system for monitoring sleeping patterns that is easy to use and inexpensive. In order to gather detailed sleep data, they used a pulse oximeter to record the user's heart rate and respiratory rate, a watch style actigraph to measure movement, in-bed thermistors for body temperature, and a microphone for audio. Their system achieves a low cost by using multifunctional sensors, but their choice of an actigraph adds a considerable amount to cost. Their approach depends on a computer for data interpretation, and the sensors themselves are not actually integrated. For instance, the watch actigraph must be

plugged into a computer to transfer data; collection is not seamless.

As part of the SENSATION Project [12] researchers have put together a system using the latest technology to detect sleep and sleepiness. They proposed using a ring that detects heart rate and wirelessly transmits the data, pressure sensitive film to measure chest and limb movement, a microcamera to make sure a driver's eyes are on the road, and (BAN) technology to have all the parts communicate wirelessly. As apparent by the choice of sensors, the consortium is more focused on preventing driver from falling asleep at the wheel than collecting data for diagnosis of sleeping disorders.

Taking a different, completely noninvasive approach to sleep monitoring, researchers at the University of Tokyo [13] have used the "surrounding sensor approach." Instead of placing sensors on the subject's body, they are using motion sensors, cameras, and microphones placed in the surrounding environment to provide noninvasive monitoring. The downside to this approach, although it is meant for home use, is that it is not very portable and therefore must be semipermanent.

Another approach to inexpensive sleep monitoring has been implemented by the University of Washington Seattle with the use of multimodal sensors. As opposed to the expensive actigraph, they investigated the possibility of using a passive infrared camera to record motion during sleep, a decision which carries the same consequences as the surrounding sensor approach, and may be more difficult to setup than sensors that simply attach to the body.

The last system we will review is the FPGA-based sleep apnea screening device for home monitoring developed by researchers at the University of Cairo. The purpose of their system is to determine whether or not a patient should undergo a full polysomnography exam, instead of being used in place of a sleep center. Also, differing from our system, the data is recorded on a Secure Digital card to be processed later by the doctor.

9. CONCLUSION AND FUTURE RESEARCH

In this paper, we have presented the details of PATHOS, a hardware- and software-based implementation for monitoring sleeping conditions and lifestyle habits related to sleeping conditions from the comfort of the user's home. It has been designed to break the field of polysomnography away from the sleep center and bring it into the patient's home by using wireless connectivity and existing hardware. By doing so, we hope that a more inviting system will lead to the diagnosis of more sleeping disorders and increase the comfort of many people.

In the future, we will be finishing the implementation in order to test the system with a real patient, collecting data and displaying output on a handheld device. We will also be running more extensive test of the graphical user interface. These analyses will allow us to assess the strengths and weaknesses of our design and modify it accordingly. Additionally, we would like to produce algorithms for calculating sleep quality scores and look at how perceived sleep quality matches up with actual sleep quality.

REFERENCES

[1] http://www.sleepfoundation.org/.

[2] X. Zhu, W. Chen, T. Nemoto, et al., "Real-time monitoring of respiration rhythm and pulse rate during sleep," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2553–2563, 2006.

[3] K. M. Al-Ashmouny, H. M. Hamed, and A. A. Morsy, "FPGA-based sleep apnea screening device for home monitoring," in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '06)*, pp. 5948–5951, New York, NY, USA, August 2006.

[4] Y.-T. Peng, C.-Y. Lin, and M.-T. Sun, "Multimodality sensors for sleep quality monitoring and logging," in *Proceedings of the 22nd International Conference on Data Engineering Workshop (ICDE '06)*, p. 108, Atlanta, Ga, USA, April 2006.

[5] S. Addison, A. Huang, T. Hui, J. Liu, and P. Nicholsen, "Sleep Monitoring," Duke University Smart House, May 2005.

[6] http://www.moteiv.com/products/tmotesky.php/.

[7] http://www.bivona.com/applications/oxiboards.htm.

[8] I. Korhonen, R. Lappalainen, T. Tuomisto, et al., "TERVA: wellness monitoring system," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 4, pp. 1988–1991, Hong Kong, October-November 1998.

[9] A. Sarela, I. Korhonen, J. Lötjönen, M. Sola, and M. Myllymäki, "IST Vivago—an intelligent social and remote wellness monitoring system for the elderly," in *Proceedings of the 4th Annual IEEE EMBS Special Topic Conference Information Technology Applications in Biomedicine (ITAB '03)*, pp. 362–365, Birmingham, UK, April 2003.

[10] J. Parkka, M. Van Gils, T. Tuomisto, R. Lappalainen, and I. Korhonen, "A wireless wellness monitor for personal weight management," in *Proceedings of IEEE EMBS International Conference on Information Technology Applications in Biomedicine*, pp. 83–88, Arlington, Va, USA, November 2000.

[11] A. van Halteren, R. Bults, K. Wac, et al., "Wireless body area networks for healthcare: the MobiHealth project," *Studies in Health Technology and Informatics*, vol. 108, pp. 181–193, 2004.

[12] T. Åkerstedt, T. Penzel, G. Ruffini, G. Dorffner, and A. Bekiaris, "New development of nano- and microsensors for monitoring of sleep and sleepiness," in *Proceedings of the International Conference on Fatigue Management in Transportation Operations*, Seattle, Wash, USA, September 2005.

[13] Y. Nishida, T. Mori, T. Sato, and S. Hirai, "The surrounding sensor approach—application to sleep apneasyndrome diagnosis based on image processing," in *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC '99)*, vol. 6, pp. 382–388, Tokyo, Japan, May 1999.

Hindawi Publishing Corporation
International Journal of Telemedicine and Applications
Volume 2008, Article ID 136054, 8 pages
doi:10.1155/2008/136054

Research Article

An Adaptive System for Home Monitoring Using a Multiagent Classification of Patterns

Ali Rammal,[1] Sylvie Trouilhet,[2] Nicolas Singer,[1] and Jean-Marie Pécatte[2]

[1] *IRIT-Centre Universitaire de information et d' Recherche (CUFR) Jean-François Champollion, site de Castres, Avenue Georges Pompidou, 81100 Castres, France*
[2] *IRIT, Université Paul Sabatier (UPS), site de Castres, Avenue Georges Pompidou, 81100 Castres, France*

Correspondence should be addressed to Nicolas Singer, Nicolas.Singer@univ-jfc.fr

Received 1 October 2007; Revised 11 February 2008; Accepted 18 March 2008

Recommended by Christine Verdier

This research takes place in the S(MA)^{2}D project which proposes software architecture to monitor elderly people in their own homes. We want to build patterns dynamically from data about activity, movements, and physiological information of the monitored people. To achieve that, we propose a multiagent method of classification: every agent has a simple know-how of classification. Data generated at this local level are communicated and adjusted between agents to obtain a set of patterns. The patterns are used at a personal level, for example to raise an alert, but also to evaluate global risks (epidemic, heat wave). These data are dynamic; the system has to maintain the built patterns and has to create new patterns. So, the system is adaptive and can be spread on a large scale.

1. INTRODUCTION

In Europe, many countries will be confronted with aging populations in the coming decades. For example, it is estimated that in 2020, 28% of the French population will be over 60 [1]. A great way to resolve partially this difficulty is to encourage old people to be cared for in their own homes. This strategy presents two main advantages:

(i) the elderly want to stay at home as long as possible; they keep the privacy they do not want to lose,
(ii) it is less expensive than a place in a collective accommodation.

Our project takes place in this context. It aims to help professional home-care teams in their job by thinking up innovative software technologies, more precisely:

(i) by increasing the number of old people looked after in their homes with an adaptive and nonintrusive remote assistance,
(ii) by reassuring family circle. The system ensures that the monitored person is secure; so, people around him feel at ease, and
(iii) by contributing towards its democratization. The use of simple elements (e.g., basic sensors) minimizes the initial cost of a monitoring system.

We made a study of systems having the same aim—the following section describes three well-known and relevant European systems in the home-care domain. These systems focus on individuals (they are user-centred): a system surveys only one person; thus, there is a duplication for each individual looked after. None of these systems collects individual monitoring for merging global behaviour patterns. Nevertheless, patterns of monitored people could be used to estimate the status of someone in relation to their community or to integrate new comings.

We propose a multiagent system that is able to generalize, which builds a classification of monitored people. An agent watches over one or more indicators of a group of people. An indicator is data about daily activities, positions, and physiological information. In a first step, the agent applies a local-classification method and obtains an incomplete patterns' partition. Next, the partial partitions are compared with each other in order to build a complete classification.

We conceived an open system: new people or/and new indicators bring in new agents or/and new patterns.

In Section 3, we present the architecture of the system and how it runs.

The system manages a set of patterns of monitored people. This dynamically updated classification has the following three main uses:

(i) to find certain similarities with the existing tools for evaluating the dependence—dependence grid of the social services, for example,
(ii) to get global statistical data about old people looked after in their own homes, and
(iii) to generate specialized alarms depending on the detected event. Once the classification is set up and people status is known, decisions can be taken to personalize the process of monitoring someone—activated sensors, generated alarms, and danger zone.

These aspects are discussed in the last section.

2. A SURVEY OF THREE HOME-MONITORING SYSTEMS

The use of computers to help people stay at home has been the subject of many research projects. Some of them are quite ambitious and regroup many partners. In this section, we describe a selection of three projects designed to assist people in their living environment. We expect to give the reader an overview of the advancement in this area and also the bases our project is laying on.

The selection shows different hardware and software problematics (communication networks, system interoperability, data analysis, emergency handling, and alerts filtering). These problems must be solved to achieve efficient monitoring. We begin by explaining the main objectives of each project. Then, we propose a table that summarizes their most relevant features.

2.1. The PROSAFE project

The PROSAFE project [2, 3] attempts to automatically identify the daily activities of the monitored person. The processing of collected data is carried out on doctor's request with an adapted interface.

The final operational objective is to detect any abnormal behaviour such as a fall, a runaway, or an accident. The research objective is to gather characteristic data about the nightly or daily activities of the patient. More precisely, the system can

(i) describe events that took place during monitoring time—time spent in bed or in the toilets, entering or leaving the bedroom, moving inside the home,
(ii) build a database with all abnormal situations detected, and
(iii) build statistics about past activities.

At the hardware level, the system configuration uses a ground network (a mobile version is also usable). Currently acquisition and data processing are local, and monitoring is both local and distant.

The PROSAFE system is primary used by the medical staff in hospitals. The interface for nurses allows them visualizing the patient state and abnormal situations (alerts and alarms) in the bedroom. As soon as an alarm is raised, a beeper calls a nurse. In the same time, doctors can access a database updated in real time with statistical data about the patient behaviour.

Experiments have been made to gather data about the daily activities of patients in hospitals, especially during the night. Experimental sites have been set up in two hospitals and three more are being installed in elderly people residences.

To conclude, let us say that one of the main features of this project is to be based on real-time analysis of data.

2.2. The AILISA project

The AILISA project [4, 5] (Intelligent Apartments for effective longevity) is an experimental platform to evaluate remote care and assistive technologies in gerontology. This ambitious project regroups specialists of smart homes, networks and computing, electronics, and signal processing. More precisely, the project sets up a monitoring platform composed of

(i) a home equipped with a set of sensors and health devices (presence detectors, wrist arterial pressure sensors, and pulse oximeter),
(ii) a smart shirt developed by the French company TAM with several sensors and electronics embedded in the textile to detect falls,
(iii) a smart assistant robot for ambulation to secure the displacements and assist the person during transfers, and
(iv) a software system to gather and analyze the sensors output.

The project aims to set up an interdisciplinary platform for the evaluation of the technologies at the three following levels: technical, medical, and ethical.

2.3. The e-Vital project

The e-Vital project [6] (cost-effective health services for interactive continuous monitoring of vital signs parameters) is a modular and ambulatory telemedicine platform. Its objective is to increase patient's feeling of safety concerning their health. Patients and caregivers feed a central database with some measuring equipment. The developed device allows staff to take measurements and data collected to be sent to the resident doctor. This doctor can remotely diagnose whether there is a problem that needs them to visit or that requires the resident to receive hospital treatment.

By way of a personal digital assistant (PDA), the e-Vital server connects monitoring devices produced by several manufacturers. The server is a multiagent system where each agent focuses on a specific task related to the medical stored data. For example, an alert manager is specialized

for the raising of alert messages, a profile manager for access management and a schedule manager for healthcare scheduling.

The e-Vital project is mainly hardware and tries to solve the interoperability problems between non compatible devices. It focuses on communication protocols and on the central database format.

These objectives (care protocol, devices interoperability) are different from ours but the approach is similar: e-Vital is an open system with several interconnected modules, one of which being a multiagent system. The difference resides on the application level: when our system is a group-centred system, e-Vital is a patient-centred system (it does not use the patient's record to develop generic profiles).

2.4. Results of the survey

We presented three systems which are able to monitor the "elderly" in their own homes. Table 1 summarizes some features that we found relevant to compare.

All three projects seek to gather information about people by the way of hardware and software solutions. They differ in the type of collected data, in the way they use it, and in their objectives. From the simple gathering of health information for caregivers, to the complete profiling of people, resources are quite different.

In all cases, the patient is an isolated person, installed in the centre of these systems; systems which have mainly a local vision of situations.

These works have inspired more recent projects; these projects are in progress so their results can not be analysed yet. This is the case of the GERHOME [7] project, led by two French research centers. This project intends to create a smart home for weak people. This objective can also be found in the European project SOPRANO [8]. Let us also talk of the European OLDES [9] project, which tackles the problem of the elderly people access to the new technologies. It tries to create low-cost hardware with very easy-to-use interfaces.

Our research is based on the progress and technologies developed in all these projects, especially those that gather information about the monitored person, whatever granularity this information can have. For example, the information can be the cardiac output, or something of higher level like behaviour information. This data is the raw material of our system and is used to generate several categories of people. Then these categories are used to make global assumptions about people belonging to the same class.

So our problematic is to collect the results of a large number of individual monitoring and to draw several categories. This classification provides several reusable classes of people.

For that, we deploy a classification framework, usable in a large-scale configuration, and based on multiagent technology. The next section describes this architecture.

3. S(MA)^{2}D SYSTEM

We propose a system able to carry out a generalization of profiles' patterns and to propose a classification of monitored people. S(MA)^{2}D (multiagent system for keeping elderly people in their own homes) is a multiagent framework in which agents use a restricted cooperation protocol to collectively perform classifications.

3.1. Multiagent and health

We chose a multiagent approach because these systems proved their adequacy in many health problems [10]. In this field, medical knowledge to solve a problem can be distributed in various places. For example, to establish the medical file of a patient, it is necessary to have analyzes and tests coming from several hospitals. Agents work in various places, each agent managing a part of the knowledge.

Multiagent architecture is particularly adequate if the problem-solving implies the coordination of various specialized people (e.g., units of a hospital must collaborate to establish patient scheduling). Then, the agents have cooperative skills to communicate and to build together a solution progressively.

Moreover, many medical problems are complex and often standard solutions are not easy to find. A multiagent problem-solving is based on decomposition in subproblems. Let us take for example organ transplants [11]: when a new organ is available, the more appropriate recipient must be found very quickly. It can be located in a very far medical center. Moreover, each hospital keeps the data of its patients; they are in the waiting list depending on the type of organ. It would be difficult to conceive and apply a complex centralized system to solve this coordination problem (e.g., a standard decision aid expert system).

Multiagent technology also proved its reliability in medical information retrieval. A great quantity of medical knowledge is available on the Internet, and it is necessary to access to the most suitable information. The agents can be employed to play the mediators between doctors and patients, or between medical resources. These agents seek information issued from various sources, analyze selected data, and choose useful information according to the profiles of the consultants.

To conclude, the agents' autonomy is an adequate paradigm to deploy systems, in which each component models the behaviour of an independent entity; this entity has its own knowledge, skills, and individual goals.

We recalled the general interest of multiagent systems in the health field. Now we are going to present the expected functionalities of our system.

3.2. Architecture and functioning

The system is based on a bunch of sensors carried by monitored people or installed in their homes. Those sensors are, for example, presence and movement sensors or medical measuring apparatus. The data coming from sensors are transformed into indicators. Some indicators can also come from human information: notes of a nurse or patient's answers to a questionnaire.

These indicators will be used by the system to generate its classification. Their abstraction from data requires asoftware

Table 1: Overview of the three projects.

Criteria	Project PROSAFE	AILISA	e-Vital
Smart home equipment	Yes Equipment is installed in hospitals and residences of elderly people	Yes Health smart homes	No
Body wear equipment	Yes Accelerometer (or GPS)	Yes Smart shirts with fall sensors	No
Medical equipment	Yes Digital entries acquisition module	Yes Wrist arterial pressure Pulse oximeter	Yes Appropriate monitoring devices
Detected emergencies and supervised risks	Accidents Falls Escapes	Some medical risks Falls	Scheduled care Vital signs defection
Target	Elderly or handicapped people Patient with Alzheimer disease	Elderly people Handicapped people	People with chronic diseases
Project range (home, living environment)	At home and in hospital	At home	Living environment (that is at home but also in mobile situations)
Risk-detection method	Sensors and statistical methods	Mainly hardware	Data management and interpretation made by a multiagent system
Scale (how many people are concerned)	The system focuses on one patient but the profiling can be used in a more large scale system	The System is providing an individual help	
Links to personal medical data	Yes	No	Patient's electronic health record are stored in a hospital database (but this database is only used by the project)
Medical validation	Tested in three hospitals (three other sites are planned)	Planned in three hospitals	Tests take place in four European pilot sites
Ethical and psychological aspects	Technical mediation between health caregivers and patients	Technical mediation between caregivers and patients Psychiatric aspects	Not mentioned
Operational or experimental	Experimental	Experimental	Experimental

layer. The set of sensors and this software layer are out of the scope of our work. It is the result of projects described in Section 2.

It is important to note that the functioning of the system is independent of the type and the number of indicators.

Indicators are collected by classification agents constituting the system. Because the system is strongly distributed, indicators of two people will not be inevitably collected by the same agent. There can also be some overlaps, if the same information is collected by several agents.

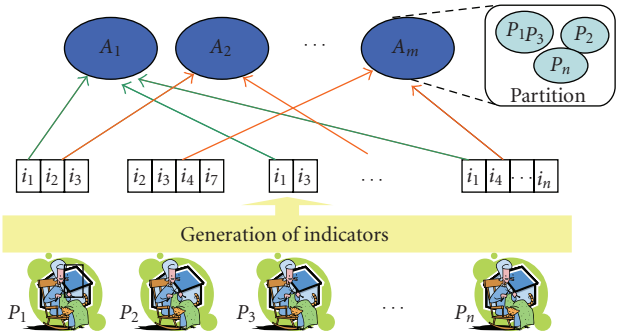


FIGURE 1: S(MA)^{2}D architecture.

Thus, classification agents A_j have indicators i_k concerning several individuals P_i (Figure 1).

With its indicators, each agent calculates a local, partial classification. This classification does not take into account all the indicators and is related to a reduced sample of the population.

Since the data inputs are numerical values, any statistical classification method is applicable.

To refine this classification, the agents communicate each other. They congregate in acquaintances network according to the similarity of the produced partitions. More precisely, each agent seeks the other agents which made a classification close to its own. To compute the classes of the collaboratively determined partition, we designed a restricted cooperation protocol in three steps: call for participation/acquaintance's group constitution/multiagent classification.

Section 3.3 gives a detailed example of this protocol.

There may be several groups of agents. They constitute parallel classifications: they are views of the same monitored people but according to various criteria (Figure 2).

3.3. Example

It is assumed that the behaviour indicators have numerical values. These values can be normalized by several methods as

(i) normalization between $[0 \cdots 1]$

$$\widetilde{I}_j = \frac{I_j - I_j^{\min}}{I_j^{\max} - I_j^{\min}}, \tag{1}$$

where $I_j^{\min}$ (resp., $I_j^{\max}$) is the minimum value (resp., maximum value) of indicator number j, and

(ii) linear normalization

$$\widetilde{I}_j = \frac{I_j - \overline{I}_j}{\sigma_j}, \tag{2}$$

where $\overline{I}_j$ is the average values of indicator number j for a given agent, and σ_j is the standard deviation of the indicator number j for a given agent

$$\sigma_j = \sqrt{V_j}, \quad V_j = \frac{1}{n}\sum_{k=1}^{n}\left(I_j^k - \overline{I}_j\right)^2, \tag{3}$$

where V_j is the variance of indicator number j for a given agent; n is the number of people monitored by an agent;

I_j^k is the indicator number j of the person number k.

Thereafter we apply our proposal on an example of 3 agents, 3 behaviour indicators, and 11 people. Suppose I_1 is the body temperature, I_2 is the number of getting up/sleeping in one night, and I_3 is the number of entries to the toilets each day.

The following table shows the distribution of people (P_i) and indicators (I_j) on the agents (A_k) of the system.

Table 2 shows that agent A_1 monitors 2 indicators I_1 and I_2 on people P_1, P_2, P_3, P_4, P_5, and P_{11}. Agent A_2 monitors 2 indicators I_2 and I_3 on people P_4, P_5, P_6, P_7, P_8, and P_{11}. A_3 monitors 2 indicators I_1 and I_3 on people P_3, P_6, P_9, P_{10}, and P_{11}.

This table also shows that people do not have the same indicators (it will often happen in real situations). For example, P_1 has only two indicators because for this person it is not necessary to test the number of entries to the toilets. The aim is that each person has the indicators suited to his case.

We assume that the sensors send data to the system on a daily basis. In reality there are indicators that are more important than others, for example, body temperature is more important than the outside temperature, so we give a weight for each indicator; this weight will help us later to form the groups of agents and to calculate the distance between classes. The most important indicator will be the one with the largest weight.

In our case we give to I_1 (body temperature) the weight 3, I_2 the weight 2, and I_3 the weight 1 (which is the default value).

By applying a local classification method (e.g., ISODATA [12]) each agent builds its partition. Each class is characterized by a midvector calculated by ISODATA.

Preliminary step: *construction of partitions* (Figure 3).

The first step is the *call for participation*. It aims to form groups of agents to generalize the classification. The agents of the system communicate with each other through the facilitator agent. The process to constitute agents groups for each agent A_i is as follows:

(i) A_i sends its indicators to other agents;

(ii) A_i receives the indicators from other agents;

(iii) for each other agent, A_i calculates the sum of the weights of common indicators (calling S_1), and the sum of the weights of noncommon indicators (calling S_2);

(iv) if $S_1 \succ S_2 A_i$ responds to the agent concerned;

(v) the agents of a group are agents who have exchanged messages between them.

