Real-Time Analysis for Intensive Care

Development and Deployment of the Artemis Analytic System

BY MARION BLOUNT, MARIA R. EBLING, J. MIKAEL EKLUND, ANDREW G. JAMES, CAROLYN McGREGOR, NATHAN PERCIVAL, KATHLEEN P. SMITH, AND DABY SOW

The lives of many thousands of children born premature or ill at term around the world have been saved by those who work within neonatal intensive care units (NICUs). Modern-day neonatologists, together with nursing staff and other specialists within this domain, enjoy modern technologies for activities such as financial transactions, online purchasing, music, and video on demand. Yet, when they move into their workspace, in many cases, they are supported by nearly the same technology they used 20 years ago. Medical devices provide visual displays of vital signs through physiological streams such as electrocardiogram (ECG), heart rate, blood oxygen saturation (SpO_2), and respiratory rate. Electronic health record initiatives around the world provide an environment for the electronic management of medical records, but they fail to support the high-frequency interpretation of streaming physiological data. Recent medical research has reported that potentially life-threatening conditions such as nosocomial infection [1], pneumothorax [2], intraventricular hemorrhage [3], [4], and periventricular leukomalacia [5] exhibit early indicators in physiological data (see "Conditions Affecting Patients in an NICU"). These indicators precede the detection of the medical conditions using existing clinical practices.

We have taken a collaborative research approach to address this need to provide a flexible platform for the real-time online analysis of patients' data streams to detect medically significant conditions that precede the onset of medical complications. The platform supports automated or clinician-driven knowledge discovery to discover new relationships between physiological data stream events and latent medical conditions as well as to refine existing analytics. Patients benefit from the system because earlier detection of signs of the medical conditions may lead to earlier intervention that may potentially lead to improved patient outcomes and reduced length of stays. The clinician benefits from a decision support tool that provides insight into multiple streams of data that are too voluminous to assess with traditional methods.

The remainder of this article summarizes the strengths of our research collaboration and the resulting environment known as Artemis, named after the Greek goddess associated with protecting child-bearing women and young children, which is currently being piloted within the NICU of The Hospital for Sick Children (SickKids) in Toronto, Ontario, Canada. Although the discussion in this article focuses on a NICU, the technologies can be applied to any intensive care environment.

Digital Object Identifier 10.1109/MEMB.2010.936454

Research Collaboration Teams

Artemis was designed, built, and deployed by a multiinstitutional, multidisciplinary research team. The research team from IBM T.J. Watson Research Center has an average of more than 15 years of industrial research experience in distributed computing, ubiquitous computing, pervasive health care, and machine learning. They also bring five years experience in building health-care solutions, including two years experience building health-care solutions using a state-of-the-art stream computing platform, which was the result of a five-year research project.

The University of Ontario Institute of Technology (UOIT) team brings to the collaboration expertise in health informatics with more than ten years of research collaboration with clinicians on information technology use in NICUs, event stream processing, and acquisition of data from medical sensors, and 20 years of expertise in data warehousing and data mining. Their research on temporal abstraction is particularly relevant for the real-time processing and temporal data-mining components of the project [6].

The team from the SickKids and the Department of Pediatrics, University of Toronto, has more than 20 years of experience in neonatology and clinical research. This team provides the clinical and medical expertise, guides in interpreting the analytic results, and leads in designing the clinical deployment of Artemis at the SickKids.

Collaborations like this involve more than just a shared technical vision and complementary research teams. The teams had to put in place intellectual property agreements and institutional Research Ethics Board applications that were mutually agreeable and that met the guidelines at each institution.

Evolution of the Artemis Design

The Artemis system design evolved over time as the research teams joined together to define the research project. The informatics research team from the UOIT has many years of experience in collaborating with clinicians on the application of information technologies in NICUs. They had started a path of research on collecting patient data from a large group of patients,

 0739-5175/10/$26.00©2010IEEE

Infection is a very common cause of morbidity and an important cause of mortality for the newborn infant.

and given the known outcomes of the patients, to perform analysis to find correlations between distinctive patterns in the physiological data streams and the onset of medical conditions. One goal of the Artemis project is to assist them in taking the steps in their research to perform temporal data mining on a broad set of physiological data from infants in a NICU. To do this, we need to capture all of the raw physiological data streams from a large number of infants over time. This means that the Artemis system must be capable of 1) interfacing with a broad set of medical devices and 2) storing the raw physiological data from multiple infants at the rate the data are generated.

A second major goal of the Artemis project is to run clinical rules, some of which are derived from the research described earlier, online and in real time. In this way, knowledge of early indicators of medical conditions can be made available to clinicians as soon as they are detected. This means the data must not

Conditions Affecting Patients in an NICU

Nosocomial Infection

Infection is a very common cause of morbidity and an important cause of mortality for the newborn infant. Although many infants acquire their infection around the time of delivery, others acquire an infection while receiving intensive care in the NICU. Nosocomial infections, also called hospital-acquired infections, are infections that are secondary to the original cause for admission into the NICU. The early diagnosis of a nosocomial infection is difficult, because the clinical signs of infection are usually subtle, vague, and nonspecific until the infection is well established. Such infections occur 48 h or more after birth and are caused by pathogens not associated with the mother. Data from the neonatal network indicate that almost 30% of infants born at 25–28 weeks gestation and more than 45% of infants born prior to 25 weeks gestation will experience a serious nosocomial infection while in the NICU (S1). Earlier detection and intervention would be expected to reduce morbidity and may reduce mortality.

Pneumothorax

One to 2% of all newborns have air or gas in the pleural cavity that separates the visceral from the parietal pleura. The lungs are surrounded by a membrane that folds back on itself, with one layer attached to the chest wall and one layer attached to the lungs. The membrane produces a fluid that acts as lubrication so that these layers move smoothly when we inhale and exhale. When air or gas accumulates between these two layers, it is called pneumothorax. Goldberg (S2) showed that the recognition of subtle clinical signs, including increased in systolic arterial blood pressure as well as an increased heart rate and pulse pressure, can lead to earlier recognition of pneumothorax, and therefore, earlier intervention.

