RepoBIT: Cloud-Driven Real-Time Biosignal Streaming, Storage, Visualisation and Sharing

Margarida Reis
Instituto de Telecomunicações
Instituto Superior Técnico
Lisboa, 1049-001, Portugal
Email: margarida.reis@tecnico.ulisboa.pt

Hugo Plácido da Silva
Instituto de Telecomunicações
Instituto Superior Técnico
Lisboa, 1049-001, Portugal
Email: hugo.silva@lx.it.pt

***Abstract*—Physiological (or biosignal) data acquisition and analysis has evolved from being intrinsically bound to medical practice or lab settings, to become a pervasive data source with numerous applications. This has been mostly leveraged by the proliferation of wearables and open hardware platforms. While the first devices offer good cloud connectivity (albeit to proprietary servers) but only provide high-level features, the latter have opposite characteristics, creating a gap in what concerns collaborative access to datasets for reuse in academic research. In this paper we describe a module devised to help bridge this gap, by enabling real-time biosignal streaming directly from an open hardware physiological data acquisition platform to a cloud-based repository over Wi-Fi. Performance tests have been made to assess the data transmission reliability, and the adopted approach to the repository enables user-friendly collaborative access, visualisation and sharing of the recorded data.**

***Index Terms*—Physiological Computing; Healthcare; Open Hardware; Internet of Things; Cloud-Based Repository.**

I. Introduction

The nascent field of physiological computing is contributing with significant advancements towards enriching the way in which people interact with digital content, through interfaces and interaction experiences mediated by physiological (or biosignal) data [1]. Wearable technologies have greatly contributed to intensify research in this field, however their sensor set is typically fixed, data collected from these devices is mostly made available as high-level features and access is offered through proprietary servers and Software Development Kits (SDKs) mostly controlled by the vendors.

As a way of overcoming the limitations associated with proprietary technologies, a diversity of low cost hardware and open source software platforms for physiological computing research is emerging. While these provide a high degree of flexibility in what concerns the sensor set configuration and access to raw data, both quite desirable for biomedical research, data acquisition and recording is done using USB, Bluetooth or Bluetooth Low Energy (BLE), hence being dependent on an intermediate device (e.g. a computer or mobile phone).

This paper proposes a methodology for real-time biosignal streaming over Wi-Fi, directly from an open hardware biosignal platform to the cloud, without the need for intermediate devices. Our work is motivated by the increasingly growing need for user-friendly, open and collaborative access to biosignal datasets for biomedical research in an Internet of Things (IoT) approach. This is particularly suited for use cases where subjects are placed in free-living environments, examples of which are a living lab, a supermarket shop floor or even creative applications (e.g. a museum or an audience). In these scenarios the aforementioned short-range communication interfaces typically exhibit multiple shortcomings.

II. Related Work

Pervasive monitoring of physiological data for biomedical research is notoriously associated with wearables, which have seen significant growth both in features and functionality. Consumer devices like the FitBit, JawBone or even the Apple Watch, are focused mostly on a fixed sensor set and high-level features derived from them (e.g. heart rate and physical activity). This has lead to the appearance of enhanced research-grade devices, such as the Empatica E4 [2], the Samsung Simband [3] or the Byteflies Sensor Dot [4].

One problem associated with these devices is the fact that measurements are mostly synced with the vendors private cloud services. Furthermore, the fad connotation and other challenges increasingly hindering wearables [5], motivated the appearance of open platforms for physiological computing, such as the OpenBCI [6] and BITalino [7], which are not limited to proprietary hardware and help leverage novel directions for physiological sensing in biomedical research [8].

In summary, the relevance of physiological data acquisition for multiple dimensions of biomedical research is widely known. Both within wearables and open platforms there is generally a dependency on intermediate devices for data retrieval, and open platforms, in particular, currently lack suitable collaborative and cloud-based support tools. This paper explores one intersection of these spaces, by conducting research on real-time streaming over Wi-Fi directly from a physiological data acquisition device to an open access cloud-based platform designed to support the publishing and access to datasets for reuse in academic research.

III. Implementation

Our research took place in three distinct stages. In the first stage the supporting hardware tools were identified. The implementation of our envisioned interaction modes was done

ISBN 978-90-827970-1-5 © EURASIP 2018

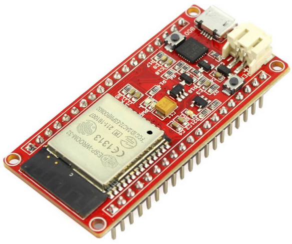

Fig. 1. ESP32 development board layout (extracted from https://www.elecrow.com/esp32-Wi-Fi-ble-board.html).

in the second stage. The third and last stage focused on the cloud interoperability. These stages are described next in more detail.

