Towards Full-disclosure: Broadening Access to SCADA Data to Improve Safety, Reliability, and Security

John Heidemann and Wei Ye—*USC/Information Sciences Institute*

The premise of this position paper is that *openness and support for disclosure* of future SCADA *data* can in improved safety, reliability, and security for networked embedded control of physical systems. Open protocols, with new approaches to access control and in-network processing are needed to make greater openness and data sharing data viable.

Today's SCADA systems are often vertically integrated, largely proprietary systems. Interoperability of systems is often limited. While closed protocols and systems may maximize the short-term benefit of systems providers, they limit the ability of the field to grow and constrain innovation by new players. There is an increasing and important trend toward open SCADA protocols, allowing third party tools to manipulate the data.

More than protocols, *data* from today's systems is often tightly controlled. In part, access may be limited because specialized software required to access the data, but more often data is constrained due to policy reasons. In some cases there are legitimate monetary, security or privacy concerns that prevent data availability. However, in many cases data is restricted due to concerns about interference (access to information might invite change or interference), worries about data quality (this data is not verified), or fears of data manipulation (can data readers alter the data), organizational impediments (data cannot leave our department), or because release is simply not considered (who would want this information?). We argue that, along with open protocols, technical and administrative steps are needed to open data.

Clearly not all data can be made public: some may reveal proprietary business processes, while other may contain personal details of employees or customers. We describe technical approaches below that support *controlled* release of information, filtered when necessary.

We argue that, for SCADA systems, open protocols and greater openness with data go hand-in-hand. The argument for open protocols, designs, and implementations is inspired by the general benefits full disclosure has brought to the security community, the strengths of open network protocols such as the Internet and world-wide web, and the prevalence of open-source software in some markets such as web and DNS servers. These systems provide strong records of reliability and security, in part due to their wide use and peer review.

In SCADA systems, the SCADA infrastructure is relatively mundane, and the real value of the system lies in the data itself. We believe that greater openness in SCADA data is needed to provide impetus for developers to build better tools and for users to explore and use those tools.

The combination of access to SCADA data with new tools that allow its manipulation, processing, and filtering, and improve reliability, safety, security and ultimately result in novel new business processes and services. Improvements in reliability come from standard approaches to manipulate data and more diverse and interchangeable tools rather than inflexible, stovepipe systems. Greater sharing and use of data can improve safety by making data available in places where it was not before—for example, by making automation process data available to worker's cellular telephones or pagers, allowing automatic alerts should they enter a dangerous area. Improvements to security seem counter-intuitive, since greater openness perhaps implies greater risk. However, some cumbersome data sharing is a fact of life even today, but it is often done ad hoc and uncontrolled (cut and paste and exporting to Excel). We believe approaches that explicitly manage sharing, access control, and anonymization, with conscious decisions about what to share and what not to share, ultimately improve security. Ultimately, we believe that all of these enable *new approaches* in industrial automation. Much as information technology has improved office productivity, we expect that greater understanding of the data and processes captured by SCADA systems and result in improved business processes, new opportunities such as data mining, and potentially whole new applications.

We are not the first to suggest protocols standardization, and in fact standards efforts for SCADA systems are prominent in industry groups and the IEEE. Instead, we focus on wider internal and public access to SCADA *data*, and protocols need to support this goal. Below we briefly consider potential benefits of greater openness of SCADA data for several application areas and the protocol and technical changes needed to make such openness feasible.

1 Application Areas

We next briefly consider two areas where relatively wide data openness is possible, and then the general case of industrial SCADA systems.

Automotive manufacturers are exploring in-vehicle automation and between-vehicle safety systems. Safe sharing between vehicles is necessary in such systems to provide cross-vehicle warnings of breaking and collision prevention. More general support for safe sharing, validation, and anonymization of data will enable applications such as automatic reporting of traffic congestion and data collection for long-term traffic planning.

Environmental systems (HVAC) see increasing use of distributed temperature monitoring and control. While HVAC information can reveal some business information (for example, how many offices are occupied), it is relatively benign. If readily available in a standard format, third parties could propose energy optimizations and utilities could better understand power usage and potentials to time-shift load.

Industrial automation and control represent the traditional use of SCADA systems, and on the surface seem to represent a clear case where data must be protected to preserve competitive advantage. While public openness of industrial SCADA data may not be possible, businesses often must communicate with partners, subcontractors, government and collaborative industrial organizations. We suggest that the same mechanisms needed for selective release of other forms of data will also be useful for managing release of industrial data. Moreover, greater internal use of SCADA data may enable process process improvements.

2 Challenges

To reach our goal of public access to SCADA systems requires advances in protocols, much richer access and anonymization control, and a shift to smart-devices in-the-field. We review each of these challenges next.

Common Protocols: Common protocols for data exchange are essential to make wide access to SCADA data possible. Several points in a SCADA system provide opportunities for open access: from the sensor in-the-field to a local data concentrator or logger, from the field to local operations headquarters, and from local facilities to regional or national operations centers. Different trade-offs arise at each level, from very simple analog interfaces at the lowest levels to wide-area network protocols at higher levels. Standardization efforts such as IEEE 1451 are important here; the key requirement is network protocols that allow data interchange between third party equipment.

To manage security and privacy, it is essential that standards allow interoperation at the *data* level, not simply at the packet or connection level. We describe the security and privacy issues below, but to the protocol this requires that intermediaries must be able to interpret SCADA data, so common protocols must specify metadata such as sensor type, calibration, data units, and similar information.

Data-level access is also important enable services that support data search and indexing, naming and linking, allowing data and phenomena to be discoverable.

Access and Anonymization Controls: Given the commercial nature of most SCADA data, several aspects security and access control must be considered.

First, one must have *access control and permissions* mechanisms to manage disclosure of data publicly and to different groups within an organization.

Second, we see the need for data-level "firewalls" that allow selective release of information based on these access controls. However, we see the need for far richer control than simple yes/no decisions on data source and destination—a firewall should be able to dispatch data at the granularity of individual readings. The ability to delay and aggregate data is also essential, since coarser data may be more acceptable for release than raw data. (For example, in the U.S., airline flight and stock trade information is available with a 15 minute time delay.) Fine control and data processing require knowledge about the data types that pass through a data-firewall, knowledge provided by standard metadata and exchange protocols described above.

Finally, anonymization is an important option. For commercially sensitive data, anonymization may reduce concerns about disclosure while still preserving some value. For example, Internet Service Providers often consider network topologies and traffic rates proprietary, but they are sometimes willing to release non-geographically identified subsets of data. Non-commercial data may still provide information about individuals that needs to be concealed. For example, while many drivers do not want their exact location publicized, they would be willing to provided anonymous information to a traffic congestion-estimation service.

In-Situ Intelligence: For common protocols and in-network access control are designed to enable *in-situ* intelligent processing of SCADA data. We believe that in-network and in-the-field processing of data is essential because it is often best place to make choices about access control and privacy and because it allows sanitization of data (aggregation and anonymization) to occur before release. The ability to move computation into the field also enables more sophisticated SCADA approaches such as report-on-exception (rather than report-always polling).