Intraventricular Hemorrhage

Intraventricular hemorrhage (IVH) is another common cause of morbidity and mortality for the newborn infant. Approximately 20% of preterm infants less than 1,500 g birthweight develop an IVH. The incidence and severity are inversely proportional to gestational age. The hemorrhages occur during the first few days of life. More than 90% of the IVHs have occurred by the third day of life. Nearly 10% of IVHs occur before delivery. Important risk factors for IVH include extreme immaturity, birth asphyxia, asynchronous breathing of ventilated preterm infants, pneumothorax, and sudden increase in arterial blood pressure (S3). Fluctuations in cerebral blood flow, and especially blood flow through the delicate, fragile blood vessels of the germinal matrix layer, a centrally located region of the brain, are considered to be the dominant cause of IVH. There are many events around the time of birth and during the first week of life, which are associated with fluctuations in cerebral blood flow. The likely cause of most IVHs is a rapid increase in cerebral blood flow occurring after a period of reduced flow.

Periventricular Leukomalacia

Periventricular leukomalacia (PVL) refers to the death of white matter near the cerebral ventricles: 3–4% of premature, very low birthweight (1,500 g or 3 lb 5 oz) infants and 4–10% of those born prior to 33 weeks gestation will develop PVL (S4). The white matter is the inner part of the brain, and periventricular refers to the part of the white matter that surrounds the ventricles. Leukomalacia refers to the softening of the white matter, which quickly leads to death of the brain tissue. PVL typically occurs when a fetus or newborn experiences oxygen deprivation during labor and delivery or at any time after birth. Variations in the oxygen and carbon dioxide content of the blood are involved in the causation of PVL. Stream analysis of physiological data has the potential to detect these variations.

References

(S1) R. Polin and L. Saiman, "Nosocomial infections in the neonatal intensive care unit," *NeoReviews*, vol. 4, no. 3, pp. e81–e89, 2003.

(S2) R.N. Goldberg, "Sustained arterial blood pressure elevation associated with pneumothoraces: Early detection via continuous monitoring," *Pediatrics*, vol. 68, no. 6, pp. 775–777, 1981.

(S3) J.J. Volpe, "Intracranial hemorrhage: Germinal matrix-intraventricular hemorrhage of the premature infant," in *Neurology of the Newborn*, 4th ed. Philadelphia, PA: Saunders, 2001, pp. 428–493.

(S4) J.J. Volpe, "Hypoxic-ischemic encephalopathy: Clinical aspects," in *Neurology of the Newborn*, 4th ed. Philadelphia, PA: Saunders, 2001, pp. 331–394.

The Artemis platform supports the acquisition and storage of patients' physiological data streams and clinical information system data for the purposes of online real-time analytics, retrospective analysis, and data mining.

only be stored in real time, but the data must be processed in real time. IBM Research entered the collaboration with the goal of exploiting a novel streaming middleware system developed in a five-year, multidisciplinary, multiparty research project. It was decided to include the streaming middleware component in the Artemis system to provide an online, real-time processing run-time environment. If we want to execute clinical rules, the neonatologist pointed out that we must incorporate patient data from a number of data sources; streaming physiological data from medical devices was necessary but not sufficient for clinical rules. Therefore, another Artemis system requirement was the need to stream the most up-to-date data from the clinical information management system (CIMS) and the laboratory system.

Artemis would have a set of components to aid in discovering clinical rules and a set of components for executing clinical rules. The Artemis team decided on a goal of creating a closed-loop system, whereby the new clinical rules, parameter values, and clinical rule refinements can be immediately deployed in the run-time component of Artemis. The key to the closed-loop system is to create an ontological relationship between the output of the knowledge extraction component of Artemis and the clinical rule execution component of Artemis. A number of Artemis system requirements are generated from this goal. We need a way to efficiently integrate newly captured patient data into the data mining repository. We need an ontology relating data mining outputs to clinical rules.

These high-level system goals generated the following Artemis design requirements:

- support real-time processing of multiple high-rate physiological data streams using a novel stream processing system
- interface to and stream data from medical devices, the clinical information management system CIMS, and the laboratory system
- store all of the raw physiological data from devices connected to an infant and selected patient data from other sources
- support temporal data mining and other data mining techniques to find relationships, particularly time-based relationships, between patterns and correlations in multiple patient data streams and medical conditions
- scale with respect to the number of data streams and the number of patients connected to the system.

The next section will describe the system that was implemented to meet these requirements.

The Artemis Framework

The Artemis platform supports the acquisition and storage of patients' physiological data streams and clinical information

