Chapter 22

AN ARCHITECTURE FOR SCADA NETWORK FORENSICS

T. Kilpatrick, J. Gonzalez, R. Chandia, M. Papa and S. Shenoi

Abstract Supervisory control and data acquisition (SCADA) systems are widely used in industrial control and automation. Modern SCADA protocols often employ TCP/IP to transport sensor data and control signals. Meanwhile, corporate IT infrastructures are interconnecting with previously isolated SCADA networks. The use of TCP/IP as a carrier protocol and the interconnection of IT and SCADA networks raise serious security issues. This paper describes an architecture for SCADA network forensics. In addition to supporting forensic investigations of SCADA network incidents, the architecture incorporates mechanisms for monitoring process behavior, analyzing trends and optimizing plant performance.

Keywords: Process control systems, SCADA networks, network forensics

1. Introduction

Process control systems (PCSs) – often referred to as supervisory control and data acquisition (SCADA) systems – are widely used in manufacturing processes, chemical plant and refinery operations, and even for automation and climate control in office buildings [7]. In a typical SCADA implementation, sensors acquire data pertaining to process behavior; this data is passed to control algorithms implemented in the SCADA system. Depending on the sensor data and control objectives, output signals are sent to actuators that adjust process inputs and move the process to the desired state.

In many industrial environments, sensors, actuators and control software are deployed in different locations, which requires the implementation of a communications infrastructure. Consequently, modern SCADA protocols, e.g., Modbus [16, 17] and DNP3 [21, 22], now include specifi-

Please use the following format when citing this chapter:

Kilpatrick, T., Gonzalez, J., Chandia, R., Papa, M., Shenoi, S., 2006 in International Federation for Information Processing, Volume 222, Advances in Digital Forensics II, eds. Olivier, M., Shenoi, S., (Boston: Springer), pp. 273–285.

cations for transporting sensor data and control signals using TCP/IP. TCP/IP is becoming the dominant transport mechanism in SCADA networks. Also, TCP/IP is facilitating interconnections between previously isolated SCADA networks and corporate IT infrastructures via gateway devices [6, 11].

The use of TCP/IP as a carrier protocol for SCADA systems and the interconnection of SCADA and IT networks raises serious security issues [9, 10, 13, 23]. Because SCADA systems were historically isolated from other networks, most SCADA protocols were designed without any security mechanisms. Transporting a SCADA protocol over TCP/IP means that an attack on the TCP/IP carrier can severely expose the unprotected SCADA protocol. Furthermore, the connectivity between IT and SCADA networks means that successful attacks on an IT network and its gateway devices could tunnel into a SCADA network, wreaking havoc on the industrial process [10].

This paper describes an architecture for SCADA network forensics. In addition to supporting forensic investigations of SCADA network incidents, the architecture incorporates mechanisms for monitoring process behavior, analyzing trends and optimizing plant performance.

2. SCADA Network Overview

SCADA networks have myriad variants ranging from small, locally centralized networks such as those used in simple manufacturing plants to vast, distributed networks used in refineries, oil and gas pipelines, and power distribution facilities. Figure 1 provides a generic view of a SCADA network, which is used throughout this paper as a framework for discussing the functional, design and interconnectivity elements of SCADA networks.

A SCADA network has two main components: the control center (top left corner of Figure 1) and the plant it controls (Site A in Figure 1). The two components are connected to each other via a SCADA server.

The control center is the hub of SCADA network operations. Its components include human machine interfaces (HMIs), engineering workstations, plant data historians, databases and various shared resources. Control center components communicate with each other using the management network, and with the plant (Site A) and other SCADA networks using the SCADA server.

The SCADA server functions as the sole interface between the control center and one or more sites for which the control center is responsible. The server is usually implemented with vendor specific software and its services are often based on the OPC standard.

