Real-Time Management of Multimodal Streaming Data for Monitoring of Epileptic Patients

Dimitrios Triantafyllopoulos, Panagiotis Korvesis, Iosif Mporas[1] and Vasileios Megalooikonomou

Multidimensional Data Analysis and Knowledge Management Laboratory

Dept. of Computer Engineering and Informatics, University of Patras

26500 Rion-Patras, Greece

Abstract: New generation of healthcare is represented by wearable health monitoring systems, which provide real-time monitoring of patient's physiological parameters. It is expected that continuous ambulatory monitoring of vital signals will improve treatment of patients and enable proactive personal health management. In this paper, we present the implementation of a multimodal real-time system for epilepsy management. The proposed methodology is based on a data streaming architecture and efficient management of a big flow of physiological parameters. The performance of this architecture is examined for varying spatial resolution of the recorded data.

Keywords: Multimodal health data; data streaming; online processing.

1. Introduction

As healthcare costs are increasing and the world population is ageing [1], the need for monitoring patients in their home environment is growing. Patients with chronic conditions such as heart failure, dementia, sleep apnea, diabetes or epilepsy, need monitoring for several days. World Health Organization predicts that chronic diseases will become the most expensive problem faced by current health care systems and sees the integration of prevention into healthcare as the main solution for this problem [2]. Information and communication technologies are expected to respond to this problem by providing personalized, low-cost, citizen centered healthcare services [3]. Recent advances in sensor technology and microelectronics have enabled the long term monitoring and management of chronic disease patients and additionally detect urgent or emergent events.

In order to monitor patients in a long term basis several systems have been developed and several products have been produced in the recent years. These systems and products aim at providing real-time feedback information about the patient's health condition [4, 5, 6, 7, 8]. The receivers of this

[1] corresponding author: Dr. Iosif Mporas, email: imporas@upatras.gr

information are the patients themselves, a medical center or the supervising physician of the patient. Alert services are also provided in case of possible imminent health threatening conditions. To achieve these goals, these systems process the data flow continuously and additionally are expected to achieve low latency and high throughput. Specifically, the data processing must keep up with data ingest rates, while providing high quality results as quickly as possible.

Our work focuses on a real time processing health system that addresses the needs of patients with epilepsy. Epilepsy is affecting approximately 1% of the world population and is the third most common neurological disorder in the United States after Alzheimer's disease and cerebrovascular events [9]. Detection of epileptic seizures is dependent upon the capture, analysis, aggregation and interpretation of large volumes of data. By utilizing a real time analysis tool, the patient's condition can be assessed without latency as is the case when manual analysis is performed. The use of continuous applications and stream processing of physiological data flow can improve the delivery of health care along several axes. The fundamental goal of stream processing is to process live data, providing real-time information results to end-users, while monitoring and aggregating information in order to provide medium- and long-term support to decision making.

The proposed system is designed to work with limited resources in a real time manner. Stream analysis over sliding windows is performed, based on a short time analysis scheme. Physiological signals are exploited in real time, while the whole design of the system is patient specific in order to adapt to each patient's characteristics. Abnormal events of higher level, such as status epilepticus, as well as high risk conditions are discovered by a continuous monitoring of the patient's condition.

The remainder of this paper is organized as follows. Section 2 generally discusses previous research. Section 3 introduces the proposed framework and describes generic architecture. Section 4 presents in detail the implementation of the online seizure detection platform. In section 5 the experiments performed for the evaluation of the system are presented, in terms of real time processing performance. Finally, the paper is concluded with a discussion of the experimental results.

2. Related work

During the past years, ambulatory monitoring of physiological parameters by wearable or implantable sensors is a research area with high interest [6, 10, 11]. Wearable health monitoring systems (WHMS) constitute a new means to address the issues of managing and monitoring chronic diseases [12, 13, 14], elderly people, postoperative rehabilitation patients, and persons with special abilities [15, 16].

WHMS's are characterized by several constraints. Wearability, unobtrusiveness, low-cost, robustness, scalability, security, privacy of medical data, low-power consumption, ease of use and embedded decision support are some among them. As shown in [17] several of these specifications are difficult to be met, such as power issues, security of private information and system bulkiness. These issues are expected to be addressed by technology's advancements in microelectronics, low-energy IC design, wireless sensor networks and big data analytics techniques.

Many systems are presented based on WHMS technology, which aim in providing real-time unobtrusive monitoring of patients' physiological parameters. Prognosis [6] is a physiological data fusion model for a multisensor WHMS. It is designed to describe the current estimated health state and context of the patient, based on a fuzzy regular language for the generation of the prognoses of the health conditions of the patient. Although this system provides decision support for the current condition of the patient it doesn't assess the patient's physiology in real time. In [18] a portable and real-time monitoring system was developed. The system implemented a seizure detection algorithm in embedded systems for online monitoring EEGs and detecting seizure events, experimented on animals. The detection algorithm is based on a linear least-squares classifier. In [19] a middleware targeted on smart-phone like healthcare applications is presented focusing on the efficient management of sensor data, where several real–time monitoring applications of physiological parameters can be designed based on the presented middleware such as fitness monitoring, telemedicine and elderly care assistance.

Several research prototypes are based on WHMS by integrating various technologies of wearable sensors. MyHeart project [20] aimed at fighting cardiovascular diseases (CVD) by prevention and early diagnosis. By adopting the use of smart clothing, the wearable system is very comfortable for the user. The developed system included an ECG and an activity sensor and was able to classify human activity. A heart belt was used for monitoring patient's heart condition. Human++ [21] has developed a body area network consisting of three sensor nodes and a base station. Each sensor is collecting and processing multichannel data from ECG, EEG and EMG, while the base station functions as a data collector in star topology, regulating the information flow. This system is improved in [22]. A small, lightweight and low power WPMS platform is developed for ambulatory and continuous monitoring for autonomic responses in real life applications. HeartToGo [23] is a cell phone-based wearable platform, continuously monitoring ECG data. Real time analysis of electrocardiogram is performed to detect of abnormal events related to cardiovascular disease.

Due to high computing and storage demands in most WHMS's, several studies have examined the solution of geo-distributed clouds. In [24] a data integration framework for mobile healthcare has been proposed. This framework is centered on the concept of ubiquitous healthcare services provided to the patients in distant from a hospital areas. In [7] the authors presented PHISP – a Public-oriented Health care Information Service Platform supporting numerous health care tasks. Among them the platform provides to individuals many intelligent and personalized services, and supports basic remote health care and guardianship.