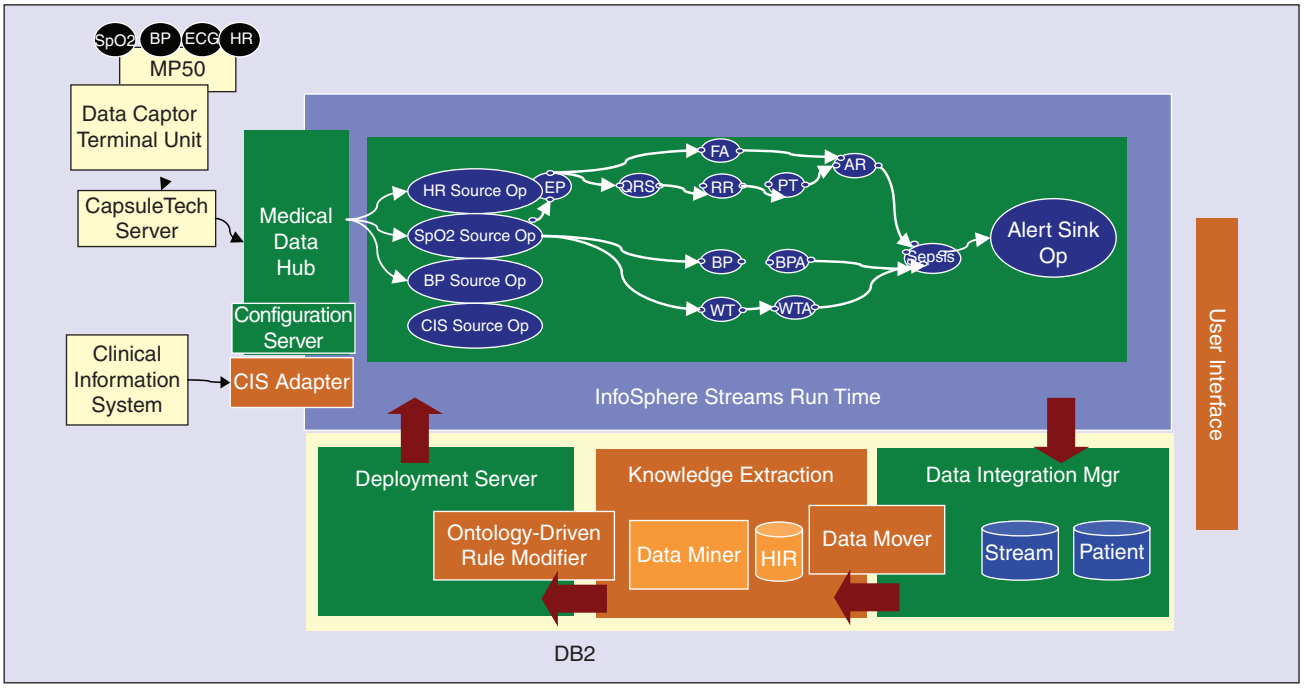

Fig. 1. Artemis system architecture.

The data integration manager allows Artemis to interact with an open database connectivity system.

system data for the purposes of online real-time analytics, retrospective analysis, and data mining. The Artemis system architecture is illustrated in Figure 1.

To meet the goal of being able to interface Artemis to the myriad of medical devices used in intensive care environments, Artemis employs a set of hardware and software elements from Capsule Tech Inc. [7]. Capsule Tech has cables and device drivers for interfacing with more than 450 different types of devices from all of the major vendors. A DataCaptor terminal unit shown in Figure 2, which is located near the medical devices, can connect to eight devices and convert the devices' RS232 output to an Internet Protocol (IP) stream. Alternate RS232 to IP converters have also been tested and found to be compatible with this configuration. By converting the data stream to an IP stream, only the terminal unit has to be located in the busy and crowded intensive care unit. The other Artemis computers can be placed in a secure area of the hospital. All data are forwarded to a Capsule DataCaptor Interface Server that can support up to 500 simultaneously connected devices [8]. Using the Capsule application programming interface (API) and software development tool kit, we implemented a server-based function to filter the data received at the server to extract only data streams necessary for the study, format it for use by Artemis, and send it to the Medical Data Hub. We implemented a configurable Medical Data Hub system consisting of a set of data hubs that receive the aggregated data item from the server and create concurrent data streams for the streaming system. Although a single Capsule Server is sufficient for our deployment, multiple servers can be used to achieve system scalability.

The clinical information system (CIS) adapter interfaces with the clinical information management system CIMS to access the SickKids CIMS patient data and stream the data to the Artemis clinical rules. Information-use protocols at the SickKids do not allow nonclinical access to the CIMS or laboratory database, of which both are Oracle based. If clinical researchers at SickKids require such data for patients satisfying their research study, a database for enrolled/qualifying patients is created in which CIMS data and laboratory database data are replicated every 30 min. Researchers then use these shadow databases for their approved research. As per this protocol, a replica subset database for Artemis was created that is populated for enrolled patients. The CIS adapter extracts the selected data for infants enrolled in the Artemis project and maintains the data in a set of database tables accessible by Artemis. We have implemented system operators to access these interface database tables and stream the updated data to the clinical rule applications.

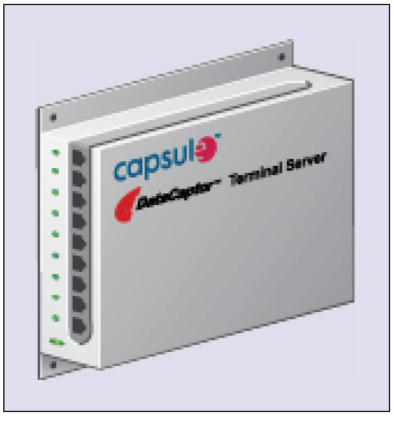

Fig. 2. Capsule DataCaptor terminal unit.

The core of Artemis is a stream computing middleware component, IBM InfoSphere Stream Computing System, which provides scalable processing of multiple streams of high-volume, high-rate data [9], [10]. The conventional approach to processing data streams is to store the data and then immediately analyze the data in near real time. If the processing cannot keep up with the rate, the data are stored and will eventually be analyzed. Stream processing systems are compute-first, store-second systems. In Artemis, processing of data streams and storing of the data are done concurrently. The stream computing system can run on a range of systems from notebooks to supercomputers; thus, it provides Artemis with a very scalable real-time execution environment. An application in streams consists of a set of operator nodes interconnected in a graph. Each operator node inputs one or more streams and produces one or more output streams.