In our case, A_1 sends I_1 and I_2. A_1 also receives from A_2 and A_3 their indicators. We find

$$\begin{aligned} A_1 \cap A_2 &= I_2, \\ A_1 \cap A_3 &= I_1, \\ A_2 \cap A_3 &= I_3. \end{aligned} \tag{4}$$

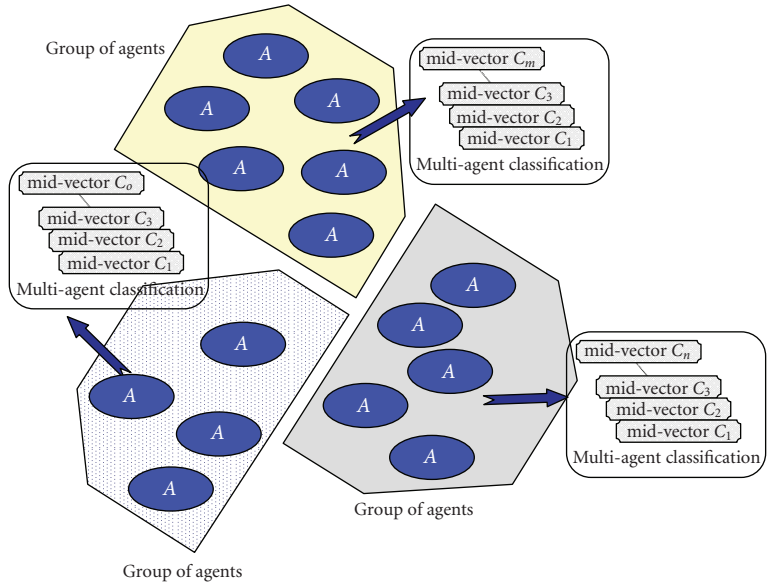


Figure 2: Multiagent classification.

Table 2: Distribution people/indicators.

People\Indicators	I_1	I_2	I_3
P_1	$\mathbf{A_1}$	$\mathbf{A_1}$	—
P_2	$\mathbf{A_1}$	$\mathbf{A_1}$	—
P_3	$\mathbf{A_1}$, $\underline{A_3}$	$\mathbf{A_1}$	$\underline{A_3}$
P_4	$\mathbf{A_1}$	$\mathbf{A_1}$, A_2	A_2
P_5	$\mathbf{A_1}$	$\mathbf{A_1}$, A_2	A_2
P_6	$\underline{A_3}$	A_2	A_2, $\underline{A_3}$
P_7	—	A_2	A_2
P_8	—	A_2	A_2
P_9	$\underline{A_3}$	—	$\underline{A_3}$
P_{10}	$\underline{A_3}$	—	$\underline{A_3}$
P_{11}	$\mathbf{A_1}$, $\underline{A_3}$	$\mathbf{A_1}$, A_2	A_2, $\underline{A_3}$

And as the weight of I_1 is greater than I_2 and I_3, A_1 chooses A_3 to form a group. The result is two groups of agents. The first group is formed by A_1 and A_3, and the second is formed by A_2.

This second step is *the acquaintance's group constitution*.

The third (and last) step is to *generalize the classification*. The agents of a group measure the distances between their classes using the weighted Euclidean distance:

$$d_w(c,c') = \left(\sum_{1 \le j \le n} w_j \cdot (d_j(c,c'))^2 \right)^{1/2}. \quad (5)$$

In which c and c' are two classes, w_j is the weight of the indicator number j, n is the number of common indicators between the two classes, and $d_j(c,c')$ is the distance between the two midvectors of the two classes according to the indicator number j.

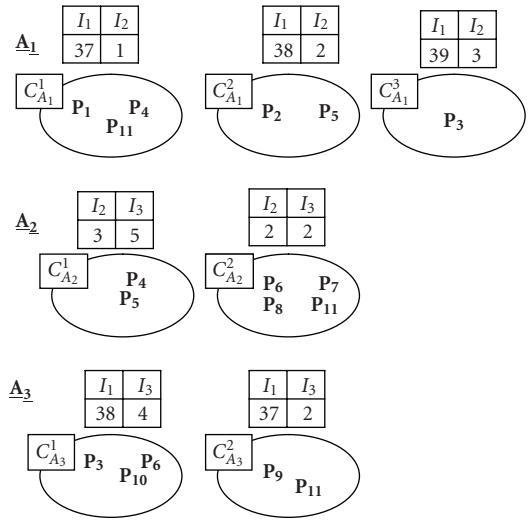


Figure 3: Local classification.

We can apply this formula on the actual values or normalized values of indicators. In this example we use the actual values. The agent A_1 seeks to each of its classes, the closest classes of its group among other agents.

Calculation of distances between classes:

$$\begin{aligned} &d_w\left(C^1_{A1}, C^1_{A3}\right) = \sqrt{3}; \qquad d_w\left(C^1_{A1}, C^2_{A3}\right) = 0; \\ &d_w\left(C^2_{A1}, C^1_{A3}\right) = 0; \qquad d_w\left(C^2_{A1}, C^2_{A3}\right) = \sqrt{3}; \\ &d_w\left(C^3_{A1}, C^1_{A3}\right) = \sqrt{3}; \qquad d_w\left(C^3_{A1}, C^2_{A3}\right) = 2\sqrt{3}. \end{aligned} \quad (6)$$

After the calculation of distances, we find that the class C^1_{A1} should be merged with C^2_{A3}, and that C^2_{A1} should be merged

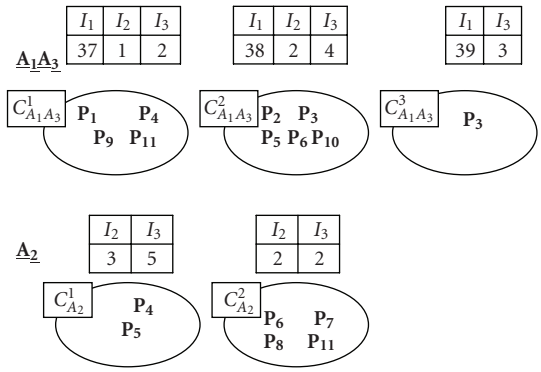


Figure 4: Result of the classification.

with C^1_{A3}. By contrast, C^3_{A1} should not be merged with C^1_{A3} because there is another class from A_1 nearest to C^1_{A3}.

The new classes thus obtained (Figure 4) have new midvectors. These vectors are the averages of the indicators values of people belonging to the same class.

A person may belong to several classes according to the indicators used. For example, P_3 is classified by A_1 and A_3 in a class by itself according to I_1 and I_2, and it is classified with P_2, P_5, P_6, and P_{10} according to I_1, I_2, and I_3.

As prospects, we intend to set a minimum threshold for the distance between classes. This threshold will be based on indicators and their weights. If the distance between two classes is greater than this threshold, they will not merge, even if they are close in the sense described above. It will be a more true-to-life approach.

3.4. Relevance of the multiagent architecture

This classification is actually multiagent because the classification result is not the work of a simple agent, as it is the case in other multiagent systems (choice of the most skilled [13]). It is really a collective work.

This multiagent classification answers to the problem of the search of patterns in an open and dynamic environment. Classical methods do not make it possible to increase the system scale: for example, when the number of entries changes (with the addition of a new indicator), all calculation and generation of classes must be made again.

Thus, our method satisfies the requirements of our application because it does not depend on the type of the indicators and does not require preliminary categories.

The management of the monitored people continues throughout the functioning of the system, as the agents collect more indicators values. Thus patterns evolve and the class of people can change.

Also an indicator can be deactivated: it corresponds to a data for which it is not essential to monitor this type of people.

4. APPLICATIONS RELATED TO HEALTH

Our system builds dynamic classifications of monitored people according to indicators that depend on the application.

This adaptability is the result of two essential characteristics. The first is the dynamic evolution of classifications—if needed, new data and new indicators can be added at any moment, and the system is able to reconfigure its classes and generate new classification patterns. The second is that the system is generic with respect to indicators and, thus, is able to function on any type of applications having strongly distributed entries.

Such a system is likely to bring solutions to several current problems in the home-monitoring field. Some of these problems are presented in this section.

Monitoring of dependent old people

Organizations of assistance to elderly people often use an evaluation grid of the dependence degree to determine the service needed by people. The result of this evaluation is also used to evaluate the cost of taking charge of someone.

The use of our classification system will make it possible to see whether there is an adequacy between the evaluation of monitored people by the grid and the produced profile classes. The matching of the two evaluation ways would validate our approach but also could consolidate the relevance of the grid criteria. In the contrary case, it will be necessary to re-examine the classification method and/or the selected indicators.

After validation, the system will be able to follow the evolution of the dependence degree of someone. Thus, somebody leaving his original pattern to enter a new one could be re-evaluated by the helper organization, and the assistance could be adapted to his new behaviours.

Detection of global medical problems

A metamonitoring will also make it possible to detect more global problems. The migration of a lot of people from a class toward another or the modification of certain characteristics of a class should indicate a collective event which affects several people; this can happen, for example, during a heat wave or an epidemic.

Remote monitoring of people suffering from chronic health problems

This help is for already detected people, suffering of cardiac and pulmonary insufficiencies, asthma, or Alzheimer disease.

The possibility of having a global vision of several monitored people can bring richer and more relevant information on the follow-up; the distribution in classes and the historic of the patterns evolution (system training) should allow new people entering the system to get a better service; in particular, more appropriate alerts according to the incurring risk should be generated.

Preventive control of high-risk people

In the long term, with the evolution of life ways, we can consider the monitoring of healthy people with personal or family antecedents relating to a disease or a medical event.

The system will make it possible to identify evolution diagrams of health parameters and life way (e.g., state-of-the-immune system, sleeping, nutrition, activity, etc.) who will indicate high risks to develop diseases.

5. CONCLUSION

We chose to tackle the home-monitoring issue in a more global way rather than in an only individual-centred way. This collective vision makes it possible to release individuals' patterns who will allow the system answering current health problems.

This large-scale and global solution (uninterrupted monitoring of hundreds of people) requires setting up a strongly distributed and dynamic system. Because classical classification methods are not adapted to this context, we had to propose a new distributed classification method.

The multiagent S(MA)^{2}D system implements this method. To evaluate its performances, we randomly generated a great number of numerical vectors of values and we observed the formation of classes.

Now, we have to define the real indicators to take into account. One of our professional partners *CVital* (platform of coordination of care and services to the person) is making a study about people whom this organism follows. This study will make it possible to define the number and the types of main indicators.

We will also request them to semantically interpret the classes.

REFERENCES

[1] K. McMorrow, *The Economic and Financial Market Consequences of Global Ageing*, Springer, Berlin, Germany, 2004.

[2] M. Chan, E. Campo, and D. Estève, "PROSAFE, a multisensory remote monitoring system for the elderly or the handicapped," in *Proceedings of the 1st International Conference on Smart Homes and Health Telematics (ICOST '03)*, pp. 89–95, Paris, France, September 2003.

[3] PROSAFE project, February 2008, http://www.laas.fr/laas/.

[4] N. Noury, "Ailisa: assessment units for medical monitoring and assistance technologies in gerontology," *Gérontologie et Société*, no. 113, pp. 97–119, 2005.

[5] N. Noury, C. Villemazet, P. Barralon, and P. Rumeau, "Ambient multi-perceptive system for residential health monitoring based on electronic mailings Experimentation within the AILISA project," in *Proceedings of the 8th IEEE International Conference on e-Health Networking, Applications and Services (HEALTHCOM '06)*, pp. 95–100, New Delhi, India, August 2006.

[6] E. Sakka, A. Prentza, I. E. Lamprinos, L. Leondaridis, and D. Koutsouris, "Integration of monitoring devices in the e-Vital service," in *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS '04)*, vol. 4, pp. 3097–3100, San Francisco, Calif, USA, September 2004.

[7] GERHOME project, February 2008, http://gerhome.cstb.fr/.

[8] SOPRANO project, February 2008, http://www.tunstall.co.uk/main.aspx?PageID=38.

[9] OLDES project, February 2008, www.oldes.eu.

[10] A. Moreno and J. Nealon, *Application of Software Agent Technology in the Health Care Domain*, Birkhauser, Basel, Switzerland, 2003.

[11] A. Moreno, A. Valls, and J. Bocio, "A multi-agent system to schedule organ transplant operation," *Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial*, no. 13, pp. 36–44, 2001.

[12] J. R. Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspective*, Prentice-Hall, Upper Saddle River, NJ, USA, 1996.

[13] S. Mukhopadhyay, S. Peng, R. Raje, M. Palakal, and J. Mostafa, "Multi-agent information classification using dynamic acquaintance lists," *Journal of the American Society for Information Science and Technology*, vol. 54, no. 10, pp. 966–975, 2003.

Hindawi Publishing Corporation
International Journal of Telemedicine and Applications
Volume 2008, Article ID 142803, 9 pages
doi:10.1155/2008/142803

Research Article

Location Estimation in a Smart Home: System Implementation and Evaluation Using Experimental Data

Youcef Rahal, Hélène Pigot, and Philippe Mabilleau

Laboratoire DOMUS, Université de Sherbrooke, 2500 boulevard de l'Université, Sherbrooke, QC, Canada J1K 2R1

Correspondence should be addressed to Youcef Rahal, youcef.rahal@usherbrooke.ca

Received 5 September 2007; Accepted 6 February 2008

Recommended by Frédérique Laforest

In the context of a constantly increasing aging population with cognitive deficiencies, insuring the autonomy of the elders at home becomes a priority. The DOMUS laboratory is addressing this issue by conceiving a smart home which can both assist people and preserve their quality of life. Obviously, the ability to monitor properly the occupant's activities and thus provide the pertinent assistance depends highly on location information inside the smart home. This paper proposes a solution to localize the occupant thanks to Bayesian filtering and a set of anonymous sensors disseminated throughout the house. The localization system is designed for a single person inside the house. It could however be used in conjunction with other localization systems in case more people are present. Our solution is functional in real conditions. We conceived an experiment to estimate precisely its accuracy and evaluate its robustness. The experiment consists of a scenario of daily routine meant to maximize the occupant's motion in meaningful activities. It was performed by 14 subjects, one subject at a time. The results are satisfactory: the system's accuracy exceeds 85% and is independent of the occupant's profile. The system works in real time and behaves well in presence of noise.

1. INTRODUCTION

The outburst of aging population in the recent and forthcoming years lays new challenges to provide assistance to the elders. Moreover, many elders may present degenerative diseases in their later years which can affect their cognitive abilities. For example, in Canada [1], people older than 65 will represent at least 25% of the population in 2030, and in 2021 the number of patients with Alzheimer and other related diseases will reach 592 000 persons (compared to 364 000 in 1992). Therefore, it becomes urgent to find a compromise between the need for constant care (at home or in institutions) and the need to lighten the load on caregivers. Also, it is highly important to ensure that the ministered care is personalized and efficient. In this regard, the Smart Assistive Home concept is an adequate solution. Indeed, it is now possible to provide a safe environment where the occupants (elders with cognitive deficiencies) can both be autonomous and feel familiar. The smart home can also be a good alternative to people who suffered cranial trauma. This population is generally young and benefits from less specialized institutions than the elders.

The DOMUS laboratory, presented in paper [2], plans to address the above issues by conceiving and testing a smart home. It is located at Université de Sherbrooke, Canada. The experimental apartment consists of a bedroom, a bathroom, a living room, a dining room, and a kitchen. The house is filled with different kinds of sensors in order to provide an accurate information about the occupant's location and activities. It can also interact with the occupant via different effectors such as touch screens, audio speakers, and controllable lights. Robust user-system interaction is ensured thanks to pervasive computing.

The occupant's location is probably one of the most important data needed to monitor the occupant's activities. Indeed, this information is useful when interacting with the occupant and in preventing dangers (by detecting falls, e.g.,). In the context of DOMUS, this information is vital in order to infer the activities already performed or those being processed, and to provide assistance where the occupant is. Our goal is to build a robust and accurate localization system using the available set of sensors already installed in the smart home. It covers the case when a single occupant is present at home, which is the most relevant case when assisting people

with cognitive deficiencies. We first analyze the current localization techniques regarding certain constraints. We then present the formalism of the method we selected and the experimental setup. Finally, we present the experiment we conducted, meant to thoroughly evaluate our system, and we discuss the results obtained. The current paper extends [3], focuses on evaluation, and thus completes the corresponding results and discussion sections. We do not discuss herein higher levels of assistance meant to detect falls, infer current activities, or interact with the occupant, and which can be implemented using the information provided by our localization system.

2. TECHNIQUES AND CONSTRAINTS

Only a few AI and robotics localization techniques can be applied in a small-scale environment [4]. Moreover, depending on the experimental constraints, the amount of available solutions is drastically reduced. These constraints are derived from two principles applied within the DOMUS technology. First, the occupant's privacy must be guaranteed. Second, the technology must be unobtrusive. This leads to the four following constraints.

(i) The use of video cameras is prohibited in order to protect the occupant's privacy.
(ii) The sensors should be dissimulated in the house to provide a familiar environment.
(iii) The use of tags worn by the occupant is avoided. This reduces anxiety and the occupant's feeling of being constantly monitored.
(iv) The most economic solutions are preferred.

The first constraint rejects video localization systems such as the one described in [5]. Considering the third constraint, solutions such as radio frequency identification (RFID) tags [6] or Wi-Fi engines [7] are also pushed aside. This is also the case with infrared (IR) or ultrasonic badges [8]. The cost constraint eliminates floor-based sensors, as described in [9]. In fact, this leaves us with inexpensive solutions which collect anonymous data. This includes devices such as IR detectors and a few other sensors which are already installed in the DOMUS home. Even in quantity, a system based on such sensors is quite affordable compared to the aforementioned solutions. We discuss the full list in Section 4.1.

Localization errors are induced when relying directly on the sensors, because the latter may sometimes send false information. This may happen because of an intrinsic error, which can be due to the sensor's error rate or to an occasional error in the experimental setup as a whole. External factors can also cause false sensor information. For instance, a draft can close a door and thus trigger a false event. So can pets sometimes. Therefore, the reliability of the localization system depends on our ability to analyze the sensor data. In this regard, recent researches show that sensor fusion is an efficient way to reinforce the validity of the location data. Whether in robotics [10] or in indoors localization [11], sensor fusion is achieved through probabilistic methods such as Bayesian filtering.

3. PARTICLE FILTERS

Bayes filters are efficiently used to estimate a person's location using a set of fixed sensors. In this method, the last known position and the last sensor event are both used to estimate a new location. The method represents an interesting compromise between accuracy and performance, and can be implemented in different ways. Fox et al. describe a few techniques to implement Bayesian filters and compare their performances [11]. Based on the results of that comparison, we decided to implement a localization system using the particle filters approach. Indeed, Kalman filters is an accurate method. However, it becomes inappropriate when different types of sensors are considered. This is the same for multihypothesis tracking. Also, grid-based approaches are robust but their poor efficiency excludes them. Finally, topological approaches do not meet the criteria, except for the robustness. On the other hand, particle filters are an efficient technique which is accurate, robust, and easy to implement. This technique is also adequate when different types of sensors are used, which is our case. We will briefly describe the technique. Refer to [11, 12] for further reading.

At any moment, the estimate of the occupant's location can be modeled as the belief in the fact that he/she is located at position x_t at instant t, given a series of previous sensor observations $o_0, \ldots, o_t$ from instant 0 to instant t : $\mathrm{Bel}(x_t) = p(x_t \mid o_t, o_{t-1}, \ldots, o_1, o_0)$. Given the Markov postulate which stipulates that only the last observation o_{t-1} is relevant, the expression of $\mathrm{Bel}(x_t)$ is simplified using Bayes conditional probabilities formula. In the discrete case, it becomes

$$\mathrm{Bel}(x_t) = \alpha p(o_t \mid x_t) \sum_{x_{t-1}} p(x_t \mid x_{t-1}) \mathrm{Bel}(x_{t-1}), \quad (1)$$

where α is a normalization constant. $p(o_t \mid x_t)$ represents the probability of observing an event o_t given that the occupant is at position x_t at instant t. It is the perceptual model. In practice, each sensor is associated to a space-dependent probability density function that represents the likelihood of receiving an event from that sensor depending on the occupant location. On the other hand, $p(x_t \mid x_{t-1})$ represents the probability that the occupant moves from x_{t-1} to x_t between instants $t-1$ and t. It is the dynamic model. In practice, it represents the motion profile of the occupant. For instance, a fast-moving occupant has a wider dynamic probability density than a slower one. This function should also include the environmental layout. In a more rigorous approach, it should even depend on time to reflect the changes in activities during the day or the week.

The algorithm based on particle filters estimates the location thanks to a set of n positions (particles). At first, these particles are drawn randomly and uniformly on the available space. An equal weight is devoted to each particle ($1/n$ for the sake of normalization). These particles model $\mathrm{Bel}(x_0)$ which, in this case, is uniform. When an event occurs (at $t = 1$), $\mathrm{Bel}(x_1)$ is computed according to (1). This operation changes the weights of the sample. The distribution obtained is then used to draw a new sample with equal weights. The new sample is from now on more centered on the latest event position. This operation is repeated every

Table 1: List of sensors per room.

	Entrance hall	Living room	Dining room	Kitchen	Bathroom	Bedroom	Total
IR	1	1	2	2	3	1	10
Tactile carpets	1	3	2	6	3	3	18
Light switches	1	1	1	3	1	1	8
Door contacts	3	0	0	30	5	10	48
Pressure detectors	0	0	0	0	0	1	1

Table 2: Density functions per sensor type.

Sensor	Density function type	Typical density function range (m^2)
IR	Square	20
Tactile carpets	Square	0.8
Light switches	Circular	1
Door contacts	Circular	1.5
Pressure detectors	Square	2

time a new event is observed. At any time, the estimate (belief) of the occupant's position in a place is simply the addition of the weights of all the particles confined in that location. Consequently, we can observe in practice a cloud of particles "following" the occupant.

Before using this technique for location estimation in the smart home, we need to make an inventory of all the sensors we use and then attribute probability density functions to them. We also need to model the occupant's motion. The next section deals with these aspects.

4. EXPERIMENTAL CONFIGURATION

4.1. The sensors

The list of sensors we consider and which are already installed and plugged in the DOMUS apartment include the following.

(i) IR presence detectors.
(ii) Tactile carpets (as seen later in Section 6.1, for the purpose of this study, these sensors are used only as location reference and not to infer the occupant's position).
(iii) Smart light switches. An event is received every time the occupant turns the lights on or off.
(iv) Electric contacts on doors (including closets, drawers, pantries, etc.). An event is received every time a door is opened or closed.
(v) Pressure detectors. These can be placed under the mattress, for example, in order to detect if the occupant is lying on the bed.

The number of installed sensors varies depending on the room and the areas of interest (see Table 1). For example, in the bedroom there is only one IR detector that covers the entire room area whereas two are installed in the kitchen: one covers the kitchen globally while the other is directed at the stove only. It is worth noting that none of these sensors would be able to give identification about the person who triggered it—as opposed to devices such as video cameras. They are also quite unobtrusive. That is clearly an advantage of our approach since it provides the required privacy imposed by our constraints.