A. Hardware Base

Due to the maturity of the existing platforms for physiological computing in what concerns sensor design and data acquisition process, we focused the work thus far on the task of creating a general-purpose hardware block capable of adding direct cloud connectivity over Wi-Fi, without requiring an intermediate device. From the currently available chipsets, we chose to base our block on the ESP32, Figure 1, a low-power programmable System-on-Chip (SoC) with dual-core and Wi-Fi connectivity, due to its superior specifications [9].

Within the physiological computing platforms, we chose to use BITalino, Figure 2, due to the fact that is has been recently reported as one of the most complete and versatile tools for biofeedback and quality of life research, even when compared with the latest wearables [10]. The platform integrates multiple individual sensors for bioelectrical and biomechanical data acquisition, linked to a Microcontroller Unit (MCU), which in turn is connected to power management and communication blocks [7].

The latter includes classic Bluetooth or BLE interfaces as options for real-time wireless data streaming to other Bluetooth-enabled devices (e.g. a computer or a smartphone). Our implementation builds upon the ESP32 chipset, introducing end-to-end Internet Protocol (IP) networking capabilities on BITalino (albeit fully generalizable for other platforms). This introduces the possibility to perform direct streaming to a cloud-based server using the maximum number of channels and at the highest sampling rate, without requiring intermediate devices.

B. Operation Modes

Our block, dubbed RepoBIT, was devised as a general-purpose accessory to receive a byte stream over the Universal Asynchronous Receiver/Transmitter (UART) and relay it over an active IP connection. The block can be configured through a purpose-built web interface made available on-board, with the settings stored in Non-Volatile Storage (NVS). Settings are dependent on the selected mode and allow the user to specify the network details (e.g. Service Set Identifier (SSID), Wired Equivalent Privacy (WEP)/Wi-Fi Protected Access (WPA) key, IP address) and the physiological data acquisition system parameters also (i.e., sampling rate, number of channels). RepoBIT can work in one of three possible operation modes:

- **Access Point:** A Transmission Control Protocol (TCP)/IP server socket is created, listening to inbound connections that client devices can use to connect to our block in order to control the acquisition and receive the data stream in real time.
- **Network Client:** A TCP/IP client socket is created, and our block attempts to connect a server address and port configured in the settings; once a connection is successfully established, commands can be sent to our module to control the acquisition and the acquired data is streamed in real time.
- **Stateless Broadcasting:** Data is streamed as Open Sound Control (OSC) messages over User Datagram Protocol (UDP) (quite convenient due to the minimalistic stateless connection communication approach); although primarily designed for audio applications, OSC is a binary format where messages consist of an address pattern, a type tag string and arguments, comparable to the structure of an URL, which for several applications involving physiological data maximizes the ease-of-use.

Regardless of the operation mode, the process of packing and dispatching data through an IP connection can take longer to complete than the rate at which the physiological data is sent by the acquisition device, leading to potential loss of data. To overcome this issue, our RepoBIT implements a UART hardware flow-control mechanism, signaling the transmitter to stop sending data (buffering it internally) while it is executing time critical operations and to resume the data transfer once such operations complete.

C. Cloud Backend

Even though our approach eliminates the need for an intermediate device for data acquisition (e.g. a computer or mobile phone), a way to handle the data stream is still needed, for which we followed a cloud-based approach. To achieve this goal, we built upon RepoVizz [11], a cloud-based repository and data handling backend for synchronous multi-modal, time-aligned recordings [12].

The RepoVizz data representation model is conceptualized by means of a `XML` structure that hierarchically organizes the contents of the multi-modal dataset, namely the metadata and pointers to the actual data files, which in our case are the physiological data time series sent by RepoBIT. Altogether, the `XML` file along with the pointed data files compose what is designated as a datapack, which are then handled as a `ZIP` file within the RepoVizz repository. To interface RepoBIT with RepoVizz, a middleware was developed for the latter that handles the data streams arriving from RepoBIT and converts them to the required RepoVizz datapack format.

Figure 3 presents a schematic representation of the datapack structure created by the RepoBIT middleware, highlighting the hierarchical organization of the `XML` file, which includes three different types of nodes: generic, description and signal nodes.