Among WHMS's constraints such as wearability, unobstractiveness, low cost, scalability have been improved by several studies performed in recent years. Less attention has been paid though, to the constraints imposed due to the real time processing of the huge amount of data in health care monitoring applications. In [5] a flexible framework that performs real-time analysis of physiological data to monitor the subject in his/her daily activities in the hospital environment is proposed. That study though focuses on the analysis algorithm needed to perform real time monitoring. Our work on the contrary focuses on data streaming characteristics, by presenting a multimodal big data streaming architecture applied in online seizure detection. Real time processing systems should fulfill a number of requirements. Particularly, low latency is essential in real time applications, as well as data fusion for being able to mine heterogeneous data sources. Additionally, the analysis algorithm should also be able to process events in real time over sliding windows. These requirements are characteristic to many streaming applications across various sectors. For example streaming applications are met in Web analytics, fraud detection, call center management, smart power meters and financial trading [25]. All these applications need to run continuous queries over high data rate streams, something that can be achieved by using a Data Stream Management System (DSMS) [25, 26, 27]. We present a health monitoring system, which by utilizing a DSMS is able to monitor and process a large continuous flow of physiological data in a real time and efficient manner.

3. The ARMOR concept

We present a flexible system performing real-time analysis of physiological data in order to detect events and high risk conditions related to epilepsy. Physiological parameters are analyzed by means of data mining in order to assess the severity of the monitored patient's condition. To enable ubiquitous analysis, real time processing of recorded data is performed. High-risk situation alert and patient information management services are also performed by the proposed system. The presented system is part of the online analysis components of the Advanced multi-paRametric Monitoring and analysis for diagnosis and Optimal management of epilepsy and related bRain disorders project (ARMOR) [28]. Fig. 1 presents the framework of ARMOR.

Fig. 1: Block diagram of the ARMOR's concept.

The first part of the ARMOR framework's chain of components consists of the physiological sensors. A variety of biosignals is measured in order to monitor several aspects of a patient's condition. In particular, electroencephalography (EEG) and electrocardiography (ECG) sensors in addition to 3-axis accelerometers (ACC) are utilized. EEG and ECG data are captured by different sensors of the Movisens GmbH-Karlsruhe [29] sensor platform. The captured data are encrypted using a FPGA-based solution and are afterwards wirelessly transmitted using Bluetooth 4.0 technology. In order to control noise, signal-to-noise ratio and DC level are measured, issuing an alarm when some preset threshold is exceeded. A wireless network is formed to transmit the data recorded from the sensors, for the sake of convenience of the patient, since a set of cables can be both cumbersome and restrictive [30]. The transmitted data are formatted in Unisens [31] format. Because a wireless protocol is followed for data transmission, an encryption phase is also required in order to avoid possible data interceptions. The encryption is based on a 128-bit block NIST-FIPS 197 Advanced Encryption Standard (AES) algorithm. The data are transmitted to an independent component of the system called Home Gateway, located in the proximity of the patient. The first action of this system is to decrypt the received packets before any further processing. Once the data are transformed to their initial form they are directed to two separate data flows. The first data flow is directed in the online detection module, while the second one is uploaded via a secure network in a data warehouse called Personal Health Record (PHR). PHR systems' role is to manage all necessary medical information for each patient. These systems are consisted of several components, which provide storage, log history and a personal health profile. The storage component is offering permanent repository to the recorded data for each patient. Medical experts should have easy access to the contents of the PHR, thus an appropriate web interface should be provided. Alarms received by the online analysis should also be managed appropriately by a notification service, in order to notify the corresponding to the patient clinicians for any detected high risk condition. Communication between the home gateway and the PHR is bidirectional. In addition to receiving the alarms produced by online analysis, the PHR provides all the necessary information to the online detection module, for adaptation to the needs of a specific patient. These parameters are configurable depending on the patient's condition and the purpose of the online analysis. The PHR has a bidirectional communication with another component that performs the offline analysis. The offline analysis server is an independent system which performs the automatic analysis of the data existing in the PHR. Since during the online monitoring there are limitations regarding processing time, memory used and complexity of the algorithms, there are certain aspects of the data that cannot be analyzed in details. The aim of the offline analysis is to perform further investigation of the recorded data, in order to discover additional knowledge of their nature.

In general, the online analysis detection module is a subsystem that is constantly processing in real time the captured physiological data in order to monitor the patient's health condition and to detect any abnormal events. Physiological measures are analyzed by means of data mining over sliding windows, in order to assess the severity of risk in the patient's condition. Efficient online analysis requires the maintenance of temporal and spatial consistency in the collected data. Time domain data fusion is achieved by synchronizing the different data flows in time, while spatial domain fusion is performed by the analysis algorithms.

The online analysis is performed on a stream of data, thus the data are split into windows. Each window corresponds to a specific time period, whose data are stored in a continuous block in memory called data block. Once a specific sample is received, it is ordered to a data block according to its time-stamp. A data block is considered the smallest unit of data that can be analyzed. Hence it should be guaranteed that before each data block is sent for analysis, it was received entirely intact. The integrity of received data could be accomplished with check-sums implemented at the hardware level; software based solutions could also be developed by combining communication protocols and managing missing values. Data blocks represent a small period of time of patient's physiological signals and environmental state. The time period each block represents should not be too small, so that the analysis can access enough data to extract valuable information. On the other hand, if the corresponding time period is too big it will just record the seizure event since small scale events will be obscured by the vast size of the period.

In data streaming applications temporal aspects of massive and continuous data streams are to be dealt with. Real-time or almost-real-time response is achieved efficiently by DSMSs. A DSMS offers all the necessary tools for performing selections from one or more data streams in the form of continuous queries. These queries are applied on specific sub-sequences (data blocks) of the incoming data and are continuously executed as new data arrive. Querying procedures in a DSMS provide operations similar to traditional database queries. Though unlike traditional static queries that operate on tables, continuous queries operate on data streams and their output can be another data stream. In this manner, combining data from multiple sensors can be expressed as selections on multiple streams, enhanced with the functionality of query operators. With this functionalities provided, it is possible to retain the streaming nature of data. Buffers and queues are required at the inputs of the system as well as between query operators in order to handle continuously arriving data. At any given time, there may be many data sub-sequences in the input and inter-operators queues, especially if the arrival rate of input data is bursty or the consumption of sub-sequences is not fast enough. Thus, the DSMS's scheduler should decide which data blocks to process next. The simplest scheduling strategies involve the allocation of a processing time slice per data block in a round-robin or first-in-first-out fashion.

