MIMIC II: A Massive Temporal ICU Patient Database to Support Research in Intelligent Patient Monitoring

M Saeed[1, 2], C Lieu[1], G Raber[2], RG Mark[1]

[1]Harvard – MIT, Cambridge, MA, USA
[2]Philips Medical Systems, Andover, MA, USA

Abstract

Development and evaluation of Intensive Care Unit (ICU) decision-support systems would be greatly facilitated by the availability of a large-scale ICU patient database. Following our previous efforts with the MIMIC (Multi-parameter Intelligent Monitoring for Intensive Care) Database, we have leveraged advances in networking and storage technologies to develop a far more massive temporal database, MIMIC II. MIMIC II is an ongoing effort: data is continuously and prospectively archived from all ICU patients in our hospital. MIMIC II now consists of over 800 ICU patient records including over 120 gigabytes of data and is growing. A customized archiving system was used to store continuously up to four waveforms and 30 different parameters from ICU patient monitors. An integrated user-friendly relational database was developed for browsing of patients' clinical information (lab results, fluid balance, medications, nurses' progress notes). Based upon its unprecedented size and scope, MIMIC II will prove to be an important resource for intelligent patient monitoring research, and will support efforts in medical data mining and knowledge-discovery.

1. Introduction

The modern intensive care unit (ICU) is undergoing a significant evolution that is being driven by rapid advancements in technology as well as changes in healthcare policy. The growth of computer networking and information storage technologies has been accompanied by an increase in the complexity and sheer number of biosensors used in the ICU; thus generating large amounts of data that must be rapidly interpreted. Moreover, there is a critical shortage of ICU clinicians and nurses that are needed for an increasingly sicker ICU patient population. Healthcare staffs are now confronted with a daunting "information-overload" challenge as they try to assimilate and interpret multiple streams of data that reside on different computer systems. Intelligent patient monitoring (IPM) systems are an active area of research. IPM systems attempt to fuse data from different data streams, detect multiparameter changes suggestive of dangerous pathophysiologies, and display alarms and hypotheses in clinically meaningful frameworks. Successful development and refinement of ICU IPM systems requires the use of a representative ICU patient database. Our initial efforts in developing the MIMIC (Multiparameter Intelligent Monitoring for Intensive Care) database resulted in the collection of approximately 100 ICU patient records [1]. Each record included several channels of real-time waveforms (multi-lead ECGs, blood pressures), hemodynamic parameters such as heart rate (HR), blood pressure (BP), and respiration. Furthermore, each patient record included: fluid balance, continuous and drip-medications, lab results, and progress notes. Patients were prospectively chosen for monitoring based on the likelihood of experiencing hemodynamic instability. Each patient record was typically 24 to 48 hours in length. IPM research is an inherently multidimensional problem; thus requiring large-scale training set. Due to the networking and storage technologies existing in 1995, the total number and individual size of each patient record was insufficient. Furthermore, a well-characterized physiologic database must also contain a sufficient number of episodes or events of interest to IPM researchers and developers. The prospective nature of MIMIC I resulted in capturing few such episodes.

We have now embarked on a new effort to develop MIMIC II; leveraging advancements in networking, storage, and relational database technologies to develop an ICU physiologic database of unprecedented size and scope. In this paper, we describe the methodology used to collect ICU patient records. The massive size of the MIMIC II database presents a significant challenge in organizing, indexing, searching, and annotating individual records. We also give an overview of a suite of novel relational database and signal processing tools that will be vital for managing and editing MIMIC II. Using these tools, we provide a brief summary of the content of MIMIC II that has been hitherto collected.