ISBN 978-90-827970-1-5

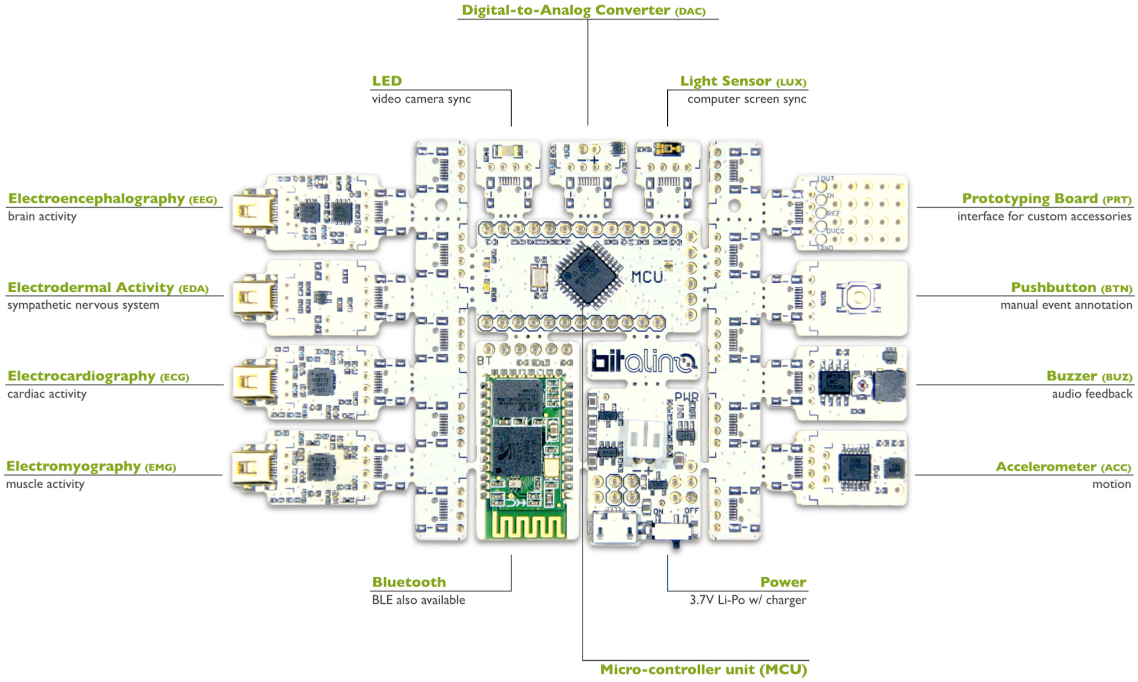

Fig. 2. BITalino (r)evolution board hardware base.

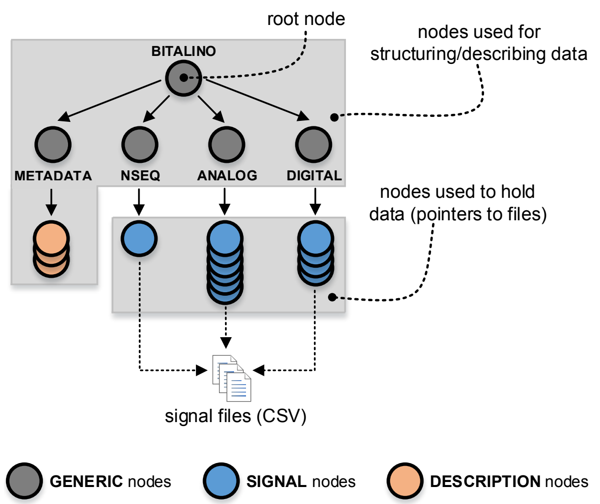

Fig. 3. Overview of the RepoVizz datapack structure used for data acquired with and streamed from the RepoBIT (adapted from [12]).

The first two node types are used for structural and description purposes, and do not hold pointers to data files. In fact, generic nodes are the basic block upon which remaining nodes can be grouped in a logical construction [12]. For this reason, a RepoBIT datapack holds a root node that has four generic children nodes: one for the recording metadata and the rest for storing the actual acquisition data: the samples' sequence numbers and the collection of acquired analog and digital values. The metadata is stored in description nodes, whose purpose is precisely to insert user-defined text information [12]. Remaining data is contained in signal nodes, with each of these nodes containing a reference to one single-channel data array, meaning that for multiple channels there have to be multiple signal nodes grouped within the corresponding generic node.

