AUGMENTING HYDROLOGIC INFORMATION SYSTEMS WITH STREAMING WATER RESOURCE DATA

S. Esswein[1], J. Hallstrom[2], C. J. Post[1], D. White[3], G. Eidson[4]

AUTHORS: Forestry and Natural Resources[1]; School of Computing[2]; Computing and Information Technology[3]; Restoration Institute[4], Clemson University, Clemson, SC USA 29634

REFERENCE: *Proceedings of the 2010 South Carolina Water Resources Conference*, held October 13-14, 2010, at the Columbia Metropolitan Convention Center.

Abstract. Access to timely and accurate hydrological and environmental observation data is a crucial aspect of an integrated approach to water resources management. This presentation describes an end-to-end system designed to support realtime monitoring and management of water resources. The main components of the hardware/software infrastructure of this system are broken into four categories and briefly described. This organization provides the basis for a synthesis of several prominent standards and software solutions relevant to the hydrologic and environmental observing communities. These standards are described in the context of their role in our end-to-end system. The presentation concludes with a case study describing a green infrastructure monitoring effort located in the City of Aiken, South Carolina.

INTRODUCTION

Population growth has increased the demand for water resources in South Carolina. Recent periods of extreme drought have highlighted the need to evaluate how water resources can be managed to meet these demands. The prevalence of hydropower and reservoirs in South Carolina means that management is possible, provided conflicting uses can be effectively balanced. Balancing instream flow requirements, reservoir pool levels, and hydropower generation schedules requires that an integrated resource management approach be used. The availability of timely and accurate hydrological and environmental observation data is a vital component of an integrated approach; particularly as water provisioning becomes increasingly fine-grained.

Acquiring observation data and incorporating it into management tools presents a host of technological challenges. The IntelligentRiver™ project is currently being developed at Clemson University to address these challenges in a water resources context. IntelligentRiver™ provides an end-to-end hardware/software infrastructure geared specifically towards handling realtime hydrological and environmental observation data. In this presentation, the components of this infrastructure are examined with an emphasis on design decisions regarding leveraging available standards and software. Insight garnered from several years of data acquisition experience is provided, along with a recent case study involving a monitoring deployment supporting the Sand River Headwaters Green Infrastructure project located in the City of Aiken, South Carolina.

There are four components or tiers of a realtime-monitoring infrastructure: (i) sensing platforms collect in situ observation data, (ii) communication and uplink technologies transmit realtime observation data, (iii) data streaming middleware provides highly distributed publication and subscription of observation data, and (iv) back-end repository and presentation services provide a means of viewing and utilizing data products. IntelligentRiver™ serves as the backbone linking these four components. A project goal includes supporting standards and data models accepted by environmental scientists and hydrologists. This presentation provides a synthesis of prominent standards relevant to hydrologic and environmental observing systems. The following standards are addressed in this presentation: the Open Geospatial Consortium's Sensor Web Enablement (SWE) initiative, the IEEE 1451 standards family, and the Consortium of Universities for Advancement of Hydrologic Science (CUAHSI) hydrologic information system. Additionally, these standards are presented in the context of our experiences operating water resource monitoring programs.

BACKGROUND

The Open Geospatial Consortium's SWE initiative provides an extensive collection of standards intended to support making sensors and instrumentation accessible and, where applicable, controllable via the web (Botts et al., 2007). Standardization began in 2005 and is ongoing, with periodic updates to existing standards and emerging standards under evaluation. The Open Geospatial Consortium's approach relies heavily on web services and data interchange using extensible markup language (XML). The strength of web services and XML technologies primarily benefits the application domain. However, the Sensor Observation Service and the Sensor Alert Service both deal with lower level sensor management (Open Geospatial Consortium, 2008). The TransducerML and Observations and Measurements standards specify observation data encodings. SensorML provides a rich metadata language for describing sensing platforms and instrumentation.