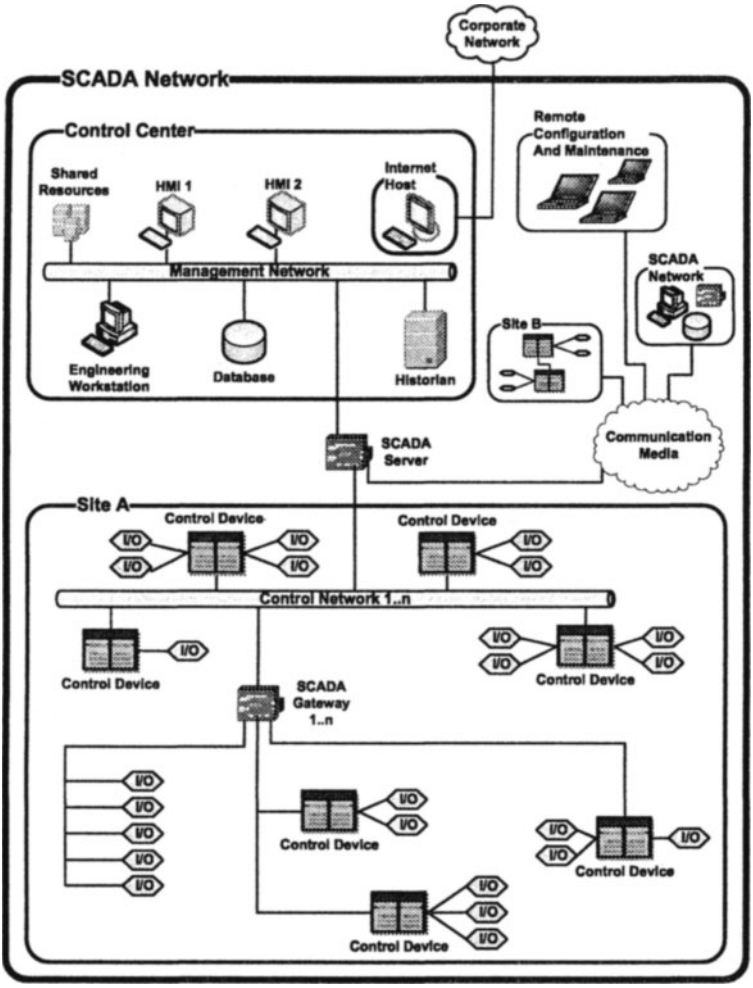

Figure 1. Generic SCADA network architecture.

The site or plant is the actual process system. From the process control point of view, it has three types of components: control devices, I/O devices (sensors and actuators) and the SCADA gateway. A control network interconnects components within a site. The 1..n notation is used in Figure 1 to express the fact that control network components may employ different protocols, e.g., Modbus and DNP3.

Control devices, which implement the process control logic, interface with and manipulate I/O devices. They include programmable logic controllers (PLCs), remote terminal units (RTUs), input/output controllers (IOCs) and intelligent electronic devices (IEDs). I/O devices

include sensors and actuators. Sensors measure specific process parameters (e.g., temperature and pressure) while actuators perform control actions (e.g., opening or closing a valve) to effect the desired changes to the process parameters.

The SCADA gateway is conceptualized at a high level of abstraction as a device that interfaces all control network components that cannot communicate directly with the SCADA server. Depending on its functionality, any control device can serve as a SCADA gateway. Special units called front-end processors (FEPs) are commonly used as SCADA gateways in industrial settings.

As mentioned earlier, control center components include human machine interfaces (HMIs), engineering workstations, databases, plant data historians and various shared resources. HMIs permit human operators in the control center to interact with process systems at a site in a limited manner. An operator at an engineering workstation has more authority over the SCADA network, and can reconfigure control devices and modify control algorithms (e.g., ladder logic) as needed.

Relevant data pertaining to process parameters and control actions is recorded in a database. Engineering workstation and HMI operators interact with the database to access and modify process data and control variables.

Historians archive data about SCADA network activities. The data includes sensor data, control actions effected by engineering workstation and HMI operators, and management network logs.

The management network may also contain various shared resources (e.g., printers, fax machines and file servers), but these are typically not considered part of the SCADA network. The generic SCADA network architecture in Figure 1 incorporates a separate Internet host that is connected to the corporate network. No physical connection exists between the SCADA management network and the Internet host. However, it is increasingly common for corporate networks to interconnect with previously isolated SCADA networks.