The programming language for the stream computing system is stream processing application declarative engine (SPADE). SPADE is a high-level declarative language for programming the streaming system [11]. It allows a programmer to specify the data streams, operators, and connections between the operators and streams. SPADE language constructs have many similarities with higher-level programming languages like stream structured query language (StreamSQL), yet provide mechanisms to interact with lower-level system programming APIs if needed. SPADE has operators that are specialized for ingesting data from varied data sources, for interacting with external entities through protocols like IP and simple mail transport protocol (SMTP), and for coordinating data streams during processing. Indeed, while SPADE provides a set of built-in stream-relational operators able to perform relational query on data streams, SPADE can also be extended. It allows the developer to specify user-defined operators typically written in C++ or Java, when the need for complex operators arises. These operators can be integrated seamlessly with built-in operators to compose applications. An application specified in the SPADE language is compiled into directed graph processing elements, which can then be instantiated on the system run time. Artemis' clinical rules are implemented as SPADE programs.

The data integration manager (DIM) consists of a set of SPADE operators that have the specialized function to interact with an open database connectivity

The intensive care unit is a very dynamic environment where the care of the patient is paramount.

(ODBC) system. Some operators are specialized to insert a steam element into the database and some are specialized to stream data from a database table to the application. The requirement to store all raw data into the database is achieved by the selective placement of DIM operators in the application graph. Periodically, the data are moved from the DIM database system to a data repository in the knowledge extraction component by the data mover. The data mover gives us the ability to control the frequency of the data movement and to perform some data transformations in preparation for data mining.

The knowledge extraction component uses new, multidimensional, temporal, data stream data mining frameworks and techniques [12]–[14]. These new data mining approaches can be automated but also enable clinicians to perform scientific method-based hypothesis research as an active participant in the multidimensional temporal data mining process. The output of the knowledge extraction component can be a new correlation between a pattern detected in the streaming physiological data and a medical condition or a new set of parameters for an existing SPADE application. We are developing an ontology that relates the output of the knowledge extraction component and the expression of the clinical rules. The ontology-driven rule modifier (ODRM) component, using the ontology, supports the translation of the new rule into a SPADE program as well as the modification of the parameters in an existing SPADE application. This closed-loop feature permits us to dynamically update a running clinical rule or replace the clinical rule with the one that has been shown through analysis and testing to be better.

The deployment server supports the deployment of the new SPADE application into the stream computing run time and maintains information about which SPADE applications are active at any given time. The deployment history can be queried to determine which version of what clinical rule was active for a given patient at a given time. The deployment server has both a graphical user interface and a programmatic interface. The ORDM uses the programmatic interface to drive the redeployment of clinical rules based on the output of the knowledge extraction component. When a patient is enrolled in Artemis, the deployment server deploys the clinical rule applications. All clinical rule applications allow per-patient parameterization.

Programming Clinical Rules in Artemis

A clinical rule is a specification as detailed within an existing clinical guideline, defined anecdotally by a clinician as part of clinical research, or proposed through data mining of a set of conditions in the physiological data streams, laboratory results, and observations of a patient, which if found to be present, holds a strong correlation as a predictor for an impending clinical event and as such should be reported by some means to a nurse or physician. Some clinical rules can be processed by the people. An example of such a rule is ''notify me when an hourly systolic blood pressure (BP) reading exceeds 140.'' The clinician who takes the patient's BP can easily monitor this clinical rule. However, when the clinical rule involves specification of second-to-second changes across multiple data sources of physiological streams, it becomes impractical and sometimes impossible for a human to detect these subtle signals of condition onset [15]. In some cases, the conditions are episodic and do not occur often, and it is unlikely a clinician will be observing when the condition occurs.

Although Artemis can easily monitor the simple clinical rules, it is ideally suited for clinical rules that specify conditions involving multiple data sources and/or high-rate data sources. Data sources such as ECGs and electroencephalograms are time sampled at 500–1,000 Hz and require nontrivial signal processing techniques for analysis. As mentioned in the previous section, Artemis clinical rules are implemented as streaming system applications encoded in the SPADE language. The following paragraphs provide an example of a clinical rule and explain portions of the program for implementing the clinical rule.

Although our testing encompasses simple and complex rules, we provide further details via the explanation of a simple clinical rule that we used to test Artemis. The clinical rule ''if mean arterial BP (MBP) is less than the neonatal patient's current gestational age (e.g., 24 mmHg for 24 weeks gestation) for 20 s or more, and if SpO_2 is less than 85% for the same period of 20 s or more, then a reportable condition is present.'' Prior to commencing our pilot study with SickKids, we utilized data from the companion Targeted Delivery Intervention Study (TARDIS) [14], a randomized control trial of volume-targeted ventilation for resuscitation of preterm babies intubated in the delivery suite. The data were collected as discontinuous segments of at least 2 h of data at a time commencing en route to the delivery suite and then at 12, 24, 36, and 48 h after delivery. For that study, each data set contains multiple concurrent streams of physiological data; however, for the purpose of this case study demonstration, we focused on six, 5-min, patient segments extracted at a reduced data set containing four of these temporal data streams, specifically: 1) ECG; 2) SpO_2; 3) MBP; and 4) BP. The team performing the test of Artemis was blinded to the characteristics of these chosen segments.

We will describe this clinical rule using the four stages of clinical rules that have been observed. Figure 3 illustrates the four stages.

The Adaptation Stage

In this stage, SPADE source operators listen for streaming events sent by the medical data hubs, which in this study are simulators which replay the TARDIS data in faster-than real time. This scenario uses two source operators: one for SpO_2 data and the other for the MBP data.

The Intrastream Analysis Stage

The SpO_2 module and the MBP module operate in parallel. The SpO_2 module contains logic to count the number of times SpO_2 events are below 85% in a 20-s sliding window and issues SpO_2 alerts if this count is greater than a predefined threshold. Meanwhile, the MBP module contains logic to count the number of times MBP events fall below the patient's gestational age, in a 20-s sliding window, and issues MBP alerts if this count is greater than a threshold. The gestational age is a parameter obtained from the CIMS. To access it within this SPADE application, we use the DIM enrich operator.