4.2. Probability densities

We assign a probability density function to each sensor. This function depends on the sensor type and, mainly, on the range the sensor covers. For instance, IR detectors have a wide probability density, taking into account the fact that they cover large portions of a room (of course, this depends on how they are installed; they can also be installed to cover small regions). Moreover, they can be triggered more easily than the other sensors. Light switch sensors are generally more reliable and are triggered in a limited area. So are the door contacts, the tactile carpets, and the pressure detectors. That is why we assigned to them more compact probability densities than to IR detectors. We list the attributes of these functions in Table 2. These functions are either 2D square functions defined as

$$p(x,y) = \begin{cases} \dfrac{1}{(x_{\max} - x_{\min})(y_{\max} - y_{\min})}, & x_{\min} < x < x_{\max}, \\ & y_{\min} < y < y_{\max}, \\ 0, & \text{elsewhere}, \end{cases} \tag{2}$$

or 2D circular functions as

$$p(x,y) = \begin{cases} \dfrac{1}{\pi R^2}, & (x - x_0)^2 + (y - y_0)^2 < R^2, \\ 0, & \text{elsewhere}. \end{cases} \tag{3}$$

We do not add noise to these functions per se. However, we make sure that, at every step, about 10% of the particles are drawn randomly from all the available space. We noticed that adding noise to density functions could result in a static particle cloud after successive sensor events in the same place. In other words, the cloud can converge in a relatively small area and remains there regardless of new sensor events. By adding the uniform noise, we ensure that there are enough particles in the environment to avoid the cloud's immobilization. As for the dynamics, the occupant's motion is modeled by a 2D normal density which standard deviation encompasses half of the apartment's area. Although this function is quite rough, it models remarkably well the fact that, between instants $t - 1$ and t, the likelihood that

(a) The system at start

(b) The occupant moves in the living room

(c) The occupant enters the kitchen

(d) The occupant opens the fridge door

Figure 1: The localization system in a real condition situation. The points show the particle cloud. The shaded areas indicate where the occupant is likely to be present.

the occupant covers a distance d quickly decreases when d increases.

5. IMPLEMENTATION

Our system is implemented in Java and connects as a client to a server which aggregates sensors and forwards events as soon as they occur. We also implemented a GUI that displays the client's state and the current prediction of the system. Figure 1 shows typical use of our system in real conditions. These are based on the observation of its behavior during the execution of a short scenario, similar to the one we describe in Section 6.1. We perform the scenario and we observe the system's predictions, knowing the occupant's location. In Figure 1(a), we start the system. The occupant is already in the living room. However, since no sensor event is detected yet, the occupant is likely to be anywhere in the house. Consequently, the particle cloud is uniformly spread. In Figure 1(b), the occupant moves towards the kitchen and the living room IR sensor is activated. The system consequently computes that the occupant is in the living room with a probability of 91%. The particle cloud is contained in the corresponding area. In Figure 1(c), the occupant enters the kitchen and the IR sensor there is activated. The system concludes that the occupant is in the kitchen at 91%. Finally, the occupant opens the fridge door and a door-contact event is received. The cloud is quickly centered on the corresponding area. Here also, the probability that the occupant is in the kitchen is high: 81%. The probability is less than in the previous case because the probability density function (the perceptual model) associated to the fridge's door-contact sensor encompasses a small zone outside the kitchen. The particle cloud is very dynamic and smoothly follows the occupant's path. The computation time (the one needed to infer the occupant's location when receiving a new event) is inferior to 1 second and thus the system is responsive in real time.

6. SYSTEM EVALUATION

While the preliminary results are encouraging, it is important to thoroughly evaluate the system's accuracy and robustness. It becomes then necessary to monitor how it behaves in real-life situations. Therefore, we conceived an experiment to collect data with people moving inside the apartment, one person at a time. The subjects have no prior familiarity with the apartment.

6.1. *The experiment scenario*

Each subject performs a scenario of about 50 minutes long. It consists of the contracted routine of getting back to home

in the evening and leaving in the following morning. This scenario maximizes the occupant's motion in every room of the apartment, all the while allowing to perform meaningful activities. It also maximizes the number of activated sensors during tasks execution. This ensures that our system is able to locate the person with accuracy in real conditions. The broad lines of the scenario are

(1) entering the house;
(2) washing hands in the bathroom;
(3) preparing a sandwich in the kitchen;
(4) eating the sandwich in the dining room;
(5) preparing coffee in the kitchen;
(6) reading a magazine in the living room while drinking coffee;
(7) going to the bathroom;
(8) lying in the bedroom (the subject is allowed to read while being on the bed);
(9) getting up and making toilette in the bathroom;
(10) leaving the house.

To avoid cognitive load, we ask the subjects to perform this scenario by periods of about 10 minutes each. For example, the three first steps of the scenario fit into such a period. At the end of each period, we stop the data collection and explain the next set of tasks. We validated the entire scenario during a preexperimentation with 3 team members. This helped us adjust the steps and make the scenario more fluent.

The accuracy of the location estimation is checked thanks to a video camera and the tactile carpets. The camera is located in the kitchen and is used to record the subject's activities during the preparation of the sandwich. Five tactile carpets are installed on the kitchen's floor. The analysis of the video validates the tactile carpets accuracy (since we use the tactile carpets in the framework of a complex acquisition system, the accuracy we measure is not rigorously equal to the one we would obtain if the carpets were used separately from this system). Knowing this accuracy, we use the tactile carpets in every room of the apartment as reliable position indicators. We compare this reference position information with the output of our system. This gives the system's accuracy, that is, the accuracy of the location information resulting of Bayesian filtering using the rest of the sensors.

6.2. *The sample*

The sample of the 14 subjects who participated in the experiment is composed of 10 females and 4 males. Their age distribution is shown on Figure 2. The age is well distributed between 22 and 73, with a mean value of 50 years old. There is however a slight majority of people older than 50. To reflect the population targeted by the research at DOMUS, we recruited at least half of the subjects from that age group.

7. RESULTS AND DISCUSSION

First of all, we observe the distribution of the system's belief. As shown in Figure 3, the mean belief is 88% with a standard deviation of 8% for all the data from our sample. There is an

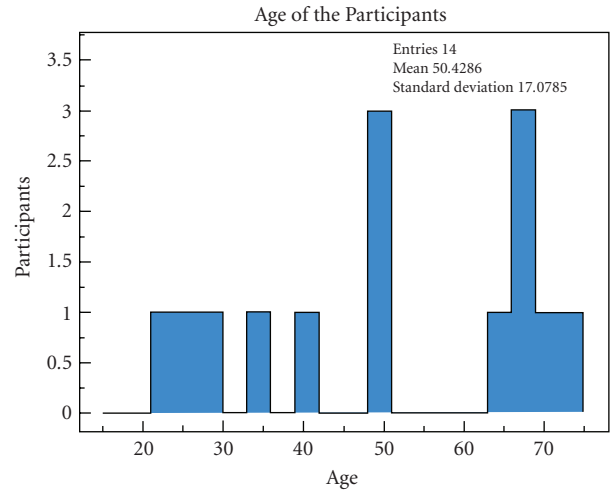


Figure 2: Distribution of the age of the subjects.

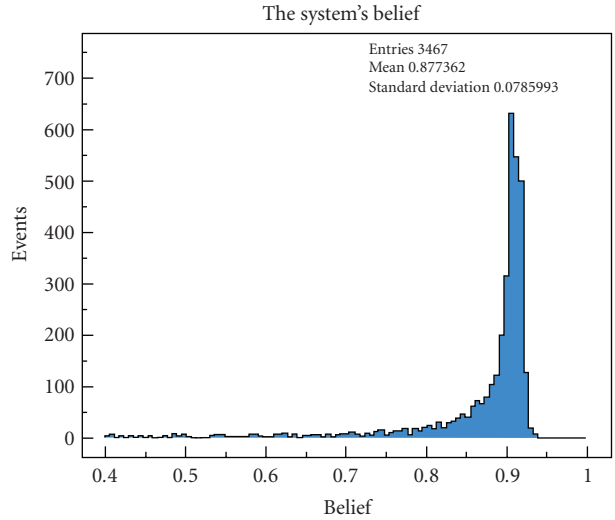


Figure 3: Distribution of the belief of our system, regrouping the data from the whole sample.

upper limit to the system's belief which is close to 90%. It is due to the 10% of noise particles that we draw randomly at every iteration. These are uniformly distributed in the smart home, and therefore the concentration of particles in a room cannot significantly exceed 90%.

The events with a low belief usually occur when the occupant goes from a room to another. With the following events, the belief gets significantly higher, showing that the particle cloud replaces itself quickly and correctly. Moreover, the global system's accuracy is 85%. This is the percentage of events where the system predicts accurately the position of the occupant, independently from the value of the system's belief. There is however an expected correlation between the system's accuracy and its belief. Events with a high belief are less prone to be false, and vice versa. We can therefore increase the system's accuracy by rejecting the events with the smallest beliefs. Figure 4 shows the variation of accuracy

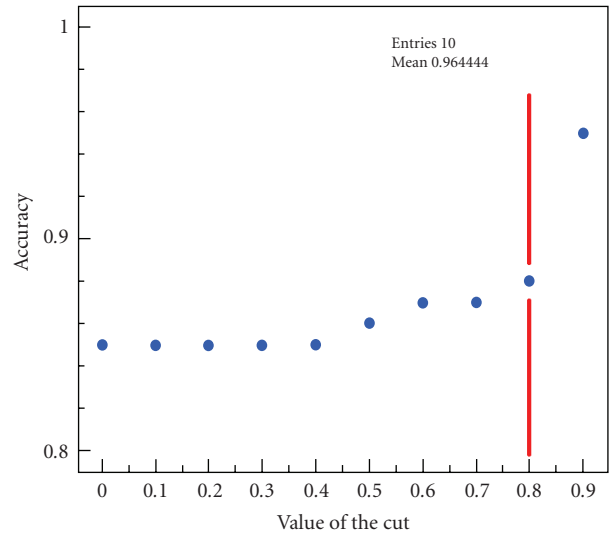


Figure 4: Variation of the accuracy depending on the value of the rejection cut on the system's belief.

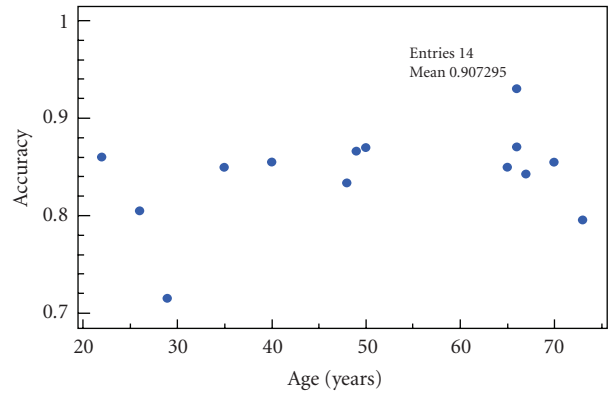


Figure 5: Variation of accuracy with the subjects' age.

Table 3: Accuracy and mean belief using different sensors.

Sensor set	Accuracy (%)	Mean belief (%)
All	87	88
IR	88	89
Light switches	50	75
Light switches and door contacts	77	80

with the value of the rejection cut on the system's belief. By rejecting events with belief less than 80%, the accuracy of the system becomes 88%. This cut rejects only 6% of true events and up to a third of false ones. The cut can of course be more drastic: by rejecting events with a belief less than 90%, accuracy can be increased to 95%. However, the cut at 80% is more reasonable, since this value is the mean belief for false events. In working conditions, one can vary the cut depending on how critical the location information has to be, that is, depending on the context and on the application that needs the information.

The most important limitation is that, as expected, our system gives incoherent results in case more than one person are present in the house. Indeed, when two people (or more) activate sensors simultaneously, the particle cloud tends to alternate between their respective locations. An upper-level solution is then necessary in order to identify and handle this trend. This limitation is a direct consequence of the fact that the information we collect is anonymous. It is because we avoid using devices that compromise the occupant's familiarity with the environment or make her/him feel monitored by wearing an RFID tag, for example (such systems also have a drawback if the occupant decides to withdraw the tag.). In this regard, our system is destined to be used to locate a single person in the house. When visitors are present, our system could be used in conjunction with another system: for example, requiring that visitors wear an RFID tag in order to differentiate them from the occupant. In this case, locating the occupant may also be regarded as less important than when the occupant is alone at home (DOMUS is designed for only one person with cognitive deficiencies per smart home. Visitors are mainly caregivers or family members).

In order to further evaluate our system, we study how it behaves regarding other experimental aspects, such as the analysis of possible correlations with the occupant's profile. We also investigate the best sensor configuration and their comparison per activity, the occupant's dynamics, the behavior in presence of noise, and finally the performance of the system.

7.1. Correlation with the occupant's profile

One of the main objectives of our study is to evaluate whether or not the system should be personalized according to the occupant's profile. The only profile-dependent parameter that influences the Bayesian filtering formula is the dynamics. Since the subjects were healthy, without notable variations, the most accurate variable that relates to the dynamics is the age. Therefore, the hypothesis is that if the system depends on the occupant's age or on their dynamics, consequently on their profile, the system would require adjustments upon deployment to fit them. Figure 5 shows that there is no significant correlation between accuracy and age (correlation coefficient $\rho = 0.36$) and that accuracy remains stable from a subject to another. This result leads to two interesting conclusions. First, the data collected shows sufficient consistency to allow us to limit the number of subjects to 14. In fact, increasing this number will not give more confidence in the results. Second, it gives us the ability to deploy this system immediately in new homes, since it is profile-independent. This is good news regarding both economic and time constraints (obviously, a configuration phase is required for all deployments, and would be of approximately equal lengths given an apartment size and a required number of sensors).

Table 4: Accuracy and mean belief per activity, using different sensors.

Activity	Sensor set	Accuracy (%)	Mean belief (%)
Walking	All	88	88
	IR	87	89
	Light switches	33	75
	Light switches and door contacts	38	72
Preparing sandwich	All	90	89
	IR	91	90
	Light switches	10	60
	Light switches and door contacts	85	87

7.2. Sensor configuration

In the purpose of addressing economic concerns, we study which sensor configurations are optimal. This information is important if one has to limit the redundancy, even though the latter could be useful to provide a robust environment. We therefore measure the system's accuracy using different sets of sensors. The complete information is in Table 3. In order to achieve this comparison, we use data from only one representative subject, whom we choose based on the mean age (50 ± 17 years old) and the duration of the experiment (48 ± 8 minutes). The values of the accuracy and mean belief using all the sensors are then slightly different than the mean ones presented in the previous section. The infrared sensors alone give comparable results with those obtained using all the sensors. That is due to the fact that their number (there is at least one such sensor in each room) and their spatial configuration (each one covers roughly the area of a room) dominate the global sensor configuration. However, it is worth noting that infrared sensors are not the most reliable type, for they are more prone to false activations, due to animal presence or a heat disturbance, for example. Consequently, deploying the system in a real setting would probably require at least one other type of sensors to ensure reliability.

7.3. Sensors per activity

The sensors are compared based on an activity criterion. We select two activities: walking and preparing a sandwich. Other activities—such as dining, reading and sleeping—do not generate enough sensor data with the present scenario to be considered for this comparison. Table 4 shows the results we obtain using different sets of sensors. For the walking activity, the IR sensors are the most accurate, because the occupant seldom activates other sensors while moving around the house. However, when preparing a sandwich, the occupant often opens the fridge and various drawers in the kitchen. Therefore, the door contact sensors are almost as accurate as the IR ones. This becomes useful for localization in case the activity being performed is known. In the previous section, we saw that IR sensors are at the core of this localization system, although they can present reliability issues. In case of preparing dinner, door contacts can be as accurate as and even more reliable than IR sensors, consequently more appropriate to locate the occupant during that activity.

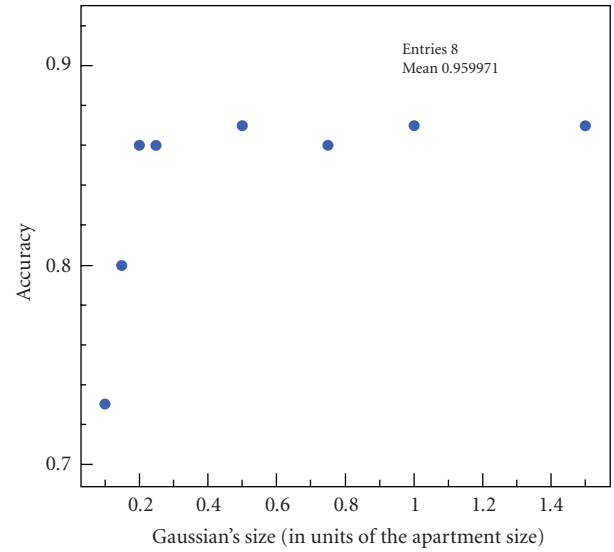


Figure 6: Variation of the system's accuracy with the occupant's dynamics, modeled by the Gaussian function's σ.

Table 5: Variation of accuracy and mean belief with noise.

Noise (%)	Accuracy (%)	Mean belief(%)
0	88	88
1	88	88
2.5	84	85
5	84	85

7.4. The dynamic model

In order to test the effects of the dynamics on the model, we analyze the system's accuracy while modifying the Gaussian dynamic function. The system's accuracy remains stable on a long range of the Gaussian's size (Figure 6). However, when the standard deviation of the Gaussian becomes too small, the particle cloud inertia increases and the model fails to reflect a natural motion. On the contrary, larger functions seem to behave very well. But this changes when coping with noise.

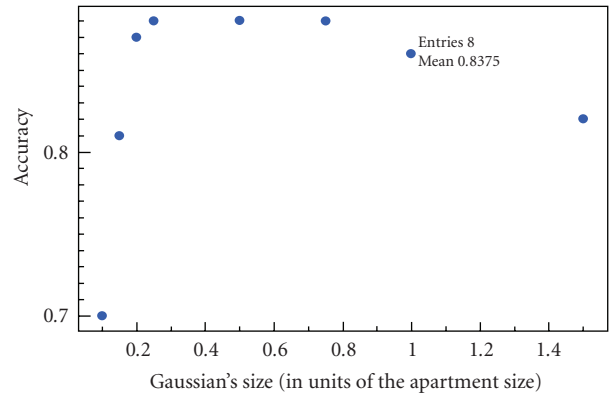


Figure 7: Variation of the system's accuracy with the dynamics, in presence of 2.5% noise.

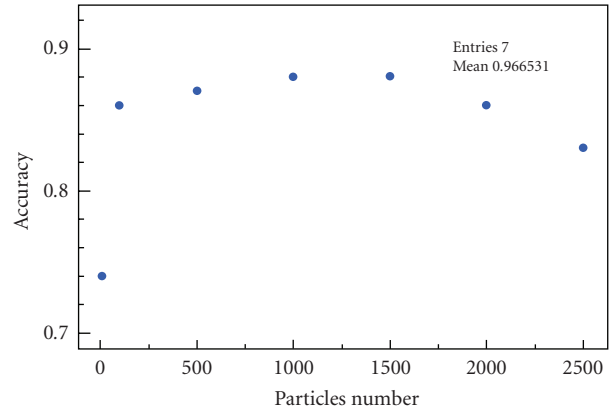


Figure 8: Variation of accuracy depending on the number of particles used in the algorithm.

7.5. Behavior in presence of noise

The experimentation took place in a laboratory. Therefore, even if the setting is similar to that of an apartment, disturbances are controlled. The subject is always alone and performing predefined activities. It is consequently important to analyze the behavior of the localization system in case of noise. First of all, since no significant noise was present in the data we collected, we generate random noise in order to complete our tests. Table 5 shows how the accuracy and mean belief are affected when contaminating the data with random sensor noise. The system is exceptionally stable and remains accurate at 84% even in the presence of 2.5% noise. Increasing the percentage of noise to 5% does not affect the localization accuracy. It is worth noting that since there was no observable noise in the data, 1% of noise is already a conservative value.

In a second step, we analyze the impact of the noise when changing the dynamic model. We thus introduce 2.5% noise while modifying the dynamics (Figure 7). Small Gaussians still fail to reproduce the occupant's dynamics. However, in the presence of noise, larger Gaussians become problematic too. The larger Gaussians model faster motion, therefore failing to reject false events even if they are significanty distant than the actual occupant's location. The Gaussian's size has then to be comparable to that of the apartment to ensure the best system's accuracy.

7.6. The ideal number of particles

Finally, for the sake of optimization, we want to infer the best parameters' settings for the algorithm. Therefore, we study how the number of particles used may affect the prediction accuracy (Figure 8). As expected, the more particles are used, the best the prediction is. However, an increase in the number of particles leads to an increase in computing time. Therefore, the cloud dynamics fails to replicate the occupant's motion. The system's accuracy is fairly stable when the particle number is in the range 500 to 2000. Therefore, the smallest value (500) becomes the best choice since it is the closest to reproduce the occupant's motion in real time.

8. CONCLUSION

We presented in this paper the localization system we put in place in the DOMUS laboratory. This system detects a single person's location by means of various anonymous sensors installed in the smart home. We set an experimental scenario in order to evaluate the accuracy, with people performing significant tasks in the smart home. The results obtained with these data are very satisfactory. The algorithm based on Bayesian filtering shows a mean localization accuracy of 85%. The system is fast and robust regarding noise. Moreover, since it is profile-independent, it can easily be deployed in the future homes that are being conceived in the laboratory. It will be interesting to observe the behavior of our system once it gets integrated in applications meant to provide high levels of assistance. Above all, it is crucial to measure the localization accuracy needed for different sets of applications. A possible continuation of this work would be to check the possibility to locate two or more people with the same experimental setup (the presence of pets is also interesting to investigate). This would enable to respect the cost and anonymousness constraints when several people are present. A multiagent approach is being investigated at DOMUS and already is giving promising results. Moreover, using data from people with cognitive deficiencies, in a real setting, would help in consolidating the results from the present study.

ACKNOWLEDGMENT

The authors would like to thank Ernest Monga from the Department of Mathematics at Université de Sherbrooke for valuable discussions about the statistical properties of the data presented here.

REFERENCES

[1] I. McDowell, G. Hill, J. Lindsay, et al., "Canadian study of health and aging: study methods and prevalence of dementia," *Canadian Medical Association Journal*, vol. 150, no. 6, pp. 899–913, 1994.

[2] H. Pigot, S. Giroux, A. Mayers, B. Lefebvre, and V. Rialle, "The intelligent habitat and everyday life activity support," in *Proceedings of the 5th International Conference on Simulations in Biomedecine*, pp. 507–516, Ljubljana, Slovenia, April 2003.

[3] Y. Rahal, P. Mabilleau, and H. Pigot, "Bayesian filtering and anonymous sensors for localization in a smart home," in *Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW '07)*, vol. 1, pp. 793–797, Niagara Falls, Canada, May 2007.

[4] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *Computer*, vol. 34, no. 8, pp. 57–66, 2001.

[5] A. M. Tabar, A. Keshavarz, and H. Aghajan, "Smart home care network using sensor fusion and distributed vision-based reasoning," in *Proceedings of the 4th ACM International Workshop on Video Surveillance and Sensor Networks (VSSN '06)*, pp. 145–154, Santa Barbara, Calif, USA, October 2006.

[6] J. Werb and C. Lanzl, "Designing a positioning system for finding things and people indoors," *IEEE Spectrum*, vol. 35, no. 9, pp. 71–78, 1998.

[7] "Ekahau Positioning Engine," http://www.ekahau.com/.

[8] H. S. Lee, B. Song, H. Y. Youn, and K. Chung, "The method of location error detection and correcting in smart home environments," in *Proceedings of the International Conference on Hybrid Information Technology (ICHIT '06)*, vol. 2, pp. 432–438, Cheju Island, Korea, November 2006.

[9] Y. Kaddourah, J. King, and A. Helal, "Cost-precision tradeoffs in unencumbered floor-based indoor location tracking," in *Proceedings of the 3rd International Conference on Smart Homes and Health Telematic (ICOST '05)*, Sherbrooke, Québec, Canada, July 2005.

[10] S. Thrun, W. Burgard, and D. Fox, "A probabilistic approach to concurrent mapping and localization for mobile robots," *Autonomous Robots*, vol. 5, no. 3-4, pp. 253–271, 1998.