Since input data flow arrives from various external sources, some samples may arrive out of order with respect to their generation time. Furthermore, data may not be received from a source for some time, which is the case when there are communication errors or the source is malfunctioning. In these cases, buffers should be maintained, which contain the previously received events of the corresponding data block. These buffers should be retained for a specific amount of time, which is usually unknown or varies through time. DSMSs incorporate a solution for this problem called punctuation [32]. Punctuation is a special event inserted to the data stream that contains a predicate, guaranteed to be satisfied by the remainder of the data stream. More specifically, punctuation guarantees that samples that have a timestamp, below the punctuation's own timestamp, will be dropped by the system. This case applies mostly on systems where data samples arrive in timestamp order. These punctuations that govern the timestamps of future events are generally called heartbeats [33, 34].

The online detection module aims in the identification of an abnormal event. Once such an event is detected then an alarm is sent to every pre specified target such as the doctor's personal email or phone or to an automatic seizure contamination system. These alarms are also sent to a permanent data warehouse (PHR). In order to be more accurate and efficient, the detection module is also configurable based on a specific patient's characteristics. These configurations regard to EEG electrodes' cardinality and placement, as well as the detectors sensitivity. All these parameters are pre specified by experts or semi-automatic analysis tools and could be unique for every patient.

4. Architecture for multimodal big data streaming

The home gateway (Fig. 2) is a system placed in the environment of the patient receiving wirelessly the transmitted data from the sensors. The online detection platform, which is part of the home gateway, is constantly receiving recorded data from the sensors and in real time is processing this data flow. As described previously the data arriving in the home gateway are encrypted packets. The decryption module is constantly receiving these packets and transforming them to their original form. After the decryption, recorded data flow is processed by the other components of the online analysis platform.

Fig.2: Block diagram of the multimodal streaming data processing architecture.

Before the online analysis, the decrypted data are aligned and synchronized in time. The data synchronization module is responsible for ensuring that the processed data are correctly placed and ordered with respect to time. Since different modalities are used, different recording devices are necessary. The sampling rate of these devices might vary upon their functional characteristics. It is therefore essential that the synchronization module is able to cope with the difference in devices' sampling frequencies. In each sample received from each sensor, a timestamp is assigned. Each timestamp can be estimated based on two different ways:

a. Timestamp sent from sensor. In this case the timestamp is sent from the recording device (provided that this functionality is available). However a significant amount of the bandwidth of the wireless channel is consumed for transferring this information. In this case the data synchronization module functions like a simple buffer, with reordering capabilities in case received samples arrive out of order. The received samples are time stamped with their generation time.

b. Timestamp calculated in the data synchronization module. This case applies when the bandwidth of the transmission channel is limited, or the recording devices do not offer a service for transmitting the timestamp of the recorded samples. The data synchronization module in this case needs to calculate the timestamp of each received channel. In this case received data are assumed to be acquired in the correct order, since the generation time is not available. Received data are time-stamped using the acquisition time of the first sample and the sampling rate of each device.

The data synchronization component should also contain a service which removes inconsistent measurements and handles missing values. Inconsistent values can be detected by imposing thresholds on the received signal. These values are marked for rejection from the online analysis. Missing values are caused by sensor failure and are also marked for rejection. An alternative strategy for handling missing values could be to resend the missing data, but due to the medical needs for real-time detection of potential seizure onsets was not adopted in our evaluation. Communication errors are assumed to be handled by communication protocols, granting data recovery in admissible time delay.

Once the synchronization problem is solved, the data flow is replicated and sent to two different subsystems, the online analysis subsystem and the data control. Fig. 3 presents the procedure followed for receiving the transmitted packets, decrypting them, marking them for rejection and reordering them if necessary and sending them to the next step of the online analysis.

Fig. 3: Operation of the synchronization and preprocessing module. Encrypted packets are decrypted, synchronized, ordered and marked for rejection if necessary.

The online analysis subsystem hosts the online seizure detection module. In this subsystem data are analyzed in order to detect the seizure events captured by the sensors. There are two main parts combined to create this subsystem. The first one is the preprocessing module. In the preprocessing step the arriving data flow is segmented in windows, either according to the number of events arrived or to the time intervals.

These two framing methods became the same in the case where all sampling frequencies are the same. Since different sensors are expected to send data in different sampling frequencies, the time interval based segmentation is more appropriate in our case. Each window in this scenario has a start and an end timestamp. Each time sample received from every device, belongs to a set of windows only if the time intervals of these windows contain the sample's timestamp. Depending on the desired detection accuracy and detection algorithm's performance, windows can be overlapping or non-overlapping. Our method is based on short-time analysis with constant length time sliding windows without overlap. For every window a data block is created. Each block's corresponding data are inserted incrementally between the start and end time of the block. At its end time a specific block stops receiving data and is transferred to the next level of the analysis. Each data block has a life expectancy, which starts when the first sample of this block arrives in the system and ends when the final step of the analysis is performed for this data block. During this period, the data block should be quickly retrieved when necessary. Therefore, the data blocks are stored in main memory.

Once the life expectancy of a data block is over, the block is dropped in order to be replaced by a new one. This cycle of reusing space in main memory is a challenging procedure, since data should be consumed faster than they are produced. The efficiency of the memory management depends on the ability of the detection algorithm to process data faster than they are received, as well as the system's physical characteristics (memory size and speed, processing capabilities, etc). Memory management can be handled efficiently by utilizing a DSMS, as the responsible service for manipulating the data flow, buffers utilized for the maintenance of the streaming nature of data, as well as continuous query operations.

Fig. 4: Operation of the framing module. Each time-stamped data sample is assigned to the set of windows it belongs to. Assignments are performed based on the timestamp of each data sample and each window's start and end time. Windows containing rejected data are dropped, while the ones whose end time has passed are sent to the detection module. New windows are created in every generation timestamp.

