First Steps Toward Scientific Cyber-Security Experimentation in Wide-Area Cyber-Physical Systems

Ryan Goodfellow
School of Electrical Engineering
and Computer Science
Washington State University
Pullman, WA
rgoodfel@eecs.wsu.edu

Robert Braden
Information Sciences Institute
University of Southern California
Marina del Rey, CA
braden@isi.edu

Terry Benzel
Information Sciences Institute
University of Southern California
Marina del Rey, CA
tbenzel@isi.edu

David E. Bakken
School of Electrical Engineering and Computer Science
Washington State University
Pullman, WA
bakken@wsu.edu

ABSTRACT

This extended abstract reports on steps towards an environment for repeatable and scalable experiments on wide-area cyber-physical systems. The cyber-physical systems that underlie the world's critical infrastructure are increasingly vulnerable to attack and failure. Our work has focused on secure and resilient communication technology for the electric power grid, a subset of the general cyber-physical problem. We have demonstrated tools and methodology for experimentation with GridStat, a middleware system designed to provide enhanced communication service for the grid, within the DETERlab cyber-security testbed. Experiment design tools for DETERlab and for GridStat will ease the creation and execution of relatively large experiments, and they should make this environment accessible to users inexperienced with cluster testbeds. This abstract presents brief overviews of DETERLab and of GridStat and describes their integration. It also describes a large scale GridStat/DETERlab experiment.

This research was sponsored by the US Department of Homeland Security and the Space and Naval Warfare Systems Center, San Diego (contract number N66001-10-C-2018), and the US National Science Foundation (contract number CNS-0831491). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of Homeland Security or the Space and Naval Warfare Systems Center, San Diego.

CSIIRW '12, Oct 30 - Nov 01 2012, Oak Ridge, TN, USA
ACM 978-1-4503-1687-3/12/10.

Categories and Subject Descriptors

D.4.6 [Security]: Protection; D.4.8 [Performance Measurements]: [Modeling and Prediction, Monitors]

General Terms

Algorithms, Management, Measurement, Performance, Design, Economics, Reliability, Experimentation, Security, Verification

Keywords

Testbed, Overlay Network, Smart Grid, Cyber-security, Cyber-Physical, Experimental Research

1. INTRODUCTION

Cyber-physical (C-P) systems are central to much of today's critical infrastructure. The increasing scale and complexity as well as the geographical distribution of these C-P systems create new vulnerabilities to attacks and failures, whether in the physical realm, in the cyber realm, or in their coupling. A broad array of research and development efforts is attempting to counter these vulnerabilities [5]. However, this research is hampered by a lack of suitable facilities for testing wide-area C-P algorithms and protocols. Such a facility would simulate or emulate wide-area communication with acceptable fidelity while scaling to medium-scale experiments. It would be readily and flexibly configurable and should be capable of supporting arbitrary protocol stacks. Our objective is to fill this gap.

Our area of interest is a large and important C-P system, the electric power grid. This abstract reports on an initial step towards an experimentation environment for wide-area communication and control architectures for the electric power grid. This environment will support large scale, repeatable experiments in a flexible and scalable manner, yet with good fidelity. In addition, it attempts to provide a user interface based on concepts familiar to experts in the particular C-P domain. We suggest that this facility will be useful for researchers developing new algorithms and for vendors designing and building products.

The electric power grid problem space can be partitioned into three almost-independent areas --- physical power system, wide-area data communication, and monitoring and control applications. Our approach is to simulate a representative portion of the power grid and to emulate the communication and application components using ISI's DETERlab cyber-security testbed [7]. The DETERlab testbed is adapted to safe and repeatable cyber-security experimentation, including containment of "risky" malware. As summarized in Section 2, DETERlab can provide a fairly realistic modeling of the cyber aspects of the problem while supporting the application software that performs monitoring and control functions.

The best choice of communication protocol stack is being debated in the power industry today, and an important application of our experimentation framework will be to investigate the strengths and weaknesses of candidates. Our initial work is based in particular on the experimental middleware package called GridStat [6], an important contender for a general-purpose wide-area communication system for C-P systems.

Section 2 outlines the properties of DETERlab while Section 3 overviews the GridStat system. Section 4 describes some technical aspects of deploying GridStat within DETERlab and Section 5 reports on a large-scale instantiation of this GridStat-over-DETERlab facility. The concluding Section 6 presents some future directions for this work.