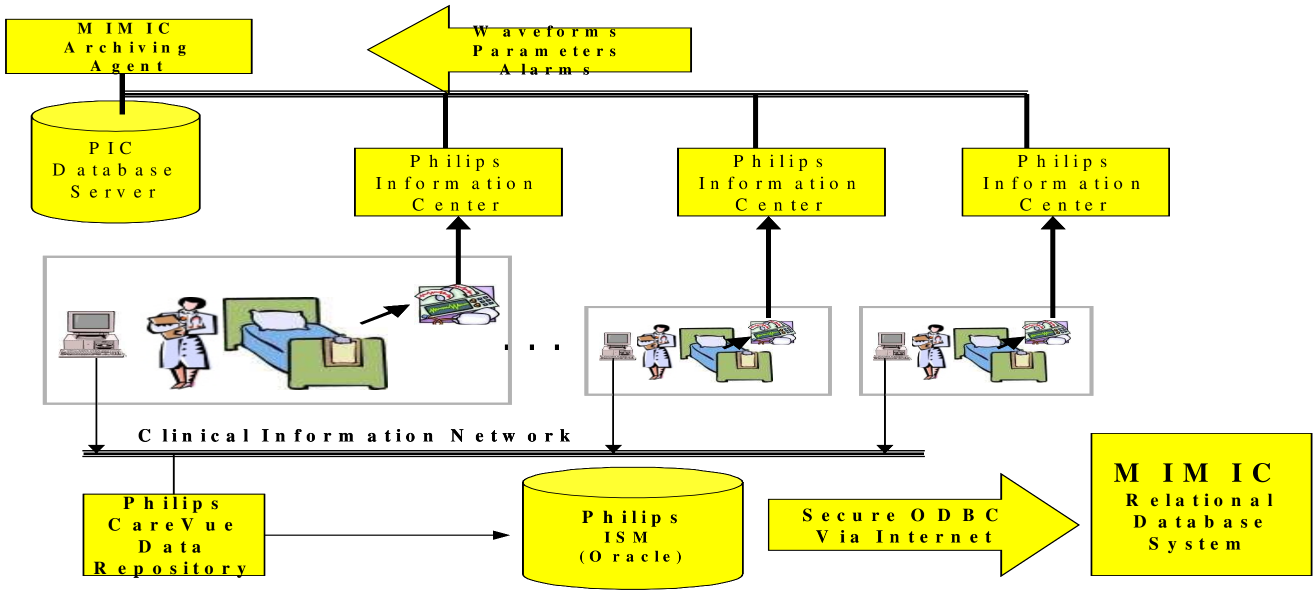

Figure 1: MIMIC-II Data Collection Architecture

2. Data collection architecture

We chose to monitor ICU patients admitted to an 8-bed Medical Intensive Care Unit (MICU) and an 8-bed Coronary Care Unit (CCU). Each patient record commenced with a patient admission and ended with a final discharge. Each record consisted of four continuously monitored waveforms (2 Leads of ECG, Arterial Blood Pressure, and Pulmonary Artery Pressure) sampled at 125 Hz, 30 1-minute parameters (HR, BP, SpO2, Cardiac Output), and monitor-generated alarms and in-ops. The waveforms and parameters were originally sourced from Philips CMS bedside patient monitors (Philips Medical Systems, Andover, MA). This data was then transmitted to a Philips Information Center Database Server (PICDBS). With the assistance and cooperation of the manufacturer, a customized archiving agent was developed to query the PICDBS. The archiving agent, equipped with a 40 Gigabyte hard drive, continuously retrieved and stored the waveform and parameter data from the MICU and CCU. Furthermore, alarms generated by the bedside monitor and nurse central station were also archived. At approximately two-week intervals, all completed records were downloaded from the archiving agent and subsequently purged from the archiving agent. The downloaded records were permanently stored onto IEEE 1394 (FIREWIRE) storage drives.

To capture the relevant clinical information for each monitored patient, we leveraged the use of the Philips Information Support Mart (ISM) that was interfaced with the units' clinical information system, CareVue (Philips Medical Systems). The ISM is a relational database that warehouses clinical information such as lab results, nurses' text notes, medications, fluid balance, and patient demographics. Customized scripts were written in SQL to query the ISM database and retrieve the clinical information for each monitored patient. The retrieved ISM records were stored in the MIMIC II database that was powered by Postgres. The data schema for the clinical information in MIMIC II was based on the schema reported in the Philips ISM User's Guide [2].

3. Data processing and fusion

The 1-minute parameter data were processed using wavelet analysis to identify potentially clinically relevant events [3]. The wavelet coefficients indicative of significant trend dynamics were stored in a relational database for subsequent retrieval and browsing. The continuous medication data and the fluid balance (urine output and IV fluid input) were abstracted from Postgres database tables and processed into time series and aligned with the 1-minute parameter data.