In Fig. 4 the framing process is shown. As time marked data arrive in this module, they are assigned to a group of windows to which they belong. This group of windows constitutes the set of ActiveWindows and is constantly updated with new data. Once a window end time expires it stops receiving data and is sent to the detector. Windows which receive a value marked as rejected should be ignored and therefore are removed from the set of ActiveWindows. Fig.5 shows the information structure after data synchronization and framing respectively. Specifically once the received packets are decrypted the received data are stored in the sensor input buffer. The received samples are stored in a structure containing the value and the sensor identification. Once the samples are processed by the data synchronization module, in the information structure is added the timestamp for each sample and the binary flag for rejection. The synchronization module in our system is working based on the timestamp of the first sample and the sampling frequency of the sensor. The first sample's time stamp is calculated by the systems internal clock. Every new data sample is attributed with a timestamp, which is calculated by the number of already received samples. The formula for calculating a data sample's timestamp is $Ttime = Tstart + S * T$, where $Ttime$ is the timestamp of the current sample, $Tstart$ is the timestamp of the first sample received, S is the number of the samples already received and T is the time difference between two different samples, which is considered as constant. The framing module results in the formation of data blocks, each one corresponding to every individual time window specified in the data flow.

Fig. 5: Synchronization and processing data flow.

The second module of the online analysis is the detector. The detection framework should be developed in such a way that a short time analysis of the received data blocks is possible. Since the data flow is constant, there are limitations in the time available to process each data block. Each data block should be processed in an amount of time smaller than its corresponding time interval. Generally in online frameworks an approximate result, with some reasonable guarantee on the quality of the approximation, is acceptable [35]. In the case of seizure detection though, the accuracy is of utmost importance. In order to perform the instantaneous detection of a seizure event all different modalities must be fused in order to the exploit all possible information for the patient. This fusion can be performed in two different ways:

1. By blending the features extracted from every sensor's data block into a single vector resulting in one decision which encapsulates information from all modalities, or

2. By performing a detection for every individual sensor stream's data block and then fusing the individual decisions to a single one. This process will also result in a synthesis of the different information captured by the sensors.

In the first case the feature space will be vast, thus either a lot of training data or an algorithm than can overcome the sparseness of the feature space (e.g. Support Vector Machines) are a necessity. In the second case each sensor's data flow is processed separately. Therefore a training set for every sensor should be available, which is, in some situations, a challenging procedure. Several detectors might be based on pre-trained models while others are developed using decision rules. In our case the models built to perform real time classification are patient specific, in order to match each patient's specific condition. As shown in Fig. 6, each patient specific model is introduced into the detector in the form of parameters.

The seizure detector developed in our study utilizes the first fusion strategy. An 21-dimensional EEG signal along with an 1-dimensional ECG signal. The dimensionality of the EEG signal can be varied based on the set of N sensors (electrodes) used, thus the EEG signal will is refered as N-dimensional signal. The data blocks of these signals are processed in parallel by time-domain and frequency domain feature extraction algorithms. More specifically, each N-dimensional data block of EEG is processed by time-domain and frequency domain feature extraction algorithms for EEG. For each individual EEG channel the spectral magnitude, autoregressive filter coefficients, the continuous and discrete wavelet transform, the energy per brain wave band (delta, theta, beta, alpha), band pass based features and phase space representation are extracted. Additionally time domain features such as zero-crossing rate and temporal statistics are calculated. Consequently, for each EEG data block a feature vector $X_{EEG} \in R^{N*F}$ is created, where N is the number of EEG channels and F the number of features for each channel. In our study F responds to 55 features per channel [36]. Each 1-dimensional ECG data block is processed by time-domain frequency algorithms for ECG. The resulting feature vector is $X_{ECG} \in R^{P}$, where P is the number of ECG features, which in our study is equal to 12. Data fusion is performed by concatenating the feature vectors in one vector, $X_{Fused} \in R^{N*F+P}$, before the classification step of the detection module. The fused feature vector is of size 1167 in our study. For further details about the method used for seizure detection which was adopted in the developed seizure detector see [36, 37].

Fig. 6: Operation of the detection module algorithm. For each window sent to the detector, a synopsis of the window's data is created. This synopsis is consisted by a set of features extracted by the *FeatureExtraction* method. The detection is performed by utilizing this synopsis with a set of parameters according to each patient's personal characteristics. If the window corresponds to a seizure, the *RiskManagementAssessment* service assesses the risk of the patient's condition.

The detection module with the combination of a risk management assessment module can provide a higher level of decision support. As shown in Fig. 6 once the classification for a specific window is performed, a new event is created containing the decision whether the window's period is classified as being a part of seizure or not. All these events, which are constantly produced as data are arriving, compose a new stream. By performing a continuous query on this stream, several information regarding the status of the patient can be extracted. The most profound of all is the detection of a seizure period. Once the onset of a seizure is detected then an alarm is produced. This alarm is sent to the data control sub system for further processing. Furthermore, this stream can be used in order to detect higher level events, such as whether the seizure is lasting for more than five minutes, a condition known as status epilepticus. For each higher level detected event an action is performed (e.g. an email is sent to the responsible physician, an alarm goes off in the proximity of the patient, etc.). This analysis is performed by the risk assessment decision support module, which performs also additional analysis to the recorded data. For example, given a body position sensor, once the patient's body position is prone and status epilepticus is detected then a high risk alarm is produced. This module is adjusted to the patient's characteristics, since each situation's risk depends on the patient's condition. It should be noted that for this higher level of decision making, a fusion of different modalities and processing in sliding windows are necessities for real time processing of these events.

The data control subsystem is a system that is responsible for manipulating the storage of data arriving to the home gateway, as well as the detected events produced by the online analysis. The data flow is segmented in epochs in order to be compressed and sent to the PHR. This way data are transferred more efficiently and possible network failures result in the loss of a smaller amount of data. The segmentation is performed by the data epoching module. The epoch segmentation can be performed in constant time intervals. The alarms produced by the online analysis should also be encapsulated in the information sent by the home gateway to the PHR. The automatic annotations manager receives each alarm produced by the online analysis subsystem and creates an annotation for each received event. Each annotation contains the information regarding the nature of a detected event, besides its timestamp of occurrence. Every annotation is then sent to the format conversion module in order to be combined with the recorded data. Each epoch along with the structured information of the detected events, belonging to that epoch, is gathered to the format conversion tool. This tool concatenates each epoch's data with the detected events and their relevant information to a file. The format of this file should be compatible with the next steps of the analysis. A popular format used for these reasons is European Data Format (EDF) [38]. Using that format medical experts can view the results of the online analysis, while these results can be used and augmented by the offline analysis services.