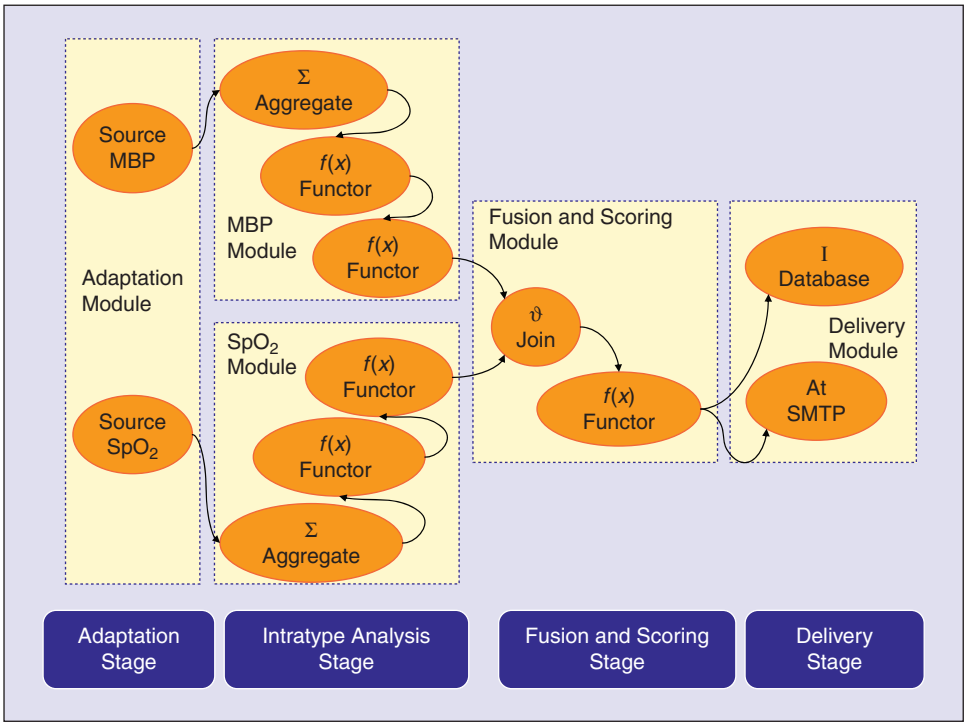

Fig. 3. Examples of stages of sample clinical rule.

The Fusion and Scoring Stage

In the fusion and storage step, the SpO_2 and MBP features are merged using a SPADE join operator. Its output is consumed by another operator that determines whether this is a reportable event for this clinical rule.

The Delivery Stage

The delivery stage externalizes reportable analytic results. In our test setup, we wrote the result to a database table and sent a short message service/e-mail message.

By replaying these traces into the SPADE application described earlier, we accurately detected which patients exhibited the clinical rule. The clinical rule was present for two patients within these six traces: patients 2 and 5. Both were correctly identified by the SPADE application. For the remaining normal patients, no alerts were generated by the SPADE application, as expected.

Figure 4(a)–(c) shows the plots of the results obtained for patient 5. We can clearly see in these graphs that the long dips in MBP and SpO_2 were accurately identified and resulted in instability alerts. For patient 5, we may note that the SPADE code also detected the number of physiological streams that are out of range, as shown in the Figure 4(a).

While clinical rules can be implemented in the SPADE language, there are approaches to integrating other rule representation languages with SPADE. The first approach leverages standard rule representation languages, thus allowing our infrastructure to be compliant with the existing rule languages. One example is the predictive modeling markup language (PMML) [16], a well-established extensible markup language (XML) dialect used to represent prediction rules. The streams team has implemented an operator that accepts PMML models and scores them in real time.

The second approach to represent clinical rules can be used by developers who have proprietary rule representation schemes. In this case, one can leverage the extensibility of the SPADE language to develop user-defined operators capable of interpreting such proprietary rules. SPADE user-defined operators can be implemented in either C/C++ or Java, thus facilitating the integration of legacy clinical rule systems.

Security and Privacy

Because the data collected is personal health-care data, we are bound by the health-care privacy laws of Canada, the United States, and the province of Ontario. The Research Ethics Boards at the three institutions all mandated a plan to ensure

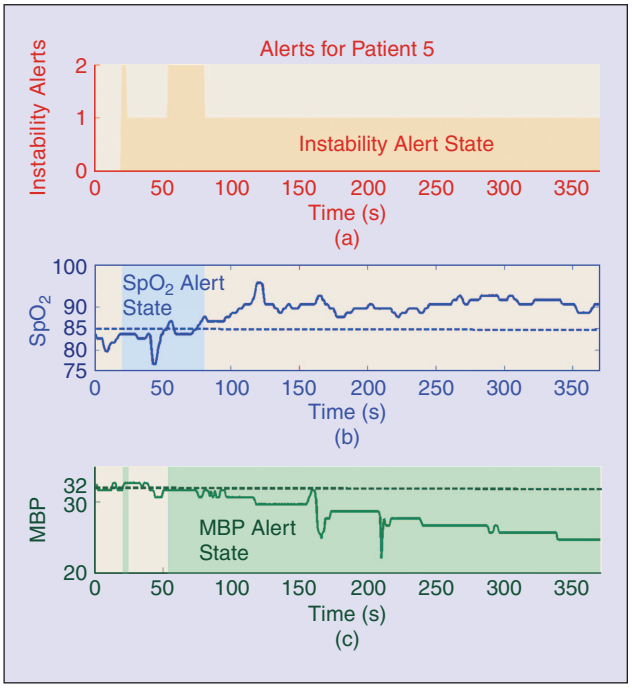

Fig. 4. Alerts generated by a sample clinical rule.