[11] D. Fox, J. Hightower, L. Liao, D. Schulz, and G. Bordello, "Bayesian filtering for location estimation," *IEEE Pervasive Computing*, vol. 2, no. 3, pp. 24–33, 2003.

[12] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2nd edition, 2002.

Hindawi Publishing Corporation
International Journal of Telemedicine and Applications
Volume 2008, Article ID 867639, 13 pages
doi:10.1155/2008/867639

Research Article

ERMHAN: A Context-Aware Service Platform to Support Continuous Care Networks for Home-Based Assistance

Federica Paganelli, Emilio Spinicci, and Dino Giuli

National Inter-University Consortium for Telecommunications, Department of Electronics and Telecommunications, University of Firenze via S. Marta 3, 50139 Firenze, Italy

Correspondence should be addressed to Federica Paganelli, federica.paganelli@unifi.it

Received 28 September 2007; Revised 19 February 2008; Accepted 20 May 2008

Recommended by Frédérique Laforest

Continuous care models for chronic diseases pose several technology-oriented challenges for home-based continuous care, where assistance services rely on a close collaboration among different stakeholders such as health operators, patient relatives, and social community members. Here we describe Emilia Romagna Mobile Health Assistance Network (ERMHAN) a multichannel context-aware service platform designed to support care networks in cooperating and sharing information with the goal of improving patient quality of life. In order to meet extensibility and flexibility requirements, this platform has been developed through ontology-based context-aware computing and a service oriented approach. We also provide some preliminary results of performance analysis and user survey activity.

1. INTRODUCTION

The growth in population ageing in most industrialized nations and the related increasing percentage of chronic disease diffusion is posing several problems at different levels of our society (political, social, and familiar/personal levels).

Statistics data provide a clear picture of the problem dimensions: according to recent results of a survey funded by the European Commission [1], Europe and Japan will experience the most pronounced ageing trends up to 2050. The share of the above 60 age group will be around 37% in Europe and even more in Japan, and slightly lower in North America (27%). The percentage of the EU population over the age of 65 is expected to reach more than 28% by 2050, with an estimated group of 80 million people that will need care and assistance services, in particular for chronic diseases [1].

New care models for chronic condition management have been proposed which define guidelines for policy planning, as well as principles for social community and health-care system organization and effort coordination. These models, usually named continuous-care models, promote home-based continuous care for chronic patients. They emphasize the fact that the effectiveness and efficiency of long-term condition care depend strongly on the capability of both the patients and their relatives to manage their cases (self-management) and on the collaboration of all involved care providers. Patients, family members, health-care teams (e.g., clinicians, general practitioner, nurses, etc.), and social community members (e.g., social workers, volunteers) should be properly informed, motivated, and prepared in order to effectively collaborate together.

One of the first examples is the Chronic Care Model (CCM), which is a conceptual, evidence-based framework developed in the USA [2]. This model proposes innovative organizational aspects aiming at improving the effectiveness and efficiency of assistance to patients affected by chronic conditions. The WHO has recently proposed the Innovative Care for Chronic Conditions (ICCC) framework, which widens the CCM in order to meet the needs of the international community [3].

Therefore, advances in information and communication technologies (ICTs) and ambient intelligence (AmI) context-aware system design are required in this framework in order to address two main objectives. The final objective is to improve the quality of life of both patients and their closest relatives. As a matter of fact, while it is obvious

that chronic conditions may cause limitations to a patient's everyday activities, it should not be overlooked that they may dramatically decrease the quality of life for a patient's family members as well [4]. In order to care for the elder, relatives may be asked to make drastic life changes, such as quitting their jobs and giving up their social lives. For such a purpose, an ulterior objective is the implementation of ICT platforms capable of supporting long-term care service provision while enabling cost savings and effective heterogeneous resource management (e.g., professionals, biomedical instruments, etc.).

From the viewpoint of AmI and context-aware system designers, home-based care models pose several technology-oriented challenges. Services which should be provided by an AmI system have been classified into the following categories [5]: (a) emergency treatment (services for emergency detection and management); (b) autonomy enhancement, that is, user assistance services for primary needs and/or everyday activities (nutrition, taking medication, monitoring vital signs, etc.); comfort services, that is, services enabling better life quality (e.g., education, socialization, etc.).

These services should be adaptive, easy to use, and strongly personalized according to user requirements, needs, and disabilities. In most existing AmI context-aware systems, the focus is on the requirements of the person needing assistance (the "patient"), whereas requirements coming from the heterogeneous community of people, who are involved to different extents in patient care and assistance (the "caregivers"), are often not taken into account when designing the system.

In this paper, we describe some results obtained through our current research on the design and development of a service-oriented architecture for continuous care provisions leveraging on context-aware and mobile technologies, called Emilia Romagna Mobile Health Assistance Network (ERMHAN). Its aim is to enable the development and delivery of an extensible set of care services which allow patients to be assisted at home in a familiar environment and support the activity and mutual collaboration of care providers who are involved to a different extent in patient care and assistance. Proper and effective information sharing for action and cooperation support by involved care providers is specifically targeted, due to their work in mobility conditions, their association to different organizational domains (e.g., hospital and municipality), and their various roles (such as general practitioner, specialist, nurse, family members, etc.).

This paper is organized as follows. In Section 2, we highlight the contribution of our work and discuss existing related research activities. Section 3 describes the specific objectives of our work with special focus on the context-aware approach and in Section 4 we describe the ERMHAN system architecture and some implementation details about the two main components, the Multichannel Health-Care Service Manager and the Context Management System. In Section 5, we discuss some preliminary system evaluation results and user testing activities. Section 6 concludes the paper by highlighting the most relevant results of our work and future research directions.

2. RELATED WORK

Several research domains can be considered of interest in delivering AmI and pervasive health services for chronic diseases, spanning from smart homes, assistive technologies and home-based health monitoring, to context-aware hospitals.

One of the first relevant contributions has been provided by researchers working at the Georgia Tech Aware Home, a prototype of a smart home, where sensing and perception technologies are used to gain awareness of inhabitant activities and to enable services for maintaining independence and quality of life for an ageing population [6]. The INHOME project [7] aims at providing the means for improving the quality of life of elderly people at home, by developing technologies for managing their domestic environment and enhancing their autonomy and safety at home (e.g., activity monitoring, simple home environment management, flexible AV streams handling, flexible household appliance access). Among more recent works, the "ubiquitous home" is a real-life test-bed for home-based context-aware service experiments [8] in Japan. A set of implemented context-aware services has been evaluated by means of real-life experiments with elderly people.

In the field of assistive technologies and home-based health monitoring systems, several examples exist: Vivago is an alarm system which provides long-term user activity monitoring and alarm notification [9]; the CareMedia system [10] uses multimedia information to track user activities.

The "hospital of the future" prototype [11] is an example of a context-aware computing system in a hospital environment. It consists of a series of context-aware tools: an electronic patient record (EPR), a pill container, and a hospital bed which displays relevant patient record information, such as the medicine schema, according to contextual information (e.g., nurse position, patient, medicine tray). Muñoz et al. [12] have recently proposed a context-aware mobile system where mobile devices are capable of recognizing the setting in which hospital workers perform their tasks, and let users send messages and access hospital services according to these contextual elements.

Despite the multitude of relevant contributions in the above-mentioned research fields, only recently research activities on pervasive services for ageing and chronic disease management have begun addressing these requirements by means of a holistic approach, taking systematically into account standard guidelines and reference models for continuous care. Consolvo et al. [4] have applied social network analysis methodology to the study of continuous care networks; they conducted a series of interviews in order to explore the space of eldercare (i.e., who was involved in the care, what types of care were needed, and what types of care were being provided); based on user study results they offer some design guidelines for the development of successful computer-supported coordinated care (CSCC) systems.

Pervasive self care is a conceptual framework for the development of pervasive self-care services [13]. This study has been promoted in the framework of self care, an initiative by the Department of Health in the UK that aims

at treating patients with long-term conditions near home. The proposed reference model, inspired by the principles of service-oriented architecture (SOA), distinguishes three main spheres: the body sphere (a body area network supported by a router which interacts with body sensors and with the home sphere); the home sphere (a home server that collects and preprocesses sensed data), and the self-care service sphere (the data processing and sharing subsystem).

Some experimentation results are given in [14], where a telemedicine system is used for the home care of patients suffering from chronic obtrusive pulmonary disease (COPD); the integrated telemedicine system provides professionals with shared and ubiquitous access to patient health records and patients with direct access to nurse case manager, telemonitoring, and televisit services.

2.1. Contribution of our work

The design of ICT tools for chronic disease management should take into account flexibility and extensibility requirements. These requirements are common to all kinds of distributed systems, but especially are applied to this application domain, due to its intrinsic characteristics, such as different national and local regulation frameworks, the heterogeneity of health centers and communities involved in care service delivery, and the patient's health status development over time. In such a complex and changing environment, cost-effective solutions should be conceived as extensible and flexible service platforms. For that reason, while the research activities surveyed above have focused on the conceptual design or implementation of applications that target specific chronic diseases and not on a specific chronic disease from the very beginning, our approach has been to design and deploy a service platform that provides general purpose services and could be easily extended and specialized in order to match specific requirements of real cases.

The aim of our research has been to design and implement ERMHAN, an extensible service platform supporting care teams in providing long-term assistance services. Extensibility and flexibility of the ERMHAN service platform are mainly achieved by means of modular and service-oriented design and the adoption of open and standardized data formats and communication protocols.

The platform design is based on the definition of basic functional blocks and their interconnection by means of web services standards. As a matter of fact, web services are recognized as an open and standardized way of achieving interoperation between different software applications, running on heterogeneous platforms and/or frameworks [15].

Semantic web technologies have been applied to data representation and processing in order to provide instruments that ease the development of pervasive and personalized care services. More specifically, semantic web is used in order to (a) represent knowledge by means of ontology-based formalisms; (b) reason over knowledge using rule-based and ontology-based engines; (c) apply reasoning techniques in order to implement personalized healthcare plans.

The prototype we have developed provides basic and general-purpose services for information sharing, distributed, and multichannel personalized access to patients' records and personalized real-time monitoring and alarm management on a per patient basis. Implemented services include access to complete and updated patient records (including patient health status, description of care provider interventions), even in mobility conditions (at the hospital or at patient's home) and through different devices (at least personal digital assistants and desktop PCs); notification of patient health status conditions and alarms, but without overwhelming care operators with too much information (this might have the drawback of disturbing users and providing them with "no information"). More effective information delivery could be achieved by routing intervention requests according to patient health status gravity and required expertise for intervention.

Based on its flexibility, the ERMHAN service platform can be specialized to address the needs of specific patient cases. To achieve this objective, the basic services provided by the ERMHAN platform can be integrated with those offered by other systems, such as assistive technologies, home automation systems, specific chronic disease prognoses, and diagnosis systems [16]. In the following sections, we provide further details about the modeling approach and system architecture of ERMHAN platform.

3. CONTEXT-AWARE MODELING AND REASONING FOR HOME-BASED CARE

The objective of this research is the design and prototypal development of a context-aware mobile service platform (ERMHAN) for long-term home-based care services delivery to chronic patients.

ERMHAN aims at supporting the main characteristics of emerging chronic care models: (a) home-centered long-term patient followup; (b) shared care provided by a network of heterogeneous care providers (named "continuous care network"); (c) adaptation of care plans on a per patient basis.

With the term "continuous care network" we refer to the broad range of people involved to different extents in the care of the elder, such as patient relatives, the multidisciplinary teams at different health-care levels (e.g., general practitioners, specialists), social operators, volunteers, and so forth. They belong to different organizations (e.g., hospitals, municipality, private institutions), may work in more than a single location, and do not usually have regular face-to-face contacts with other network members.

Due to the complexity of the care network, its ad hoc nature, and loosely coupled structure, we have identified a set of basic services to be provided to care providers: ubiquitous access to patient-health records (via mobile device or desktop PC); role-based information sharing among different network members belonging to different organizations; capability of adapting health plans on a per-patient basis. More specifically, health-plan personalization is implemented in ERMHAN in terms of personalized context-aware rules for alarm detection and management, as described in the next paragraph.

Even if the aim of our work is to provide services to care providers, and consequently developing assistive

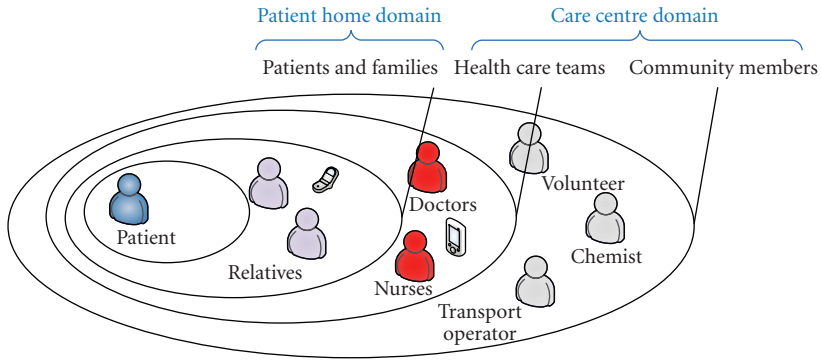


Figure 1: ERMHAN applications scenario.

technologies are not the main target of this work, ERMHAN provides some basic services for patient assistance, for example, medication reminders, manual alarm activation, and communication request services. As the focus of this work is on the architecture for service integration and delivery, human computer interaction (HCI) issues, such as usability and accessibility, have not been taken into account at this stage of our study.

The ERMHAN reference application scenario consists of two main domains (see Figure 1):

(a) the patient home, that is, the environment where the patient lives, often with some relatives, and where patient monitoring and assistive technologies would be deployed;

(b) the care centre, that is, the domain of the organization responsible for care service integration and delivery (e.g., hospitals, nursing homes or municipalities). For the sake of simplicity, in our scenario this is presented as a single logical point of care, providing ICT services to care operators who can belong to different organizations. Actors considered in our scenario are doctors (general practitioners and specialists) and nurses. More complex scenarios involving other organization domains (e.g., voluntary associations, pharmacies, etc.) will be analyzed in future research activities.

3.1. Context awareness

This paragraph describes what we do intend for "context awareness" in the framework of home-based continuous care and the context model that has been adopted in ERMHAN.

Context awareness is a general concept that refers to the capability of a system to be aware of its physical and logical environment and to intelligently react according to this awareness. The definition of context is likely to change according to the application domain and related purposes [17].

In our application scenario, context awareness is meant as the capability of the system to acquire and interpret relevant information with respect to the main goal (patient quality of life) and to perform actions aiming at achieving such a goal. Thus, the context may include information about patient health status, inferred by sensed biomedical parameters (e.g., vital signs), and home environment (e.g., temperature and relative humidity). Even social context (i.e., information regarding people populating the patient care network) may be considered as relevant information to be exploited for context-aware assistance service delivery. The acquisition of these context data can be profitably exploited by the system in order to infer relevant situations related to patient status, (such as critical condition and alarm detection) and trigger proper actions in order to facilitate the care providers' operations (i.e., alerting care providers for intervention).

3.1.1. ERMHAN context model

An ontology-based context model, written in Web Ontology Language (OWL [18]), is employed throughout the entire process of sensing, interpreting, managing, and exchanging context information. We have extended a general purpose ontology-based context model [19] with concepts and relations describing the home care setting.

The OWL-based context model is used in order to represent system knowledge about patient health status by means of ontology-based formalisms. The context model provides a uniform representation of data coming from heterogeneous sources, such as biomedical sensors, home monitoring systems and users' manual input. Ontology and rule-based reasoning are applied over the context model in order to check the consistency of context information and to infer further knowledge. Rule-based reasoning is applied in the ERMHAN system in order to infer patient health conditions, detect alarm conditions originating from a patient health status or other events (e.g., the patient does not take medication), and select the alerting policy which is most suitable according to alarm level and other context information (e.g., availability of care providers for intervention).

The main advantage of this approach consists in representing the business logic of the system in an explicit and declarative way. Thus, adaptation of health plans on a per patient basis can be realized by allowing care providers to specify rules tailored to the patient case.

Table 1: Alert policy examples.

Alarm level	Alert policies
VERY LOW	(i) SMS to patient relative, no acknowledgement
LOW	(i) SMS and mail to general practitioner, no acknowledgement (ii) SMS and mail to relative, no acknowledgement
MEDIUM	(i) SMS and mail to general practitioner or nurse, acknowledgment required (ii) Send SMS and mail to relative, no acknowledgment
HIGH	(i) Message to emergency operator, acknowledgment required (ii) SMS to patient relative, acknowledgement required

Hereafter we describe the main features of the context modeling and reasoning capabilities implemented in the ERMHAN system. A complete description of our ontology-based context model for home health monitoring and alerting in patient care networks is out of the scope of this article and is fully described in [20].

The context model is composed of the following parts.

(1) The *patient-domain ontology* includes context items about patient biomedical parameters, location, and activity. This information is used by the system to automatically infer patient health status and detect alarm situations by means of rule-based reasoning. Biomedical parameters instances are represented together with some relevant properties, such as measurement values and parameter ranges. Each range is specified in terms of upper and lower thresholds and related alarm level; when a measured value falls out of the thresholds, an alarm of the corresponding level is detected. Patient health status is thus determined by comparing biomedical parameters' measured values with a set of parameter ranges. We have specified four basic alarm levels (very low, low, medium, high), but the model can be easily extended to include further levels.

(2) The *home environment ontology* includes context data that describes the patient home environment. For instance, monitoring environmental parameters (such as temperature and relative humidity) are needed in order to maintain a healthy environment and detect alarm situations. Alarm triggering based on environment parameters is analogous to the one based on biomedical parameters, described above.

(3) The *social context ontology* represents care network resources (health teams, social community members, etc.) and their relations. Care network members that are represented as concepts in this ontology include the patient, patient relatives, who take active part in implementing the assistance plan, health operators (medical and paramedical staff) who are in charge of patient care; social community organizations which offer assistance services (e.g., transport and companion services). The social context ontology is also used to represent other information which is considered relevant to the ERMHAN application purposes, such as the availability of a health operator for an intervention and the distance of a relative from patient location (travel time). At present, users can provide this information via a web form provided by the profile management services, by choosing among predefined values (e.g., "available," "busy," "not available"). As part of future work, we will investigate the use of context-aware systems that can help in semi-automatically updating and managing such information (e.g., GPS positioning and inference mechanisms).

(4) The *alarm management ontology* represents the policy that should be instantiated to manage an alarm. Such policy describes the steps that should be performed in the event of an incoming alarm. Rule-based reasoning over this ontology model is used to select the suitable alarm handling policy according to different context items, such as the incoming alarm level, the patient identifier, and the availability of care team members (i.e., the information represented by the social context ontology).

3.1.2. Rule-based reasoning examples

As mentioned above, the ERMHAN system uses rule-based reasoning over the context ontology instances for triggering alarms and defining policies for handling alarms.

Alarm triggering is based on the analysis of context data represented in the Patient and home-domain ontologies. We have defined a set of first-order rules so as to determine if an alarm has to be triggered and which level should be activated according to measured context data and corresponding thresholds. For instance, the following rule triggers an alarm classified as "high" when the heart rate frequency is less than 40 beat/minute and the systolic blood pressure is higher than 160 mm/Hg:

IF (HeartRateFrequency<40 AND
SystolicBloodPressure>160) THE
HealthStatus has AlarmLevel "HIGH."

Likewise, when an alarm has been triggered, rule-based reasoning is used to determine which alert policy should be actuated for handling the alarm situation. In ERMHAN, an alert policy is represented as a set of instructions specifying whom should be alerted and how (via SMS, mail, phone call) and if acknowledgment is required. In Table 1, we define a basic example policy for each alarm level.

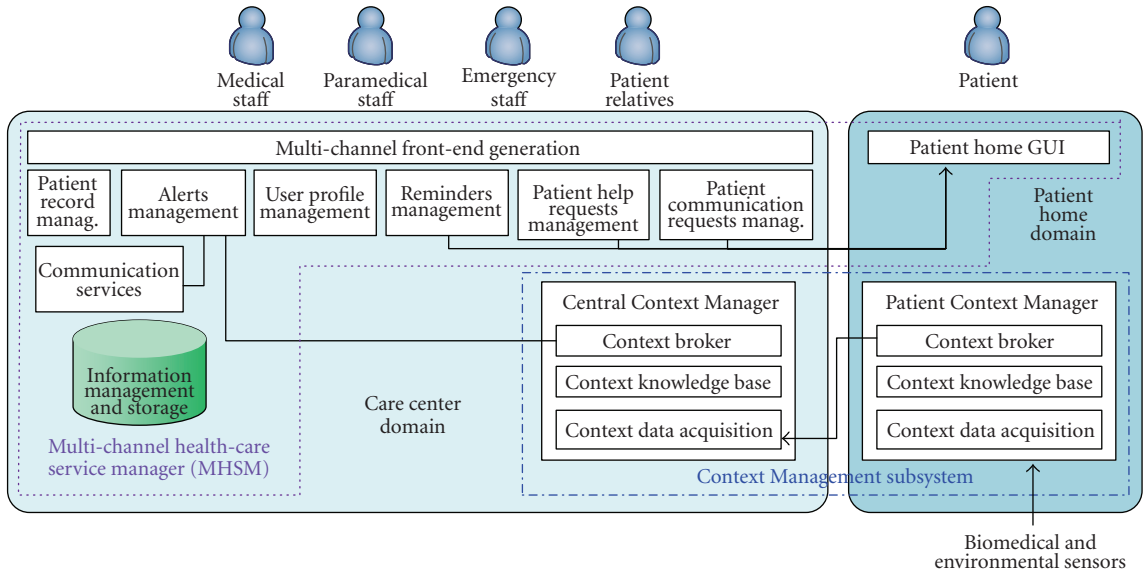


Figure 2: ERMHAN architecture.

For instance, we suppose that when a "MEDIUM" level alarm is detected, the system should perform the following actions (see Table 1).

(a) Alerting an operator via SMS or email. First, a general practitioner is alerted. If he/she does not send the acknowledgment within a fixed-time interval, other operators (e.g., nurses) are alerted until one of them sends an acknowledgment. At the end, if no acknowledgment is received, the system alerts an emergency operator, available 24/24 and 7/7. The list of operators to be contacted is dynamically created by the system by selecting the caregivers which are assigned to that patient and are "available" for intervention.

(b) At the same time at least one of family members should be alerted, but acknowledge is not required.

Adaptation of system behavior on a per-patient basis can thus be achieved by care providers by specifying personalized alarm levels, parameter thresholds, and alert policies tuned to the individual's conditions.

4. ERMHAN SYSTEM ARCHITECTURE

The ERMHAN system architecture, shown in Figure 2, is made of the following main components.

The Multichannel Health-Care Service Manager (MHSM). This component delivers mobile health-care context-aware services to patients, relatives, and care providers, targeting a variety of user devices. It is deployed in the care centre domain and includes a front-end component deployed in the patient-home domain.

The Context Management system. This component deals with the acquisition of context data from heterogeneous sources, the management, and storage of ontology-based context instances, reasoning over context knowledge, and the delivery of relevant contextual knowledge to context-aware applications (e.g., MHSM).

MHSM and Context Management systems have been designed as completely separate and independent modules. This has been done in order to strongly decouple context data acquisition, management, and context-aware service delivery. Communication among these components is based on web service interfaces.