3. SCADA Network Forensics

The industrial use of SCADA systems has traditionally emphasized human and plant safety followed by the continuity of operations (availability). These requirements were maintained by isolating SCADA systems. However, the relatively recent encroachment of corporate intranets and even the open Internet into plant environments has made SCADA security a major issue. SCADA security, which was emphasized in Presidential Decision Directive 63 [24], has manifested itself in numerous

guidelines and standards promulgated by industry and government organizations such as AGA [1, 2], ANSI/ISA [3, 4], API [5] and NIST [18]. These guidelines and standards have been embraced by industry, and strong efforts are being undertaken across all sectors to secure SCADA systems [8, 14].

Forensics becomes relevant when security is breached [15]. SCADA network forensics involves the detailed examination of evidence related to security incidents. The goal is to discover the *modus operandi* and the identity of the perpetrator, as well as what went wrong so that the system can be hardened to prevent future incidents.

An architecture that supports SCADA network forensics can enhance industrial operations. Network forensics cannot be performed without mechanisms that systematically capture relevant traffic and state information throughout the network [15, 19, 20]. In the context of SCADA networks, the capture and subsequent analysis of sensor data and control actions assists in monitoring process behavior, examining trends and ultimately optimizing plant performance.

SCADA networks differ from corporate IT networks in several respects.

- **Security Principles:** Availability is the primary concern in most SCADA networks, followed by integrity and confidentiality. Corporate networks often emphasize confidentiality.

- **Protocol Isolation:** SCADA protocols, even those employing a TCP/IP carrier, are not used in an IT network. However, SCADA networks use traditional network protocols and are, therefore, vulnerable to attacks originating from or targeting IT networks.

- **Protocol Variations:** Most SCADA protocols are based on open standards. However, vendors often augment protocols or employ proprietary out-of-band mechanisms to provide additional functionality. On the other hand, protocol standards are enforced more strictly in traditional IT networks to support interconnectivity and interoperability.

- **Traffic Uniformity:** IT networks mainly have user-generated traffic. Consequently, the communication patterns are difficult to predict. SCADA network traffic is routine and predictable.

- **Traffic Volume:** IT networks have huge traffic volumes. SCADA networks have very light traffic. The messages tend to be shorter (e.g., Modbus messages never exceed a few hundred bytes) and fewer messages are transmitted.

- **Network Forensics:** Forensics in large-scale IT networks is extremely complicated and expensive [19, 20]. On the other hand, SCADA network forensics may be considerably simpler. Traffic uniformity and lower traffic volume in SCADA networks makes it possible to log relevant process/control data associated with every message and to subsequently analyze the data.

4. Incorporating Forensic Capabilities

This section describes an architecture that supports *post mortem* analysis of SCADA network traffic. The architecture employs "forensic agents" at strategic locations within a SCADA network. These agents forward relevant portions of network packets to a central location for storage and subsequent retrieval. A complete SCADA network history is obtained by reconstructing network events from packet fragments captured from locations throughout the SCADA network.

The architecture presented in Figure 2 incorporates two main entities: agents and a data warehouse. An agent captures SCADA network traffic in its local network segment and forwards a synopsis [19, 20] of each packet to the data warehouse. The data warehouse analyzes each packet synopsis and creates a data signature that it stores along with the synopsis in a storage area designated for the appropriate agent. The data warehouse also supports queries on the stored data. Note that an isolated network is used for all communications between agents and the data warehouse.

Generally, a SCADA network will have several types of agents and one data warehouse. A Level 1 agent is connected directly to the management network. Level 2 agents are located directly on the control networks. Level 3 agents are positioned downstream from the SCADA gateway. Note that the positioning of Level 3 agents is harder to categorize as the structures of SCADA networks differ considerably based on the industrial processes being controlled.