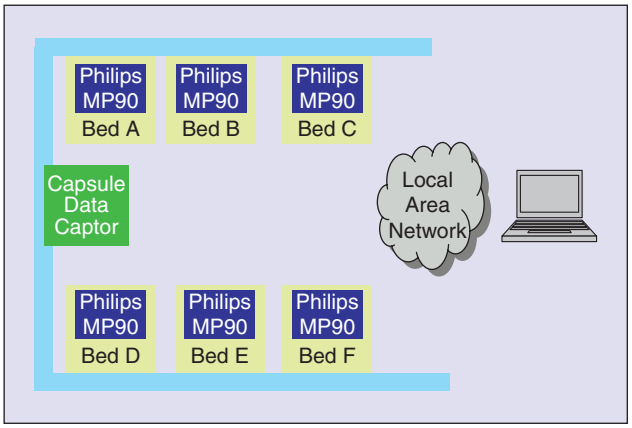

Fig. 5. NICU bay floor layout.

compliance with the laws. We do not collect or store any data that could directly identify the infants. We deidentify the data prior to it entering Artemis; Artemis processes and stores only deidentified patient data. An Artemis identifier is generated for each infant when the infant is enrolled. An unique identifier of the patient monitor associated with each data element is transmitted from that patient monitor. The hospital tracks an association between a bed and the patient monitor associated with that bed. During enrollment processing, an association between the patient monitor identifier and the Artemis identifier is placed in a mapping database table. As the data streams into the Artemis system, one of the initial operators has the task to pick up the patient identifier from the data elements and insert the associated Artemis identifier, thereby, deidentifying the data.

The Artemis system operates on the hospital network that has been secured for the transmission of all types of medical data. The processing components and Artemis database system are located in a physically secure location accessible only by Artemis team members. Using the built-in authentication system of the operating system and the authorization mechanism and access control system of the database system, Artemis controls the operations a user can perform on the physiological data streams stored in the database. Because the system is separate from all other hospital systems, we can administer access to the Artemis database to ensure there is no unauthorized access, even from within the hospital's network.

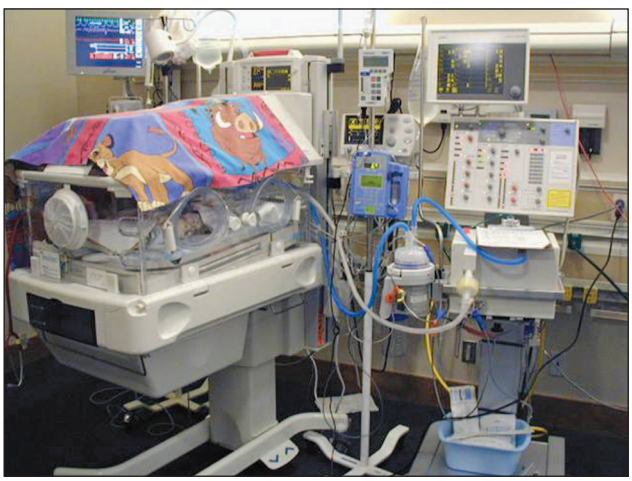

Fig. 6. Example of devices and equipment attached to an infant in an NICU (photo used with permission from Terry Tremethick).

The interactions between the deployment server and the streaming system are done over the network, so steps have been taken to close the security exposure. First, access to the deployment server is password protected. Furthermore, the deployment server authenticates with the streaming system for each remote operation it attempts to perform on the streaming system.

The UOIT creates a mirror copy of the Artemis database by doing incremental downloads on a periodic basis. The connection between the computers at the SickKids and UOIT is made using a secure tunnel. The tunnel is implemented using the tunneling capabilities of the secure shell 2 (SSH2) protocol. The tunnel is encrypted with a 4,096-b Rivest, Shamir, Adleman public encryption algorithm (RSA) key. At the UOIT, the mirror copy is maintained on a dedicated computer in a secure locked room accessible to members of the UOIT team only.

The IBM team maintains a mirror by performing incremental periodic downloads from the UOIT mirror. The same secure tunnel mechanisms are used to securely transmit the data. The IBM mirror is stored on a password-protected machine stored in a secure laboratory environment. The data are available only to members of the IBM team.

Deployment

Artemis has been deployed in the NICU of the SickKids in Toronto, Ontario, Canada, since early August 2009. In this phase of deployment, we are capturing physiological data streams and electronic health record information forwarded from the CIMS data for the infants within the study. The physiological data being collected contains ECG, heart rate, respiratory rate, and blood SpO_2; BP may be streamed or obtained as CIMS observations. We also monitor these data using Artemis and an initial version of a clinical rule for the early detection of nosocomial infection, which represents a complex rule that will be reported in future research publications. In this phase, we have monitored as many as four infants at the same time. Because the stream computing middleware is scalable, we believe our approach could easily be used to monitor all patients within the NICU, but we have not tested this configuration. As of early December 2009, 19 infants have been enrolled in the research study.

A Capsule DataCaptor terminal unit is located in each of two NICU bed bays. The placement of a unit in a bed bay is illustrated in Figure 5. Figure 6 shows a typical bed in a NICU with the associated equipment. The DataCaptor terminal unit transmits data streams from Philips IntelliVue MP70 patient monitors to the Capsule server. Each DataCaptor terminal unit can handle up to eight patient monitors. The dimensions of the unit are 9.3 in by 10.6 in by 2.1 in. To evaluate the processor and memory requirements of the Artemis server components, we partitioned the Artemis server among a Windows PC and two Linux notebooks.