Each data stream is stored in the matching referenced `CSV` file, that contains comma-separated-values, representing the samples of the signal, with a header in the first line specifying the sampling rate and the minimum and maximum amplitudes of all the samples [12]. The acquisition lasts until the connection to RepoBIT is closed (e.g. by turning off the block). Figure 4 shows the RepoVizz cloud-based interface through which the user can visualize, annotate and share the recorded data with others.

IV. Performance Tests

Given the underlying motivation for this work, geared towards providing a more user-friendly infrastructure to support the creation of biosignal datasets for biomedical research, an important aspect is to characterize the data transfer reliability. As such, we evaluated the performance of the RepoBIT infrastructure for each operation mode. We tested different configurations of the number of acquired channels and sampling rate, as this makes the packet size and the rate at which data is sent vary.

ISBN 978-90-827970-1-5 © EURASIP 2018

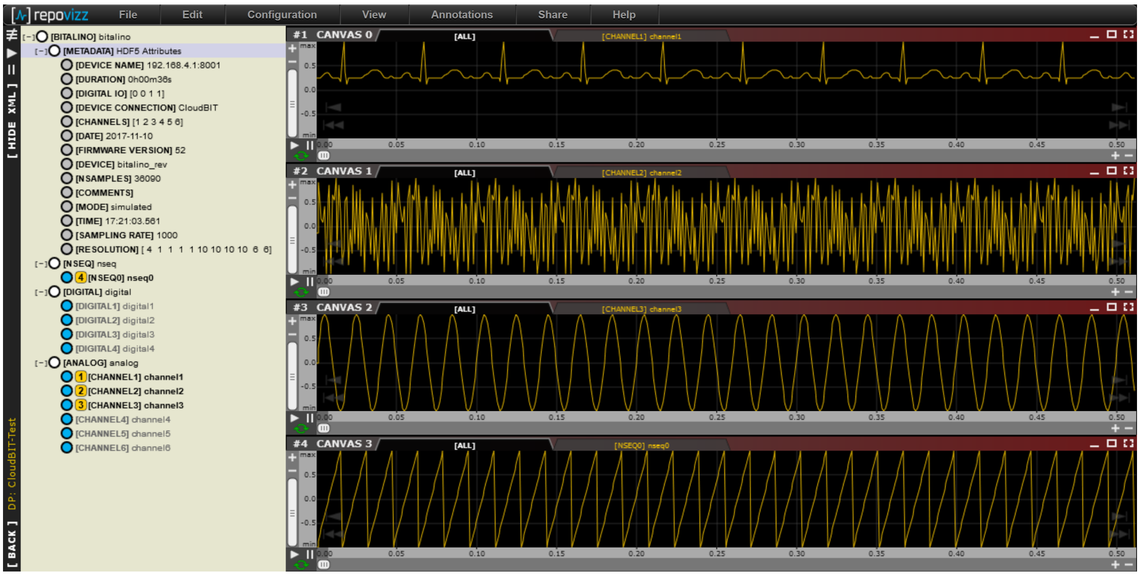

Fig. 4. RepoVizz visualizer for a RepoBIT recording.

The evaluated metrics were the total number of Cyclic Redundancy Check (CRC) fails and the total estimated number of lost packets associated with those fails (based on analyzing the sample sequence numbers), for continuous acquisitions with a total duration of 8 h. The CRC fails enable us to evaluate the occurrence of overflow events on the transmission buffer of the physiological data acquisition system (as a result of the flow control mechanism), which are indicative of issues in the communication between RepoBIT and the acquisition device at the UART level. The lost packets enable us to evaluate networking issues.

Table I summarizes the experimental results. As can be observed, for both TCP-based modes, i.e. Access Point and Network Client, data is reliably streamed without any data loss even in the scenario of 6 analog channels acquired at 1 kHz (which is the most demanding configuration on the BITalino device used as test bed for the physiological data acquisition component). In the UDP mode, i.e. Stateless Broadcast, on the other hand, regardless of the chosen configuration, even when using the less demanding configuration for BITalino where only 1 channel is sampled at 1 Hz, data may be loss both at the UART and at the networking level.