5. Evaluation of the online seizure detector

The proposed system was validated by means of real processing time latency under different workloads. More specifically, the system's real time performance has been tested under a) different number of EEG electrodes and window sizes and b) different number of extracted features. The first case applies when spatial resolution of measured signals is sacrificed, in exchange for higher detection frequency. One scenario where this case may find practice is when the seizure events of the patient are focal, meaning they occur in a specific brain region. The second case applies when the online module's performance is favored (by processing less features), against the detector's accuracy. As shown in [36] by removing the worst ranked features, the detector does not suffer from significant accuracy losses, which supports the whole online operation.

5.1. Experimental setup

In our study the seizure detection algorithm as well as the feature extraction algorithms were implemented in Matlab. The developed detector was integrated in Microsoft's StreamInsight, which is the DSMS used in our study. The system's characteristics where the experiments were performed are summarized in Table 1.

Table 1: Simulation system's characteristics

For evaluating the performance of the developed algorithm under different workloads, we examined the real processing time latency (RPTL) of the algorithm under two different case studies. As RPTL we consider the processing time the detector needs to produce a result for each corresponding second of real input data. RPTL is defined as:

$$RPTL = \frac{P}{N} * \frac{1}{1-overlap} \tag{1}$$

where P is the processing time for a window in seconds, N is the window's size in seconds and overlap $\in [0, 1)$ is the percentage of overlapping between two consecutive windows.

RTPL should be less than one for every second on average, since otherwise the system will not be able to handle the continuous streaming nature of the data and have real-time response. Spontaneous increases of this time for a small number of data blocks are acceptable, since it is assumed that it can be handled by the DSMS. Our experiments included tests in windows of various sizes.

The data used in our study consiste of 21 EEG electrodes (C3, C4, Cz, F3, F4, F7, F8, Fp1, Fp2, Fz, O1, O2, P3, P4, Pz, T1, T2, T3, T4, T5, T6) and one electrocardiographic channel. Data were recorded with sampling frequency equal to 250 Hz for EEG and 256 Hz for ECG. In our study we excluded EEG reference electrodes O1 and O2, for not offering a useful amount of information. Since the ECG stream was composed of one only stream and the corresponding feature extraction algorithm [36] is not computationally heavy, the ECG features used in every experiment were always the same.

5.2. Evaluation

To examine the performance of the detector for different spatial resolutions of the recorded data, average real processing time was evaluated for different numbers of EEG electrodes. This case applies when a smaller than the original set of electrodes is used for monitoring the patient's brain functionality. Defining the most appropriate set for every patient is not part of this study, so the choice of each electrode set was random. The ECG channel was part of every simulation. The sliding windows applied to the data stream were of size of 1, 2 and 4 seconds with no overlap. A number of 100 windows were extracted from experiment data in each case, in order to estimate the average RTPL for every different window size. For each case the average execution time corresponding to each window is shown in Table 2.

Table 2: Average execution time per window for various electrode sets and window sizes

As expected reducing the number of electrodes results in significantly shorter time needed for processing each data block. As it is observed from Table 2 using the whole set of electrodes (except O1-O2) results in best than one second of processing time per window in almost every case. The difference is negligible though for different window sizes, when the same number of electrodes is used. Subsequently, as shown in Fig. 7 the larger the size of the window the shorter real processing time that corresponds to each second of actual data. This is justified by the fact that the feature extraction algorithms are scalable in the size of the data block and the feature vector's size extracted for each window is invariable regardless of the window size. Thus, larger window sizes lead to shorter real processing time per second which in turn makes possible the usage of more overlap between consecutive windows.

Fig. 7. Real time processing corresponding to one second of actual time, for various window sizes and number of electrodes.

6. Discussion and Conclusion

In this paper a multimodal big data streaming architecture is presented, which utilizes an online seizure detector. Real time analysis of patient's physiological data is performed in order to monitor the health condition of the patient and discover abnormal events and high risk conditions. Patient specific models are used, in order to provide suitable decision support to the patient's condition. The system performs streaming data analysis over sliding windows applied to physiological data flow and is using big data mining techniques to instantaneously analyze the condition of the patient. Experiments were performed to evaluate the real time processing of the developed detector in terms of spatial resolution of recorded signals and amount of information assessed from physiological signals.

The large amount of information produced by the sensors should be processed in a specific amount of time, before the system overflows with unprocessed data. Specifically each segment of data corresponding to one second of actual time should be processed for one second at most. We showed that this problem can be tackled either by reducing the number of electrodes used to record the data, or alternatively by increasing the size of the windows, in which the data are segmented. These goals can be achieved because the feature extraction and detection algorithms utilized in our study are independent from the window size the data are segmented to and depend only on the number of electrodes used to record the data.

Given sufficient bandwidth the timestamp information of each sample can be sent by the recording devices. In our case there were bandwidth constraints for lower consumption, therefore synchronization of different sources was achieved by using timestamp information of the first sample sent from each source to calculate the timestamp of each following data sample, given the fact that the sampling frequency of each data source was stable. Fusion of different sources can also be achieved by extracting features from every data source in parallel and then utilizing a processing algorithm that incorporates all these features together. Alternatively, each data source can be mined separately and in parallel with the others, thus the fusion is achieved in the decision level. In our study, we approached the EEG and ECG fusion problem with concatenating the features extracted corresponding to each data source's time window, while a SVM classifier used in the detection module achieved robust performance. The modular structure of the proposed architecture allows the use of it, as a basis framework, to other similar applications with multimodal and/or heterogeneous streaming data.

7. Acknowledgement

The research reported in the present paper was partially supported by the ARMOR Project (FP7-ICT-2011-5.1 - 287720) "Advanced multi-paRametric Monitoring and analysis for diagnosis and Optimal management of epilepsy and Related brain disorders", co-funded by the European Commission under the Seventh' Framework Programme.