Before Artemis could be deployed at the SickKids, we had to perform some tests to provide assurances to the hospital staff. All interfaces to the existing NICU devices and hospital systems had to be designed so that no data or process used by the NICU staff would be disrupted. Hardware deployed in the NICU could not interfere with clinical processes and had to be effectively invisible to the medical staff. In the NICU, there is a patient monitor (Philips IntelliVue MP70) per patient, and

the devices connected to the patient are connected to the patient monitor. There are two ports on the patient monitor for externalizing the data streams. One port is currently being used for clinical purposes, and we were able to demonstrate that we can get the stream from the alternate port without interference.

We also performed extensive testing to ensure that the deployment would not interfere with the existing network traffic or with the existing slower feeds of data to the CIMS database. The initial development testing was all performed away from the NICU setting utilizing a Philips IntelliVue MP50 running in demonstration mode located within the Health Informatics Research laboratory located at UOIT. Once alpha component testing and integrated system beta testing were completed, we commenced testing utilizing an Philips IntelliVue MP70 located within the NICU in demonstration mode and then later with a patient simulator. Our initial tests show that the network traffic generated is below 0.2% (based on a 100 Mb/s network) for each Philips IntelliVue MP70 connected to the Artemis system.

The deidentified data are collected by the deployed Artemis system at the SickKids into the local Artemis relational database. Periodically, the UOIT team, using secure connections, performs an incremental download from the Artemis database to a database at UOIT.

The nosocomial infection clinical rules will be refined over time with both new feature sets and fusion rules based on data mining results and review of the correlation of computed features with infants who actually develop nosocomial infection.

Lessons Learned

Even in this early phase of deployment, we have learned some lessons that should be of interest to other researchers and to us as we move forward.

Expect Unforeseen Situations in NICUs

We realized early that it was imperative to develop mechanisms that allow us to dynamically change the configuration of the system. In particular, mechanisms have been built to allow the dynamic deployment of data sources and the redeployment of parts of applications on a running instance of the Artemis system. Even though it was not a primary design requirement, the decision to implement these mechanisms has been quite important for the success of the deployment. On several occasions, we had to either change the configuration of the system or modify a running application and redeploy it.

Minimize Data Source Dependency for Clinical Rules

The intensive care unit is a very dynamic environment where the care of the patient is paramount. Patients are disconnected and reconnected to devices for various reasons. The staff is more concerned with the generation of data they need to care for the patient than they are in making sure the research system is receiving the data it needs. A clinical rule should be designed to generate analytic results as long as the requisite data streams are flowing. The temporary loss of some data streams may reduce the set of features available, but, the presence of the requisite streams ensures the generation of a set of critical features required to generate results. This causes some complication in the implementation of clinical rules in general. The initial nosocomial clinical rule design assumed the presence of all seven data streams; however, we determined that four of the streams were always present and the other three were less likely to be present. We reimplemented the clinical rule to work under these conditions.

Conclusions

We have described the design, implementation, and initial deployment of Artemis. Artemis has the potential to revolutionize intensive care medicine, because it gives clinicians a way to discover early indicators of medical conditions and to encode these in clinical rules. Artemis also provides a way for clinicians to have online, real-time execution of the clinical rules in an intensive care environment. We have also shown the need for a diverse collaborative team to address the broad set of issues involved in such a system.

While many research and industry-based initiatives exist to propose clinical rules for specific condition onset or propose methods for detecting features in certain streams, these approaches do not apply a systems-based approach for the support of multiple streams, from multiple patients relating to multiple diagnoses [17].

The UOIT and SickKids-based teams have membership with the Canadian Neonatal Network (a network of 30 level 3 and above NICUs in Canada) and envision an ultimate national rollout of the technology introduced here.

Marion Blount received his B.S. degree in electrical engineering from Duke University and his M.S. and Ph.D. degrees from Stanford University. Over the course of his career at IBM T.J. Watson Research Center, he has conducted research in the areas of distributed computing, memory coupled systems, pervasive computing, and utility computing. In recent years, his interest has been in the application of stream computing technologies to the health-care domain.

Maria R. Ebling received her B.S. degree from Harvey Mudd College and her M.S. and Ph.D. degrees in computer science from Carnegie Mellon University. She is a research staff member and senior manager at the IBM T.J. Watson Research Center. She manages a team, building systems capable of detecting events in the real world and reacting to those events in a timely and appropriate manner. Her interests are in distributed systems supporting mobile and pervasive computing, privacy, human–computer interaction, and applications of these technologies to health care.

J. Mikael Eklund received his B.Sc. and M.Sc. degrees in engineering in 1989 and 1997, respectively, and Ph.D. degree from the Queen's University, Kingston, in 2003. Between his bachelor's and master's studies, he was a simulator flight control systems engineer with CAE Electronics in Montreal, Quebec. He is an assistant professor and program director for electrical and software engineering at the UOIT. Until August 2006, he was a visiting postdoctoral scholar in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. His research is in the areas of time-series analysis, nonlinear system identification, intelligent control, and data knowledge

discovery with applications in smart medical systems, robotics, and other autonomous systems.

Andrew G. James received his graduate degree in biomedical informatics from the Oregon Health and Science University, Portland, Oregon, and a medical degree from the University of Auckland, Auckland, New Zealand. He is a staff neonatologist at the SickKids and an associate professor of pediatrics at the University of Toronto, Toronto, Canada. He is interested in the application of information technologies for the care of the newborn infant. His primary focus is clinical decision support and biomedical knowledge representation. He is the site investigator at SickKids for the Canadian Neonatal Network's evidence-based quality improvement research program. He is chair of the Informatics Research Committee and Canadian Neonatal Network and cochair of the Concept Model Special Interest Group, and International Health Terminology Standard Development Organization, Copenhagen, Denmark. He is the foundation chief medical editor of AboutKidsHealth.ca, an evidence-based Web site for children, parents, and families.

