BioStream: A System Architecture for Real-Time Processing of Physiological Signals

Amir Bar-Or, David Goddeau, Jennifer Healey, Leonidas Kontothanassis, Beth Logan, Alex Nelson, JM Van Thong
Cambridge Research Laboratory
HP Laboratories Cambridge
HPL-2004-128
August 2, 2004*

scalable, streaming, physiological monitoring, electrocardiogram, multi-sensor

Continuous monitoring of physiological signals has the potential to greatly improve the quality of life of patients with chronic diseases. Recent advances in sensor technology make the capture of such signals possible. In this paper, we present BioStream, a real-time, operator-based software solution for managing physiological sensor streams. It is built on top of a general purpose stream processing software architecture. The system processes data using plug-in analysis components that can be easily composed into plans using a graphical programming environment. The architecture is scalable, allowing implementation on systems ranging from desktops to server farms. It guarantees real-time response and data persistence in a distributed environment. We apply this architecture to the problem of multi-patient, real-time, physiological signal monitoring, analysis, indexing and visualization.

* Internal Accession Date Only
To be published in and presented at the IEEE Engineering in Medicine and Biology Society Conference (EMBS), 1-5 September 2004, San Francisco, CA
Approved for External Publication

1 Introduction

Recent advances in sensor technology allow continuous, real-time ambulatory monitoring of multiple patient physiological signals including: electrocardiogram (ECG), body temperature, respiration, blood pressure, oxygen levels, and glucose levels [1, 2]. These systems are conveniently packaged as a single product and could be used to give a complete picture of the patient health. However they generate large amounts of patient data that must be intelligently analyzed and archived to be most useful.

In previous studies, home telecare systems using non-continuous monitoring and telephone data transmission [3, 4] have already proven effective in the management of chronic diseases such as congestive heart failure (CHF) and hypertension. Continuous monitoring would allow telecare systems to accommodate a larger number of pathologies including asymptomatic conditions like atrial fibrillation for which intermittent monitoring is not sufficient. Technology that would allow healthcare providers to deploy, configure, and manage such monitoring systems, would provide a tremendous service to the healthcare industry while at the same time improving the quality of life for thousands of patients.

BioStream is a system for real-time processing of physiological signals, built on a general purpose, novel streaming data processing software architecture. The system integrates signal analysis and visualization with a searchable database archive, similar to the approach proposed in [5].

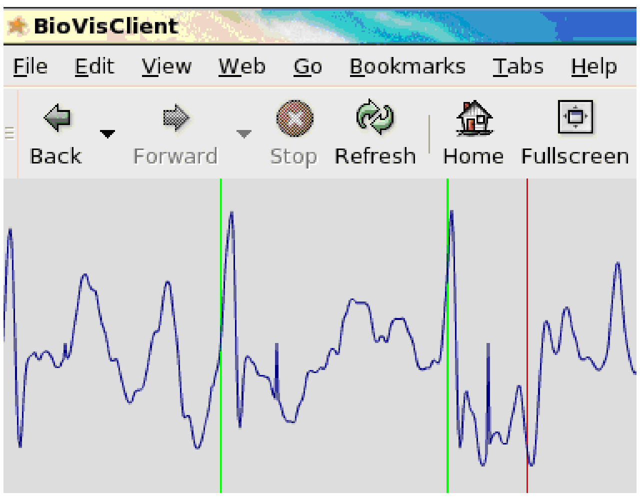

Figure 1. *The BioStream system allows streaming physiological data to be visualized in real-time along with QRS peak detections. This figure shows two normal sinus rhythm beats and one pre-mature ventricular beat.*

The BioStream system can be used for real-time remote monitoring and alert applications, long-term studies, and data mining. Using real-time analysis on incoming streams, the system can be used to detect life-threatening events, such as heart attacks and strokes, allowing an alarm to be sent to a call-center where medical staff could take appropriate action. In addition, web-based visualization of the streaming sensor data allows caregivers to perform remote assessments if patients call in to report feeling symptomatic. Figure 1 shows an example of web-based remote access to cardiac signals using the BioStream system. This figure shows two normal and one premature ventricular beat from an ambulatory ECG recording.

