Environmental Modelling & Software 84 (2016) 505–517

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Real-time environmental sensor data: An application to water quality using web services

Brandon P. Wong*, Branko Kerkez

Civil and Environmental Engineering, University of Michigan, United States

ARTICLE INFO

Article history:
Received 15 February 2016
Received in revised form
5 June 2016
Accepted 30 July 2016

Keywords:
Real-time data
Web services
Sensor networks
Adaptive sampling
Internet of things (IoT)

ABSTRACT

While real-time sensor feeds have the potential to transform both environmental science and decision-making, such data *are rarely* part of real-time workflows, analyses and modeling tool chains. Despite benefits ranging from detecting malfunctioning sensors to adaptive sampling, the limited number and complexity of existing real-time platforms across environmental domains pose a barrier to the adoption of real-time data. We present an architecture built upon 1) the increasing availability of new technologies to expose environmental sensors as web services, and 2) the merging of these services under recent innovations on the *Internet of Things (IoT)*. By leveraging recent developments in the IoT arena, the environmental sciences stand to make significant gains in the use of real-time data. We describe a use case in the hydrologic sciences, where an adaptive sampling algorithm is successfully deployed to optimize the use of a constrained sensor network resource.

Software and data availability

The use case was implemented using the Xively Internet of Things platform and a Flask web-server (written in Python 2.7) running on an Elastic Beanstalk t2.micro instance provided by Amazon Web Services. All experimental data from this study are hosted on a secure Xively feed and available upon request. The source code and implementation parameters are available on a public repository: https://github.com/kLabUM/IoT. As of 2015, all of these tools are available at no cost for a project of the scale discussed in this paper. Web connectivity is required of all hardware and software. For additional information, contact bpwong@umich.edu.

1. Introduction

Recent advances in sensing, computation and communications have enabled a massive suite of low-cost, low-power connected devices. This is particularly true for modern wireless sensor networks (Christodoulou et al., 2010, Jin et al., 2010), which now support the reliable, low-cost, near-instant transmission of measurements from field-deployed sensors. For enterprise-scale web applications, *RESTful* web services have also witnessed a surge in popularity (Kübert et al., 2011) while advances in the hardware realm have been accompanied by new architectures and protocols that exploit the bidirectional communication and Internet-connectivity of embedded devices. As such, libraries and *application programming interfaces* (*APIs*) enable users to quickly deploy RESTful web services on almost any software or hardware platform. This is significant, as most new devices from popular hardware and datalogger manufacturers increasingly support web communication via Wi-Fi, Ethernet, cellular, and other physical channels. Through these efforts, the *Internet of Things* (*IoT*) has recently been proposed as the backbone that will route and manage the vast quantities of data collected by these sensor networks (Atzori et al., 2010, Christin et al., 2009). In many environmental applications, however, these technological advances merely serve as a convenience to reduce field visits, provide data visualization, and simplify data collection by streaming sensor data to central repositories for subsequent analysis. Real-time data are rarely used as part of automated workflows, analyses and modeling tool chains.

In the computer science communities, in particular the area of embedded systems, the definition of real-time carries with it explicit performance guarantees, such as deadlines and timing constraints (Lee and Seshia, 2011). Such a strict definition, however, may be too technical to appeal to the broader environmental communities. While an actual definition may be out of reach

---

* Corresponding author.
E-mail addresses: bpwong@umich.edu (B.P. Wong), bkerkez@umich.edu (B. Kerkez).

http://dx.doi.org/10.1016/j.envsoft.2016.07.020
1364-8152/

considering the diversity of applications in the environmental sciences, an underlying principle persists: *real-time data* are data available for use as soon as they are collected to make a decision within a constrained time window, independent of sampling frequencies. This principle does not seek to distinguish between notions of real-time or near real-time, as is often the case in many studies (Christodoulou et al., 2010, Zhang et al., 2012, Hefeeda and Bagheri, 2009).

While not ubiquitous across the broader environmental domains, the use of real-time data for decision-making is not novel in some fields. For example, in the atmospheric sciences, satellite data is assimilated daily into advanced models which are used by various scientists and decision makers (Lakshmanan et al., 2007), while across meteorology, real-time radar feeds and terrestrial sensors inform stake holders across agriculture, transportation and disaster response (Zhang et al., 2011). However, despite the availability of low-cost, low-power hardware and data platforms, the benefits of these real-time resources have yet to be leveraged broadly across the remaining environmental sciences.

Scientific data analyses are more commonly conducted after an experiment has been completed, which for many studies could last months or years. A reason for the lack of real-time data adoption relates to the fact that most scientists may simply be satisfied with *continuous,* rather than *real-time,* data. The use of sensors across the environmental sciences thus appears to be retroactive, rather than adaptive. This would suggest that the major benefit of real-time data relates to decision-making, where assimilation of sensor information into models will enable rapid response to extreme events such as floods, wildfires and earthquakes.

