Distributed Data Management Architecture for Embedded Computing

Hans-Werner Braun[1],Todd Hansen[1], Kent Lindquist, Bertram Ludäscher[1], John Orcutt[2], Arcot Rajasekar[1], Frank Vernon[2]
San Diego Supercomputer Center[1] , Scripps Institution of Oceanography[2]
University of California, San Diego

Keywords: real-time data management, sensors, embedded computing, virtualization, metadata, data grids

Introduction

Real-time data management is faced with the problems of disseminating large collections of data to users and applications, providing a collaborative environment for analyzing and performing analysis-intensive computing, and at a basic level with problems of providing access, managing, curating, storing and moving large quantities of real-time data. The evolution of the data grid technologies provides solutions to these problems by developing middleware that can seamlessly integrate multiple data resources and provide a uniform method of accessing data across a virtual network space. The "Grid" is a term used for defining a variety of notions linking multiple computational resources such as people, computers and data [Foster & Kesselman 1999]. The term "Data Grid" has come to denote a network of storage resources, from archival systems, to caches, to databases, that are linked of data grids can be found in the physics across a distributed network using common interfaces. Examples of implementations include physics community [PPDG 1999, Hoschek 2001, GriPhyN 2000, NVO 2001], biomedical applications (BIRN 2001) , ecological sciences [KNB 1999] and earthquake and multi-sensor systems [NEES 2000], etc. In this paper we describe a real-time data grid architecture, called the VORB, that we are developing as part of an NSF-funded project called ROADNet [ROADNet 2001]. Because of rapidly evolving capabilities for observing the Earth, especially with network of sensors providing continuous real-time data, data volumes will reach petabytes in scales. Moreover, with emerging technological capabilities it will be possible to deploy multiplicity of sensor networks that can generate data spanning many disciplines.

ROADNet

ROADNet (http://roadnet.ucsd.edu) is a multi-disciplinary project performing research and building infrastructure to enhance our capacity to monitor and respond to changes in our environment by developing both the wireless networks and the integrated, seamless, and transparent information management system that will collect and stream seismic, oceanographic, hydrological, ecological, and physical data to a variety of end users in real-time. The ROADNet team is building upon currently deployed autonomous field sensor systems, including sensors that monitor fire and seismic hazards, changing levels of environmental pollutants, water availability and quality, weather, ocean conditions, soil properties, and the distribution and movement of wildlife.

The software tools being developed will extend existing concepts in connectivity using wireless and internet (HPWREN) , object ring buffers (ORB) for collecting disciplinary data to virtual ORBs (VORB) for managing multiple connections to multiple ORBs. The VORBs will not only provide data to multiple users in real time, but will provide interfaces with archival ORB and more traditional databases. Many of these interactions will be mediated through XML wrappers which will provide the basis for data discovery. A rule-based programmable interface is being developed to dynamically reconfigure and prioritize data capture, event detection and analysis from sensor networks.

Sensor Integration Strategies

A large portion of this project involves sensor integration. Researchers are by trade more interested in the data they collect then the complexities of gathering the data. As a result many have resorted to simple interfacing measures that do not make themselves easily available for real-time data collection. Part of ROADNet's task is to develop methods to integrate these sensors into seamlessly fused sensor networks. Currently, the ROADNet team is working on integrating a large number of sensors and data loggers;, a brief (sample) list is shown in the table below:

Datalogger type	Usage
Video cameras	for visual analysis (wild-life migration, river monitoring, tide monitors)
Campbell data logger	weather,strain monitor, general purpose sensor interface and logging
Handar	weather logger
Solinst	Stream level conductivity,temp , etc
ADCP (Acoustic Doppler Current	Measure ocean currents

Profiler)	
Ashtech GPS	RTCM corrections for Differential GPS
CODAR	Surface Current Profile Radar
Reftek, Quanterra, Kinemetrics	Broadband seismic and strong motion sensors

A few of the sensors are IP-addressable and many have RS-232 ports which we connect through an Avaya Wavelan/EC-S. This allows us to convert RS-232 to IP via either Ethernet or wireless which connect to ORBs at the other end of the network. For more information on sensor integration and an up to date listing of our current strategies see: http://roadnet.ucsd.edu/tech-advice.html. The sensors are connected through IP protocol into a network of sensors. We utilize the skills and experience acquired through the HPWREN (http://hpwren.ucsd.edu) project to engineer the most optimal connection. In many cases we utilize existing HPWREN network. In other methods of interconnectivity we are exploring connectivity through satellite systems and CDPD (IP over cell phones) systems.

