SCADA systems: Challenges for forensic investigators

Irfan Ahmed, Sebastian Obermeier, Martin Naedele, Golden G. Richard III

Abstract

SCADA systems run 24/7 to control and monitor industrial and infrastructure processes. In case of potential security incidents, several challenges exist for conducting an effective forensic investigation. This paper discusses these challenges and investigates potential solutions. It shows the limitations of traditional IT-based approaches and also presents research challenges for initiating and continuing research in SCADA systems. Furthermore, it discusses how the research community has tackled these problems so far.

Keywords: Security, Process control systems, digital forensics

Introduction

An industrial automation and control system is a set of devices used to regulate the behavior of physical processes. For example, a thermostat is a simple control system that senses the temperature and turns on or off the heater to maintain the temperature at a set point. Control systems are used to monitor and control industrial and infrastructure processes, such as chemical plants and refinery operations, electricity generation and distribution, and water management. If a control system is spread over a wide area and can supervise its individual components, it is often called a supervisory control and data acquisition (SCADA) system [1]. However, for brevity, we will use the term SCADA in this paper to stand for all kinds of control systems, which share a common key characteristic that such systems are connected to physical processes and thus, need to be continuously available and react with deterministic response time.

Early SCADA systems were intended to run as isolated networks, not connected to the Internet, and customers did not require any specific cyber security mechanisms. They were comprised of simple input/output devices that transmitted the signals between master and remote terminal units. Due to advances in technology, they have evolved and adopted current technologies (such as wireless IP communication). In recent years, more and more SCADA systems have evolved to communicate over public IP networks [2]. Some are also connected to the corporate intranet or directly to the Internet, in order to seamlessly integrate SCADA information and external information, e.g. corporate mail systems or weather data. The reachability of SCADA systems from a much wider network brings threats that were unimaginable at the time when those systems were conceived. Vendors, asset owners, and regulators have realized this increasing threat over the past decade, and have started to address it by means of security mechanisms, processes, standards, and regulation [3].

The discovery of Stuxnet in June 2010 was an additional eye opener for SCADA owners and operators. It is the first discovered malware that is specifically written to work against an automation system and has infected an estimated 50,000 to 100,000 computers worldwide [4]. More recently, another malware called “Flame” has been discovered, which is an espionage tool and is considered yet another order of magnitude more sophisticated than Stuxnet [5].
Figure 1 illustrates a typical architecture of a SCADA system showing its most common components. A SCADA system (for controlling infrastructures like power, gas, oil, or water) is generally comprised of a control center and field sites. The field sites are distributed over a wide geographical area, which are all connected to the control center by different communication media such as satellite, Wide Area Network (WAN) or radio/microwave/cellular networks. Field sites are equipped with field devices such as Programmable Logic Controllers (PLCs) or Remote Terminal Units (RTUs) that control the on-site machines and/or periodically send information about the state of the field equipment to the control center. The control center is the hub of the SCADA system. Its major components include Human Machine Interface (HMI), database management system (Historian) and Server or Master Terminal Unit (MTU) components. The MTU initiates all communication with field sites and receives the data sent from the field devices. It then, if necessary, preprocesses the data and sends it to the historian for archiving. The HMI presents the information to the human operator.

![Figure 1: A typical SCADA Architecture in a simplified logical view.](image)

**Forensics for SCADA Systems**

Digital Forensics is an essential part of cyber defense and becomes relevant when there is a security breach [6]. It can generally be defined as the collection and analysis of data from different relevant sources (such as storage devices and network streams) in a manner that is admissible to a court of law. It is performed on digital devices and usually used to investigate the cause and consequence of an incident where if the traces of a crime such as unauthorized network access or theft of a digital file are found, it may further lead to the admissibility of evidence to a court. However, not all the application of digital
forensics necessarily involves presenting the data or evidence in a court of law. For instance, digital forensics is also used for internal corporate investigation to find the cause of an incident in order to limit the possibility of the incident occurring again in the future.

Today, the reliability of many SCADA systems is not only dependent on safety, but also on security [3]. The recent attacks against SCADA systems by sophisticated malware (such as Stuxnet and Flame) demands forensic investigation to understand the cause and effects of the intrusion on such systems so that their cyber defense can be improved. Not just this, but when the news of unleashing the cyber dogs [7] to attack enemies is prevalent, forensic practice becomes essential to find the traces of an attack and gather evidence against the entity that has tried to sabotage the critical infrastructure of a country. Clearly, SCADA systems need to be protected from internal and external actors who have malicious intent. Forensic investigation can play a vital role in a protection strategy for SCADA systems and may assist in the prosecution of attackers, but also in a deep analysis of the underlying SCADA IT system, for example, in the case of non-malicious events such as malfunctioning hard disks or other hardware.

A forensic investigation can answer several intriguing questions about an incident. For instance, consider a scenario of a SCADA system recently hit by malware, which has caused the system to malfunction. A forensic investigation can be an effective way to answer questions such as:
- Is the SCADA system still compromised by malware?
- A virus scan revealed that the Java cache contains a known exploit. Was the exploit successful? What payload does it have, and has that compromised the system?
- How can the SCADA system operator clean the system after an infection, and reliably bring it back into a known good state, without having to shut down the complete system?
- An operator has installed a suspicious, untrusted application downloaded from the Internet. Did that application change components that are relevant for the stable operation of the SCADA system?

