Data Aggregation in SCADA Networks

Hoang Nguyen (hnguyen5@uiuc.edu) and Klara Nahrstedt (klara.cs.uiuc.edu)
Department of Computer Science
University of Illinois at Urbana-Champaign

Abstract

Supervisory Control and Data Acquisition (SCADA) systems have been used in various critical infrastructures such as power grid and water treatment systems. A SCADA system consists of a number of remote terminal units (RTUs) collecting sensor data from Intelligent Electronic Devices (IEDs) and sending them to a control center, via a WAN (Wide Area Networks) communication system. It has been known that traditional SCADA systems are vulnerable to cyber attacks. In this work, we propose an architecture and first set of algorithms to support malfunction detection in SCADA networks, caused by software update from vendors, faulty devices or malicious attacks.

First, we discuss an architecture for malfunction detection in SCADA networks containing three major components: semantic data aggregators, topology resource aggregators and alarm aggregators. Semantic aggregation component collects raw data from devices and stores them into an efficient representation for timely alarm detection. Topology aggregation component monitors resources of all devices and aggregates their QoS into a simple representation for timely and secure admission control. Based on the output of the first two components, alarm aggregation component will detect, correlate and report most important alarms to the operators.

Second, within our architecture, we explore, at this point, two major issues

1) Sampling methods, important for any data aggregation and
2) the semantic data aggregation

We propose three different sampling methods for data aggregation. The first scheme works by letting the center samples data directly, sequentially and periodically from each sensor. The second scheme lets a single intermediate aggregator aggregates data from sensors and reports to the center. The last scheme has multiple intermediate aggregators collecting data from sensors.

Finally, we focus on semantic data aggregation problem. We show that this problem can be considered as a mean-shift and variance-shift detection problem. Essentially, the goal is to minimize the worst detection delay while keeping false alarm rate less than a pre-determined threshold. We modify and apply a non-parametric CUSUM (cumulative sum) algorithm to solve the semantic data aggregation problem. We implement the three schemes in ns2 and compare them with various topologies and communication settings. The preliminary result shows the relations between false alarm rate and average detection delay under different sampling methods.

This material is based upon work supported by the National Science Foundation under Grant No. CNS-0524695.