4.1. Multichannel health-care service manager

The Multichannel Health-Care Service Manager (MHSM) is the ERMHAN component that provides mobile context-aware health-care services to end-users [21]. MHSM provides end-users with proper services according to their roles and related requirements. ERMHAN services are specifically targeted to care-provider requirements. Thus, end-user roles which have been taken into account are medical, paramedical, and the emergency centre staff. Nonetheless, we have also considered some basic requirements for patients and patients' relatives.

Medical staff

The medical staff must be constantly informed about the patient's conditions. Each member can access ERMHAN services remotely, via a desktop PC or a PDA. Available services include the following.

- Patient record management. Medical staff members can access and update the patient record using a personal digital assistant (PDA) or desktop PC. Information sharing among care professionals is achieved by providing them with the capability of accessing a uniform view of patient information. Permission for reading and modifying patient record

fields is tailored according to the user's role (e.g., general practitioner, specialists, etc.)

- Alerts management service. When an alarm is triggered (manually by the patient or automatically by the Context Management system), the alerts management service contacts the staff through the appropriate channels (e.g., mail and/or SMS). As specified by the alert policy, alerts can be sent as informative messages or as intervention requests, depending on the alarm gravity level. In the latter case, the caregiver is asked to acknowledge the alert reception and to confirm the intervention request. More details about the alert policy are provided in the next subsection.
- Profile management service. The medical staff members can specify their own availability for the intervention (e.g., available, not available, busy). This information is taken into account by the rule-based reasoning process used to select the operators which should be contacted in the alert policy.

Paramedical staff

The paramedical staff (e.g., general assistance operators, nurses, etc.) can access the system remotely, by means of a PC or a PDA. Similarly to the medical staff, but with different role-based access rights, they can access the following services: patient record management, alerts management services, and profile management services.

Emergency staff

The emergency staff is composed of "always available" operators. They are notified by the alerts management service in order to respond to emergency situations (highest critical alarm level) or when other health operators have not responded to lower level alarms (e.g., medical staff).

Patient

We considered the following patient requirements for eliciting a basic set of services: the capability of easily communicating with relatives or health operators, manually activating alarms, and having support for implementing the health assistance plan. Requirements depending on specific patient conditions (e.g., physical impairments, cognitive disabilities, etc.) have not been considered in this study. The patient is equipped at home with a Tablet PC and uses services through a Front-End available on the Tablet PC touch screen. More specifically, services which are available to patients at home include the following.

- Reminders service: the system alerts the patient that medication has to be taken. The alert signal is an audio and video alarm which is remotely triggered on the front-end deployed at patient home.
- Help requests: the patient can manually generate an alarm, in order to request an urgent intervention by the emergency staff.
- Communication requests: the patient can request to be contacted by some relatives.

Patient relatives

The patient's family members provide general assistance to the patient during daily routine activities at home and should be kept informed about patient conditions when outside the home (via the alert management service and profile management service).

Services targeting relatives can be accessed via mobile devices with minimal technical requirements (i.e., cell phones with an XHTML browser).

Multichannel health-care service manager implementation

The MHSM has been designed and developed as a multichannel web application written in Java language, leveraging the J2EE JSP/Servlet technology framework. The MHSM has been developed as a 3-tier architecture including the following.

(1) A back-end tier, named MHSM information management and storage, handles and stores the following information resources: the patient record, containing patient personal data, personal assistance plan, vital sign thresholds, medical, sanitary and assistance diaries, and prescriptions (e.g., taking medication, vital sign measurement scheduling); user profiles, containing end-user personal data, organizational roles, caregiver availability status (for medical and paramedical staff and patient relatives); organization models, representing the care network structure, (i.e., care network members and patients under care).

(2) An intermediate tier hosts the business logic of the ERMHAN Services (patient record management, alerts management, profile management, reminder management, patient help, and communications request management).

(3) An upper tier, composed of a multichannel front-end generation service, generates user interfaces tailored to the capabilities of the different end-user mobile devices (e.g., display size, resolution, memory and processing capabilities, etc.). The target devices for this scenario are PDAs, Tablet PCs, cell-phones, and laptops. The model view controller design paradigm [22] has been adopted in order to properly manage user interface generation, business and navigation logic, and content generation.

The platform uses the web services technology standards (SOAP, WSDL) to interact with external components, such as the Context Management system. The underlying DBMS is MySQL. A communication management service implements an SMS gateway and the SMTP protocol in order to provide SMS and e-mail alerts to care network members (the adoption of further communication channels and protocols will be evaluated in the future). This component has been designed and developed by the industrial partner of the KAMER Project (HP Italy).

4.2. Context Management system

This section describes the Context Management (CM) system architecture that we have developed to ease the

implementation of context-aware assistance services for chronic patients.

The CM system is based on a general-purpose framework providing basic features for context data acquisition, reasoning, and delivery [19]. It is composed of two different node types.

(a) The Patient Context Manager (PatientCM). This node is deployed at the patient site. It acquires data retrieved by biomedical and environmental sensor networks. These data are preprocessed by the PatientCM in order to detect abnormal conditions and are transmitted to the MHSM in order to update patient records with new measurement values. Sensed data are injected into the patient and home-domain ontology instances and rule-based reasoning processes are performed over this context base in order to detect incoming emergency situations and trigger corresponding alarms. The Central Context Manager is then notified of the occurrence of an alarm.

(b) The Central Context Manager (CentralCM). This component is deployed in the care centre domain. When an alarm situation has been detected by a PatientCM, the CentralCM processes context information about patient health status and care team member availability, in order to define an alert policy for informing care team members and request their intervention in case of emergency. It handles the alarm management and social context ontologies by creating related instances updated with context information about patient health status, triggered alarm levels and information about care team member availability. Based on this context information, reasoning rules are applied in order to define the proper alert policy for alarm handling. The resulting alert policy is then sent to the MHSM for its implementation.

Both the Patient and the Central Context Managers include the following main blocks.

Context data acquisition: this component receives context data from context providers via SOAP messages (e.g., web service-enabled sensor networks), or via sensor adapters.

Context knowledge base: this is a knowledge base composed of ontology-based context models and instances stored in a database, a rule-based reasoner and rule files. This component is based on Jena, an open source Semantic Web framework [23].

Context broker: this component distributes context data to external components by notifying the interested applications when the context has changed or by providing updated information in a request/response chain. Both notification and request/response paradigms are implemented through a SOAP message exchange [24].

In each Context Manager, the communication between internal components has been implemented according to the observer design pattern, that is, through suitable EventListeners that listen for events and encapsulate the information to be processed [25]. Therefore, such decomposition allows each internal component to process its own specific information independently.

The CentralCM and PatientCM have been implemented as J2EE web applications and their external interfaces are exposed as web services.

4.2.1. Alarm detection and management scenario

This paragraph provides further insight into the Context Management system architecture (see Figure 3) by describing the system behavior in an alarm detection and management scenario. Figure 4 provides a simplified graphical representation of this scenario as a sequence of numbered steps.

General practitioners and nurses can access and modify patient health information through the MHSM interface (Step 1). In particular, general practitioners can update remote monitoring parameters, such as scheduling sensed data acquisition and related alarm thresholds. For instance, a practitioner can specify that the body temperature should be measured at 7 am and 5 pm every day, that a measured value higher than 38°C should trigger a MEDIUM-level alarm if reached at 7 am, and a high-level alarm if reached at 5 pm. The MHSM sends this information to the CentralCM which routes the message to the proper PatientCM (Step 2).

The *CentralCM.PatientServiceManager* component offers a web service interface which can be invoked by the MHSM in order to communicate messages such as updates of patient diagnostic test scheduling and vital sign alarm thresholds. The CentralCM manages a message queue for each PatientCM and messages that come from the MHSM are then routed to the proper message queue. The PatientCM.CentralCMClient component periodically polls the *CentralCM.PatientServiceManager* to request new messages.

This periodical polling is also used to monitor the status of both the PatientCM operation and the connection between the PatientCM and the CentralCM. When the CentralCM does not receive the expected periodical poll by a PatientCM (this condition may be caused by connection failures or PatientCM malfunctions), it triggers a corresponding alarm and the emergency staff is alerted to handle this technical problem by activating the required procedures.

The *PatientCM.CentralCMClient* parses the message payload and produces suitable events, such as TimingEvents for scheduling biomedical data acquisition.

New timing events are consumed by the *PatientCM.DataAcquisitionScheduler*. According to the prescriptions contained in the TimingEvents objects, the scheduler allocates corresponding timers triggering the patient's biomedical data acquisition (e.g., body temperature, heart rate frequency) from web service-enabled body sensor networks (Step 3).

The *PatientCM.DataCollector* exposes a web service interface which is invoked by the sensor networks to communicate new acquired vital sign values. These data are then passed on to the DataDispatcher, that notifies the

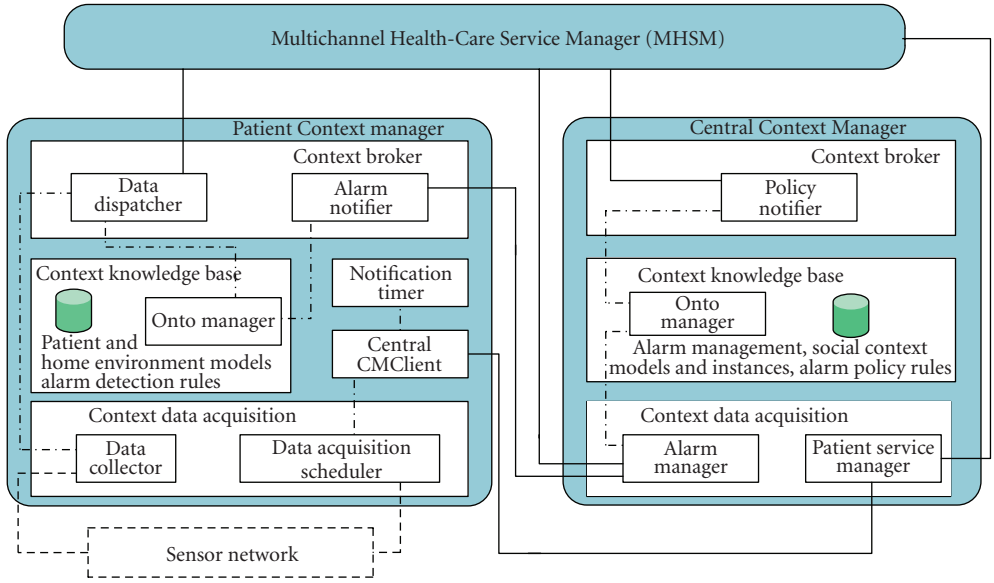


Figure 3: ERMHAN Context Management system.

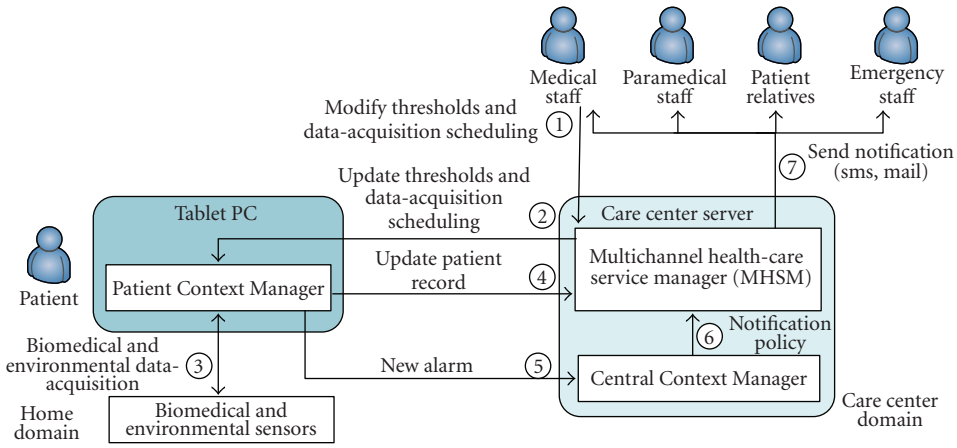


Figure 4: Alarm detection and management scenario.

MHSM of these data for patient record updating (Step 4) and the OntoManager module for updating the Patient and home environment ontologies instances. Rule based reasoning is then applied to the knowledge base in order to detect possible alarms and produce AlarmEvents. For instance, if the sensed body temperature is higher than 38°C a MEDIUM-level alarm is triggered. The *PatientCM.AlarmNotifier* notifies a new AlarmEvent to the CentralCM, by invoking the *CentralCM.AlarmManager* component (Step 5).

The *CentralCM.AlarmManager* receives new incoming alarms notified by the PatientCM and routes them to the *CentralCM.OntoManager* which has to create an alert policy for alarm handling. Firstly, it feeds the knowledge base (composed of the alarm management and social context ontologies) with instance data, such as attributes of the AlarmEvent (patient identifier and alarm level) and updated information about the patient social context (health operators availability) which are made available by the MHSM. Then, it applies rule-based reasoning to produce the alert policy.

The *CentralCM.PolicyNotifier* is the component which sends the alert policy to the MHSM, by invoking a specific web service (Step 6). The MHSM then implements the policy by sending proper alerts to the contact lists (Step 7).

Filtering techniques are implemented by the *CentralCM.AlarmManager* to handle sequences of alerts related to the same emergency situation. As a matter of fact, while an alarm is being handled, notification related to the same situation can be generated again by the PatientCM (e.g., because of new vital sign measurements). The filtering technique is thus applied to avoid overwhelming health operators with redundant alerts.

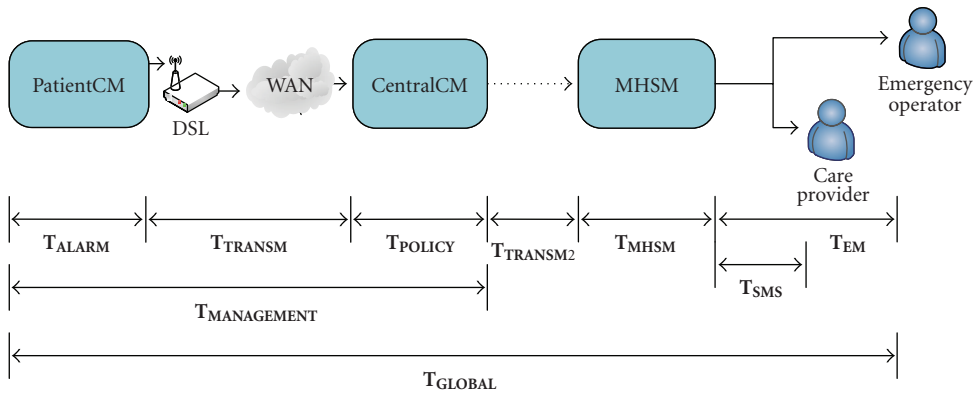


Figure 5: Time intervals for ERMHAN performance evaluation.

5. SYSTEM EVALUATION AND TESTING ACTIVITIES

This section discusses the ERMHAN system results in terms of performance evaluation and user testing activities.

5.1. Context Management system performance evaluation

The following considerations are concerned with a performance estimation of the ERMHAN Context Management system obtained by collecting average alarm generation and transmission times in a basic configuration. This configuration is composed of one PatientCM and one CentralCM. The PatientCM is geographically located 40 Km away from the CentralCM. The server hosting the CentralCM has an external IP and accesses the Internet through the university wide area network. The PatientCM can access the Internet through a standard low-band with DSL connection (640 Kbps). More precisely, the PC hosting the PatientCM has a wireless connection to an 802.11g DSL gateway. In particular, we have considered the following hardware configuration.

(1) PatientCM: AMD Athlon64 3200+, 1024 Mb RAM.

(2) CentralCM: Intel Pentium 4 2.0 GHz, 768 Mb RAM.

The performance measurement is based on the following parameters: the alarm detection time ($\mathbf{T_{ALARM}}$) which is the time interval between the acquisition of an out-of-range biomedical parameter value and the triggering of the corresponding alarm obtained by means of rule-based reasoning (Section 3.1.2); the transmission delay ($\mathbf{T_{TRANSM}}$) which is the time needed for transmitting a message through the connection between PatientCM and CentralCM; the alert policy time ($\mathbf{T_{POLICY}}$), which is the time elapsing at the CentralCM side between the reception of an incoming alarm originating from the PatientCM and the generation of the corresponding alert policy. The sum of these time intervals determines the overall time needed for the alert policy generation from time the out-of-range biomedical parameter is acquired ($\mathbf{T_{MANAGEMENT}}$):

$$T_{\text{MANAGEMENT}} = T_{\text{ALARM}} + T_{\text{TRANSM}} + T_{\text{POLICY}}. \quad (1)$$

The value of $\mathbf{T_{MANAGEMENT}}$ does not vary significantly with respect to alarm levels. Nonetheless, alarm levels influence the kind of alert policy generated by the CentralCM and implemented by the MHSM. An evaluation including also policy implementation and the notification to at least one care provider should also take into account the following time intervals (see Figure 5): the time needed for transmitting the alert policy from the CentralCM to the MHSM ($\mathbf{T_{TRANSM2}}$); the time needed by the MHSM for instantiating the alert policy ($\mathbf{T_{MHSM}}$); the SMS latency ($\mathbf{T_{SMS}}$) which is the delay for transmitting SMSs to health operators; the delay for transmitting alert to the care centre in order to notify emergency operators ($\mathbf{T_{EM}}$).

The formula for calculating the global time interval ($\mathbf{T_{GLOBAL}}$) from the alarm detection to the alert notification to care providers differs according to the alarm level and corresponding alert policy.

For VERY-LOW- and LOW-level alarms, the following formula is applied:

$$T_{\text{GLOBAL}} = T_{\text{MANAGEMENT}} + T_{\text{TRANSM2}} + T_{\text{MHSM}} + T_{\text{SMS}}. \quad (2)$$

For MEDIUM-level alarms, the calculation has to take into account the fact that health operators (general practitioners and nurses) are notified in sequence: a first operator is alerted via SMS, if he/she does not respond within a specified time interval ($\mathbf{T_{TIMEOUT}}$), another operator is notified. After $\mathbf{N}$ failed attempts ($\mathbf{N}$ is specified in the alert policy), the alarm is communicated to an emergency operator ($\mathbf{T_{EM}}$).

If a care provider confirms the SMS reception after n attempts ($n \leq N$),

$$T_{\text{GLOBAL}} = T_{\text{MANAGEMENT}} + T_{\text{TRANSM2}} + T_{\text{MHSM}} + n(T_{\text{SMS}} + T_{\text{TIMEOUT}}) \quad (3)$$

otherwise

$$T_{\text{GLOBAL}} = T_{\text{MANAGEMENT}} + T_{\text{TRANSM2}} + T_{\text{MHSM}} + N(T_{\text{SMS}} + T_{\text{TIMEOUT}}) + T_{\text{EM}}. \quad (4)$$

For HIGH-level alarms the following formula is applied:

$$T_{\text{GLOBAL}} = T_{\text{MANAGEMENT}} + T_{\text{TRANSM2}} + T_{\text{MHSM}} + T_{\text{EM}}. \quad (5)$$

Table 2: Mean and standard deviation values of alarm detection ($\mathbf{T_{ALARM}}$), alert policy definition ($\mathbf{T_{POLICY}}$) and management, and overall alarm management times ($\mathbf{T_{MANAGEMENT}}$) in a basic Context Management configuration.

Alarm level	$\mathbf{T_{ALARM}}$		$\mathbf{T_{POLICY}}$		$\mathbf{T_{MANAGEMENT}}$	
	Time slot 1	*Time slot 2*	*Time slot 1*	*Time slot 2*	*Time slot 1*	*Time slot 2*
	Average		*Average*		*Average*	
Very low	1567 ± 220	1358 ± 213	869 ± 203	843 ± 175	7728 ± 951	6814 ± 1303
	1449 ± 238		854 ± 186		6196 ± 1247	
Low	1457 ± 207	1360 ± 217	879 ± 218	869 ± 203	8235 ± 618	6963 ± 1247
	1409 ± 217		874 ± 210		7604 ± 1168	
Medium	1455 ± 178	1576 ± 220	865 ± 215	869 ± 232	8433 ± 652	7837 ± 680
	1515 ± 208		867 ± 223		8137 ± 728	
High	1402 ± 201	1576 ± 222	859 ± 221	855 ± 225	7481 ± 610	8848 ± 635
	1489 ± 229		857 ± 222		8164 ± 925	

As mentioned above, this paragraph focuses on performance estimation of the Context Management system in delivering the most critical service (alarm detection and handling). Thus, our evaluation is based on measured values of T_{ALARM}, T_{TRANSM}, and T_{POLICY}. Values of further parameters ($\mathbf{T_{TRANSM2}}$ $\mathbf{T_{MHSM}}$, $\mathbf{T_{SMS}}$ and $\mathbf{T_{EM}}$) are not available for dissemination as they have been collected by the industrial partner (HP Italy). Moreover, values of $\mathbf{T_{SMS}}$ and $\mathbf{T_{EM}}$ are dependent on the characteristics of the network infrastructure. In a real implementation scenario, these parameters might be optimized according to appropriate technological choices, such as using public or professional radio-mobile networks for SMS delivery (i.e., GSM and TETRA, resp.) and a dedicated fixed line for communicating with the emergency operators, as well as establishing proper service level agreements with network operators.

To estimate the performance of the implemented system in detecting and managing alarms, we have simulated the generation of 60 sample alarms for each of the four alarm levels ("Very low," "Low," "Medium," and "High"). Such samples have been acquired by monitoring the activity of the Context Management system upon an interval of five working days, during working hours, in order to stress the system during worst traffic conditions (i.e., peak traffic hours). In more detail, experiments were performed during two time slots: 10:00 A.M.–2:00 P.M. (*Time Slot 1*) and 2:00 P.M.–6:00 P.M. (*Time Slot 2*).

Table 2 reports the mean and standard deviation values in milliseconds of alarm detection time ($\mathbf{T_{ALARM}}$), alert policy time ($\mathbf{T_{POLICY}}$), and the overall alarm management time ($\mathbf{T_{MANAGEMENT}}$), which is calculated as the sum of $\mathbf{T_{ALARM}}$, $\mathbf{T_{POLICY}}$, and the transmission delay ($\mathbf{T_{TRANSM}}$).

Measured values of $\mathbf{T_{ALARM}}$ as well as $\mathbf{T_{POLICY}}$ do not vary significantly across alarm level. $\mathbf{T_{POLICY}}$ measured values are higher than $\mathbf{T_{ALARM}}$ ones as the major part of the processing load is leaning to the PatientCM side, according to the architecture of the Context Managers.

Measured values of the overall $\mathbf{T_{MANAGEMENT}}$ vary significantly according to network traffic conditions which directly influence $\mathbf{T_{TRANSM}}$ values. In this experiment, we can estimate an overall alarm time swinging between 8 and 10 seconds, with a transmission delay that can be evaluated in about 5 seconds. The higher values of the standard deviation for alarm times can be charged on interferences in the PatientCM wireless connection and on the network conditions in the restricted observation window (five working days) considered for this evaluation.

We expect that a more performant connection would significantly improve the Alarm Management Time, and a wider observation interval (i.e., one month) would produce more uniform statistics for this indicator.