8. References

1. Hao Y. and R. Foster, "Wireless body sensor networks for healthmonitoring applications," Physiol. Meas., vol. 29, pp. R27–R56, Nov.2008
2. World Health Organization, "Integrating prevention into health care," http://www.who.int/chp/about/integrated cd/en/, 2012, [03-23-2014]
3. World Health Organization. (2009, Oct. 9). The World Health Report 2008, Primary Health Care, Now More Than Ever. World Health Org., Geneva, Switzerland. [Online]. Available: www.who.int

4. Joseph M. Woodside, "EDI and ERP: A Real-Time Framework for HealthCare Data Exchange", Journal of Medical Systems, June 2007, Volume 31, Issue 3, pp 178-184.
5. Apiletti, D.; Baralis, E.; Bruno, G.; Cerquitelli, T., "Real-Time Analysis of Physiological Data to Support Medical Applications," IEEE Transactions on Information Technology in Biomedicine, , vol.13, no.3, pp.313,321, May 2009
6. Pantelopoulos, A.; Bourbakis, N., "Prognosis—A Wearable Health-Monitoring System for People at Risk: Methodology and Modeling," Information Technology in Biomedicine, IEEE Transactions on , vol.14, no.3, pp.613,621, May 2010
7. Wang PengWei; ZhiJun Ding; Changjun Jiang; MengChu Zhou, "Design and Implementation of a Web-Service-Based Public-Oriented Personalized Health Care Platform," Systems, Man, and Cybernetics: Systems, IEEE Transactions on , vol.43, no.4, pp.941,957, July 2013
8. Gyoun-Yon Cho, Seo-Joon Lee, Tae-Ro Lee, "An optimized compression algorithm for real-time ECG data transmission in wireless network of medical information systems", Journal of Medical Systems, December 2014, 39:161.
9. Hauser A. (1997) Epidemiology of seizures and epilepsy in the elderly. In: Rowan A, Ramsay R, eds. Seizures and epilepsy in the elderly. Boston: Butterworth-Heinemann, 7-18.
10. Gatzoulis, L.; Iakovidis, I., "Wearable and Portable eHealth Systems," Engineering in Medicine and Biology Magazine, IEEE , vol.26, no.5, pp.51,56, Sept.-Oct. 2007
11. Lymberis, A.; Dittmar, A., "Advanced Wearable Health Systems and Applications - Research and Development Efforts in the European Union," Engineering in Medicine and Biology Magazine, IEEE , vol.26, no.3, pp.29,33, May-June 2007
12. Seyfettin Noyan Oğulata, Cenk Şahin, Rızvan Erol, "Neural Network-Based Computer-Aided Diagnosis in Classification of Primary Generalized Epilepsy by EEG Signals", Journal of Medical Systems, Volume 33, Issue 2, pp 107-112.
13. Ahmet Alkan, M. Kemal Kiymik, "Comparison of AR and Welch Methods in Epileptic Seizure Detection", Journal of Medical Systems, Volume 30, Issue 6, pp 413-419.
14. M. Kemal Kiymik, Abdulhamit Subasi, H. Rıza Ozcalık, "Neural Networks with Periodogram and Autoregressive Spectral Analysis Methods in Detection of Epileptic Seizure", Journal of Medical Systems, Volume 28, Issue 6, pp 511-522.

15. Bonato P., “Advances in wearable technology and applications in physical medicine and rehabilitation,” Journal of NeuroEngineering and Rehabilitation, vol. 2, p. 2, Feb. 2005
16. Bonato P., “Wearable sensors/systems and their impact on biomedical engineering,” IEEE Eng. Med. Biol. Mag., vol. 22, no. 3, pp. 18–20, May/Jun. 2003
17. Pantelopoulos, A.; Bourbakis, N.G., "A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on , vol.40, no.1, pp.1,12, Jan. 2010
18. Chang, Da-Wei; Sheng-Fu, Liang; Chung-Ping, Young; Fu-Zen, Shaw; Su, A.W.Y.; You-De Liu; Yu-Lin, Wang; Yi-Che, Liu; Jing-Jhong, Chen; Chun-Yu, Chen, "A Versatile Wireless Portable Monitoring System for Brain–Behavior Approaches," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol.1, no.4, pp.440,450, Dec. 2011
19. Seeger, C.; Van Laerhoven, K.; Buchmann, A., "MyHealthAssistant: An Event-driven Middleware for Multiple Medical Applications on a Smartphone-mediated Body Sensor Network," Biomedical and Health Informatics, IEEE Journal of ,vol.PP, no.99, pp.1,1, 2014
20. Habetha, J., "The myheart project - fighting cardiovascular diseases by prevention and early diagnosis," Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE ,vol.Supplement, no., pp.6746,6749, Aug. 30 2006-Sept. 3 2006
21. Gyselinckx, B.; Penders, J.; and Vullers, R. “Potential and challenges ofbody area networks for cardiac monitoring,” J. Electrocardiol., vol. 40,pp. S165–S168, Nov. 2007.
22. Brown, L, Grundlhner B, Penders J, and Gyselinckx B, “Body Area Network for Monitoring Autonomic Nervous System Responses”, in Proc. of 2009 3rd International Conference on Pervasive Computing Technologies for Healthcare - Pervasive Health 2009, Article number 5173603, April 2009.
23. Jin, Zhanpeng; Oresko, J.; Shimeng Huang; Cheng, A.C., "HeartToGo: A Personalized medicine technology for cardiovascular disease prevention and detection," Life Science Systems and Applications Workshop, 2009. LiSSA 2009. IEEE/NIH , vol., no., pp.80,83, 9-10 April 2009