Object Ring Buffer

Object Ring Buffers have been successfully used for capturing real-time data in seismic networks [Harvey, et al., 1998a; Al-Amri and Al-Amri, 1999; von Seggern, et al., 2000]. ORBs provide a means to capture data in real-time and store them for immediate use as well as archive them for later analysis. ORBs expose the data they are holding through a well-defined Application Programming Interface (API) and also encapsulate their raw-data with additional metadata information. Systems such as Antelope [Vernon and Wallace, 1999; Vernon, 2000; von Seggern, et al., 2000] provide additional utilities for analysis using the ORBs data such as event-detection using waveform analysis.

ORBs are used by the seismic community in many individual networks for capturing data from multiple sensors connected together in autonomous networks [Harvey, et al., 1998b, Vernon and Wallace, 1999]. Inter-ORB communication is used for exchanging data between networks in near-real-time. Each sensor is statically linked to one or more ORBs.

Virtual Object Ring Buffer (Fig. 1)

In order to enhance the capabilities of ORBs for real-time data exchange and dynamic reconfiguration of data flow in federated sensor networks, we implement '*Virtual ORB*'s (or VORB) that can be used to provide transparency and independence from physical ORB implementations. VORBs can be used as nodes in a federated sensor network to act as data exchange points. The VORB will also provide a means of developing a test bed for testing new kinds of virtual real-time sensor network implementations.
At the simplest level, a VORB will provide a simple store-and-forward capability that can receive data from a sensor and distribute it to multiple physical ORBs. At a more complex level a VORB can perform real-time filtering, data subsetting and/or multiplexing in order to optimize the data distribution to multiple ORBs. Users and client applications will communicate through VORBs instead of directly attaching to ORBs. This will provide physical independence of ORB location and access. The ORB operation is dictated by a rule Base (acted upon by a rule Engine). The rules can be used to dynamically change the activity of the ORB. For example, if a VORB senses that an ORB is off-line, it can redirect the data going to that ORB to another ORB designated by the rules. The rules are triggered by sensory events (sensed by simple filters and event detectors in the VORB), and by external and administrative commands.

Adaptive Rules for Real-time Data Capture

The VORB network will be dynamically programmable in a rule-based language, which can adapt to the changing needs of the sensor networks and user/administrative criteria. The rule-based system will also be useful for dynamically reconfiguring sensor networks. As an example, consider that there is a fire in a region with a sparse network of sensors. Then, agencies can quickly move in several mobile sensors to the area (e.g. by helicopters) in order to monitor the spread of fire. These mobile sensors can be dynamically configured to send data to one or more VORBs that are 'nearer' to them and, using the rules, disseminate the information to specific ORBs for real-time analysis.

Event detection and event processing is a central component in real-time environmental monitoring systems. For example, a typical chain of activities in a seismic network includes signal detection from waveform data packets, followed by triggering of processing routines to determine whether a relevant seismic event has been detected, in which case procedures are triggered that calculate the location and magnitude of the seismic event. The rule-based VORB will provide capabilities for complicated event detection and reaction., combining an active rules approach [LLM98] with efficient XML stream processing [LMP02].

Conclusion

The VORB system provides a working prototype for distributed real-time data management. The ROADNet project is deploying the system to connect sensors from ships (R/V Revelle http://quakeinfo.ucsd.edu/cgi-bin/revelle.cgi) to seismic networks across the world to hydrological sensors on snow packs and estuaries.

Acknowledgement:
This research has been sponsored by the National Science Foundation under Award No. OCE-0121726.

References

Foster, I., and Kesselman, C., (1999) "The Grid: Blueprint for a New Computing Infrastructure," Morgan Kaufmann.

PPDG, (1999) "The Particle Physics Data Grid", (http://www.ppdg.net/).