The data warehouse is typically housed in a secure location within the control center, but is not connected to the management network. Thus, the data warehouse is in close proximity to human operators, but is not vulnerable to exploits on the network that it monitors.

Some large industrial systems incorporate multiple interconnected SCADA networks. In such architectures, it is necessary to log traffic across different SCADA networks. This is accomplished by placing agents between the SCADA networks (Figure 3). The agents forward data to a common data warehouse that is not associated with a particular SCADA network. These units are referred to as Level 0 agents and

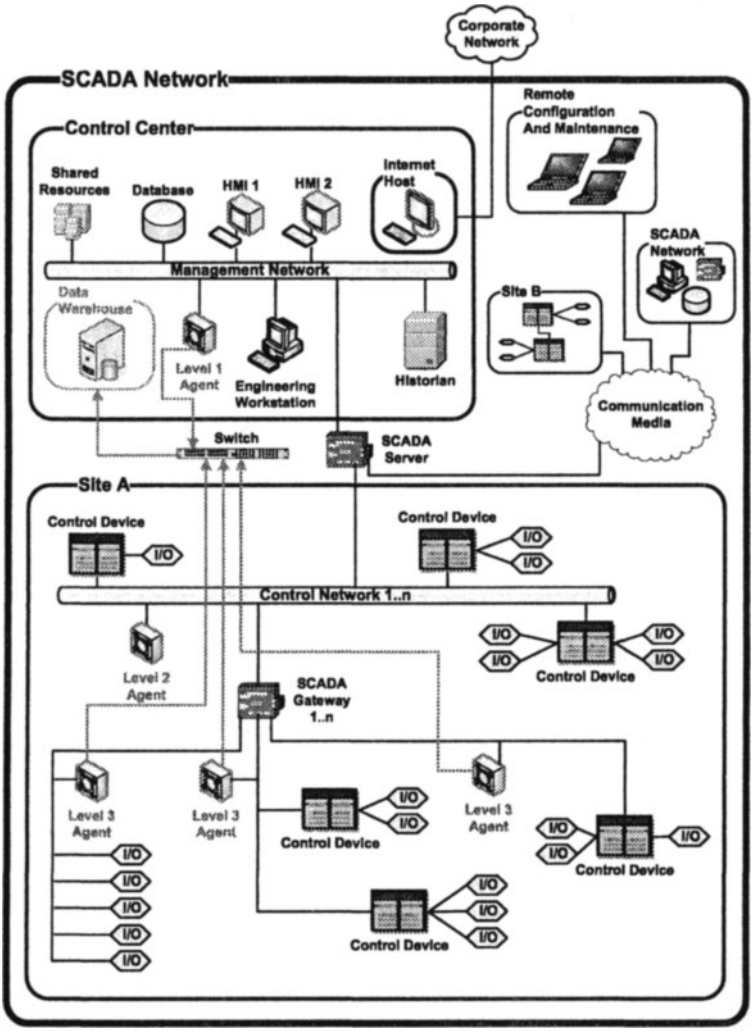

Figure 2. Generic SCADA network architecture with forensic capabilities.

Level 0 data warehouses. Note that depending on the interconnected network topology, there may be one or more Level 0 agents. We will see later that, in order to facilitate robust querying, the Level 0 data warehouse must be connected via a closed network to the data warehouses of the individual SCADA networks.

5. SCADA Traffic Collection and Analysis

Forensic agents capture and analyze SCADA network traffic. They create synopses of network packets that contain information relevant

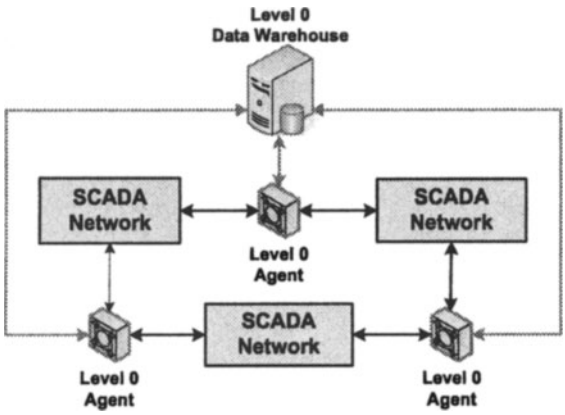

Figure 3. Communication between SCADA networks.