24. Constantinescu, L.; Jinman Kim; Feng, D.D., "SparkMed: A Framework for Dynamic Integration of Multimedia Medical Data Into Distributed m-Health Systems," Information Technology in Biomedicine, IEEE Transactions on , vol.16, no.1, pp.40,52, Jan. 2012
25. Ali, M.; Chandramouli, B.; Goldstein, J.; Schindlauer, R., 2011. "The extensibility framework in Microsoft StreamInsight." In *Proceedings of the 2011 IEEE 27th International Conference on Data Engineering* (ICDE '11)
26. Arasu, A.; Babu, S.; Widom, J., "CQL: A Language for Continuous Queries over Streams and Relations" DBPL 2003: 1 -19
27. StreamBase Inc. http://www.streambase.com/.
28. ARMOR [Online]. Available: http://armor.tesyd.teimes.gr/welcome
29. Hey S., Anastasopoulou P., Bideaux A., Antonopoulos C., Voros N., Fernandez A., Megalooikonomou V., Krukowski A., "System Architecture", Cyberphysical Systems for Epilepsy and Related Brain Disorders, Voros N., Antonopoulos C. (Eds), Springer, pp. 127-136
30. Townsend, K.A.; Haslett, J.W.; Tsang, T.K.K.; El-Gamal, M.N.; Iniewski, K., "Recent advances and future trends in low power wireless systems for medical applications," System-on-Chip for Real-Time Applications, 2005. Proceedings. Fifth International Workshop on , vol., no., pp.476,481, 20-24 July 2005
31. Unisens www.unisens.org
32. Tucker, P.; Maier, D.; Sheard, T.; Faragas, L. "Exploiting punctuation semantics in continuous data streams". IEEE Trans. Knowl. and Data Eng., 15(3):555–568, 2003.
33. Johnson, T.; Muthukrishnan, S.; Shkapenyuk, V.; Spatscheck, O. "A heartbeat mechanism and its application in Gigascope." In Proc. 31st Int. Conf. on Very Large Data Bases, pages 1079–1088, 2005
34. Srivastava U. and Widom J. "Flexible time management in data stream systems." In Proc. ACM SIGACT-SIGMOD Symp. on Principles of Database Systems, pages 263–274, 2004
35. Chaudhry, N.; Shaw, K.; Abdelguerfi, M. "Stream data management", Advances in Database Systems, Vol. 30, Springer 2005
36. Mporas, I.; Tsirka, V.; Zacharaki, E.I.; Koutroumanidis, M.; Megalooikonomou, V. "Online Seizure Detection from EEG and ECG signals for Monitoring of Epileptic Patients," 8th Hellenic

Conference on Artificial Intelligence (SETN 2014), Lecture Notes in Computer Science, vol. 8445, pp 442-447, 2014

37. Mporas, Iosif;Tsirka,Vasiliki;Zacharaki,Evangelia I.;Koutroumanidis, Michalis; Richardson,Mark;Megalooikonomou,Vasileios "Seizure detection using EEG and ECG signals for computer-based monitoring, analysis and management of epileptic patients", Expert Systems with Applications, Vol. 42 (6), 2015, pp. 3227–3233
38. Kemp Bob; Jesus Olivan "European data format ‘plus’ (EDF+), an EDF alike standard format for the exchange of physiological data", Clinical Neurophysiology, Volume 114, Issue 9, September 2003, Pages 1755-1761, ISSN 1388-2457.

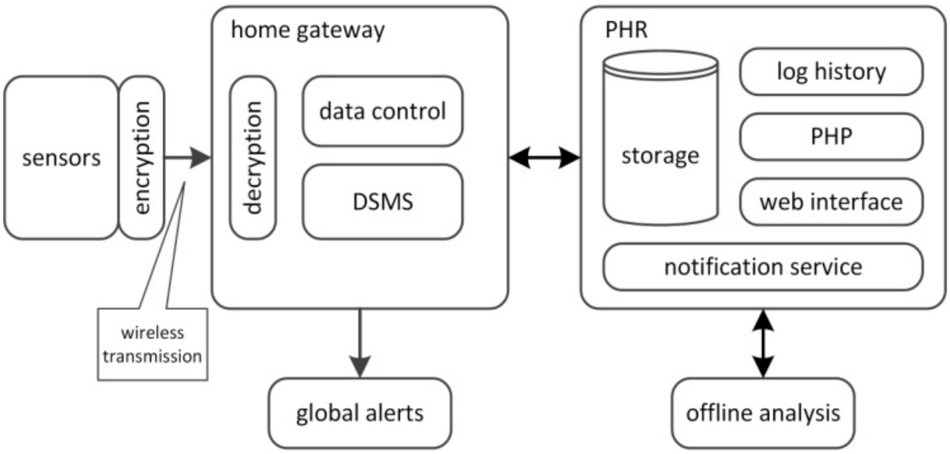

Fig. 1: Block diagram of the ARMOR's concept.

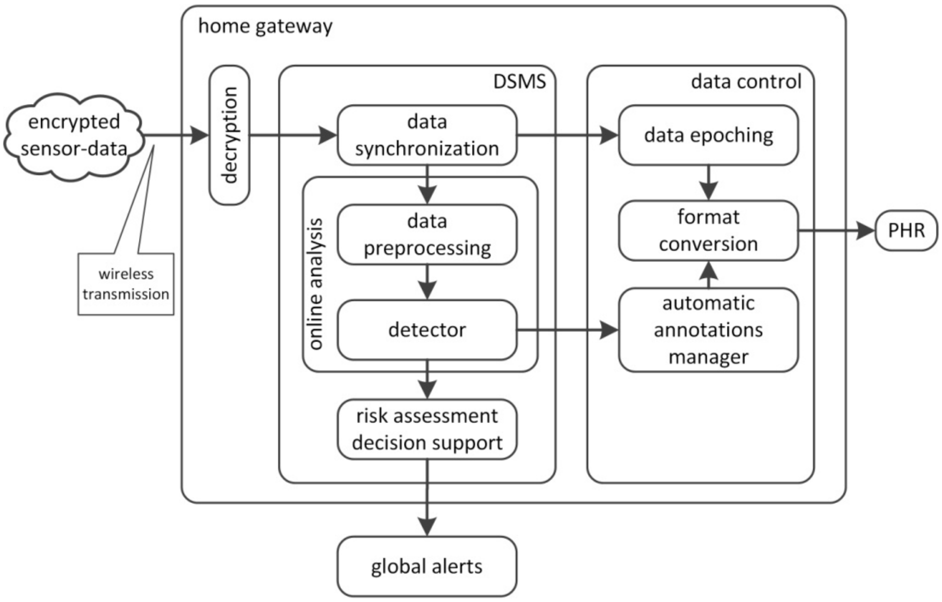

Fig.2: Block diagram of the multimodal streaming data processing architecture.

```
Syncronization and preprocessing module

WHILE encryptedPackets are received
      DecryptedData = decryptionModule(encryptedPacket);
      TimeStampData = dataSynchronization(DecryptedData);
      IF (encryptedPacket arrived in wrong order)
            ORDER TimeStampData based on <TimeStamp>;
      ELSEIF  (missing values or errors are detected)
            MARK TimeStampData for rejection;
      ENDIF
      FramingModule(TimeStampData);
ENDWHILE
```

