Near-Real-Time Spatiotemporal Precipitation Virtual Sensor Creation based on NEXRAD Level II Data in a Semantically-enhanced Digital Watershed

Yong Liu*, David J. Hill*, Alejandro Rodriguez*, Luigi Marini *, Rob Kooper* , Joe Futrelle*, Barbara Minsker§, James D. Myers*

*National Center for Supercomputing Applications (NCSA)
University of Illinois at Urbana-Champaign
1205 W. Clark St.
Urbana, IL 61801, USA
{yongliu, djhill1, alejandr, lmarini, kooper, futrelle, jimmyers@ncsa.uiuc.edu}

§NCSA and Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign
205 North Mathews Ave
Urbana, IL 61801, USA
minsker@illinois.edu

ABSTRACT

In this demonstration paper, we describe the technologies and implementations that allow near real-time creation of new virtual precipitation sensors using NEXRAD Level II streaming data from the National Weather Service at user-specified point locations and time intervals in an integrated digital watershed with a Google Map-based web interface. The spatiotemporal and thematic transformation workflows to produce such new time series data stream are created using NCSA's CyberIntegrator tool, which are run periodically upon the arrival of new radar data. A streaming data ontology is developed to handle temporal proximity concepts such as "previous" and "next" for irregular temporal data streams. Data and metadata management is provided by Tupelo, a semantic content management middleware. The new point-based virtual sensor can lower the barriers of using NEXRAD data for many hydrological applications as well as can be used for anomaly detection and calibration. The technologies of this demonstration can be applied beyond NEXRAD data streams, since the NEXRAD data include challenging characteristics common to many other types of environmental sensors such as irregular temporal frequency, irregular spatial coverage, and missing data.

Keywords

virtual sensor, NEXRAD, precipitation, spatiotemporal and thematic transformation, workflow, ontology, digital watershed

1. INTRODUCTION

The rapid development and deployment of environmental sensor networks provides opportunities to observe the earth and environment in unprecedented ways [1]. However, environmental sensor networks typically exhibit behaviors such as irregular temporal frequency, irregular spatial coverage, and missing data, which present significant barriers for the use of their data by the general scientific community. An example of such a sensor network is the Next Generation Radar (NEXRAD) system, which is a network of approximately 160 high-resolution Doppler weather radars operated by the National Weather Service.

In this demonstration paper, we describe technologies and implementations that can drastically lower the barriers to the adoption of NEXRAD Level II data (hereafter refer to as NEXRAD data) by environmental researchers through the creation of point-based virtual precipitation sensors in near real-time at user-specified locations and time intervals in a semantically enhanced digital watershed. Our approach includes general spatiotemporal and thematic workflow capability, streaming data and virtual sensor management middleware, and an intuitive Google Map-based web interface. A virtual sensor is defined as an abstraction of a physical sensor created through aggregation and transformation of the raw sensor data (see e.g., [3]). Creation of such virtual sensors will enable future applications such as near-real-time detection of sensor anomalies as well as other environmental and hydrologic watershed modeling and applications.

2. BACKGROUND AND RELATED WORK

The NEXRAD system measures reflectivity, radial velocity and spectrum width of the radar echoes returned from volumes within the atmosphere at a frequency of 5, 6, or 10 minutes (but never exact) depending on the operation mode of the radar. The reflectivity is correlated with the precipitation rate. NEXRAD data is available through a variety of means as downloadable data files. Recently, Krajewski *et al.* [2] presented the Hydro-NEXRAD prototype, a system that will provide researchers with NEXRAD data at the watershed level via a web-based map. It is important to note that, currently, none of these mechanisms are designed for near-real-time transformation and aggregation of the NEXRAD data, nor they allow researchers to implement their own transformation of reflectivity to rainfall measurements. The demonstration case in this paper illustrates our ability to handle near-real-time irregular temporal streaming data with a user-composable workflow that is run periodically, triggered by the newly arriving data.

3. TECHNOLOGY AND IMPLEMENTATION

In this section, we describe the details of architectural components of the integrated digital watersheds and processing steps for the creation of virtual sensors.

3.1Architectural Components

The semantically-enhanced digital watershed consists of the following components: a Google Map-based web interface front end, a virtual sensor middleware layer, a streaming data/metadata ontology and related middleware (Tupelo[5] and streaming data API), and workflow tools and services (CyberIntegrator [4]). Figure 1 shows a schematic view of the architectural components.

The Google Map interface interacts with the NCSA's Digital Synthesis Environment (DSE) using RESTful (REpresentational State Transfer) services to schedule and modify periodic

CyberIntegrator workflow executions. Users can register and de-register virtual sensors in the Tupelo-managed registry through the virtual sensor management middleware, and subsequently retrieves streaming data via a streaming data API. Tupelo data/metadata semantic middleware is implemented using RDF (Resource Definition Framework), a W3C semantic web standard. The streaming ontology [6] is an OWL-DL ontology that allows the semantic representation of arbitrary time-series data.