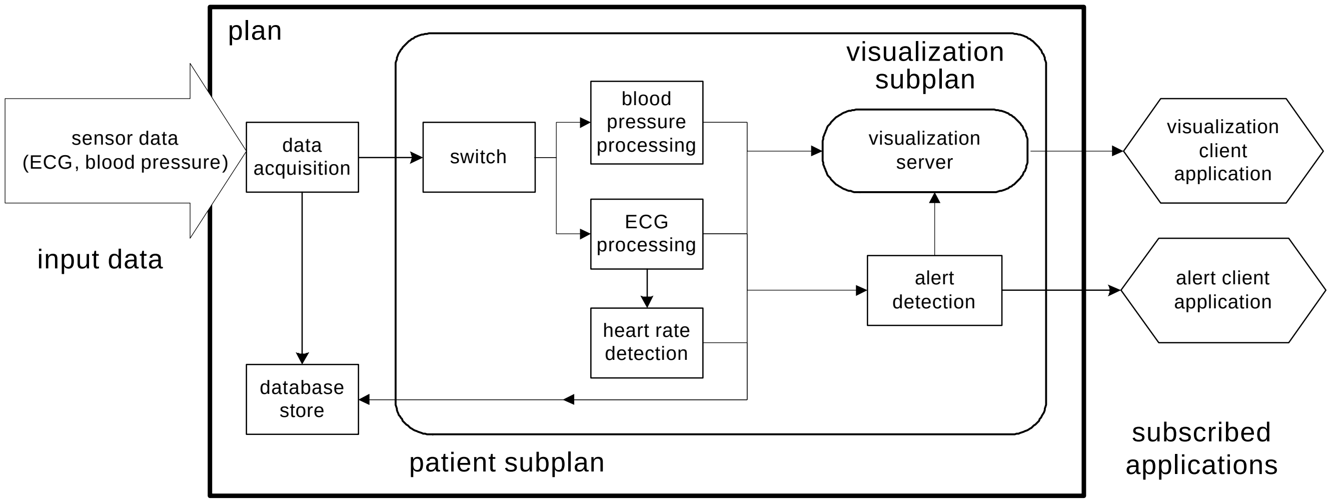

Figure 2*: A BioStream patient plan for processing multiple sensor channels. The plan acquires data and duplicates the stream sending one stream to the database back end and a second to the patient subplan where real-time analysis is performed. The visualization server is activated on alert or on request.*

The BioStream system also allows long-term home patient monitoring using pre-programmed data plans. These plans can be used to detect specific conditions such as atrial fibrillation that can occur infrequently and have non-specific symptoms. Short visits to the hospital or non-continuous telecare monitoring can not reveal the true state and nature of such diseases. A data plan can also be used to track changes in a patient's condition over time. This feature can be used to measure how much a patient's condition is deteriorating due to a known pathology or improving in response to a new therapy. Continuous monitoring allows problems to be detected early so that preventive actions can be taken. In addition, clinicians can use BioStream as a research tool to test hypotheses through the mining of past data by creating a test plan for their hypothesis and "re-playing" archived data through the plan. The re-play feature can also be used to check a current patient's historical record for instances of a pathology that was not originally anticipated.

2 Methodology

BioStream is an operator-based system that collects, analyzes, correlates, and stores streaming information. The system is highly customizable and can be tailored to the special needs of each patient by trained medical staff. The system can handle a large number of patient plans, each with its own set of sensors and analysis algorithms. It can also simultaneously handle patients with varying medical conditions that require different levels of guaranteed quality of service. The system is designed to be distributed and scalable. It allows doctors to access the power of multiple server grade computers for data mining searches and can scale from monitoring one to thousands of patients.

In order to provide a solution that is suited for a wide range of medical scenarios both in terms of cost and functionality, the BioStream system has been designed to have the following properties:

Flexibility –The system easily integrates multiple types of sensors and analysis algorithms in a common framework.

Persistence – Patient physiological data is reliably stored in searchable data archives, allowing more thorough analysis.

Real-time and latency guarantees – The system ensures that all the data streams are processed in real-time at the cost of bounded latency[1].

Scalability – The architecture allows implementation on hardware systems ranging from a single PC handling a few individuals (e.g. elders in their homes), to large clusters or parallel systems handling thousands of patients (e.g. a hospital, medical centers, telemonitoring services).

Fault tolerance – The system is designed to be resilient to server and network failures, ensuring that data is processed in a timely fashion and backed up. For high-risk patients, the monitoring service must be highly available and provide life saving alerts reliably.

The BioStream system supports a programming paradigm based on computational elements termed *operators* in a similar way to [6, 7, 8]. Operators can range from general-purpose data manipulation and control tasks that resemble database operators such as *filter* or *join*, to domain specific algorithms for analyzing medical data, such as heart beat classification. Operators are required to conform to a well-defined programming interface which allows the system scheduler to manage them efficiently, without posing too much of a burden on the operator developer. It is therefore easy to add new operators into the system as new algorithms for the analysis of physiological signals are developed.