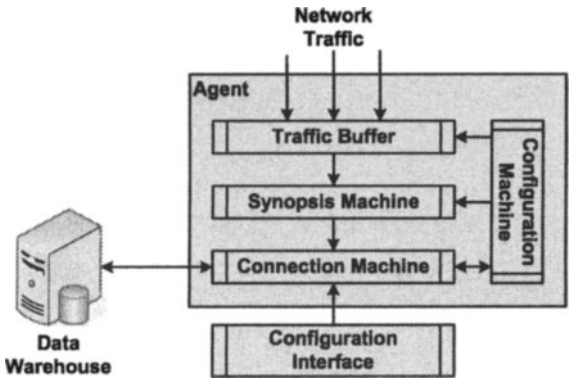

Figure 4. SCADA packet synopsis generation.

to forensic analysis [19, 20]. An agent incorporates a network traffic buffer, synopsis machine, connection machine and configuration machine (Figure 4).

The traffic buffer stores unprocessed network traffic. It employs a multi-threaded implementation of the standard producer/consumer algorithm and a bounded buffer. The `libpcap` [12] interface provides a mechanism for capturing raw Ethernet packets. Currently, development is focused on SCADA protocols (e.g., Modbus and DNP3) that can be transported using Ethernet frames. Future development plans also include support for serial communications (e.g., RS 232/485).

The synopsis machine constitutes the core of a forensic agent. It examines each packet from the traffic buffer and generates a packet synopsis

in the manner specified by the configuration rules. Note that partial synopses are produced for each encapsulating protocol. For example, for the OSI model, agents could produce layer 3 (network) and layer 4 (transport) synopses. Partial synopses are combined with location information and timestamps to produce a synopsis that is stored.

The connection machine supports agent communication. Secure communication is achieved by requiring architectural components to register with an authentication engine and using access control lists (ACLs) to implement mutual authentication.

Agent configuration is critical to the success of the forensic architecture. The configuration machine provides a mechanism to regulate operation of the agent at various levels (Figure 4), e.g., security settings, device registration, synopsis settings and protocol recognition modules. External devices attempting to configure an agent must be registered with the authentication engine and must use a common configuration interface. Some security settings are similar to those employed in IT networks, while others, such as synopsis settings, are unique to this architecture.

Proper configuration of the synopsis engine is important because of its role in the architecture. Two methods may be employed: level-based configuration and manual configuration. The level-based method configures agents according to their location, allowing agents to be configured with pre-defined synopsis algorithms. Agents are then configured as Level 0 (between SCADA networks), Level 1 (management network), Level 2 (control network) or Level 3 (behind a SCADA gateway). Configuration by level is simple and convenient. Manual configuration of agents may be performed to fine tune agent behavior and packet analysis.

Note that numerous SCADA protocols are used in industrial environments. Moreover, in addition to standard protocols, e.g., Modbus and DNP3, some deployments implement variations or subsets of standard protocols or proprietary protocols. The requirement to deal with diverse SCADA protocols has motivated the design of modular agents with configurable synopsis engines.

6. SCADA Traffic Storage and Querying

Forensic agents submit their SCADA traffic synopses to a designated data repository for storage. The design uses a relational database and query mechanisms to support forensic investigations. The SCADA traffic storage and querying facility incorporates a connection machine, data