The Institute of Electrical and Electronic Engineers (IEEE) provides a family of standards collectively known as IEEE1451 to support the integration of transducers into a networked environment (Song, E.Y.; Kang Lee 2008) (IEEE, 1999). Standardization by the IEEE Instrumentation and Measurement Society's Sensor Technology Technical Committee began in 1997. Three standards under IEEE1451 have received approval with numerous standards still under review. Unlike SWE, IEEE1451 is geared towards physical hardware and protocol specification. However, it does provide a network communications model and a low-level transducer metadata specification that overlap with portions of SWE. Recent efforts by the Open Geospatial Consortium have explored incorporating aspects of IEEE1451 into SWE (Botts et al., 2007).

The Consortium of Universities for Advancement of Hydrologic Science (CUAHSI) Hydrologic Information System (HIS) provides an "information infrastructure for the support of the hydrologic science throughout the hydrologic community (Hooper et al., 2004)." The core of HIS is a comprehensive hydrologic data model and the WaterML Web Services providing "an Internet-based system for sharing hydrologic data" (CUAHSI, 2010). This system also includes a host of analysis, visualization and data management tools. Like the SWE, the HIS suite is geared toward application level functionality rather then sensing platforms and communications.

METHODOLOGY

IntelligentRiver™ technology can be organized into four components or tiers: (i) sensing platform, (ii) communications and uplink, (iii) data streaming middleware, and (iv) repository and presentation services. The IntelligentRiver™ architecture is described here with an emphasis on the use of third-party software and standards. A more detailed exploration of the core architecture is available (Eidson et al., 2010).

At the sensing platform tier we utilize a variety of sensor instrumentation including both commercially available equipment (e.g. multi-parameter sondes, soil moisture probes and weather stations) as well as custom developed sensors and assemblies (e.g. dendrometer, buoys, water table measurement). Newer IntelligentRiver™ deployments rely heavily on the Motestack in situ sensing platform, an ultra low-power embedded computing device developed at Clemson University. The Motestack provides a variety of sensor interface methods (SDI-12, Analog, 1 wire) and communications options (802.15.4/ZigBee and 802.11 WiFi). Long battery life and remote configurability provide longer maintenance intervals, reducing monitoring costs. Motestacks can compliment existing instrumentation, e.g. providing advanced communications capabilities to the SDI-12 compatible Yellow Springs Instruments (YSI) 6600 multi-parameter sonde.

The communications and uplink tier of IntelligentRiver™ acts as a bridge between the sensing platform tier and the data streaming middleware tier. Our installation sites have spanned the spectrum from low-lying coastal sites with extremely dense vegetation to heavily trafficked urban centers. In a typical installation, the low power 802.15.4/Zigbee radio communications protocol is utilized to link sensing nodes with special purpose nodes that bridge traffic onto either WiFi or Cellular "backhauls". These "backhaul" technologies are too expensive or power consumptive to install at individual nodes, but provide a necessary link to the data streaming middleware tier.

The data streaming middleware tier plays a key IntelligentRiver™ role by linking observation data produced by sensing platforms with virtually any number of potential consuming applications. Relational databases, web based presentation tools and quality control/assurance tools are examples of consuming applications. Any number of applications can elect to interact directly with the data streaming middleware to get access to realtime observation data. Incorporating a middleware simplifies application development by allowing consuming applications to "decouple" themselves from potentially unpredictable sensing platforms. Decoupling means that producers and consumers do not interact or depend directly on each other. Other requirements of the middleware include high performance and no single point of failure, two characteristics provided by the distributed configuration utilized by IntelligentRiver™. Current deployments rely on an open-source messaging system called NaradaBrokering. This messaging software provides brokered publish/subscribe communications. This approach employs intermediaries or "brokers" to handle message routing between observation data sources/publishers and observation data sinks/subscribers (Pallickara and Fox, 2003). A new version of our data streaming middleware is underway that replaces NaradaBrokering with a newer, higher performance messaging software called RabbitMQ (RabbitMQ, 2010).