Carolyn McGregor received her bachelor of applied science in computer science honors degree and a Ph.D. degree in computing science from the University of Technology in Sydney, Australia, where she developed new ways to use intelligent decision support systems (IDSSs) to assist organizations in the monitoring of business performance. This research included an innovative IDSS framework that she extended, resulting in the solution manager service (SMS). The SMS contained pioneering research in Web services, event stream processing, multiagent data mining, and near–real-time data warehouse loading. Her recent research focus has been the improvement and extension of that research within the context of health and medicine for advanced support for clinical management and research, together with new remote patient care frameworks and patient journey modeling and reengineering approaches. She is the Canada research chair in health informatics based at the UOIT, Canada.

Nathan Percival received his bachelor of mathematics in computer science, information systems option from the Faculty of Mathematics at the University of Waterloo and a master's degree in information technology security from the Faculty of Business and Information Technology at the UOIT. He is currently a staff member at UOIT supporting the Faculty of Engineering and Applied Science and the Health Information Research Group.

Kathleen P. Smith is a master's of health science (health informatics) candidate and a research associate at the UOIT under the supervision of Dr. Carolyn McGregor and Dr. J. Mikael Eklund. Her research interests include real-time event processing, temporal data mining, clinical rule ontologies, and the application of services computing to critical care and women in computing and engineering.

Daby Sow received his bachelor of science in electrical engineering from the Universite Laval, Quebec, Canada, and his master of science and Ph.D. degrees in electrical engineering from Columbia University. He was a graduate research assistant at Columbia University in the Department of Electrical Engineering from 1996 to 2000, working on lossy source coding theoretical problems with applications to image and video processing. He is a research staff member at IBM T.J. Watson Research Center. His research interests are at the intersection of several fields of computer science and electrical engineering, including stream computing, complex event processing, applied machine learning, signal processing, and information theory.

Address for Correspondence: Marion Blount, IBM T.J. Watson Research Center, 19 Skyline Drive, Hawthorne, NY 10532 USA. E-mail: mlblount@us.ibm.com.

References

[1] P. Griffin and R. Moorman, "Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis," *Pediatrics*, vol. 107, no. 1, pp. 97–104, 2001.
[2] N. McIntosh, J.-C. Becher, S. Cunningham, B. Stenson, A. Laing, A. J. Lyon, and P. Badger, "Clinical diagnosis of pneumothorax is late: Use of trend data and decision support might allow preclinical detection," *Pediatric Res.*, vol. 48, no. 3, pp. 408–415, 2000.
[3] J. Fabres, W. A. Carlo, V. Phillips, G. Howard, and N. Ambalavanan, "Both extremes of arterial carbon dioxide pressure and the magnitude of fluctuations in arterial carbon dioxide pressure are associated with severe intraventricular hemorrhage in preterm infants," *Pediatrics*, vol. 119, no. 2, pp. 299–305, 2007.
[4] V. Tuzcu, S. Nas, U. Ulusar, A. Ugur, and J. R. Kaiser, "Altered heart rhythm dynamics in very low birth weight infants with impending intraventricular haemorrhage," *Pediatrics*, vol. 123, no. 3, pp. 810–815, 2009.
[5] S. Shankaran, J. C. Langer, N. Kazzi, A. R. Laptook, and M. Walsh, "Cumulative index of exposure to hypocarbia and hyperoxia as risk factors for periventricular leukomalacia in low birth weight infants," *Pediatrics*, vol. 118, no. 4, pp. 1654–1659, 2006.
[6] M. Stacey and C. McGregor, "Temporal abstraction in intelligent clinical data analysis: A survey," *Artif. Intell. Med.*, vol. 39, no. 1, pp. 1–24, 2007.
[7] Home page of Capsule Tech Inc. (2009). Medical device connectivity with capsule [Online]. Available: http://www.capsuletech.com/index.htm
[8] Web page describing solutions of Capsule Tech Inc. (2009). Point of care connectivity [Online]. Available: http://www.capsuletech.com/our-solution-products.htm
[9] L. Amini, H. Andrade, R. Bhagwan, F. Eskesen, R. King, P. Selo, Y. Park, and C. Venkatramani, "SPC: A distributed, scalable platform for data mining," *Proc. SIGKDD 2006, Workshop Data Mining Standards, Services, and Platforms*, Aug.20, 2006pp. 27–37.
[10] N. King. (2008). Instant information: Stream computing [Online]. Available: http://www.ibmdatabasemag.com/story/showArticle.jhtml?articleID=211300227
[11] B. Gedik, H. Andrade, K. Wu, P. S. Yu, and M. Doo, "SPADE: The system S declarative stream processing engine," in *Proc. Int. Conf. Management of Data, ACM SIGMOD*, 2008, pp. 1123–1134.
[12] M. Stacey, C. McGregor, and M. Tracy, "An architecture for multi-dimensional abstraction and its application to support neonatal intensive care," in *Proc. 29th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC'07)*, Lyon, France, 2007, pp. 3752–3756.
[13] C. McGregor and M. Stacey, "High frequency distributed data stream event correlation to improve neonatal clinical management," in *Proc. Inaugural Int. Conf. Distributed Event-Based Systems (DEBS'07)*, Toronto, ON, Canada, CD-ROM, 2007, p. 6.
[14] C. Catley, K. Smith, C. McGregor, and M. Tracy, "Extending CRISP-DM to incorporate temporal data mining of multi-dimensional medical data streams: A neonatal intensive care unit case study," in *Proc. 22nd IEEE Int. Symp. Computer-Based Medical Systems (CBMS2009)*, pp. 1–5.
[15] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychol. Rev.*, vol. 63, no. 2, pp. 81–97, 1956.
[16] Data Mining Group. (2009). Data mining group—Welcome to DMG [Online]. Available:http://www.dmg.org
[17] M. Stacey and C. McGregor, "Survey paper: Temporal abstraction in intelligent clinical data analysis," *Artif. Intell. Med.*, vol. 39, no. 1, pp. 1–24, 2007.