In order to minimize reasoning tasks execution time (inference on ontology instances), we perform reasoning on small ontologies populated by a small number of instances, since ontological reasoning execution times grow at least linearly with respect to the ontology size [26]. This time could be further optimized by performing part of the reasoning task before the service request. Future research activities will thus focus on analyzing which reasoning tasks could be performed prior to service requests and consequently on reengineering the PatientCM and CentralCM.

5.2. User trials

As ERMHAN services have been designed to support care networks and thus to address requirements of care providers, testing is primarily to be focused on evaluating health operators' acceptance of implemented features.

A trial has been performed in a nursing home in Piacenza for system evaluation by professional caregivers. Further trials are planned in a nursing home in Florence in the near future, for more extensive evaluations by chronic condition patients. For these future tests, a sensing system for vital signs and environmental monitoring will also be deployed. Testing activities will thus focus also on evaluating patient acceptance with regard to deployed services (i.e., monitoring, medicine reminder, help, and communication request services).

In the testing stage already conducted in Piacenza, biomedical and environmental sensing was simulated by a web application. Biomedical parameters that were represented in the model included heart rate frequency, pulse

oxymetry, systolic and diastolic blood pressure, body temperature, and glycemia. The web application provides services for manual input or predefined scenario simulation for context data acquisition.

During each session, practitioners were equipped with mobile devices and PCs. The project staff took care of simulating the acquisition of biomedical and environmental parameter values and the occurrence of alarms through the dedicated web interface; health operators were asked to react to such events by using the ERMHAN system as they were in the "real world," such as ignoring alerts or confirming the intervention request reception and taking charge of the case. Practitioners were also asked to perform day-by-day operations through the system, such as accessing and updating patient records, modifying the scheduling of biomedical parameters acquisition and related alarm thresholds. At the end of these trials, the testers' opinions about system features have been collected through interviews and questionnaires.

This first trial session was conducted with 11 test users (including both general practitioners and nurses). As a consequence, we are far from having statistically significant data available, but we are able to illustrate some preliminary results which can be drawn from questionnaire responses. A large majority of users were quite satisfied by the system's overall features (more than 60%). The capability of accessing patient health records has been judged useful and easy to use (but some users had already tested similar features in other experimental systems). The alarm management service and the capability of specifying availability for intervention were judged especially useful (42% of users) and useful (50% of users). Most users appreciated the capability of remotely modifying the alarm thresholds and data acquisition scheduling on a per-patient basis (18% expressed high appreciation, 73% good appreciation, 9% were neutral). A group of users (30%) complained about some misalignment between MHSM patient record presentation and paper-based records in use in their nursing home (especially in terms of use, information organization and classification). This aspect will be further investigated in future testing activities and properly analyzed when defining a methodology for customized deployment if ERMHAN is applied industrially.

6. CONCLUSIONS

In this paper, we have presented ERMHAN, a context-aware mobile service platform supporting mobile caregivers in their daily activities. ERMHAN has demonstrated its capability of providing an extensible set of services aiming at supporting care networks in cooperating and sharing information for the goal of improving a chronic patient's quality of life.

ERMHAN has been designed as a modular system, and its components have been implemented by adopting standard technologies (e.g., Internet protocols, XML, Web services). This approach makes the system easily extensible to match specific patient requirements within an ambient intelligence environment.

Future work will concentrate on security and dependability issues as well as on extending the features provided by the service platform. In the home domain, this will mainly consist in deploying body and home sensor networks, and integrating input/output devices for patients (e.g., alarm button and TV displays). As for the care centre domain, the main developments will include designing graphical user interfaces for system configuration and customization (e.g., customization of predefined alarm management policies and patient case sheets for health centers providing specialized services); adopting existing standards (e.g., HL7) for assuring interoperability with hospital back-end systems.

Leveraging on the ERMHAN modular design, future work will also deal in analyzing and integrating existing user interfaces and applications designed according to HCI (Human Computer Interaction) principles, especially interfaces designed for people with cognitive [27] and physical impairments [28].

Such advancements, together with more extensive testing activities, including patients' evaluation of system features, are needed for a final assessment of the proposed platform.

ACKNOWLEDGMENTS

The KAMER Project is funded by the Regione Emilia Romagna, Italy. The partners are Hewlett Packard Italy, Cattolica University of Piacenza, and the Italian National Inter-University Consortium for Telecommunications (CNIT). Technical assistance from Luca Capannesi, Department of Electronics and Telecommunications of the University of Florence, is gratefully acknowledged.

REFERENCES

[1] A. Börsch-Supan and H. Jürges, "The Survey of Health, Ageing and Retirement in Europe-Methodology," Mannheim Research Institute for the Economics of Aging, MEA, Mannheim, 2005, http://www.share-project.org/new_sites/Documentation/TheSurvey.pdf.

[2] J. E. Epping-Jordan, S. D. Pruitt, R. Bengoa, and E. H. Wagner, "Improving the quality of health care for chronic conditions," *Quality & Safety in Health Care*, vol. 13, no. 4, pp. 299–305, 2004.

[3] World Health Organization (WHO), "Innovative Care for Chronic Conditions: Building Blocks for Action," 2002, http://www.who.int.

[4] S. Consolvo, P. Roessler, B. E. Shelton, A. LaMarca, B. Schilit, and S. Bly, "Technology for care networks of elders," *IEEE Pervasive Computing*, vol. 3, no. 2, pp. 22–29, 2004.

[5] J. Nehmer, M. Becker, A. Karshmer, and R. Lamm, "Living assistance systems: an ambient intelligence approach," in *Proceedings of the 28th International Conference on Software Engineering (ICSE '06)*, pp. 43–50, Shanghai, China, May 2006.

[6] G. D. Abowd, I. Bobick, I. Essa, E. Mynatt, and W. Rogers, "The aware home: developing technologies for successful aging," in *Proceedings of the 18th National Conference on Artificial Intelligence (AAAI '02)*, Edmonton, Canada, July-August 2002.

[7] I. Korhonen, P. Paavilainen, and A. Särelä, "Application of ubiquitous computing technologies for support of

independent living of the elderly in real life settings," in *Proceedings of the 2nd International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications (UbiHealth '03)*, Seattle, Wash, USA, October 2003.

[8] T. Yamazaki, "The ubiquitous home," *International Journal of Smart Home*, vol. 1, no. 1, pp. 17–22, 2007.

[9] A. Särelä, I. Korhonen, J. Lotjonen, M. Sola, and M. Myllymaki, "IST Vivago—an intelligent social and remote wellness monitoring system for the elderly," in *Proceedings of the 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine*, pp. 362–365, Birmingham, UK, April 2003.

[10] A. J. Bharucha, "CareMedia: automated video and sensor analysis for geriatric care," in *Proceedings of the 19th Annual Meeting of the American Association for Geriatric Psychiatry*, San Juan, Puerto Rico, USA, March 2006.

[11] J. E. Bardram, "Applications of context-aware computing in hospital work: examples and design principles," in *Proceedings of the Annual ACM Symposium on Applied Computing (SAC '04)*, pp. 1574–1579, Nicosia, Cyprus, March 2004.

[12] M. A. Muñoz, M. Rodríguez, J. Favela, A. I. Martinez-Garcia, and V. M. González, "Context-aware mobile communication in hospitals," *Computer*, vol. 36, no. 9, pp. 38–46, 2003.

[13] G. Roussos and A. Marsh, "A blueprint for pervasive self-care infrastructures," in *Proceedings of the 4th Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom '06)*, pp. 479–484, Pisa, Italy, March 2006.

[14] P. de Toledo, S. Jiménez, F. del Pozo, J. Roca, A. Alonso, and C. Hernandez, "Telemedicine experience for chronic care in COPD," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 3, pp. 567–573, 2006.

[15] D. G. Katehakis, S. G. Sfakianakis, G. Kavlentakis, D. N. Anthoulakis, and M. Tsiknakis, "Delivering a lifelong integrated electronic health record based on a service oriented architecture," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 6, pp. 639–650, 2007.

[16] M.-J. Huang, M.-Y. Chen, and S.-C. Lee, "Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis," *Expert Systems with Applications*, vol. 32, no. 3, pp. 856–867, 2007.

[17] A. K. Dey and G. D. Abowd, "Towards a better understanding of context and context-awareness," in *Proceedings of the CHI Workshop on What, Who, Where, When, and How of Context-Awareness (CHI '00)*, The Hague, The Netherlands, April 2000.

[18] Web Ontology Language, W3C recommendation, March 2005, http://www.w3.org/TR/owl-guide.

[19] F. Paganelli, G. Bianchi, and D. Giuli, "A context model for context-aware system design towards the ambient intelligence vision: experiences in the eTourism domain," in *Proceedings of the 9th ERCIM Workshop on User Interfaces for All*, vol. 4397, pp. 173–191, Königswinter, Germany, September 2006.

[20] F. Paganelli and D. Giuli, "An ontology-based context model for home health monitoring and alerting in chronic patient care networks," in *Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW '07)*, vol. 2, pp. 838–845, Niagara Falls, Canada, May 2007.

[21] F. Paganelli, E. Spinicci, A. Mamelli, R. Bernazzani, and P. Barone, "ERMHAN: a multi-channel context-aware platform to support mobile caregivers in continuous care networks," in *Proceedings of the 2nd IEEE International Workshop on Services Integration in Pervasive Environments*, pp. 355–360, Instanbul, Turkey, July 2007.

[22] Model-View-Controller, May 2006, http://java.sun.com/blueprints/patterns/MVC.html.

[23] "JENA: A Semantic Web Framework for Java," February 2005, http://jena.sourceforge.net.

[24] SOAP Version 1.2, March 2006, http://www.w3.org/TR/soap.

[25] E. Gamma, R. Helm, R. Johnson, and J. Vissides, *Design Patterns: Elements of Reusable Object-Oriented Software*, Addison-Wesley Professional, Reading, Mass, USA, 1994.

[26] A. Agostini, C. Bettini, and D. Riboni, "A performance evaluation of ontology-based context reasoning," in *Proceedings of the 5th Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom '07)*, pp. 3–8, White Plains, NY, USA, March 2007.

[27] B. Boussemart and S. Giroux, "Tangible user interfaces for cognitive assistance," in *Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW '07)*, vol. 2, pp. 852–857, Niagara Falls, Canada, May 2007.

[28] K. Z. Gajos, J. J. Long, and D. S. Weld, "Automatically generating custom user interfaces for users with physical disabilities," in *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '06)*, pp. 243–244, ACM Press, Portland, Ore, USA, October 2006.

Hindawi Publishing Corporation
International Journal of Telemedicine and Applications
Volume 2008, Article ID 678302, 5 pages
doi:10.1155/2008/678302

Research Article

Building Application-Related Patient Identifiers: What Solution for a European Country?

Catherine Quantin,[1] François-André Allaert,[2] Paul Avillach,[3,4] Maniane Fassa,[1] Benoît Riandey,[5] Gilles Trouessin,[6] and Olivier Cohen[7]

[1] *Service de Biostatistique et Informatique Médicale, CHU de Dijon, INSERM EMI 0106, 21079 Dijon Cedex, France*
[2] *Department of Epidemiology and Biostatistics, Mc Gill University, Montreal, Canada QC H3G1Y6*
[3] *Laboratoire d'Epidémiologie, Statistique et Informatique Médicales (LESIM), Université Victor Segalen Bordeaux 2, 146 rue Léo-Saignat, 33076 Bordeaux Cedex, France*
[4] *Laboratoire d'Enseignement et de Recherche sur le Traitement de l'Information Médicale, Faculté de Médecine, Université de la Méditerranée Marseille, 13284 Marseille Cedex 07, France*
[5] *Institut National d'Etudes Démographiques (INED), 133 Boulevard Davout 75980 Paris Cedex 20, France*
[6] *OPPIDA Sud, Batiment F 78 allée Jean Jaurès, 31000 Toulouse, France*
[7] *HC Forum, Les Jardins de Maupertuis, 7 Chemin de la Dhuy, 38240 Meylan, France*

Correspondence should be addressed to Catherine Quantin, catherine.quantin@chu-dijon.fr

Received 28 September 2007; Accepted 14 February 2008

Recommended by Christine Verdier

We propose a method utilizing a derived social security number with the same reliability as the social security number. We show the anonymity techniques classically based on unidirectional hash functions (such as the secure hash algorithm (SHA-2) function that can guarantee the security, quality, and reliability of information if these techniques are applied to the Social Security Number). Hashing produces a strictly anonymous code that is always the same for a given individual, and thus enables patient data to be linked. Different solutions are developed and proposed in this article. Hashing the social security number will make it possible to link the information in the personal medical file to other national health information sources with the aim of completing or validating the personal medical record or conducting epidemiological and clinical research. This data linkage would meet the anonymous data requirements of the European directive on data protection.

1. INTRODUCTION

In the majority of industrialized countries, patient identification lies at the heart of many of the concerns relating to electronic processing of health information. In August 2004, the French government decided by law to start a national project for an electronic health record called the "dossier médical personnel," the personal medical record (PMR) [1]. It is designed to promote health care coordination, enhance communication of health information, and reduce iatrogenic accidents. The information system corresponding to this project is still under construction. Regarding the patient identifier, one of the proposals has been to use the existing social security number (SSN).

However, civil society and bodies representing citizens (patients' associations, those defending individual liberties, national councils of medical associations) are worried, quite rightly, about the security of medical information [2] should the SSN be used as an identifier in the field of health. They are afraid that medical information may be linked with other information (social, economic, financial, employment) often identified by the SSN. This is one of the reasons why the government decided to postpone this project (November 2007) [3]. Nevertheless, in December 2007, a working group was mandated by the government to propose, as soon as possible, a solution for the HIN, which will be used not only for the DMP project but also for health in general.

Through this paper, we first propose a method which aims both to reassure citizens' representatives regarding the security of medical information and to give the opportunity to the French government to utilize a derived SSN with the

same reliability as the SSN. A solution based on the utilization of anonymity techniques is proposed.

Secondly, we analyse the conditions necessary to make this first solution interoperable at the European level.

2. BACKGROUND

In an increasing number of countries, patients have direct access to their medical records (MR), and some countries including France [1] have decided to provide a personal medical record (PMR) to the patient. We define MR [4] as patients' medical information recorded by the medical practitioner under his or her own responsibility and ideally electronically signed by him/her in order to authenticate the provider (health professional) and to prevent any modification of its content.

In contrast, the PMR is personally supervised by each patient, who has the right to mask any information he/she does not want to be read. As the owner of this PMR, the patient determines who can gain access to his/her record. Of course, patients are very worried that information concerning their private lives may be disclosed. Such concerns are increasing as more sensitive medical details, such as psychiatric records, HIV status, and genetic information, are stored in their PMRs. In France, to enhance the coordination of care, the PMR's data will be stored in national shelters. So patients need to be sure that medical information may not be linked with other information (social, economic, financial, employment) also identified by the SSN, if the SSN was used as an identifier for the PMR.

Many other countries have chosen a homogenous national patient identification system (e.g., Northern European countries such as Denmark, Finland, Luxembourg, the Netherlands, Belgium, the United Kingdom, and Ireland) [5], and the former countries in this list are even using the same identification number for other fields than health care. Apart from Belgium, the above-mentioned countries implemented a unique patient identifier some time ago, and the citizens are used to it and to the creation of national health databases. However, in the United Kingdom, patients are reassured because there is no national shelterer and medical records are stored by the general practitioner who cannot transmit or receive any information regarding a patient without his/her consent.

In Southern European countries identification is often organized at a regional level (Spain and Italy) and patients are more familiar with the creation of large health databases organized at a federal level as in other countries (Canada and the USA).

To reassure French patients regarding the security of their medical data which will be stored at a national level, we propose in this paper the creation of a secure patient identifier which should be different from the SSN to avoid linkage with other data, but which should be as reliable as the SSN. This proposed solution could be used by any country needs to implement a unique health identification number.

3. MATERIALS AND METHODS

3.1. A solution to the security problems

As we said before, it is perfectly possible to preserve the confidentiality expected by the patient by putting in place anonymity procedures [6, 7] such as those adopted by the Institut de Veille Sanitaire (the French Institute for Public Health Surveillance) [8] on the recommendation of the French National Commission for Data protection, in the context of followup procedures for the 30 diseases subject to mandatory reporting (including AIDS).

Unlike encryption methods that must be reversible to allow the legitimate recipient to decode the message, unidirectional hashing techniques, such as the standard hash algorithm in its modified version SHA-2 (SHA-2 is considered "significantly stronger" than SHA-1, although somewhat slower: NSA and NATO recommend it in the SUITE-B package (ECC, AES and SHA-2.)), are irreversible. Hashing produces a strictly anonymous code (it is not possible to retrace the patient identity) that is always the same for a given individual and thus ensures that patient data can be linked. There are many medical applications including the creation of national databases (such as those relating to the national followup of infected persons—approximately 100 000 patients—which are an excellent example of how it can help epidemiological research [9], with complete patient approval [10]) as well as regional and interregional databases in many areas (cancer, perinatal diseases, genetic diseases). This system, which is based on the hash coding of the social security number [11], the gender and the date of birth, has also made it possible to link all standardized hospital discharge abstracts, classified into French diagnosis-related groups at the French national level, and to link the data of the national medical insurance information system. An anonymity procedure based on hash coding is also used to chain patient files in Switzerland [12]. Similar solutions [13], also derived from the irreversible encryption of the unique social security number, have been proposed in Belgium [14] and New Zealand.

In the case of the PMR, the situation is no more complex because similar requirements must be met.

(i) The French Commission for data protection, and associations of patients and healthcare professionals demand the confidentiality of personal information contained in the PMR.

(ii) Public health bodies or individuals need to have access to these data, particularly when the patient has given express consent.

Ideally, hashing the social security number would help to meet these requirements. Regarding confidentiality, insofar as the social security number (SSN) could not be reconstituted using the health identifier number (HIN), the link between them would be broken. Another advantage of using hash coding is that it meets the criterion of being focused (created and maintained solely as a support for health care). This is one of the main criteria of a universal health care identifier that were published by the American Society for

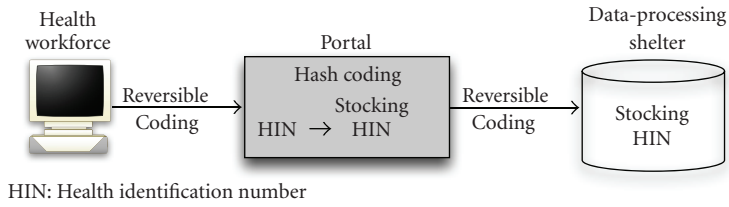


Figure 1: Schema of the proposed solution.

Testing and Materials, a standards development organization accredited by the American National Standards Institute, in the standard guide for properties of a universal healthcare identifier [15].

This solution satisfies the demands of French Commission for data protection in that the SSN would be rendered anonymous by a trustworthy third party and would thus permit the generation of a health identifier number (HIN) known by the patient. This supposes that the HIN would be included in the patient smart card, used for administrative and financial purposes (reimbursement of medical care).

For French patients associations, the interest and importance of printing the health identifier number (HIN) on the patient smart card lie in the fact that it would mobilize and sensitize patients, and increase their awareness that the identifier is different from the SSN. In fact, if the SSN was rendered anonymous using a hash algorithm, there would be no link between the new number and the personal details of the individual, which is not the case for the SSN. Nevertheless, it seems to us that this solution is an illusion in terms of security and reliability because the correspondence between the SSN and HIN would be set out on the patient smart card and would therefore also be known by the health workforce: the SSN is needed in the smart card for reimbursement purposes.

With regard to methodology, to provide anonymity safely requires double hashing, because anonymity is not guaranteed by a single hashing process [6, 7]; two hashing keys must be used to obtain complete anonymity. As a consequence, this solution must be completed by the following principles. According to Figure 1, the health professional can provide to the portal of the application a coding HIN obtained through reversible encryption. The HIN hashing transformation is carried out by the portal. The HIN will itself be transformed into a stocking HIN in the same way that the SSN was transformed into an HIN, but with a different key (see Figure 1). This portal can then transmit this code to the data-processing shelterer safely and reliably.

If a person manages to get the SSN of a patient and uses the hash function, he has to know the key used by the trustworthy third party to obtain the HIN of the patient. This is very difficult. However, as the SSN and the HIN of the patient will both be included in the smart card of the patient and used by health professionals, there is a risk of a breach of privacy in health structures. That is why we propose to apply a double-hash procedure. The aim of hash coding of the HIN by the Portal (with a more different key than that used for hash coding the SSN) is to obtain an anonymised identifier (stocking HIN) before sending it to the data processing shelter.

It also seems fundamental to us that prerequired conditions be imposed in terms of security and confidentiality [18] regarding access to medical data in health structures. This raises several concerns, the first of which is related to the authentication of health-care professionals. The solution would be a smart card attributed to professionals in both the private sector and public hospitals. In France, the use of such cards was imposed in May 2007. Our second concern is the archiving of health data by health professionals. It is not only about controlling the right of access to different applications, but also about ensuring the security of local hard disk storage, preventing direct access to the databases, and guaranteeing the security of external archives (Band and CD-ROM). Lastly, we are also very concerned about the security of exchanges among health professionals. We must point out that many health-care professionals require access to medical records from home, for example, or when on call. A similar situation arises when records are forwarded to doctors during vacations through the internet. We also know that some data processing specialists or computer science experts need access to health structure data for application maintenance. We then strongly recommend the use of networks like virtual private network, in addition to the professional authentication card, as a solution to the problem of the security of data exchange.

3.2. Conditions for making this solution interoperable

Mobility must not compromise healthcare for European nationals. Residents of one member state travelling to or working in another member state should have the same right to high-quality healthcare as all other Europeans (see, e.g., Regulation (EC) No 883/2004). It is thus very important to ensure interoperability between the patients identifiers in the different European countries. However, in Europe, each country has its own patient identifier with very different structures and contents, and which may not always be derived from an SSN. As a consequence, the use of the SSN alone (or a derived SSN) to generate the HIN in European countries would not solve the interoperability problem.

One solution would be to add personal patient characteristics such as family name, first name, date of birth (separately hashed) to be able to link data even in the case of field errors or other local errors to the SSN, which would help to comply with the recommendations of the International

Association of Medical Information Technology, and where possible to ensure interoperability of this identifier with a European identifier [16].

We propose to add to the European health card, as well as the national social insurance number of the patients in each country, a family-based identifier which could contribute to the harmonisation of patients' identification at the European level. This solution would lead to the creation of a European health identifier which would allow patient data to be gathered anywhere in Europe, whatever their location, even if their social insurance number changes according to their country of residence. This will be useful, at the individual level, to provide higher quality health care due to a better followup of the patient and to facilitate the reimbursement of health care costs. At the community level, this will increase the reliability of public health statistics. The great advantage of this European health identifier based on the family component is that it is built from very basic information, which is available to everybody, easily checkable and permanent throughout the patient life. This is not the case for the social insurance number. Our proposal to use the family-based identifier to create a unique European identifier will make it possible to link the data of a patient even when he or she resides outside his or her home country in Europe. It will also contribute to the establishment of European public health statistics by matching healthcare data of the patients records with other administrative data (mortality, social information, etc.) after anonymisation of these data in accordance with the European directive on data protection.