The final tier of the IntelligentRiver™ infrastructure is termed the Repository and Presentation tier. This tier is comprised of a number of software solutions for storing, visualizing and sharing observation data. Existing repositories include an *Oracle* database, the *Thematic Realtime Environmental Distributed Data Service* (THREDDS) and a simple text-base web delivery system. The Oracle database uses the Xenia model for storage of scalar observation data. THREDDS, developed by Unidata, provide metadata and data access standards, including *OpenDAP, HTTP file access, the Open Geospatial Consortium's Web Coverage Service, and the NetCDF subset service*. The text-based delivery system supports downloading observation data directly into familiar tools like *Microsoft Excel.* Work is underway to support the CUAHSI observation data model in addition to the *Xenia* model. Supporting the CUAHSI model will allow us to offer a WaterML web services method for sharing hydrologic data with other hydrologic research efforts. Presentation services include web-based applications for the visualizing observation data spatially (e.g. *Google Maps*) and temporally (e.g. charts and tables). The *GeoServer* Internet mapping engine provides additional data access pathways, including *GeoRSS, Open GeoSpatial Consortium's Web Feature Service,* and the *Keyhold Markup Language (KML)*.

Sand River Headwaters Case Study

The most recent addition to the IntelligentRiver™ monitoring program is a project to evaluate sustainable development practices. Rain gardens, bioswales, underground cisterns and pervious pavements have been installed to capture and treat stormwater in downtown Aiken. Monitoring includes measurement of stormwater flow and retention times using Isco Automated Samplers and Area Velocity Flow modules. Ancillary data such as meteorological (*Campbell Scientific, Inc*) and soil moisture conditions (*Decagon Devices, Inc*) are also collected and incorporated into IntelligentRiver™. Soil moisture measurements are acquired using the Motestack platform and transmitted to aggregate nodes over a 2.4GHz 802.15.4 based network. Custom aggregate nodes bridge 802.15.4 traffic onto an 802.11g WiFi network linked with the Clemson University Network Operations Center (NOC). Upon arrival at the NOC, observations are passed to the data streaming middleware and distributed to consuming applications including the IntelligentRiver™ website.

DISCUSSION

Of the three standards described, the IEEE1451 family of standards has the greatest potential to contribute to the sensing platform and communication tier of the IntelligentRiver™ infrastructure. In particular, IEEE1451.1 provides a low-level metadata specification intended to be integrated directly into the sensing layer, allowing the storage of sensor description and calibration data directly on a sensor. Manufacturers are beginning to ship sensors from the factory with this capability (e.g. *National Instruments Corp.*). Coupled with other IEEE1451 hardware standards, "plug & play" sensor implementations are possible. An advantage of the IEEE1451 is its emphasis on low computing power requirements. The metadata is concise enough to fit within the strict constraints of our sensing and communications tiers. Work on integrating SensorML into the Motestack is underway, but hampered by the verbose nature of markup languages, requiring compression technologies or custom template approaches. IEEE1451 communications models are built on simple request/response or publish/subscribe semantics making them better suited for embedded computing devices versus heavier weight web services used by SWE and CUAHSI HIS.

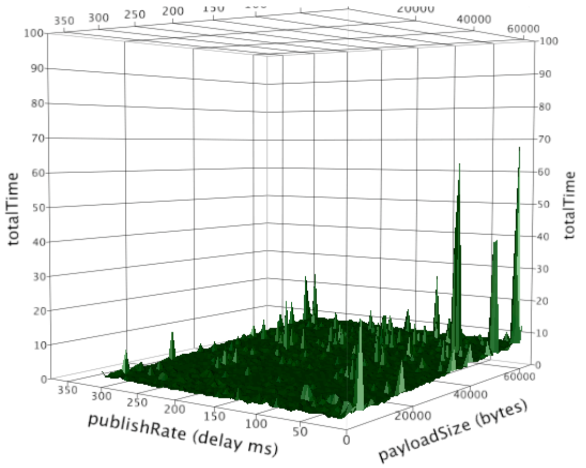

Figure 1 - Benchmark Results. Total time represents transit delays incurred between data acquisition and data delivery at consuming application.