2. DETERLAB OVERVIEW

The DETER project at the USC Information Sciences Institute operates the DETERlab cyber-security testbed in support of a large world-wide community of researchers in cyber security and other computer science subfields. The DETER project also executes its own research program in testbed technology. This program is aimed at substantially advancing the infrastructures, tools, and methodologies underlying DETERlab.

DETERlab uses the cluster testbed architecture developed at the University of Utah and implemented in their cluster testbed called Emulab [13]. The DETERlab testbed is implemented as two clusters, at ISI and at UC Berkeley. The facility currently has pools of roughly 500 high-performance PC nodes and 10 FPGA-based reconfigurable hardware elements. These nodes are interconnected by dynamically-configurable VLAN Ethernet switches. The testbed control plane allocates these resources upon user request to multiple simultaneous and independent experiments. Users receive exclusive hardware-level access to the machines allocated to them, and they can set up network topologies, operating systems, and applications of their choice. The testbed control plane is an extended version of Utah's Emulab [13] testbed control software.

PC nodes may be used as hosts, routers, traffic shapers, traffic generators, firewalls, etc. The system automatically inserts traffic shaping nodes ("delay nodes") to emulate wide-area link characteristics such as bandwidth, delay, and loss. To create a DETERlab experiment, a user prepares a description of the desired network infrastructure as a set of nodes inter-connected by links and LANs, and the specific software to be loaded. From this description, the testbed control plane allocates the requested nodes, sets up VLANs to achieve the requested topology, and loads the requested operating systems and application software.. After the nodes have been successfully initialized, the experimenter may log into the experiment's nodes (as root, if desired) to set up and run the experiment. Much of this set up process can be automated with startup scripts executed by the testbed control software.

The testbed research program of the DETER project has developed tools to generate large and complex experiment descriptions and to automate experiment execution. These tools enable practical experimentation with large distributed systems. For example, DETERlab's Magi framework [9] allows experimenters to coordinate testbed activities – configuring and starting distributed systems in sequence, monitoring execution, and subsequent analysis of results.

3. GRIDSTAT MIDDLEWARE

Monitoring and control systems operate in local and wide areas to ensure stability and reliability of the power grid. These high-level functions are implemented by low-level sensing and actuation functions of intelligent electronic devices (IEDs). IEDs distributed throughout the power system are linked by a communication system to monitoring and control applications in centralized control centers. The focus of the work presented here is providing an experimentation environment for this communication system.

This environment has been demonstrated using GridStat, a data delivery middleware package developed at Washington State University. GridStat provides a reliable and efficient communication service for control center applications and IEDs, by building an overlay network on top of the normal Internet and Transport protocol layers and by performing in-network processing. More specifically, GridStat can perform five important functions for the electric grid C-P system.

(1) GridStat provides the strong quality-of-service (QOS) guarantees demanded by control systems. The next generation of wide-area monitoring and control [8] will be enabled by advanced IEDs that are capable of high data rates and use high-speed packet-switched communication. As a result, they will be sensitive to QoS attributes such as latency, packet drop rates, etc [6] [12].

(2) Monitoring data is streamed over the network in real time and is typically delivered to multiple control centers. GridStat therefore provides (middleware-layer) multicasting to minimize bandwidth for monitoring data.

(3). Many monitoring applications require streamed data at a rate that is only a sub-multiple of the rate at which the data is produced by IEDs. GridStat takes advantage of this to economize on link bandwidth by "down-sampling" the data stream whenever possible. While this requires processing within the GridStat network that is specific to the format of the data, this down-sampling has significant benefit for the scalability of the network [10]

(4) GridStat guards against lost data (e.g., due to link congestion) by sending each data stream multiple times across redundant paths. Redundant data is discarded at the receiver.

(5) GridStat implements a data-centric publish/subscribe paradigm. For each logical data stream ("variable"), GridStat constructs multiple redundant and disjoint data paths (”virtual circuits") between publisher and subscriber, while ensuring that

the data rate and latency requirements of the subscription are met.

The GridStat data delivery middleware uses four logical entity types: publishers, subscribers, forwarding engines, and QoS brokers. Publishers and subscribers are collectively called “participants” in a GridStat overlay network. Forwarding engines push data from upstream publishers to downstream subscribers. The set of all forwarding engines in a GridStat network is called the data plane.

A QoS broker is a centralized controller for the distributed data plane. When a new subscriber request is passed to a QoS broker, the broker creates the necessary paths (virtual circuits) through the data plane to satisfy the QoS parameters of the request. QoS brokers can exist in a hierarchy; where each leaf broker manages some subset of the overall data plane, while the broker at the top would have complete visibility.