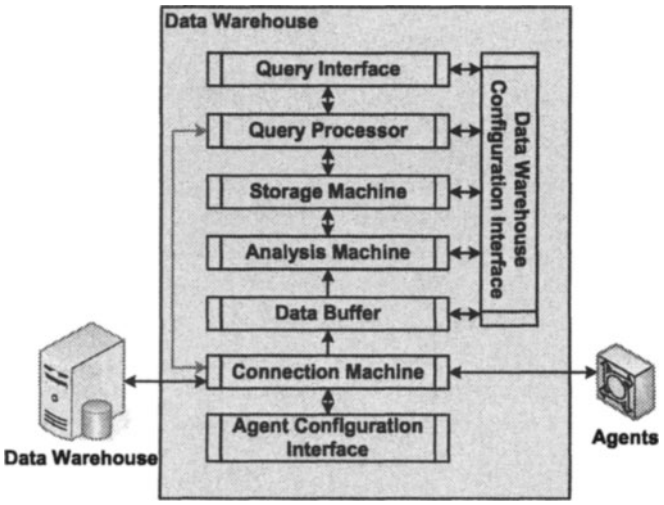

Figure 5. SCADA traffic storage and querying.

buffer, analysis machine, storage machine, query interface, query processor and agent configuration interface (Figure 5).

The connection machine supports communications between data warehouses and registered agents. Connections with registered agents are used to receive synopses and configure agents. Connections to other data warehouses facilitate the processing of queries that span multiple SCADA networks.

Synopses submitted by agents for storage are placed in the data buffer. Packet synopses are then passed to the analysis machine using a producer/consumer algorithm.

The analysis machine produces signatures associated with synopses that are used for event reconstruction as well as to analyze and correlate SCADA traffic patterns. Signatures reduce storage requirements while maintaining forensic capabilities. To illustrate the concept, consider a PLC that communicates with specific field devices. Therefore, only the relevant addresses must be stored and associated with this PLC. The corresponding device-based signature is generated by correlating synopses from all the agents that observe traffic associated with the PLC.

Pattern analysis capabilities are currently being developed for the forensic architecture. For example, PLCs execute repetitive control loops that have well-defined communication patterns. These patterns can be analyzed to produce network-based signatures for forensic investigations and anomaly detection.

The storage machine uses hash tables and a relational database. Each registered agent has a set of hash tables, which are used to index repetitive signature data associated with an agent. For example, the partial synopses generated for communication between two devices with largely static-address-oriented data need not be stored more than once. Instead a pointer to the proper entry is used as the signature, which is stored in the database, to identify the communication. The number of tables associated with an agent depends on the type and quantity of synopses generated by the agent.

The query interface enables forensic investigators to reconstruct incidents, perform system checks and analyze process trends. The current interface provides two SQL-based querying mechanisms. The first uses a GUI and a pre-defined set of elements and options to support routine analyses. The second provides a console that gives analysts complete freedom to specify queries. Results are provided in the form of reports augmented with graphical information about the SCADA network, devices and agents.

The query processor fields queries received from a local query interface or from another SCADA network via the connection machine. The query processor determines if the resolution of the query involves information from a different SCADA network. In this case, a query is sent to the appropriate data warehouse, which in turn dynamically generates and processes a query that is returned to the sender.

7. Conclusions

The use of TCP/IP as a carrier protocol and the interconnection of IT and SCADA networks expose industrial processes to attack over the Internet. Novel architectures are required to facilitate forensic investigations of SCADA network incidents. An architecture that supports SCADA network forensics can also enhance industrial operations. Network forensics requires mechanisms that systematically capture relevant traffic and state information throughout the network. In the context of SCADA networks, the capture and subsequent analysis of sensor data and control actions assists in monitoring process behavior, examining trends and ultimately optimizing plant performance.

The architecture proposed in the paper is specifically designed for SCADA networks. The architecture is scalable and flexible. Also, it can handle multiple interconnected SCADA networks and diverse protocols.

References

[1] American Gas Association, *Cryptographic Protection of SCADA Communications; Part 1: Background, Policies and Test Plan*, AGA Report No. 12 (Part 1), Draft 5 (www.gtiservices.org/security/AGA12Draft5r3.pdf), April 14, 2005.

[2] American Gas Association, *Cryptographic Protection of SCADA Communications; Part 2: Retrofit Link Encryption for Asynchronous Serial Communications*, AGA Report No. 12 (Part 2), Draft (www.gtiservices.org/security/aga-12p2-draft-0512.pdf), May 12, 2005.