Data from the sensors is streamed into the system through a directed graph of operators, which performs the appropriate analysis sequence on the data. The graph of operators that processes the signals of a single patient is called the patient plan. Operators communicate via streams. Streams provide flexibility in operator scheduling, hide the details of communication from the operators, and allow operators to be unaware of the physical location of their upstream and downstream operators. Stream communication mechanisms range from shared memory, to formatted messages over TCP/IP, to XML messaging over HTTP. The BioStream execution engine performs the operator scheduling and runtime load balancing over the nodes in a distributed environment. Parts of a plan, or sub-plans, can be replicated and dynamically instantiated to achieve data parallelism and flexibility.

Figure 2 shows an example of a plan for processing ECG and blood pressure sensor data. On the left side, data from the sensors are acquired by the system. The outer rectangle is the plan that is executed by the BioStream engine. The inner rounded rectangles are subplans, which provide a convenient grouping mechanism for commonly occurring tasks that require multiple operators. For example the patient subplan would be instantiated multiple times, once for every patient. Finally the small rectangles are the operators performing the analysis. The arrows connecting the operators are the data streams responsible for inter-operator communication. In this particular plan, blood pressure and ECG data are stored in a database and analyzed by domain-specific operators. The annotations produced by the analysis are then stored in the same database and also forwarded to alert and visualization operators. This allows the

---

[1] In this context, latency is the amount of elapsed time between the occurrence of an event and its detection by the system.

medical personnel monitoring the patient to view both the annotations and the original signal data to make their diagnosis.

New patient plans can be easily designed with a graphical programming environment, as shown in Figure 3, or generated by a program. Plans are saved as XML files, which can be subsequently loaded, compiled and executed by the BioStream engine. Runtime scheduling and optimization details are hidden from the user. Plan execution is guaranteed to be real-time at the cost of bounded latency. This implies that while the system can keep up with the flow of data from sensors without falling behind, the time between the occurrence of an event and its detection by the system can sometimes be significant due to buffering and operator processing constraints. Even though the BioStream vision is primarily to provide a health monitoring solution, it has been designed as a general-purpose infrastructure for real-time stream processing applications and thus can be easily adapted to target applications from other domains such as media processing and sensor networks. This allows a patient's physiological signals to be augmented with additional channels of audio and video to keep track of the wearer's context and augment patient diaries.

3 Discussion

Most commercial products for home health care monitoring have focused on either ambulatory Holter monitoring [9] or in-home blood pressure, weight and questionnaire readings [2, 10]. Currently, 24-hour Holter recordings are analyzed by trained technicians who select small segments of data from the recording and submit them to physicians along with a summary report. The majority of the recording is never seen by the physician and the data is usually only stored on tape for a requisite holding period before being discarded. The data are not stored in an active database where they can be easily re-analyzed. This prevents doctors from finding patterns that may have been developing over time, perhaps in response to a new therapy. In contrast, the BioStream system allows physicians full access to both archived and real-time data. The same graphical interface that was used to specify plans for real-time analysis can also be used to create plans for the analysis of historical data. This makes it easy to re-examine a patient record against new hypotheses, and to mine data across patients for trends.

The success of systems like BioStream is critical if we are to constrain the rapid increase in the cost of healthcare. Most of this cost is due to the fact that current healthcare systems focus more on curing disease than preventing it. Cost savings from systems like BioStream stem from three sources. First is the use of computers to analyze large volumes of data to detect pathologies that would have otherwise remained unnoticed, allowing early diagnosis and treatment. Second is the use of remote monitoring to allow physicians to adjust patient medications in a timely manner without the need for frequent hospital visits. Finally, the use of commodity components and of an intuitive graphical programming environment reduces the training, maintenance, and administrative costs of running the BioStream system.