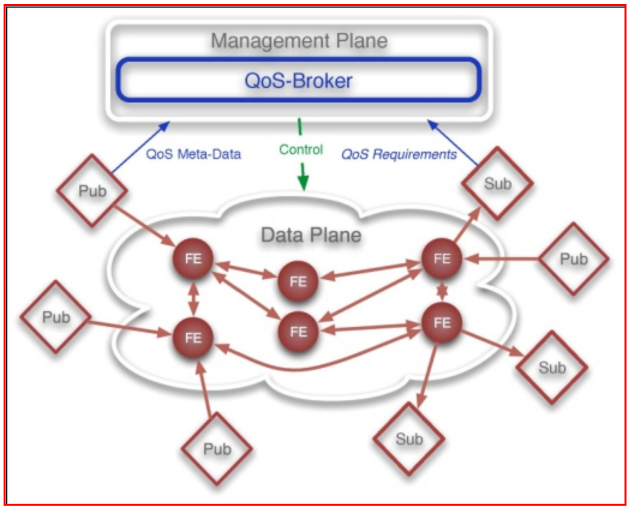

Figure 1: Simple GridStat Network

Figure 1 shows a simple GridStat network. When a publisher comes online, it sends QoS meta-data to its QoS broker. The QoS broker is thus aware of both the data variables (streams) available for subscription and the maximum rate at which this data may be provided. When a subscriber requests a variable at a given rate, the broker(s) in the management plane construct a path for the specified variable through the data plane from the publisher to the subscriber, at the rate that the subscriber has requested. If this is not possible. the management plane rejects the subscription request. The management plane will construct data paths in such a way that publishers and forwarding engines move data across each link at the minimum rate as dictated by subscriptions.

Configuring a GridStat network requires specification of many parameters, such as the data plane topology, how participants hook into the data plane, publication variable names, and QoS parameters for subscribers. GridStat users can conveniently set these configuration variables using an intuitive higher-level configuration language called GridStat Structured Text (GSST). The GSST compiler outputs the configuration files for all components of the GridStat overlay network.

4. GRIDSTAT IN DETERLAB

Our initial goal was to instantiate an experimentation environment specifically using GridStat and studying the data communication aspects of the C-P system. We have not yet incorporated actual simulation of the power grid or control center processing of the data.

As a middleware system, GridStat is invoked by data sources and sinks using GidStat’s upward-facing API. Using this API, applications can push (or pull) data over wide-area networks without worrying about how the transport actually takes place. We wanted to enable use of the GridStat system without burdening users with writing custom code to interface data sources and applications to the GridStat API. C-P system experimentation is certainly not limited to programmers.

To meet these objectives, we provide “canned” data sources and sinks. The canned sources (publishers) accept simple command-line arguments pointing to synchrophasor data archives that the publisher will ’play out’ at a specified constant rate. Similarly, subscribers are pre-programmed to archive data into an SQL database. This avoids the need to write a custom subscriber against using the GridStat API. However, the programming option is available if real data streams are desired.

The GSST compiler output configuration files for the “canned” sources and sinks. For users who wish to perform experiments but have no data we have built a small API and embedded domain specific language in Haskell for generating synchrophasor data frame archives and the associated configuration frames.

For our purposes it would have been possible to send arbitrary binary data, but for credibility in the power community it seemed worthwhile to use a real data format. While GridStat can be used as a general purpose data delivery system for rate-based traffic, our DETERlab instantiation is restricted to synchrophasor data in the widely-used IEEE C37.118-201 format [4].

To perform an experiment with GridStat in DETERlab, a power engineer would:
(1) define the topology and configuration of the Internet underlay network, using the normal DETERlab tools,
(2) define the GridStat configuration in GSST and compile it into configuration files,
(3) start the GridStat components, and
(4) observe and perhaps measure the communication.

For security experiments for example, an experimenter can take down one or more links or invoke DDoS attack generators or other malware facilities in DETERlab.

5. A LARGE SCALE EXPERIMENT

Using the tools presented in the previous sections, we have created a large scale GridStat experiment that runs on DETERlab. The design of the experiment is inspired by the real world transmission system operated by Bonneville Power Administration [1]. This system includes approximately 500 substations and 600 transmission lines. We deploy a Publisher of mock synchrophasor data to correspond to each end of every electrical transmission line. These data sources stream pre-recorded data at a maximum rate of 60 Hz.