We could also propose to use biometric technologies (and then apply hash coding to ensure anonymity). Biometric technologies are sometimes proposed as a way to associate a patient with his or her medical data, as they do not require the patient to bring any documents or remember any information. Though this technology represents real progress both in the identification and in the authentication of the patient, there are still many questions [17] regarding the accuracy and reliability of each biometric technology and the associated costs. But the main problem lies in the acceptability of such systems by organisations concerned with ethical considerations such as patients' associations, national ethics committees, human rights associations, and national committees for data protection. For example, in France, the use of biometric solutions for identification in the field of health has not been approved by the National Ethics Committee.

4. DISCUSSION

Initially, the government planned to use the SSN as an HIN, as health professionals already know the SSN for health reimbursement reasons. Patients' associations, those defending individual liberties, national councils of medical associations, have asked for an identity number HIN that is different from the SSN, their requirements are the following.

(i) First, that patients are made aware that the HIN used for medical care is not the same as the SSN.

(ii) Second, that there are two different databases for health professionals (located in hospitals or in other structures): a database for health reimbursements which will exclusively use the SSN for administrative reasons, and another medical care database for medical purposes which will not use the SSN because of its private aspect, but an anonymous identity number HIN different from the SSN.

We had to comply with these requirements and are thus obliged to provide tables to show the link between the SSN and the HIN for all patients in all health structures for administrative purposes, which can be considered as the most important disadvantage of the proposed solution.

Patients' associations and national councils of medical associations have been aware of the need for a correspondence table between administrative databases (for health reimbursements) and personal databases (for medical purposes). This correspondence table was not a real problem for them as such tables currently exist in hospitals because of the need to link medical personal data (hospital discharge abstracts referenced by the first and second name, date of birth, etc.) with billing data referenced by the SSN. This particularly arises from the fact that administrators need to have access to medical information in order to determine the levels of hospital activity and justify hospital budgets. The solution for the existing correspondence table between administrative databases (for health reimbursements) and personal databases (for medical purposes) was to put it under the responsibility of the medical information department. That is why the associations concerned consider that using an HIN obtained through irreversible transformation of the SSN provides more security than the current situation where the medical data are stored with personal characteristics (first and second name, date of birth, etc.).

We consider that our proposal must be accompanied by the reinforcement of measures regarding the security of medical information inside hospitals, as proposed at the end of the Section 3. In particular, we focused on the need to code (reversible encryption) both the HIN and SSN in hospitals when they are related to health-care databases.

As a consequence, the main concern regarding the personal medical record (PMR) and centralised storage of patient medical information is the security of data, not only at the level of the health structures, but also at the data-processing shelterer. Knowing that a correspondence table between the SSN and the first HIN would exist in health structures, we propose transforming the first HIN into a storing HIN by using an irreversible encryption method to secure storage (data-processing shelterer). As only the manager of the portal will know the key of this second hashing, no dictionary attacks can be used to try to obtain the first HIN, and thus the corresponding SSN through the correspondence table (if hashed at the health structure level).

Concerning the proposed solution to security problems, this method has several advantages. First, in terms of organisation, it avoids implementation of the hashing function at the level of the health professional. Second, it also avoids supplementary costs, implementation delays, and the disclosure of the technical hashing system to other actors.

The second advantage is that it provides the possibility to use different hash-coding keys to generate a different HIN for each major application such as the PMR, regional, and/or national applications such as health care networks, health administration, diagnosis and healthcare, and epidemiological research without delivering the corresponding application keys to all of the health professionals. Moreover, hashing responsibility and key management can be delegated to each application.

If the concept of specific HINs for different applications is accepted, it will be necessary to create a structure similar to the Data Matching Agency created in Australia to manage all of the identifiers. This agency could be in charge of matching with regard to legal, deontological, organisational, and technical aspects, since the trustworthiness of such an agency is based on a complex combination of legal obligations (in order to provide justifiable legal reliability) and organisational-and-technical security measures (in order to provide justifiable robustness with regard to accuracy of data).

Management of HINs consists in not only ensuring the linkage of the different databases, but also guaranteeing the security of the different identifiers. For example, the real risks raised by using the same identifier in regional and national applications have to be considered: it is important to preclude the possibility of unauthorized linkage by indirect means. This agency could also manage the distribution of encryption keys among the different applications and actors. To ensure that legislation is respected and the interests of the patients protected, the agency would liaise with national commissions for data protection.

5. CONCLUSION

Our proposal for a French health identification number would make it possible to uniquely identify and link a patient to his or her specific medical data. Hashing the social security number will allow linkage between the information of the personal medical file and other national sources of health information with the aim of completing or validating the personal medical record or performing epidemiological research.

Adding personal patient characteristics such as first name and family name, date of birth and/or biometric identifiers (separately hashed, then merged) to the hashed social security number could also be proposed in the discussions about the creation of a European health care identifier. It would thus contribute to the establishment of European public health statistics by matching healthcare data of the patients records with other administrative data. These data linkage systems would meet the requirements for anonymous data issued in the European directive on data protection.

REFERENCES

[1] Dossier Médical Personnel, September 2007, http://www.d-m-p.org.

[2] A. Iversen, K. Liddell, N. Fear, M. Hotopf, and S. Wessely, "Consent, confidentiality, and the data protection act," *British Medical Journal*, vol. 332, no. 7534, pp. 165–169, 2006.

[3] http://lesrapports.ladocumentationfrancaise.fr/BRP/074000713/0000.pdf.

[4] C. Quantin, F. A. Allaert, M. Fassa, B. Riandey, M. Fieschi, and O. Cohen, "How to manage a secure direct access of European patients to their computerised medical record and personal medical record?" *Technology and Informatics*, vol. 127, pp. 246–255, 2007.

[5] C. Quantin, O. Cohen, B. Riandey, and F.-A. Allaert, "Unique patient concept: a key choice for European epidemiology," *International Journal of Medical Informatics*, vol. 76, no. 5-6, pp. 419–426, 2007.

[6] C. Quantin, H. Bouzelat, F. A. Allaert, A. M. Benhamiche, J. Faivre, and L. Dusserre, "Automatic record hash coding and linkage for epidemiological follow-up data confidentiality," *Methods of Information in Medicine*, vol. 37, no. 3, pp. 271–277, 1998.

[7] C. Quantin, H. Bouzelat, F. A. A. Allaert, A. M. Benhamiche, J. Faivre, and L. Dusserre, "How to ensure data security of an epidemiological follow-up: quality assessment of an anonymous record linkage procedure," *International Journal of Medical Informatics*, vol. 49, no. 1, pp. 117–122, 1998.

[8] "Monitoring system for the 30 diseases subject to mandatory reporting (including AIDS) MDOinfo 01," http://www.invs.sante.fr/publications/2003/mdo_infos/mdo_infos_01.pdf.

[9] C. Quantin, C. Guinot, A. Tursz, J. L. Salomez, C. Rogier, and R. Salamon, "Should epidemiological data be extracted from personal medical files?" *Revue d'Epidémiologie et de Santé Publique*, vol. 54, pp. 177–184, 2006.

[10] C. Verity and A. Nicholl, "Consent, confidentially and the threat to public health surveillance," *British Medical Journal*, vol. 324, no. 7347, pp. 1210–1213, 2002.

[11] G. Trouessin and F. A. Allaert, "FOIN: a nominative information occultation function," *Studies in Health Technology and Informatics*, vol. 43, part A, pp. 196–200, 1997.

[12] F. Borst, F. A. Allaert, and C. Quantin, "The Swiss solution for anonymously chaining patient files," *Medinfo*, vol. 10, part 2, pp. 1239–1241, 2001.

[13] T. Churches and P. Christen, "Some methods for blindfolded record linkage," *BMC Medical Informatics and Decision Making*, vol. 4, article 9, 2004.

[14] F. H. Roger France, E. De Clercq, and M. Bangels, "Purposes of health identification cards in Belgium," in *Connecting Medical Informatics and Bio-Informatics: Proceedings of the 19th International Congress of the European Federation for Medical Informatics*, vol. 116, pp. 415–420, Amsterdam, The Netherlands, IOS Press, 2005.

[15] American Standards for Testing Materials, Standard Guide for Properties of a Universal Healthcare Identifier (UHID), Designation: E1714-95, vol. 14.01, June 2000.

[16] C. Quantin, F. A. Allaert, B. Gouyon, and O. Cohen, "Proposal for the creation of a European healthcare identifier," *Studies in Health Technology and Informatics*, vol. 116, pp. 949–954, 2005.

[17] M. Vaclav and R. Zdenek, "Biometric authentication systems," Tech. Rep. FIMU-RS-2000-08, Faculty of Informatics, Masaryk University, Brno, Czech Republic, April 2006, http://www.ecom-monitor.com/papers/biometricsTR2000.pdf.

[18] "Decree Number 2007-960 issued on 15 may 2007 concerning the confidentiality of medical information stored on informatics' support or electronic transmission," http://www.admi.net/jo/20070516/SANP0721653D.html.

Hindawi Publishing Corporation
International Journal of Telemedicine and Applications
Volume 2008, Article ID 763534, 9 pages
doi:10.1155/2008/763534

Research Article

A Tamper-Resistant and Portable Healthcare Folder

Nicolas Anciaux,[1] Morgane Berthelot,[2] Laurent Braconnier,[3] Luc Bouganim,[1] Martine De la Blache,[3] Georges Gardarin,[4] Philippe Kesmarszky,[5] Sophie Lartigue,[6] Jean-François Navarre,[3] Philippe Pucheral,[1,4] Jean-Jacques Vandewalle,[7] and Karine Zeitouni[4]

[1] *Institut National de Recherche en Informatique et en Automatique (INRIA), Rocquencourt, 78153 Le Chesnay Cedex, France*
[2] *SANTEOS SA, Tour Manhattan, 5,6 Place de l'Iris, 92926 Paris la Défense Cedex, France*
[3] *Conseil Général des Yvelines, Hôtel du Département, 2 Place André Mignot, 78012 Versailles Cedex, France*
[4] *PRISM Laboratory, University of Versailles, 45 avenue des Etats-Unis, 78035 Versailles Cedex, France*
[5] *Association Locale de Développement Sanitaire (ALDS), 25 avenue des Aulnes, 78250 Meulan, France*
[6] *Coordination Gérontologique Intercommunale du Territoire Est Yvelines (CoGITEY), 6 avenue du Maréchal Franchet d'Esperey, 78004 Versailles, France*
[7] *Gemalto, 6 rue de la Verrerie, 92190 Meudon, France*

Correspondence should be addressed to Luc Bouganim, luc.bouganim@inria.fr

Received 1 October 2007; Revised 21 March 2008; Accepted 27 May 2008

Recommended by Frédérique Laforest

Electronic health record (EHR) projects have been launched in most developed countries to increase the quality of healthcare while decreasing its cost. The benefits provided by centralizing the healthcare information in database systems are unquestionable in terms of information quality, availability, and protection against failure. Yet, patients are reluctant to give to a distant server the control over highly sensitive data (e.g., data revealing a severe or shameful disease). This paper capitalizes on a new hardware portable device, associating the security of a smart card to the storage capacity of a USB key, to give back to the patient the control over his medical data. This paper shows how this device can complement a traditional EHR server to (1) protect and share highly sensitive data among trusted parties and (2) provide a seamless access to the data even in disconnected mode. The proposed architecture is experimented in the context of a medicosocial network providing medical care and social services at home for elderly people.

1. INTRODUCTION

Since the early days of medicine, and before the advent of computers, people have managed healthcare data manually, accumulating drug prescriptions, examination results, and other medical documents, all of which were inscribed on paper and stored in physical folders at home or at the family doctor office. Although archaic by today's standards, this manual information sharing scheme provided the data owner (i.e., the patient) with control over the sharing and usage of his or her information with the advised assistance of his family doctor and under the protection of the Hippocratic Oath. The patient control was not compromized by the digitization of medical documents in a first stage, simply because the information was scattered among several incompatible information systems in hospitals, clinics, and practitioner's offices. No one knows where the information is, how to access it, whether it is complete and accurate, and under which format it has been produced.

During the last decade, several countries launched ambitious electronic healthcare record (EHR) programs with the objective to increase the quality of healthcare while decreasing its cost [1]. For example, the national *Connecting for Health* (http://www.connectingforhealth.nhs.uk) program in the UK, the *National Switch Point* (http://www.nictiz.nl) managed by Nictiz in the Netherlands, or the healthcare system *Infoway* (http://www.infoway-inforoute.ca/en/home/home.aspx) in Canada are all running projects aiming at building a wide scale EHR. In a recent report [2], French Deputy J.-P. Door identified more than 100 EHR running projects worldwide at the scale of a country or region in a recent report. The objective

of centralizing medical information in database systems is manifold (centralization refers to the fact that the data is stored, organized, made available,and controlled by database servers, whatever the computer system infrastructure is): *completeness* (i.e., to make the information complete and up to date), *availability* (to make it accessible through the internet 24 hours/7 days a week), *usability* (to organize the data and make it easily queryable and interpretable), *consistency* (to guarantee integrity constraints and enforce atomicity and isolation of updates), *durability* (to protect the data against failure), and *security* (to protect the data against illegal accesses).

On the other hand, studies in different countries [3, 4], show that several patients and even practitioners are reluctant to use EHR systems arguing increasing threats on individual privacy. This suspicion is fuelled by computer security surveys pointing out the vulnerability of database management systems (DBMSs) against external and internal attacks [5]. Indeed, centralizing and organizing the information make it more valuable, thereby motivating attacks, and facilitates abusive usages. In consequence, EHR providers must comply with very stringent legislation regarding the storage and preservation of medical data. Regardless of the security procedures put in place at the server and their effectiveness, the patient has the sense of losing control over his or her data. There are four main reasons for that.

(i) *Guidance of the patient consent*: the patient is usually asked to give his consent to an access control policy specifying who (individuals or roles) is granted access to which part of his folder. Even with the help of a practitioner, it is difficult to ensure that this consent is fully enlightened. This is due to the high number of people interacting with the folder, the diversity of their roles, the complexity of the medical information, and the intrinsic difficulty to determine which data (or data association) reveals a given pathology. Consequently, the patient usually adheres to a predefined access control policy that he does not really master. Complementary to access control, audit trails can help the patient tracking a posteriori who accessed which part of his folder and when. However, audit trails exploitation is fairly complex and may require a dedicated query language [6]. With respect to the free expression of the patient consent, EHR systems cannot compete with the archaic manual information sharing scheme.

(ii) *Unbounded data retention*: limited data retention is one of the central principles of laws related to the safeguard of personal data [7]. Limited data retention attaches a lifetime to a data (e.g., 10 years for a given court sentence) after which it must be withdrawn from the system [8]. Unfortunately, limited data retention conflicts with the primary objective of an EHR, that is building a complete medical history of each patient. In addition, [9] highlighted the difficulty to physically destroy information stored in existing DBMSs, showing that it can be recovered by a forensic analysis in many ways. This reinforces the patient's perception that his complete history is recorded forever. A side effect is that the patient may choose not to store some information in his folder (synonym of incompleteness and lower quality of healthcare) because he cannot assess a priori the sensitiveness of this information.

(iii) *No security guarantee outside the server domain*: healthcare data is likely to be extracted from the server and hosted in a client device (e.g., the doctor's or the patient's device). Typically, healthcare folders need to be extracted for use in a disconnected mode, for example, to provide healthcare at home. This situation will remain the rule for a while, that is until every point of the territory be connected through a secure, fast, reliable, and free of charge network. Unfortunately, the hosting device is much more spyware, trojan, and virus prone than the server, introducing an important security breach in the architecture.

(iv) *No disconnected access to the folder:* EHR has been designed with online usage in mind. This may constitute a real barrier for a large category of patients (e.g., elderly, disabled, and needy people), the prerequisite to get access to their folder being either to use a terminal at some public place or to own a PC, to master its use (including the computer administration burden) and to pay for an internet connection. If these latter conditions are not satisfied, a practitioner providing healthcare at home will have to download on his mobile device the folders of all visited patients, a complex and time-consuming task, beside the security breach mentioned above.

In this paper, we propose a novel organization of EHR aiming at circumventing these four drawbacks. This organization capitalizes on a new hardware device called *secure portable token* (*SPT*). Roughly speaking, an SPT combines a secure microcontroller (similar to a smart card chip) with a large external flash memory (gigabyte sized) on a USB key form factor [10, 11]. An SPT can host onboard data and run onboard code with proven security properties thanks to its tamper-resistant hardware and a certified operating system [11]. Embedding a database system and a web server in an SPT gives the opportunity to manage a healthcare folder (or a part of it) outside the EHR server with no loss of security. Accessing the onboard folder while being disconnected from the network requires a simple rendering device equipped with a USB port and running a web browser. More, the embedded DBMS can be made self-administered so that the patient keeps a full control over the onboard data, with no external intervention of a database administrator. The data retention period and the sharing of onboard data can be organized similarly to the previous manual scheme, under the patient control. The resulting architecture is decentralized, with a central server managed by a public or private database service provider and one embedded local server per patient. Splitting the folder in a centralized part and a local part remains under the patient's responsibility.

However, the local server cannot provide properties like availability and durability on the local data on its own. We will show that combining the capacities of the central server and the local servers restores these fundamental properties.

Our project is not the first to promote the use of secure tokens. A growing number of initiatives are using smart cards (e.g., the French *Sesam Vitale* [12], the *Patient Health Smart Card* in New York [13], the *National Health Card System,* (24 millions of smart cards, see http://www.gi-de.com/portal/page?_pageid=42,55000&_dad=portal&_schema=PORTAL.), in Taiwan, etc.) to carry patient's national security number and practitioner's certificate in order to implement a strong identification process. Also, the German EHR initiative plans to use smart cards to carry recent prescription information [14]. As our project, the German initiative underlines the interest of holding part of the patient's folder locally. However, this project uses a traditional smart card technology thereby proscribing storing a significant volume of patient's records locally (the smart card is endowed with less than 100 KB of stable memory, (Wiki in German: http://www. de.wikipedia.org/wiki/Elektronische_Gesundheitskarte), versus 256 MB for the preliminary SPT we are currently working on). To the best of our knowledge, our project is the first to tackle the technical challenges related to the management (storage, querying, and secure sharing) of large folders embedded in secure token.

The paper is organized as follows. Section 2 introduces the functional architecture of the proposed EHR system and discusses different scenarios, showing the benefit of combining a central EHR server with secure portable folders. Section 3 sketches important technical challenges and outlines the solutions proposed. Section 4 describes an application of our solution to a medical network providing healthcare at home and Section 5 concludes.

2. FUNCTIONAL ARCHITECTURE AND SCENARIOS

The proposed EHR architecture, pictured in Figure 1, is built around a central DBMS server providing the functionalities mentioned above (completeness, availability, usability, consistency, durability, and security) on all patient's folders. This server offers an internet access to all authorized users (e.g., doctors, nurses, etc.) and enforces access control policies defined by the patients. These policies are based on default policies promoted by EHR groups like the GIP DMP (the GIP DMP is in charge of setting up the French medical folder project called DMP (dossier médical personnel). See http://www.d-m-p.org/docs/EnglishVersionDMP.pdf and http://www.d-m-p.org/) and can be refined with the help of practitioners. To allow disconnected accesses to any patient folder, a folder replica is managed on a secure portable token (SPT) provided to each patient. Once endowed with the appropriate embedded software (typically a web server and a DBMS) and timely synchronized with the central server, SPT servers successfully to surrogate the central server. The access control policy defined by the patient is enforced uniformly on the central DBMS and on the SPT. Note that the synchronization between central and SPT servers must occur in all situations (e.g., even if an SPT remains at patient's home and is never connected to the network). We detail this synchronization process later.

SPT is the means by which a patient can recover a full control over his folder, similarly to the manual sharing scheme. Let us assume that a patient is willing to hide information he judges as highly sensitive (e.g., because it may reveal a serious or shameful disease). In a traditional EHR system, if the patient does not fully trust the server or does not fully understand the access control policy (for the reasons mentioned above), he has no other choice than discarding this information, thereby producing an incomplete folder. Here, the patient can store this information, called *hidden data* (*HD*), locally on his SPT without replicating it on the central server. A data which is not hidden is called *regular data* (*RD*) and is replicated on the SPT and the central server. If the patient changes his mind afterwards (e.g., following the advice of his doctor), he still has the opportunity to change the status of a data from hidden to regular. Note that the reverse conversion (from regular to hidden) is uncertain since the data could have been queried and/or copied beforehand. Hiding data matches the privacy objective but the durability property is lost for this data since a portable token is by nature not durable (it might be lost or destroyed). Durability can be restored by using the central server as a repository for cryptoprotected data (i.e., encrypted and signed). To this end, hidden data is encrypted by the patient's SPT and stored encrypted on the central server but the encryption keys are never revealed to the central server. (Note that encryption techniques are sometimes used by central servers to protect the database footprint [15]. With such server-based encryption solutions, data is encrypted and decrypted on the fly by the DBMS. Server-based encryption is thus orthogonal to the SPT-based encryption applied to hidden data). Encryption keys stay under the SPT control and make themselves durable thanks to a trusted depositary. The central server guarantees the durability and availability of hidden data without being able to interpret its content.

Let us now assume that the patient is willing to share hidden data among a restricted *trusted circle* of persons. The patient can define appropriate access control rules on this data so that it becomes accessible to these people in the presence of the patient and of his SPT (as in a paper-based scheme). The SPT allows an even smarter way of exchanging sensitive information. The patient may grant a trusted circle of participants to access his hidden data even if it is not together at the same physical place. This may be helpful in case of emergency or if a remote diagnosis is required. This can be implemented by sharing the encryption keys of the hidden data among the SPTs of people participating in the same trusted circle so that only these SPTs are able to free the hidden data from its cryptoprotection. Note that those keys are never externalized from the SPT, thus allowing enforcing access control rules locally (those rules are stored encrypted along with the hidden data on the central server). To distinguish them in Figure 1, regular (resp., hidden) data and its related access control rules are pictured in white (resp., grey). As mentioned above, the access control policy

defined by a patient must be enforced uniformly on the central DBMS and on the SPT, independently of the status of the data. With hidden data, the access control policy is strengthened by the obligation of physically sharing the encryption keys, and this sharing is totally under the patient control. Once decrypted, hidden data is still protected by the access control policy enforced by the SPT.

Let us illustrate the behavior of the system through a scenario involving three participants: an elderly patient named Bob, his family doctor Jim, and a nurse Lucy. Every participant owns an SPT. Several medical examinations are prescribed to Bob who designates a subset of them as hidden (the others being considered as regular). The medical lab performing the examination pushes the results on the central server. Results corresponding to hidden data are cryptoprotected using Bob's public key before being pushed. (Bob's public key is delivered by a PKI server while Bob's private key is replicated on every SPT belonging to Bob's trusted circle. The management of private keys is under the control of the secure chip and even the SPT holder cannot interfere or tamper it. For the sake of conciseness, we do not detail further the key exchange protocol among SPT. For efficiency, asymmetric encryption is used only to encrypt symmetric keys used to protect the hidden data).

Lucy frequently visits Bob at home. Bob has no internet connection and seldom leaves home. Thus, Lucy acts as a synchronization means for Bob's folder (as any other person visiting Bob and owning an SPT). Before the visit, Lucy downloads from the central server only the latest updates, either hidden or regular, performed in Bob's folder. This includes the recent examination results. During the visit, Lucy's and Bob's SPTs are synchronized. The latest updates from the central server are integrated in Bob's local folder. Conversely, the latest updates performed in Bob's local folder, if any, are loaded on Lucy's SPT. This allows refreshing the central server replica the next time Lucy will connect to the server. Synchronized data is only delivered to the recipient (i.e., central server or patient's SPT) using a secure protocol (authentication and encryption). Lucy cannot get access to this data, protected by the tamper resistance of the SPT.