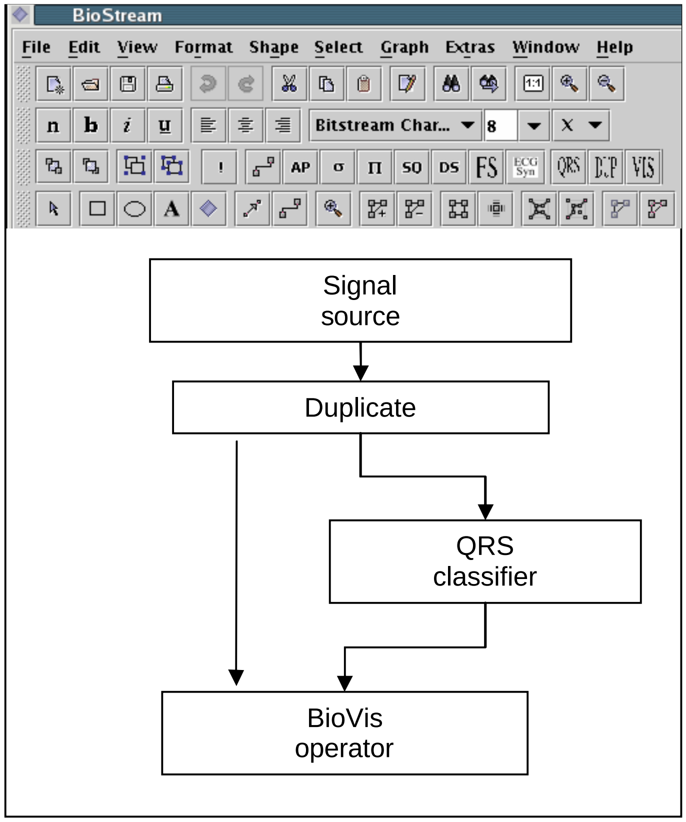

Figure 3. *The BioStream graphical programming environment and a sample data plan. This plan duplicates the data stream from an ECG sensor, uses an operator to detect and classify the QRS complex to create a beat annotation stream and pipes both streams to the BioVis operator so that the signal and annotations can be viewed concurrently.*

4 Current Status and Future Work

We have completed a first prototype that allows monitoring of ECG signals and detection of simple pathologies by classifying QRS complexes. The classification of each heartbeat generates a stream of R-R interval time series metadata that is consumed by the visualization client along with the raw signal to produce an annotated display. This R-R interval time series stream can also be consumed by pathology detection algorithms, such as those for atrial fibrillation which analyze inter-beat interval variation. The visualization server allows Macromedia Flash client applications to monitor the annotated ECG signals on various platforms, ranging from workstations to PDAs.

The goal of the BioStream project is to support multiple sensors and patients, and to automate analysis by incorporating known and novel algorithms as operators in a generic stream processing framework. Our system is intended to be used with an ambulatory monitoring system which will transmit data back to a central aggregator using a low-power wireless protocol. We are currently in the process of developing a pilot study to detect atrial fibrillation in collaboration with a medical institution. In preparation for this study we are developing HIPPA compliant protocols for secure data transmission and storage.

Acknowledgment

We would like to acknowledge the contribution of Nitzan Peleg and Yuval Sherman (HP Haifa Research Lab) for their significant work on system design and implementation in the early stage of the project.

References

[1] Marculescu et al., "Ready to Wear," *IEEE Spectrum.* Vol. 40, No. 10, Oct. 2003, pp.28-32.

[2] A. F. Cardenas, R. K. Pon and Robert B. Cameron, "Management of Streaming Body Sensor Data for Medical Information Systems," in Proc. *METMBS 2003,* Las Vegas, NV, pp. 186-191.

[3] B. G. Cellar, N. H. Lovell and J. Basilakis, "Using information technology to improve the management of chronic disease. *Med. J. of Australia*, 2003; 179 (5): 242-246.

[4] N. Maglaveras, I. Lekka et al., "Congestive Heart Failure Management in a Home-Care System through the CHS Contact Center," *IEEE Computers in Cardiology* 2003, vol. 30, pp.189-192.

[5] G. C. Harfst, A. Qureshi, J. V. Guttag, D. Curtis, "Querying Medical Sensor Streams," *MIT Laboratory for Computer Science Research.* Abstract.

[6] D. Abadi et. al. "Aurora: a new model and architecture for data stream management". *VLDB Journal*. Vol. 12(2):120-139. 2003

[7] D. Tanguay, "A software framework for developing high-performance streaming media applications". *Private communication*, HP, June 2003.

[8] J. Widom, R. Motwani et. al. "Query Processing, Resource Management, and Approximation in a Data Stream Management System". In Proc. *CIDR 2003.* pp 245-256.

[9] J. Cai, S. Johnson, G. Hripcsak. "Generic data modeling for home telemonitoring of chronically ill patients". *Proc AMIA Symp 2000:116-120*.

[10] http://www.healthhero.com/