Fig. 3: Operation of the synchronization and preprocessing module. Encrypted packets are decrypted, synchronized, ordered and marked for rejection if necessary.

```
Framing module

ASSIGN TimeStampData to corresponding ActiveWindows;
GET CurrentTimeStamp;
FOR (every Window IN ActiveWindows)
      IF (Window CONTAINS Rejected TimeStampData)
            REMOVE Window FROM ActiveWindows;
      ELSEIF (Window. EndTime < CurrentTimeStamp)
            DetectionModule(Window);
            REMOVE Window FROM ActiveWindows;
      ENDIF
ENDFOR
IF (CurrentTimeStamp >= NextGenerationTimestamp)
      AddNewWindow(ActiveWindows);
      update(NextGenerationTimestamp);
ENDIF
```

Fig. 4: Operation of the framing module. Each time-stamped data sample is assigned to the set of windows it belongs to. Assignments are performed based on the timestamp of each data sample and each window's start and end time. Windows containing rejected data are dropped, while the ones whose end time has passed are sent to the detection module. New windows are created in every generation timestamp.

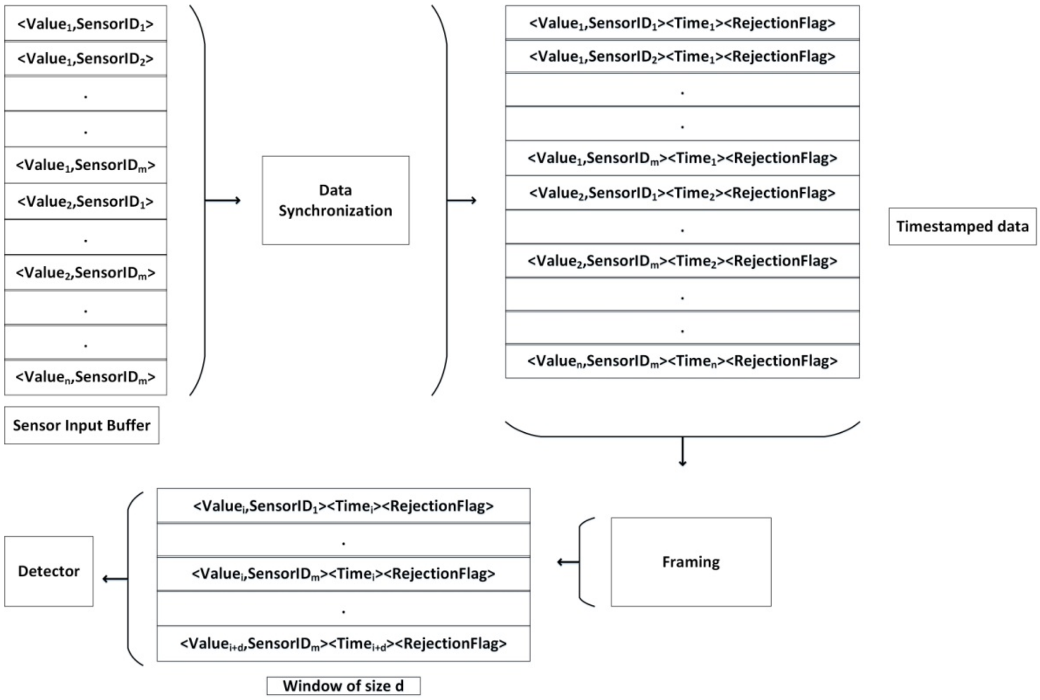

Fig. 5: Synchronization and processing data flow.

Detection module algorithm

```
Synopsis = FeatureExtraction(Window)
Detection = Detector(Synopsis,Parameters)
IF (Detection == "Seizure")
      RiskManagementAssessment(Window,Sinopsis,CurrentTime)
ENDIF
```

Fig. 6: Operation of the detection module algorithm. For each window sent to the detector, a synopsis of the window's data is created. This synopsis is consisted by a set of features extracted by the *FeatureExtraction* method. The detection is performed by utilizing this synopsis with a set of parameters according to each patient's personal characteristics. If the window corresponds to a seizure, the *RiskManagementAssessment* service assesses the risk of the patient's condition.

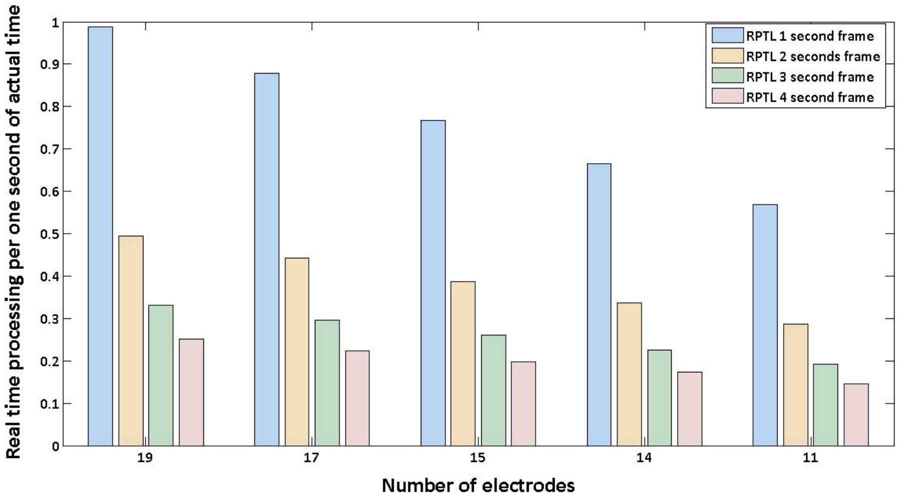

Fig. 7. Real time processing corresponding to each second of actual time, for various frame sizes and number of electrodes.

CPU	Intel Xeon E5 @ 3.7 GHZ
Cores	4
Threads	8
Cache	Level 3
RAM	16GB DDR3 @ 930 MHZ

Table 1: Simulation system's characteristics

EEG channels	w = 1 sec	w = 2 secs	w = 3 secs	w = 4 secs
19	0. 960	0. 960	0. 960	0. 960
17	0. 852	0. 852	0. 852	0. 852
15	0. 751	0. 751	0. 751	0. 751
13	0. 654	0. 654	0. 654	0. 654
11	0. 551	0. 551	0. 551	0. 551

Table 2: Average execution time per window for various electrode sets and window sizes