Jim participates in Bob's trusted circle. At his office, he can connect to the central server and view Bob's up-to-date folder, including the results of the recent examinations and possible updates carried back by Lucy (in the limit of Jim's access rights granted by Bob). When visiting Bob at home, Jim can get the same level of information by connecting locally to Bob's SPT.

To summarize, any authorized people (or people playing an authorized role) can connect to the central server or to an SPT local server and retrieve the regular data he is granted access to by the access control policy. No people outside the trusted circle can get access to the hidden data, whatever their role(s) and privilege(s). Indeed, this hidden data is cryptoprotected on the central server and the encryption keys are only known by the patient's SPT and by the SPTs of people belonging to her trusted circle. Hidden data stored on the patient's SPT benefits from the SPT's tamper resistance. Encryption keys (symmetric and private keys) are transmitted from the patient's SPT to the SPTs of the trusted circle using a secure protocol (based on symmetric encryption). This transmission may happen either in a connected mode or via another SPT as part of the synchronization described above. Finally, synchronized data (regular, hidden, or encryption keys) is never disclosed to anyone except the recipient and is also protected during transmission by a secure protocol.

This mode of operation provides stronger privacy preservation guarantees than any traditional EHR. First, attacks conducted at the server can only reveal regular data, hidden data being absent from the server or encrypted with keys that are let under the control of the clients' SPTs. Most advanced server-based security solutions, even those using encryption [15], cannot offer such a level of protection because encryption keys remain accessible at the server side to enable query execution. Second, attacks conducted at the client terminal cannot reveal more than the data displayed by the application at runtime, but no data is ever stored on client terminals in the clear. Third, the SPT inherits its security from a tamper-resistant hardware and a certified embedded code (certified CC EAL4 or 5, FIPS or using other relevant scheme [11]). We do not argue that SPT is provably unbreakable because ultimate security does not exist but it makes the attacks so complex and costly to implement as they become meaningless in practice. To make the analysis complete, let us consider anyway the impact of an SPT attack. Breaking a patient's SPT will lead to disclose her medical folder stored locally (hidden and regular data); breaking a doctor's SPT will lead to disclose the encryption keys of the patients having registered this doctor in their trusted circle; finally, breaking an SPT serving for synchronization (typically a nurse's SPT) will not disclose any information.

3. TECHNICAL CHALLENGES

Many technical challenges are introduced by the proposed EHR architecture, like the access control definition and enforcement, the management of encryption keys, the data synchronization between the central server and the embedded local servers, and so forth. Due to space limitation, this section focuses on the challenges related to the SPT and the embedded data management techniques which are central to the architecture.

3.1. SPT hardware and operating system

An SPT aims at combining in the same hardware platform a secure chip and a mass storage NAND flash memory (several gigabytes soon). The secure chip is of the smart card type, with a 32 bit RISC CPU clocked at about 50 MHz, memory modules composed of ROM, tens of KB of static RAM, a small quantity of internal stable storage (NOR flash) and security modules. A first obvious challenge is to produce this hardware platform and assess its tamper resistance and performance. In this platform, the mass storage NAND flash memory is outside the secure chip (connected by a bus) and does not benefit from the chip hardware protection. A second challenge is then to enforce

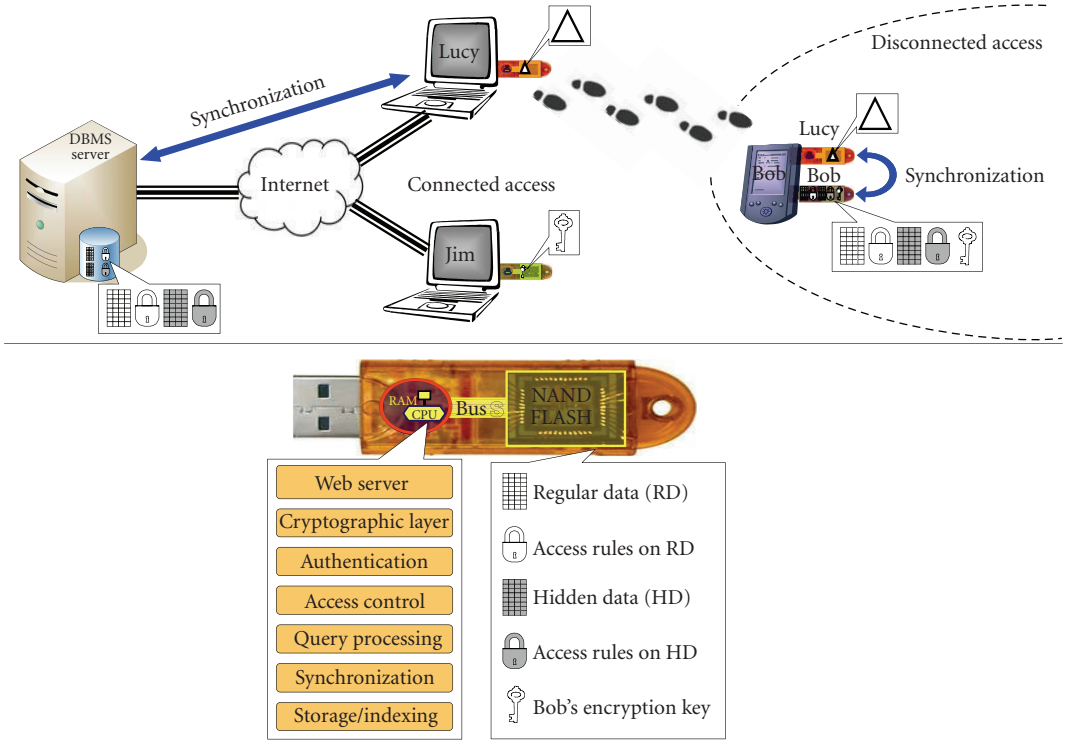


Figure 1: Functional architecture.

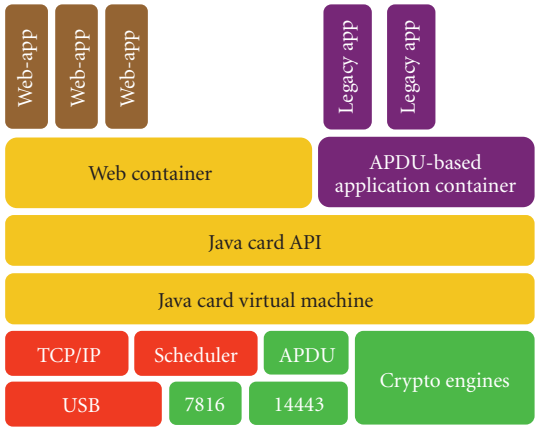


Figure 2: SPT system architecture.

the integrity and the confidentiality of the data stored in mass storage thanks to the cryptographic capability of the secure chip, with a minimal impact on read/write performance. Considering the increasing storage capacity, computation power and communication throughput of this new generation of smart tokens, integrating them in a distributed architecture as regular computers is a third challenge.

Gemalto, the smart card world leader, is developing an experimental SPT platform. This platform includes a new multitasking operating system allowing the development of web applications based on Java and Servlet technology, and thus offering a standardized means to integrate services or embedded web applications to the SPT. The operating system supports natively:

(i) the USB 2.0 protocol and the internet protocol IP for communicating with the external world [16];

(ii) multithreaded Java applications;

(iii) cryptographic primitives (some of which being implemented in hardware);

(iv) memory management and garbage collection;

(v) servlet management and web server.

The internal architecture of the SPT in described in Figure 2. For more technical details, we refer the reader to [17].

3.2. *Embedded database system*

DBMS designers have produced light versions of their systems for personal assistants (e.g., Oracle-lite, DB2 everyplace, SQL Server for Window CE) but they never addressed the more complex problem of embedding a DBMS in a chip. Initial attempts toward a smart card DBMS were ISOL's SQL Java machine [18], the ISO standard SCQL [19], and the MasterCard Open Data Storage [20]. All these proposals concerned traditional smart cards with few resources and therefore proposed basic data management functionalities (close to sequential files). Managing embedded medical folders requires much more powerful storage, indexation, access

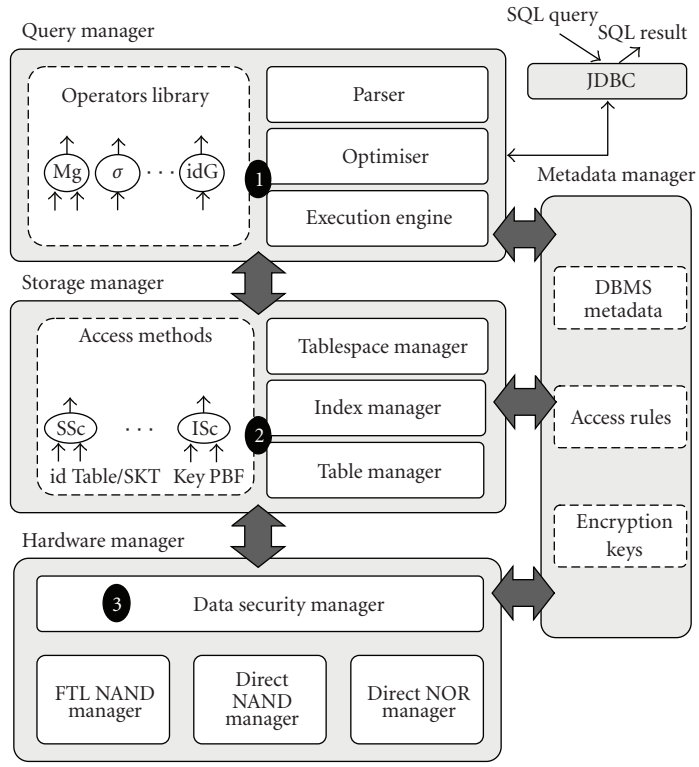


Figure 3: DBMS kernel architecture.

control, and query capabilities. PicoDBMS was the first full-fledged relational DBMS embedded in a smart card [21] and was implemented on top of Gemalto's smart card prototypes [22]. The peculiarities of secure chip environments compel to deeply revisit existing DBMS techniques like storage and indexing models, query execution strategies and transaction management. PicoDBMS has been designed for managing databases stored in a (megabyte sized) EEPROM stable memory integrated in the secure chip and protected by the chip tamper resistance.

The SPT framework introduces important new challenges [23].

(1) How to support complex access rights and queries over a gigabyte-sized onboard database (compared to PicoDBMS, the database size grows by three orders of magnitude while the RAM resource was multiplied roughly by 5)?

(2) How to organize the data storage and the indexes with an acceptable insert/update time considering the peculiarities of NAND flash memory (fast reads, costly writes, block-erase-before-page-rewrite constraint)?

(3) How to protect the onboard database against confidentiality and integrity attacks without degrading the query performance; this challenge is related to the flash protection problem mentioned in Section 3.1 but solutions specific to database management can be devised to provide optimal performance?

Figure 3 depicts the main software modules of the embedded DBMS, tagged with numbers indicating the related technical challenge.

The *Query Manager* is in charge of parsing the incoming SQL query, building an optimal query execution plan, and executing it. This module must consider peculiar execution strategies to answer complex SQL queries over a large quantity of data (gigabytes) while coping with the SPT hardware constraints (challenge 1). To tackle this challenge, we designed a massive indexing scheme presented in [24], which allows processing complex queries while consuming as little RAM as possible and still exhibiting acceptable performances. The idea is to combine in the same indexing model generalized join indices and multitable selection indices in such a way that any combination of selection and join predicates can be evaluated by set operations over lists of sorted tuple identifiers. The operator library (algorithms for the operators of the relational algebra, e.g., select, project, join, and aggregate) and the execution engine integrate those techniques.

The *Storage Manager*, on which the query manager relies to access the database content (index and tables), is directly concerned with challenge 2. Indeed, the proposed massive indexation scheme causes a difficult problem in terms of flash updates, due to the severe read/write constraints of NAND flash. Therefore, we designed a structure which manages data and index keys sequentially so that the number of rewrites in flash is minimized. The use of summarization structures (based on bloom filters [25]) and vertical partitioning reduce

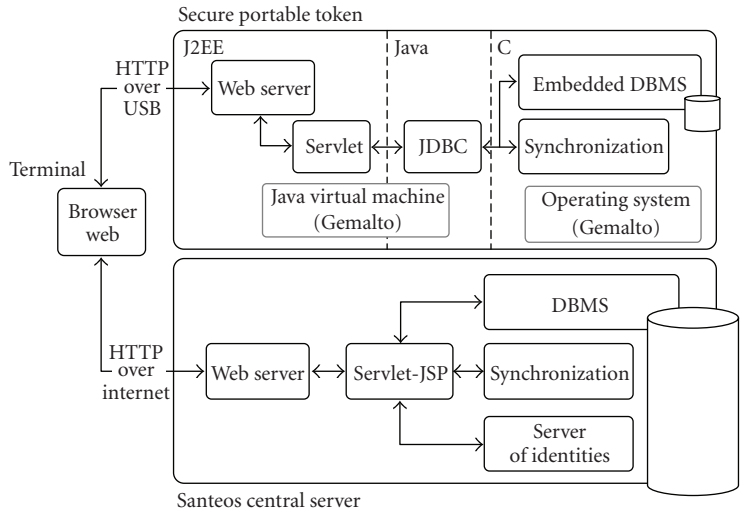


Figure 4: Software platform.

the cost of index lookups. These additional structures are also managed in sequence. A first implementation of this principle has been patented jointly by INRIA and Gemalto [26] and is integrated in the current DBMS prototype.

The *Hardware Manager* embeds the methods for accessing the different memory modules of the SPT (through the Flash Translation Layer (FTL) [27] or with direct access). It includes techniques associated with challenge 3, to protect the confidentiality and the integrity of the data, in an efficient way with respect to DBMS access patterns. Indeed, our massive indexation technique leads to numerous, random, and fine grain accesses to raw data. We conducted preliminary studies [28], in which we combine encryption, hashing, and timestamping techniques with query execution techniques in order to satisfy three conflicting objectives: efficiency, high security, and compliance with the chip hardware resources.

Finally, the *Metadata Manager* manages the DBMS metadata, the access control rules regulating the access to regular and hidden data and the encryption keys.

4. EXPERIMENTAL PLATFORM

4.1. Experiment in the field

The functional architecture presented in Section 2 will be experimented in the context of a medicosocial network providing medical care and social services at home for elderly people in the Yvelines district, France. Today, the coordination among the participants of this network (doctors, nurses, physiotherapists, social workers, etc.) is organized around a paper-based folder. This folder stays at home and is queried and updated by every participant. This solution suffers from two main drawbacks. First, instead of providing a natural and controlled way of sharing medical and social information, the paper-based folder is the primary source of confidentiality breach in this setting. Indeed, this folder must be shared by all participants but there is no means by which an effective access control policy can be implemented. Second, the folder cannot be accessed and filled remotely, precluding remote diagnosis, and is often incomplete due to its physical centralization. Replacing this paper-based folder system by a traditional EHR would introduce the drawbacks mentioned in Section 1. The proposed experimentation will combine a central database with medicosocial folders embedded in SPT, according to the architecture presented in the previous section.

This experimentation will be conducted with a population of about 100 volunteer patients and 25 practitioners and social workers in the Yvelines district in France. It involves the following partners: the French National Research Institute in Computer Sciences (INRIA), University of Versailles, SANTEOS (EHR provider for the prefiguration phase of the French national healthcare folder system), Gemalto (world leader in the smart card domain), ALDS (a home healthcare association), and CoGITEY (a clinic with a section dedicated to elderly people). This project is partly funded by the Yvelines district council and by the French National Agency for Research (ANR). The design phase started in January 2007 and the experimentation in the field will be conducted fall 2009.

4.2. Software platform

In the experiment, the final user (either practitioner or patient) will be able to connect to a server (either central or embedded) with a web browser running on any terminal (fixed or mobile). A web-based interface (GUI) is provided to browse the patient folder. By manipulating the GUI, the user generates HTTP requests to the server, thereby activating Servlets which in turn generate database queries and build the next page of the interface. Whatever the server it is connected to, the GUI provides similar functions, for example, access to patients' folders, authorization management, and so forth. The software platform enabling this behavior is presented in Figure 4.

The central server is equipped with commercial software including a web server (the GUI is generated using Servlets and JavaServer Pages), a relational DBMS (to store, index, and retrieve patients' folders), and a server of identities (to manage certificates and identifiers for medical and social workers and patients).

The SPT embeds a proprietary web server and Servlets communicating with a lightweight DBMS engine via a JDBC-like driver. A synchronization module is also embedded in the SPT to synchronize the embedded folder with the copy stored in the central server. Thus, the software deployed on the central server and the SPT provide similar functionalities while relying on highly different technology to cope with the SPT hardware constraints. For more information about the software platform, we refer the reader to [17].

5. CONCLUSION

In this paper, we presented an alternative to centralized EHR systems, relying on a new hardware portable device called *secure portable token* (*SPT*). This architecture is being implemented in the scope of the DMSP and PlugDB projects started in November 2006 and will be experimented in the fall of 2009 in the context of a medical social network providing medical care and social services at home for elderly people.

The objectives pursued are

(i) to build a shared medicosocial folder providing the highest degree of availability, whatever the mode of operation (disconnected or not);

(ii) to reestablish a natural and powerful way of protecting and sharing highly sensitive information among trusted parties.

The expected outcome of this project is to demonstrate that these two objectives can be reached with a positive impact on the coordination of medical and social workers and on the acceptation of patients of an electronic usage of their medical history.

ACKNOWLEDGMENTS

This work is partially funded by the French National Agency for Research (ANR) under the RNTL-PlugDB project and by the Yvelines district council under the DMSP contract.

REFERENCES

[1] S. H. Brown, M. J. Lincoln, P. J. Groen, and R. M. Kolodner, "VistA—U.S. Department of Veterans Affairs national scale HIS," *International Journal of Medical Informatics*, vol. 69, no. 2-3, pp. 135–156, 2003.

[2] J.-P. Door, "Information report (in French), Rapport d'information N°659, Enregistré à la Présidence de l'Assemblée nationale le 29 janvier 2008, déposé par la Commission des Affaires Culturelles, Familiales et Sociales, sur le dossier médical personnel, et présenté par M. le député Jean-Pierre Door," 2008.

[3] Markle Foundation, "National Survey on Electronic Personal Health Records," September 2007, http://www.phrconference.org/assets/research_release_101105.pdf.

[4] IPSOS, "Opinions et attitudes des médecins et des patients à l'égard du Dossier médical partagé, Rapport Ipsos Santé/CNAM," 2003.

[5] Computer Security Institute, "CSI/FBI Computer Crime Survey," 2006, www.gocsi.com.

[6] R. Agrawal, R. J. Bayardo Jr., C. Faloutsos, J. Kiernan, R. Rantzau, and R. Srikant, "Auditing compliance with a hippocratic database," in *Proceedings of the 30th International Conference on Very Large Data Bases (VLDB '04)*, vol. 30, pp. 516–527, Toronto, Canada, August-September 2004.

[7] European Directive 95/46/EC, "Protection of individuals with regard to the processing of personal data and on the free movement of such data," *Official Journal of the European Communities*, no. L. 281, p. 31, November 1995.

[8] N. Anciaux, L. Bouganim, H. van Heerde, P. Pucheral, and P. M. G. Apers, "InstantDB: enforcing timely degradation of sensitive data," in *Proceedings of the 24th International Conference on Data Engineering (ICDE '08)*, Cancun, Mexico, April 2008.

[9] G. Miklau, B. N. Levine, and P. Stahlberg, "Securing history: privacy and accountability in database systems," in *Proceedings of the 3rd Biennial Conference on Innovative Data Systems Research (CIDR '07)*, Asilomar, Calif, USA, January 2007.

[10] InspireD integrated project, www.inspiredproject.com.

[11] Eurosmart, Smart USB Token, White Paper, April 2008, http://www.eurosmart.com/Update/08-05/Eurosmart_Smart_USB_Token_WP_April08.pdf.

[12] J. Smith, "Smart health care—the next generation," *Smart Card News*, vol. 15, no. 12, 2006.

[13] Smart card trends, "Infineon provides security chips for largest US patient smart card project," www.smartcardstrends.com/det_atc.php?idu=5960&PHPSESSID=9cc1e5da7e5535ae326c8210def413fb.

[14] B. Blobel and P. Pharow, "A model driven approach for the German health telematics architectural framework and security infrastructure," *International Journal of Medical Informatics*, vol. 76, no. 2-3, pp. 169–175, 2007.

[15] U. Mattsson, "Secure.Data Functional Overview," Protegity Technical Paper TWP-0011, 2000, http://www.protegrity.com/white-papers.htm.

[16] J.-J. Vandewalle, "Smart card research perspectives," in *Proceedings of the International Workshop on Construction and Analysis of Safe, Secure, and Interoperable Smart Devices (CASSIS '04)*, Lecture Notes in Computer Science, pp. 250–256, Marseille, France, March 2004.

[17] PlugDB and DMSP Projects, www-smis.inria.fr/~DMSP.

[18] L. C. Carrasco, "RDBMS's for Java cards? What a senseless idea !," ISOL corp. white paper, 1999.

[19] International Standardization Organization, "Integrated Circuit(s) Cards with Contacts—part 7: Interindustry Commands for Structured Card Query Language (SCQL)," ISO/IEC 7816-7, 1999.

[20] MasterCard International, MasterCard Open Data Storage. Technical Specifications, Version 2.0, 2002, http://www.mastercardintl.com.

[21] P. Pucheral, L. Bouganim, P. Valduriez, and C. Bobineau, "PicoDBMS: scaling down database techniques for the smartcard," *The International Journal on Very Large Data Bases*, vol. 10, no. 2-3, pp. 120–132, 2001.

[22] N. Anciaux, C. Bobineau, L. Bouganim, P. Pucheral, and P. Valduriez, "PicoDBMS: validation and experience," in

Proceedings of 27th International Conference on Very Large Data Bases (VLDB '01), pp. 709–710, Roma, Italy, September 2001.

[23] N. Anciaux, L. Bouganim, and P. Pucheral, "Future trends in secure chip data management," *IEEE Data Engineering Bulletin*, vol. 30, no. 3, pp. 49–57, 2007.

[24] N. Anciaux, M. Benzine, L. Bouganim, P. Pucheral, and D. Shasha, "GhostDB: querying visible and hidden data without leaks," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 677–688, Beijing, China, June 2007.

[25] B. Bloom, "Space/time tradeoffs in hash coding with allowable errors," *Communications of the ACM*, vol. 13, no. 7, pp. 422–426, 1970.

[26] P. Pucheral and S. Yin, "System and Method of Managing Indexation of Flash Memory," European patent by Gemalto and INRIA no. 07290567.2, May 2007.

[27] Intel Corporation, "Understanding the Flash Translation Layer (FTL) specification," 1998.

[28] N. Anciaux, L. Bouganim, and P. Pucheral, "Data confidentiality: to which extent cryptography and secured hardware can help," *Annals of Telecommunications*, vol. 61, no. 3-4, pp. 267–283, 2006.