Jay Hatcher
Article 6

An Adaptive Cyberinfrastructure for Threat Management in Urban Water Distribution Systems

Summary:

Urban water distribution systems allow rapid spreading of a contaminant from a single location because of design aspects that allow the system to adjust to varying conditions. Systems usually have a high level of network interconnectivity and fluctuation in flow that allows the systems to handle varying water demand and bypass portions of the network that are not working. However, these same properties allow a contaminant to quickly reach other parts of the network. Contamination is usually not detected quickly be consumers or operators because of uncertainty in the state of the system. By the time the contamination is detected via reports by consumers or the water quality sensor network, it can be difficult to determine the source of the contamination, when it began, where additional measurements need to be taken to better locate the source, and the extent of the contamination.

This article proposes a model that can help answer these questions in real time using simulation models that use both Bayesian and optimization methods and using special-temporal data management methods on a grid computing framework. Components of the framework are arranged in a hierarchy, and each component mutually adjusts to the needs of the other components. The three levels of the hierarchy are Sensors and Data, Algorithms and Models, and Middleware and Resources.

The category Sensors and Data contains the components for both mobile and stationary wireless radio frequency network sensors as well as the static water quality sensor network already in place in urban water systems. These components send data to a data center that forwards the collected data to a secure data store via a separate network to avoid interfering with the wireless network. The final component of this layer of the hierarchy is the Adaptive Data Receptor and Controller. It simulates the water system's behavior based on stochastic models of individual customer water usage. The testing environment will also simulate contaminant introduction and the response of the installed sensors for evaluating the performance of the system.

The Models and Algorithms layer deals with trying to find the source of a contamination. The time-series of sensor data is analyzed to try to determine likely sources of contamination and release histories that minimize error between simulation predictions and sensor observations. A future application is to also try and control the contamination to meet threat management objectives by adjusting hydraulics in the network.

The Optimization Engine component provides a set of optimization procedures that search for a likely source based on measurements, calibrate the model with new data, and find an effective control strategy for managing contamination. The complexity of the methods used is dynamically based on available computational resources. Data is added to the search in real time as it arrives from the sensors.

The Bayesian/Monte-Carlo Engine component uses procedures to identify efficient sampling plans for the sensors to reduce prediction uncertainties. Samples are taken where they are believed to yield the most useful information using Bayesian analysis. Markov chain Monte Carlo methods for Bayesian analysis are used for efficiency. Adaptive approaches for balancing run-time and accuracy are to be explored.

The Simulation Engine component is performed by EPANET, a water distribution network hydraulic and water quality modeling tool. EPANET uses well-established methods to model the water network. It is interfaced with the simulation controller to dynamically adapt to resource availability and adjust to new data. The Simulation Engine will be coupled to real network models to develop algorithms and evaluate performance and scalability.

The Optimization and Simulation Controllers select appropriate optimization algorithms or Monte-Carlo simulation given the state of the system, including available resources and communications from other components. Information from these communications allows the model controller to select the optimal model resolution and resource requirements to launch new simulations, alter model data, or terminate simulations.

The final layer of the hierarchy is Middleware and Resources. This layer coordinates processes to manage workflow between the sensor network, model calculations, available Grid resources, and human operators. The Grid Resource Broker and Scheduler components deal with two scenarios, requests to access available resources for a calculation, and acquisition of resources in order to predict a very accurate calculation in an emergency situation. Several middleware packages help to meet the first goal, and evaluation of an on-demand emergency ticket system is planned with assistance from the TeraGrid community.

The Adaptive Workflow Portal allows access to the system to the user community. The workflow portal can browse through dynamically changing workflows. A workflow component repository will allow dynamic changing of components at runtime and the development of automatic adaptive algorithms.

This entire system works together to dynamically address the needs for threat analysis and management in urban water distribution systems. The system provides real time feed of data from automatic meter readers and sensors to assist urban authorities in effectively managing water contamination occurrences.

Application of DDDAS:

This system is a complex DDDAS. The simulation interacts dynamically with the sensors, determining how much data to read and when to read it based on predictions made by the simulation and the current available resources. The data collected further informs the simulation and the resource allocation system of new developments which affect what parts of the sensor network will be examined more closely and what resources will be allocated to different parts of the system. Information coming from the simulation also inform the selection of new models for the simulation, which can replace existing models or run with current simulations in parallel, allowing alternative predictions that further inform the system as long as resources are available. The entire system is very dynamic and adaptable. Each component of the system both affects and is affected by others in a valiant attempt to maximize the use of dynamically changing resources and provide an appropriate response to a large number of possible scenarios.

Ways to Improve DDDAS Application:

The use of DDDAS in this system is very well developed. However, the nature of communications between these different components was not discussed in much detail, and communication overhead could prevent the system from behaving as well as it seems to on paper. Communication optimization is certainly surmountable and has most likely been done to some extent in actual implementations of this system. If additional types of sensors or newer prediction methods where added in the future it is somewhat uncertain how well the system would scale. It is very modular, so swapping out an old method for a better one should not be a problem, but adding more components might put a strain on the system. Even so, the dynamic nature of the system would almost certainly put up with some additional components before possibly running into problems. The system seems very well designed.