4. Relational database tools

The transfer, storage, and management of clinical information extracted from the CareVue clinical information system was enabled through our customized MIMIC relational database. The web-enabled relational database system was comprised of a Postgres Database Server. The web server chosen for use was modAolServer and the ARSDigita Community System with Linux [4]. SQL scripts were written in SQL and TCL for the purpose of de-identifying the clinical data. Patient information such as names, medical record number, and social security number were automatically removed or modified in their respective fields in the database tables as well as in the nurses' text notes. The patient records were downloaded from the hospital system and simultaneously de-identified.

A text search engine was created to allow users to query the MIMIC database for key words. Queries are supported for searching the text notes as well as specified tables. Complex queries are supported to search for patterns of interest within patient records.

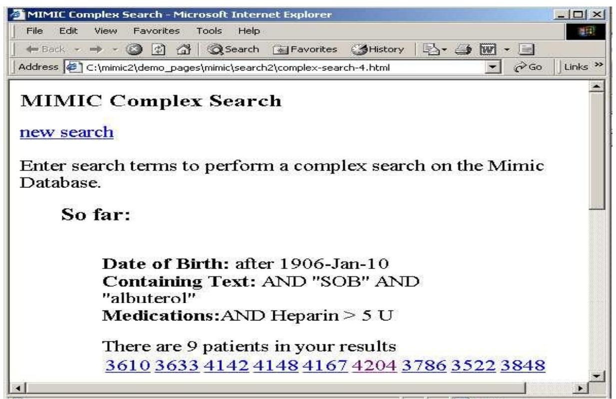

Figure 2: Example of complex text search.

To allow for future modifications of the MIMIC database, administrators can dynamically add new tables to be incorporated into the overall data schema. For example, as we eventually hope to include physician-overviewed annotations of MIMIC records, tables can be created for annotations and seamlessly integrated with the MIMC database.

Users are also allowed to customize how they choose to view the clinical data. For example, if a user is interested in a temporal view of a patient record, the data from each table (i.e. medications table, fluid balance, text notes) is displayed in a format based upon chronological order. However, if a user is only interested in a certain field, such as medications, then a user may select a view of the medication data only.

5. Results

We have thus far captured approximately 800 ICU patient records. The MIMIC II database includes nearly 120 Gigabytes of data. Typical records are approximately 100 hours in length. We include sample records from the MIMIC II database in figures 3 and 4. These samples illustrate the richness of the data as well as the potential insight gained by fusing physiologic waveforms and parameters with clinical information. Figure 3 is a plot of the longest MIMIC II record captured thus far. The capture of long-term multiparameter physiologic data and clinical information (fluid balance, medications) may allow for more quantitative techniques to be developed for ICU decision-support systems. Figure 4 includes the recording of vital parameter time series and information extracted from clinical notes, such as the occurrence of 3 episodes of supraventricular tachycardia (SVT). By correlating changes in hemodynamic variables to changes in heart function or medications (such as lasix), a greater insight into the physiology of a patient can be gained.

6. Conclusions

As the MIMIC II database grows in size and is refined, we believe that it will be an important resource to IPM researchers. However, in order to facilitate its use by the research community, there are several important tasks that must be completed. ICU patient records must be annotated so that clinically significant events are identified and can be correlated to changes in acquired data streams. We are hopeful that the nurses' text notes contain sufficient information to allow for the identification of hemodynamically significant events. We will investigate automated techniques of extracting useful information from text notes to accelerate the annotation process.

We believe that we have identified several powerful database and signal processing tools that will be of great value in the management and indexing of MIMIC II. As these tools are refined and integrated with annotations, MIMIC II may become a valuable tool for research and medical education.

Acknowledgements

The National Aeronautics and Space Administration supported this work through the NASA Cooperative Agreement NCC 9-58 with the National Space Biomedical Research Institute. Mr. Saeed was supported through a fellowship from Philips Medical Systems.

References

(1) Moody G, Mark RG. A Database to Support Development and Evaluation of Intelligent Intensive Care Monitoring. In: Computers in Cardiology 23:657-660 1996.
(2) CareVue Clinical Data Management Information Support Mart User's Guide. Agilent Technologies, 2000.
(3) Saeed M, Mark RG. Efficient Hemodynamic Event Detection using Relational Databases and Wavelet Analysis. In: Proc. Computers in Cardiology, 2001.
(4) Greenspun P. Phil and Alex's Guide to Web Publishing. http://www.arsdigita.com/books/panda/

Address for correspondence.