Hoschek, W., Jaen-Martinez, J., Samar, A., Stockinger, H., and Stockinger, K. (2000) "Data Management in an International Data Grid Project," *IEEE/ACM International Workshop on Grid Computing Grid'2000*, Bangalore, India 17-20 December 2000.

GriPhyN, (2000) "The Grid Physics Network", (http://www.griphyn.org/proj-desc1.0.html).

NEES, (2000) "Network for Earthquake Engineering Simulation", (http://www.eng.nsf.gov/nees/).

KNB, (1999) "The Knowledge Network for Biocomplexity", (http://knb.ecoinformatics.org/).

NVO, (2001) "National Virtual Observatory", (http://www.srl.caltech.edu/nvo/).

RoadNet, (2001) "Wireless Access and Real-time Data Management", (http://roadnet.ucsd.edu/).

Harvey, D., D. Quinlan, F. Vernon, and R. Hansen (1998), "ORB: A New Real-Time Data Exchange and Seismic Processing System", ," Seis. Res. Lett. 69, 165.

Harvey, D., F. Vernon, G. Pavlis, R. Hansen, K. Lindquist, D. Quinlan, and M. Harkins(1998), "Real-Time Integration of Seismic Data from the IRIS PASSCAL Broadband Array, Regional Seismic Networks, and the Global Seismic Network," EOS Trans. AGU 79, F567. Processing System," Seis. Res. Lett. 69, 165.

A-Amri, M.S., and A.M. Al-Amri. (1999) "Configuration of the Siesmographic Networks in Saudi Arabia," Seis. Res. Lett. 70, 322-331.

Von Seggern, D., G. Biasi, and K. Smith, (2000), Network Operations Transitions to Antelope at the Nevada Seismological Laboratory," Seis. Res. Lett. 71, 444.

Vernon F., and T. Wallace (1999), "Virtual Seismic Networks," IRIS Newsletter, 18, 7-8.

Vernon F., (2000) "Overview of the USARRAY Real-Time Systems", EOS Trans AGU 81, S318.

Lausen, G., B. Ludäscher, and W. May. (1998) "On Active Deductive Databases: The Statelog Approach," In Transactions and Change in Logic Databases, H. Decker, B Freitag, M Kifer, A Voronkov, eds. LNCS 1472, 1998.

Ludäscher, B., P. Mukhopadhyay, Y. Papakonstantinou (2002) "A Transducer-Based XML Query Processor", Intl. Conference on Very Large Databases (VLDB).

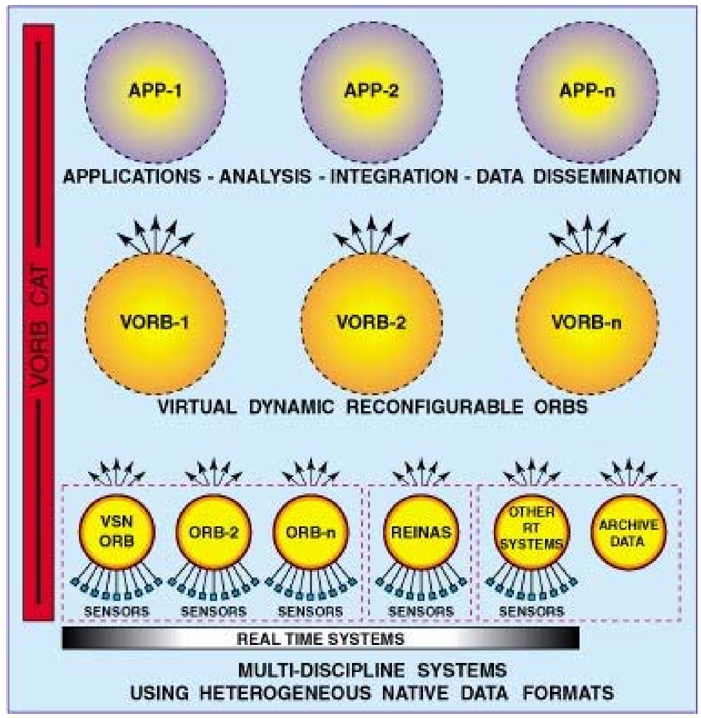

Figure 1. Virtual Object Ring Buffer