The data streaming middleware tier is largely outside the scope of IEEE1451, SWE & CUAHSI HIS. IEEE1451 includes a general recommendation for a publish/subscribe messaging approach, but intentionally excludes implementation specifics beyond message encodings (IEEE, 2007). As this area of IEEE1451 matures, it may warrant additional consideration. From a communications perspective, the SWE & CUAHSI specifications and publish/subscribe messaging are not mutually exclusive, however both approaches would introduce additional complexity and a loss of flexibility over our current approach. A custom middleware tier built on existing messaging software allows us to focus efforts on middleware performance and reliability. Figure one shows the performance of a single messaging node under a range of publishing rates and payload sizes using a RabbitMQ based messaging layer. Transit delays between observation data publication and client data access average between 5-6ms. As demonstrated in Figure one, the messaging layer provides stable performance with reasonable transit delays (<70ms) under very high observation publication rates and large payload sizes (60Kbyte). The data repository and presentation services are where SWE and CUAHSI HIS are strongest. The SWE Sensor Observation Service and SensorML specifications provide the core of the sensor querying and data retrieval tools. SensorML is utilized for application level metadata supporting discovery of sensors and sensing platforms. Testing of a GeoServer based Sensor Observation Service is underway. Integration of HIS offers opportunities to share data with other hydrologic research efforts and use existing CUAHSI developed analysis and visualization tools.

Acknowledgements.

This work is supported through the *Clemson PSA Remote Sensing Initiative and Clemson Computing and Information Technology.*

LITERATURE CITED

Botts, M. et al., 2007. OGC Sensor Web Enablement: Overview and High Level Architecture. OpenGIS Whitepaper. Version 3. Open Geospatial Consortium Inc.

CUAHSI 2010 (last access). CUAHSI Hydrologic Information System. http://his.cuahsi.org.

G. W. Eidson, S. T. Esswein, J. B. Gemmill, et al. 2010. The South Carolina Digital Watershed: End-to-End Support for Real-Time Management of Water Resources. International Journal of Distributed Sensor Networks, vol. 2010, Article ID 970868, 8 p.

Hooper, R., Maidment, D., Helly, J., Kumar, P., Piasecki, M 2004. CUAHSI Hydrologic Information Systems, 20th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography and Hydrology.

IEEE 2007. Standard for a Smart Transducer Interface for Sensors and Actuators—Common Functions, Communication Protocols, and Transducer Electronic Data Sheet (TEDS) Formats, *IEEE STD 1451.0-2007*, IEEE Instrumentation and Measurement Society, TC-9. The Institute of Electrical and Electronics Engineers, Inc., New York, NY

IEEE 1999. Standard for a Smart Transducer Interface for Sensors and Actuators— Network Capable Application Processor (NCAP) Information Model, *IEEE STD 1451.1-1999*, IEEE Instrumentation and Measurement Society, TC-9. The Institute of Electrical and Electronics Engineers, Inc., New York, NY.

Open Geospatial Consortium Inc. 2008. OpenGIS: Sensor Web Enablement Architecture. OGC 06-021r4, Open Geospatial Consortium Inc.

Pallickara, S. and Fox, G., 2003. NaradaBrokering: A middleware framework and architecture for enabling durable peer-to-peer grids. In The ACM/IFIP/USENIX International Middleware Conference, pages 41-61, Berlin, Germany, Spring-Verlag.

RabbitMQ 2010 (last access). Messaging that just works. http://www.rabbitmq.com/.

Song, E.Y., Kang, L., 2008. "Understanding IEEE 1451-Networked smart transducer interface standard - What is a smart transducer?," *Instrumentation & Measurement Magazine, IEEE* , vol.11, no.2, pp.11-17.