Mohammed Saeed
E25-505, MIT
77 Massachusetts Avenue
Cambridge, MA, USA 02139
Email: msaeed@mit.edu

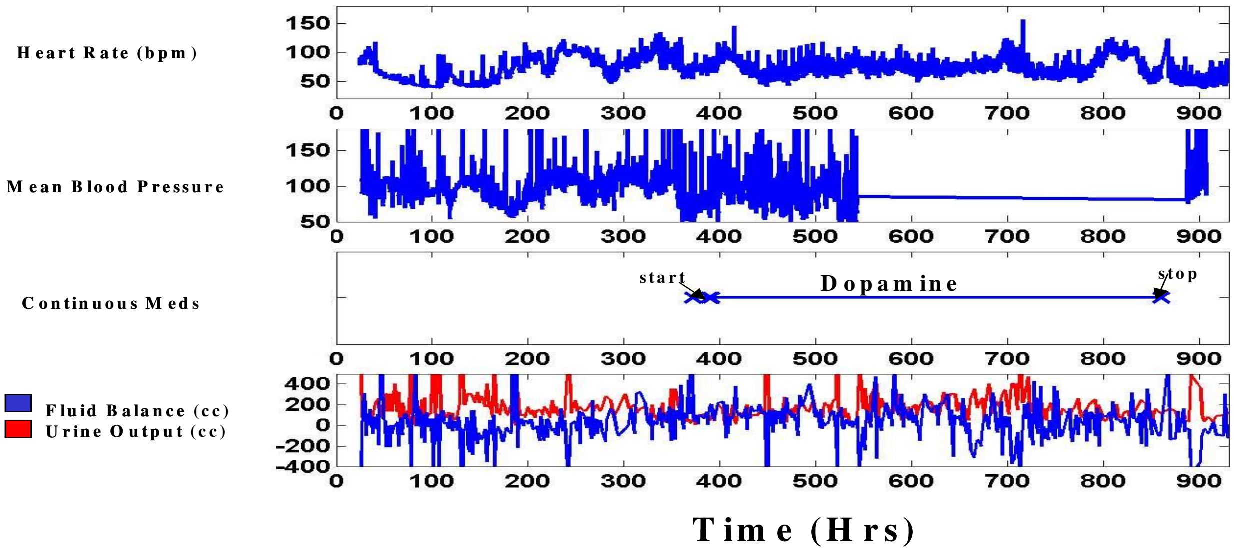

Figure 3: Example record exceeding 900 hours with multiparameter data and clinical information (fluid balance and medications).

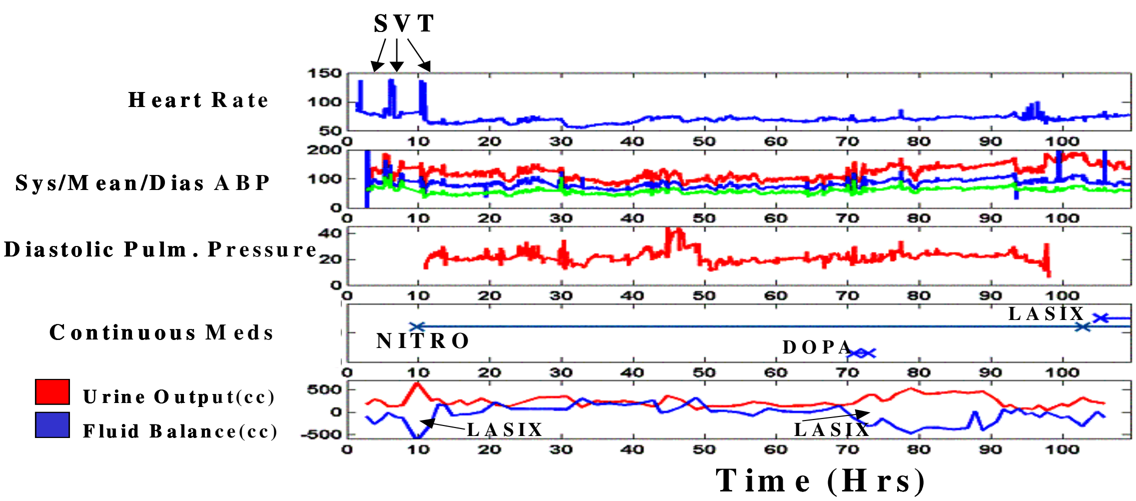

Figure 4: Record with significant hemodynamic events such as episodes of SVT and fluid balance shifts.