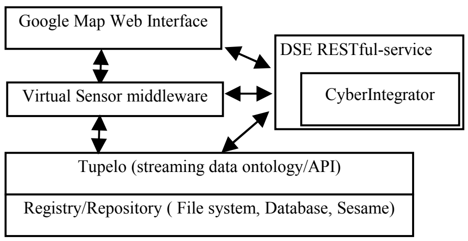

Figure 1. Components of the Integrated Digital Watershed

3.2 Steps for the Virtual Sensor Creation

To create a virtual precipitation sensor, spatiotemporal and thematic transformation of the NEXRAD data is implemented as a set of scientific workflows. These workflows are created in the CyberIntegrator toolkit [4]. The first workflow retrieves the latest data from the NEXRAD stream, extracts the reflectivity measurements from user-specified points in the radar coverage region using spatial interpolation, and performs a thematic transformation of the reflectivity data into rainfall rate (see Figure 2, steps 0, 1, 2). The result of the first workflow can be published as multiple data streams (i.e., one data stream per virtual sensor; step-3). For those virtual sensors where rainfall accumulation needs to be computed, a second workflow is launched at specified times, which performs temporal aggregations of the step-3 data, producing "virtual sensor" data streams that consist of *n*-minute rainfall accumulations (steps 4 and 5). The virtual sensor middleware (not shown in Figure 2) provides tools to create and update virtual sensor coordinates and stream IDs in the Tupelo registry, so that both workflows can obtain the necessary input information. Specifically, step 1 needs the coordinates of the virtual sensor, and step 4 needs the ID of the relevant stream The reasons that our implementation uses two workflows for the processing are two-fold: first, users might be interested in only the rainfall rates virtual sensor, which does not require further processing; secondly, workflow 2 is triggered at different time intervals than that of workflow 1.

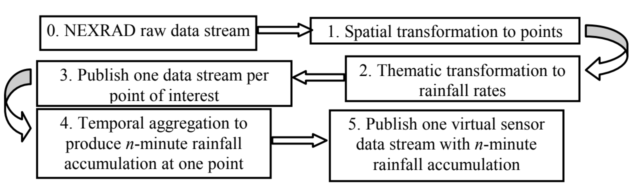

Figure 2. Virtual Sensor Creation Process

4. A DEMONSTRATION CASE

The demonstration of this prototype system's user-level functionality is shown in the context of a use case of the Chicago watershed where near real-time precipitation virtual sensor can be created based on data from one NEXRAD radar station. A Google Map-based web browser front-end enables users to "point-and-click" any point in the watershed that's covered by the radar and create virtual sensors on-the-fly. A user specified time-interval is used to define the desired rainfall accumulation resulting in a new time-series data stream. Users can use the map interface to add, update, remove and see the locations of existing virtual sensors. The new virtual sensor data stream can be accessed from both API level and visualization level. Additionally, each virtual sensor generates a time series plot of precipitation data.

5. CONCLUDING REMARKS

This paper explores the near-real-time streaming data processing based on the semantic streaming data and virtual sensor management middleware and workflow tools. The demonstration highlights the system's effectiveness in creating precipitation virtual sensors from NEXRAD data and presents the capabilities of the general toolkit. We will also discuss potential extensions to the system including handling data streams from multiple NEXRAD stations, creating polygon-based virtual sensors, as well as linking to continuously running data-driven modeling tools. Potential performance and efficiency enhancements that will allow our approach to scale to many heterogeneous sensors (i.e. the sensor web) will also be outlined, as will the role of the technologies presented here in a near-real-time combined sewage overflow forecasting and control project in Chicago.

6. ACKNOWLEDGMENTS

The authors would like to acknowledge the Office of Naval Research, which supports this work as part of the Technology Research, Education, and Commercialization Center (TRECC) (Research Grant N00014-04-1-0437) managed by NCSA. We also thank UIUC/NCSA Adaptive Environmental Sensing and Information Systems (AESIS) initiative for funding part of the work through Grant 1-200250-251024ICR. Finally, we would like to acknowledge the entire TRECC team at NCSA working on the Digital Synthesis Framework for Virtual Observatories.

7. REFERENCES

[1] Gross, Neil (1999). The Earth Will Don an Electronic Skin, Business Week, August 1999.

[2] Krajewski ,W. F. et al. (2008) Hydro-NEXRAD: community resource for use of radar-rainfall data, CUAHSI CyberSeminar April 25, 2008. Available at http://www.cuahsi.org/cyberseminars/Krajewski-20080425.pdf

[3] Liu, Yong., et al.(2008), Virtual Sensor-Powered Spatiotemporal Aggregation and Transformation: A Case Study Analyzing Near-Real-Time NEXRAD and Precipitation Gage Data in a Digital Watershed, Accepted at Environmental Information Management Conference 2008, September 10 - 11, 2008, University of New Mexico, Albuquerque, NM

[4] Marini, L., et al. (2007), Supporting exploration and collaboration in scientific workflow systems, in AGU, Fall Meet. Suppl., Abstract IN31C-07. 2007: San Francisco, CA.

[5] Tupelo semantic content repository. http://dlt.ncsa.uiuc.edu/wiki/index.php/Main_Page, 2008.

[6] Streaming data ontology. http://ar1.ncsa.uiuc.edu/StreamingData/DataStreamDL.owl, 2008.