[3] American National Standards Institute/Instrumentation Systems and Automation Society, *Security Technologies for Manufacturing and Control Systems (ANSI/ISA-TR99.00.01-2004)*, October 2004.

[4] American National Standards Institute/Instrumentation Systems and Automation Society, *Integrating Electronic Security into the Manufacturing and Control Systems Environment (ANSI/ISA-TR99.00.02-2004)*, October 2004.

[5] American Petroleum Institute, *API 1164, SCADA Security*, American Petroleum Institute, September 1, 2004.

[6] M. Berg and J. Stamp, A Reference Model for Control and Automation Systems in Electric Power, Technical Report SAND2005-1000C, Sandia National Laboratories, Albuquerque, New Mexico, 2005.

[7] S. Boyer, *SCADA: Supervisory Control and Data Acquisition (Third Edition)*, Instrumentation, Systems and Automation Society, Research Triangle Park, North Carolina, 2004.

[8] British Columbia Institute of Technology, *Good Practice Guide on Firewall Deployment for SCADA and Process Control Networks*, National Infrastructure Security Coordination Centre, London, United Kingdom, 2005.

[9] E. Byres, J. Carter, A. Elramly and D. Hoffman, Worlds in collision: Ethernet on the plant floor, *Proceedings of the ISA Emerging Technologies Conference*, 2002.

[10] E. Byres, M. Franz and D. Miller, The use of attack trees in assessing vulnerabilities in SCADA systems, *Proceedings of the International Infrastructure Survivability Workshop*, 2004.

[11] E. Byres and T. Nguyen, Using OPC to integrate control systems from competing vendors, *Proceedings of the Canadian Pulp and Paper Association Technical Conference*, 2000.

[12] B. Fenner, G. Harris and M. Richardson, The `libpcap` Project (sourceforge.net/projects/libpcap).

[13] J. Graham and S. Patel, Security Considerations in SCADA Communication Protocols, Technical Report TR-ISRL-04-01, Intelligent System Research Laboratory, Department of Computer Engineering and Computer Science, University of Louisville, Louisville, Kentucky, 2004.

[14] D. Kilman and J. Stamp, Framework for SCADA Security Policy, Technical Report SAND2005-1002C, Sandia National Laboratories, Albuquerque, New Mexico, 2005.

[15] K. Mandia, C. Prosise and M. Pepe, *Incident Response and Computer Forensics*, McGraw-Hill/Osborne, Emeryville, California, 2003.

[16] Modbus IDA, *MODBUS Application Protocol Specification v1.1a* (www.modbus.org/specs.php), June 4, 2004.

[17] Modbus IDA, *MODBUS Messaging on TCP/IP Implementation Guide v1.0a* (www.modbus.org/specs.php), June 4, 2004.

[18] National Institute of Standards and Technology, *System Protection Profile – Industrial Control Systems v1.0*, Gaithersburg, Maryland, 2004.

[19] K. Shanmugasundaram, H. Bronnimann and N. Memon, Integrating digital forensics in network architectures, in *Advances in Digital Forensics*, M. Pollitt and S. Shenoi (Eds.), Springer, New York, pp. 127-140, 2005.

[20] K. Shanmugasundaram, N. Memon, A. Savant and H. Bronnimann, Fornet: A distributed forensics system, *Proceedings of the Second International Workshop on Mathematical Methods, Models and Architectures for Computer Network Security*, 2003.

[21] M. Smith and M. Copps, *DNP3 V3.00 Data Object Library Version 0.02*, DNP Users Group, September 1993.

[22] M. Smith, and J. McFadyen, *DNP V3.00 Data Link Layer Protocol Description*, DNP Users Group, June 2000.

[23] J. Stamp, J. Dillinger, W. Young and J. Depoy, Common Vulnerabilities in Critical Infrastructure Control Systems, Technical Report SAND2003-1772C, Sandia National Laboratories, Albuquerque, New Mexico, 2003.

[24] The White House, *Presidential Decision Directive 63: Critical Infrastructure Protection*, National Security Council, Executive Office of the President, Washington, DC, 1998.