While the ability to respond to natural disasters is invaluable, significant benefits of real-time data arise to environmental researchers as well, especially in the detection of faulty sensors and data acquisition systems. This is particularly true for experiments in harsh or remote environments where site visits may be infrequent and equipment outages can result in significant lapses in continuous data streams. For such experiments, real-time alerts will go a long way toward improving the quality of continuous data sets.

Perhaps the most compelling benefit of real-time data relates to the ability to usher in a new generation of adaptive scientific experiments. By adding real-time functionality to non-real-time studies, scientists will be able to perform innovative studies that respond to dynamic experimental conditions. As illustrated in this paper, this includes the ability to guide an experiment in real-time to adaptively sample signals or locations of interest during the most relevant intervals, which will significantly improve the use of constrained experimental resources and thus the quality of scientific experiments.

Across many domains, the notion of real-time is often complicated by operational requirements, which drive a lack of consensus around the definition of the actual term. Regardless of application, however, the utility of real-time data is governed by constrained time windows during which decisions have to be made. These time windows can range from days (e.g. climate modeling and data assimilation (Zhang et al., 2012)) to minutes or seconds (e.g. flood or wildfire forecasting (Hefeeda and Bagheri, 2009)). Outside of these time windows the data can be classified as *historical*, thereby limiting their utility for immediate decision-making. A wealth of tools have been developed to store, process and visualize historical sensor data (Argent et al., 2009, Castronova et al., 2013, Horsburgh et al., 2009, Gronewold et al., 2013, Goodall et al., 2008), but these frameworks have yet to be extended to provide real-time functionality.

In this paper, we present a summary of existing efforts to enable the use of real-time data across a broad set of domains, showing that the complexity and limited number of these existing real-time data platforms limits their adoption by the environmental sensing community. The majority of these platforms requires persistent expert support and cannot always be easily ported to existing field equipment and sensor networks, even by experienced researchers who readily operate continuous sensing campaigns. With real-time data systems also come different operational requirements, including the ability to continuously update and operate on new data, communicate with remote sites, monitor the operational status of devices, and manage user privileges throughout the system. We discuss these barriers to adoption and present a solution built upon two cornerstones: 1) the shift of environmental sensors and actuators[1] to a more generic web service model, and 2) the merging of these services under the recent architectural innovations on the Internet of Things.

To that end, we introduce a web service-centric approach to enable a flexible, reliable and powerful means by which to store, transmit and analyze real-time data. By focusing on recent advances in the IoT arena, we will show that the environmental sciences stand to make rapid gains in the use of real-time data while simultaneously improving flexibility related to implementation and maintenance. Rather than building a new platform, we will show how existing IoT platforms already provide a backbone to integrate real-time data from web-enabled environmental sensors and devices to meet requirements of interoperability, support, reliability, and security. By leveraging the services provided by these platforms, these web-enabled sensors and devices can also seamlessly interact with a multitude of web resources, including powerful cloud computing services and web-based models. A use case from the hydrologic sciences illustrates how a script can be deployed as a web service within this framework to enable low-power sensor networks to adaptively sample dynamic water quality parameters during storm events. While not a one-size-fits-all solution, our approach is expected to conform well to the requirements of most environmental applications, particularly for those where large sensor networks are deployed.

2. Existing platforms and real-time data efforts

Data systems employed across the environmental domains may be broadly classified into two groups: 1) systems used for the storage, retrieval and visualization of data, and 2) data systems designed explicitly for real-time operations. While the former do not explicitly treat real-time data, they do provide powerful mechanisms by which to standardize data retrieval and storage (Christodoulou et al., 2010, Horsburgh et al., 2009, Argent et al., 2009, Gronewold et al., 2013, Goodall et al., 2008). Some of these platforms conform to a set of community standards (e.g. *WaterML*, *DelftFEWS*, etc., see (Taylor et al., 2010)) that reduce operational overhead and enable the seamless use of standard-compliant tools for scalable storage, management and visualization of data. However, interactions with data are often carried out through direct user queries to the system, with no or limited mechanisms in place to automatically notify users of new readings or events as they occur. Furthermore, such architectures are not typically designed to enable alerts or the discovery and access to field-deployed sensors or actuators, thus limiting their use in control-centric and decision-making applications.

A number of these systems are also designed for domain-specific applications, thus limiting their use across a broader set of domains. In most cases, end-users are required to implement and

---

[1] Sensors generate an electrical signal in response to stimuli from the environment. Actuators respond to an electrical signal and act upon their environment (e.g. a gate that opens or closes).