Motivation

In most process monitoring and control systems, data is typically sampled at the workstations and compressed before being archived in historians. Feedback is provided through alarm and control messages, and visualizations to both the system itself and to human operators. Analytics applications typically read data from these historians, process the data, and provide results and visualization off-line or in a time period that is considerably slow in comparison to the performance of the manufacturing process. Because of the focus on control, only a minimal set of process conditions and plant equipment is monitored. There needs to be a way to both improve the performance of the analytics system and scale the system to monitor a much larger set of plant resources in real-time.

Contributions

- Designed an efficient and scalable data management and analytics framework for large-scale industrial process monitoring and control systems.
- Large-volume distributed real-time data flows collected from heterogeneous data sources are streamed, stored, processed, and visualized in the proposed systems in near real-time.
- Designed a generic real-time streaming protocol, and a novel database schema on HBase for efficient data loading and retrieving.
- Implemented various decision making algorithms in process control applications into logical computation topologies on Storm and mapped them to physical machines in the cluster to fully utilize the parallel computational capability.

Data Flow and Real-time Processing Framework

We use a combination of Kafka and Storm frameworks to support real-time messaging and processing on time series data.

Efficient Database Schema Design

TAG_TABLE: This table applies a two way lookup trick to store the <Tag ID, Tag Name> and <Tag Name, Tag ID> pairs.

DATA_TABLE: Every row in this table holds an hour’s raw data for a tag.

AGG_HH_DATA_TABLE: Every row in this table holds an hour’s aggregated data for a tag with the scale of minutes.

Real-time Data Visualization

We implemented a rich set of web services by accessing the HBase through Thrift interface. This enables the users to visualize time series data and their statistics in a real-time manner.

Architecture of the Real-time Analytics Platform

Heterogeneous Data Sources

Hardware equipment and testbed setup in UConn Wireless Sensing and Control Lab

WirelessHART mesh network simulator

RPI-based imaging systems for plant monitoring

An overview of the crude simulator

I am working on designing low-power and high-speed real-time wireless platforms for pervasive industrial monitoring and control. I am also working on the development of a scalable data management framework for large-scale industrial sensing and control applications, such that large-volume distributed real-time data flows collected from heterogeneous sensors and actuators can be streamed, stored, processed, and feedback to the systems in real-time.