The CRC fails (loss of packets at the UART level) is explained by the time required to build the OSC messages on the RepoBIT block, which occasionally leads to an overflow in the transmit buffer on BITalino, while the loss of packets at the network level is explained by the fact that the UDP protocol is not connection oriented (hence sample delivery is not guaranteed). Considering that 8 h of data acquired at 1 kHz generates approximately 480000 packets, the data loss can be considered to be fairly acceptable, as expressed by the CRC fail of 0.34% and 1.66% of loss packets. However,

TABLE I
PERFORMANCE OF THE REPOBIT INFRASTRUCTURE WHILE ACQUIRING DATA FOR 8 h IN LIVE MODE.

		Operation mode / # of channels			
		Access Point	Network Client	Stateless Broadcast	
	Sampling Rate [Hz]				
# of CRC fails	1	-	-	5	10
	10	-	-	-	286
	100	-	-	-	1019
	1000	0	0	-	1622
# of lost packets	1	-	-	58	49
	10	-	-	-	1558
	100	-	-	-	5324
	1000	0	0	-	7945

when choosing the operation mode the trade-off between the simplicity of the communication protocol and the reliability needs to be taken into account.

V. CONCLUSIONS AND FUTURE WORK

In this paper we present RepoBIT, an infrastructure to support cloud-based physiological data acquisition and handling without requiring an intermediate device such as a computer or a mobile phone. Multiple operation modes have been implemented, targeting both high performance (TCP-based) and broader interoperability (OSC over UDP) use cases. The latter, in particular, introduces a stateless connection and minimalistic approach, which is quite user friendly. In addition, data is streamed as OSC messages, well supported across programming languages, making the block easily interfaceable with third-party tools and usable in a multitude of operating environments.

Experimental evaluation allowed us to conclude that both TCP-based modes are reliable for long-term data acquisitions (8 h), even in the most demanding configuration of the physiological data acquisition system used as a testbed. As for the UDP-based mode, despite the data loss, it has shown acceptable performance. Overall, RepoBIT appears to provide a suitable infrastructure to address the goals that motivated our research. Future work will focus on the validation in real-world use cases, namely through the deployment in environments such as living labs, museums or supermarkets, and exploring the possibility to interface with RepoVizz directly at the hardware level.

ACKNOWLEDGMENT

The authors would like to thank Panos Papiotis and Sebastián Mealla from the Universitat Pompeu Fabra Music Technology Group for their support in the topics related with the RepoVizz repository developed at their institution. This work is funded by FCT/MEC through national funds and when applicable co-funded by FEDER PT2020 partnership agreement under the project UID/EEA/50008/2013 "SmartHeart".

REFERENCES

[1] S. Fairclough and K. Gilleade, *Advances in Physiological Computing*, ser. Human-Computer Interaction Series. Springer-Verlag London, 2014.

[2] M.-Z. Poh, N. Swenson, and R. Picard, "A wearable sensor for unobtrusive, long-term assessment of electrodermal activity," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 5, pp. 1243–1252, 2010.

[3] Samsung, "Simband digital health device," (last accessed on Feb-2018), simband.io/documentation/simband-documentation/.

[4] Byteflies, "Byteflies wearable health," (last accessed on Feb-2018), byteflies.com/science.

[5] W. Kim and J. Choi, "Challenges for wearable healthcare services," *Int'l Journal of Web and Grid Services Archive*, vol. 12, no. 4, pp. 407–417, 2016.

[6] J. Frey, "Comparison of an open-hardware electroencephalography amplifier with medical grade device in brain-computer interface applications," in *Int'l Conf. on Physiological Comp. Syst.*, 2016, pp. 105–114.

[7] H. Silva, A. Fred, and R. Martins, "Biosignals for everyone," *IEEE Pervasive Computing*, vol. 13, no. 4, pp. 64–71, 2014.

[8] H. P. da Silva *et al.*, "Off-the-person electrocardiography: Performance assessment and clinical correlation," *Health and Technology*, vol. 4, no. 4, pp. 309–318, 2015.

[9] Espressif Systems, "ESP32 Tech. Ref. Manual," 2017.

[10] P. N. *et al.*, "A review between consumer and medical-grade biofeedback devices for quality of life studies," in *Recent Adv. in Inf. Syst. and Tech.* Springer, 2017, vol. 570, pp. 275–285.

[11] Universitat Pompeu Fabra (UPF) - Music Technology Group (MTG), "Multimodal online database and visualization tool," (last accessed on Feb-2018), repovizz.upf.edu/.

[12] O. M. *et al.*, "repovizz: a framework for remote storage, browsing, annotation, and exchange of multi-modal data," in *ACM Int'l Conf. on Multimedia*, 2013.