For the data plane, the publishers and subscribers are co located with their publishing and subscribing devices or applications. In our system, a publishing "device" is a program playing out archived synchrophasor data, while a subscriber application is a database connection.file. The overlay network consists of five forwarding engines that are assumed to be sited at key substations along major transmission paths.

A Tech Report [in preparation] describes in detail the design and implementation of this transmission system test network using the GSST and Magi tools described here.

6. FUTURE DIRECTIONS

This work shows that DETERlab can provide an intermediate-scale testing environment for real-world distributed C-P technology built around GridStat middleware, and that tools can render this testing facility accessible to power system engineers as well as computer scientists. We seek to attract power engineers developing synchrophasor applications for wide-area protection, monitoring and next-generation control.

So far, our work has side-stepped the physical aspects of the C-P system by playing out pre-recorded data files The next major advance will be to connect a power system simulator with monitoring hooks into the emulated wide-area network.This will mean installing a simulation engine of sufficient power to create the physical grid dynamics and provide the C-P coupling into the cyber emulation in DETERlab. The interesting problem here will be the synchronization of the simulation (which must run in real time or faster) with the emulation (which necessarily runs in real time.)

Another step forward could be to provide the DETERlab testbed with precision timing, a key piece of additional infrastructure that will enable a much wider range of C-P system experimentation. Access to accurate time as a first class resource in the DETERlab testbed will allow experimenters to move beyond canned data sources to simulate/emulate actual IEDs.

Finally, we plan to utilize the experiment capability to learn how GridStat behaves under stress. One research area to be investigated is security of the GridStat management plane. A malicious broker or an attack on the management plane can have detrimental effects on the GridStat system. Preventing this will require tight management of the interactions between QoS-Brokers and participants. As a cyber-security testbed, DETERlab provides an optimal environment to experiment with defending the management plane, at a scale that models the real target operating environment.

REFERENCES

[1] Bonneville power administration. http://www.bpa.gov/. Accessed: Aug 10 2012.

[2] IRIGB standard. http://www.irigb.com/. Accessed: Aug 10 2012.

[3] IEEE standard for a precision clock synchronization protocol for networked measurement and control systems. IEEE Std 1588-2008 (Revision of IEEE Std 1588-2002), pages c1 –269, 24 2008.

[4] IEEE standard for synchrophasor measurements for power systems. IEEE Std C37.118.1-2011 (Revision of IEEE Std C37.118-2005), pages 1 –61, 28 2011.

[5] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr. Basic concepts and taxonomy of dependable and secure computing. Dependable and Secure Computing, IEEE Transactions on, 1(1):11 – 33, jan.-march 2004.

[6] D. Bakken, A. Bose, C. Hauser, D. Whitehead, and G. Zweigle. Smart generation and transmission with coherent, real-time data. Proceedings of the IEEE, 99(6):928 –951, june 2011.

[7] T. Benzel. The science of cyber security experimentation: the deter project. In Proceedings of the 27th Annual Computer Security Applications Conference, ACSAC ’11, pages 137–148, New York, NY, USA, 2011. ACM.

[8] A. Bose. Smart transmission grid applications and their supporting infrastructure. Smart Grid, IEEE Transactions on, 1(1):11 –19, june 2010.

[9] A. Husain, B. Wilson, and G. Lawler. Scalable workflows for networking and cyber security experiments. 2012.

[10] V. Irava. Low-Cost Delay-Constrained Multicast Routing Heuristics and their Evaluation. PhD thesis, Washington State University, 2006.

[11] S. Schwab, B. Wilson, C. Ko, and A. Hussain. Seer: a security experimentation environment for deter. In Proceedings of the DETER Community Workshop on Cyber Security Experimentation and Test on DETER Community Workshop on Cyber Security Experimentation and Test 2007, DETER, pages 2–2, Berkeley, CA, USA, 2007. USENIX Association.

[12] J. Stahlhut, T. Browne, G. Heydt, and V. Vittal. Latency viewed as a stochastic process and its impact on wide area power system control signals. Power Systems, IEEE Transactions on, 23(1):84 –91, feb. 2008.

[13] B. White, J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M. Newbold, M. Hibler, C. Barb, and A. Joglekar. An integrated experimental environment for distributed systems and networks. In Proc. of the Fifth Symposium on Operating Systems Design and Implementation, pages 255–270, Boston, MA, Dec. 2002. USENIX Association