From a forensic perspective, a SCADA system can be viewed in different layers (as illustrated in Figure 2 as an example) based on the connectivity of the various SCADA components and their network connectivity with other networks such as the Internet [1]. In Figure 2, layer 0, which is the lowest layer, contains the individual field devices connected via a bus network. Layer 1 has controllers that receive input signals from the field devices and other controllers upon which they perform operations to steer the individual field devices, by sending output signals to them. Layer 2 consists of the supervisory network - typically a local network connected to the lower layers for specific operations such as showing current monitoring state at the HMI. Layer 3 is typically the operation DMZ, in which historians, domain controllers and application servers are located. The upper layers correspond to the enterprise IT networks, in which the regular enterprise desktops and business servers operate. Most of the forensic analysis in SCADA systems involves the first three layers (i.e. layers 0, 1 and 2) as they contain the special SCADA components and are crucial for controlling the underlying industrial processes. However, the analysis may further extend to the other, higher layers (i.e. layers 3, 4 and 5) if needed. This paper focuses on the first three layers mentioned in the diagram.
Live Forensics

A SCADA system has a critical requirement of being continuously operational and therefore a forensic investigator cannot turn off the SCADA system for data acquisition and analysis [8]. In this case, live forensics is a viable solution for digital investigation in SCADA systems. Live forensics [9] is a relatively new and emerging field in digital forensics where the data acquisition and analysis are performed on a running system. However, the critical nature of SCADA systems and the 24/7 availability requirement entails forensic investigators spending as little time on a live SCADA system as possible, necessarily performing live data acquisition and then subsequent offline analysis of the acquired data.
Live data acquisition involves acquiring both volatile data (such as the contents of physical memory) and non-volatile data (such as data stored on a hard disk). It is different from traditional dead disk acquisition, which involves bringing the system offline before the acquisition, where all volatile data is lost. Increasingly, volatile data plays a significant role in an effective and efficient investigation. For instance, volatile data in physical memory contains the current state of the system, such as the number of open network connections, process information, and encryption keys. However, given that volatile data in a running system changes continuously, a live data acquisition approach faces the following challenges in capturing data for effective investigation.

- **Early data acquisition after incident**: Live data acquisition needs to be performed as quickly as possible after an incident in order to capture any traces of the incident on volatile data before the processes or services on the running system overwrite useful data, such as information about recently unloaded kernel modules or drivers.

- **Digital evidence validity**: The admissibility of digital evidence in court may be affected if the evidence’s integrity is violated. The intention is to prevent the malicious manufacturing of evidence against an innocent person or avoid errors while handling evidence in the course of the forensic investigation. Forensic investigators normally prove the integrity of evidence by computing a cryptographic hash of the actual evidence on the compromised system and its acquired copy, which is used for all the examination and analysis. If, however, the compromised system remains live, the state of the data may change between the copying and the hash calculation, rendering hashing ineffective as an integrity check.

Moreover, this also creates an inconsistent data image that neither accurately represents the state of the data when data acquisition starts nor the state of the data after the acquisition ends, which may cause difficulty in analyzing the acquired data. For instance, due to the data inconsistency, sometimes an operating system in the disk image cannot boot for experimental analysis.

Despite the importance of live data acquisition, it is still unclear how contemporary live data acquisition tools should be run on a SCADA system so that they minimize risk to SCADA system services. To the best of authors’ knowledge, no such guidelines are available to date. However, there is a situation where safe data acquisition should be possible under many circumstances. Specifically, SCADA systems typically have a primary and a backup system. When the primary system is broken or malfunctioning, the SCADA system is switched to its backup system [10]. Forensic investigators might use this capability by switching from the system that is under investigation to the backup and performing live data acquisition without worrying about the availability of the SCADA services. However, switching to the backup system approach may not be feasible in some cases where the incident may have also affected the backup. For instance, the malware that has infected the primary system may also infect the identical backup system. That may even demand immediate recovery if the SCADA owners and operators decide that the incident can jeopardize the normal functionality of the system. This usually results in flushing all the infected system components and bringing them back to their normal state, which would not provide sufficient time to the investigator for data acquisition.
Forensic Challenges in SCADA systems

Forensic investigators have to deal with the problems arising from the unique features of SCADA systems, which prevent them from directly applying contemporary forensic tools and techniques.

The unique features of SCADA systems include:

**Deterministic network traffic**: Network traffic in SCADA systems is deterministic due to the fact that a system component communicates with other system components in a pre-defined manner. This contrasts with office IT systems, where desktop machines and servers communicate based on requests in a non-deterministic way [19]. Based on this deterministic behavior, stringent rules can be applied to harden the security of the system, with any non-deterministic behavior flagged as an anomaly. For instance, an intrusion detection system might be configured to consider a specific communication pattern as normal [11]. Security tools that expect such deterministic behavior may raise false alarms or prevent forensic tools from operating properly. For instance, a firewall might be configured with strict rules allowing communication between specific SCADA components, but disallow communication between the investigator’s machine and SCADA components during data acquisition.

**Customized operating system kernel**: SCADA system can have customized kernels running on its components despite that such kernels are hard to patch for new updates. Kernels are customized to achieve better performance, support critical applications etc. For instance, PatriotSCADA [12] is a firewall solution for SCADA networks that uses a customized Linux kernel to enforce access control and role-based security for every request in the kernel. However, it may not be possible to run the data acquisition tools on a customized kernel unless they are compatible with each other. For instance, the data acquisition tool DD may require loading the fmem kernel module (in Linux) to access the physical memory through the device /dev/fmem (that is created by the module), if the regular /dev/mem device in Linux is not accessible. Until the module is compiled with the customized kernel, it is possible that the module may not load into the kernel.

**Resource constrained devices**: The availability of SCADA system services is also subject to the availability of adequate system resources (i.e. CPU, memory, IO etc.). SCADA system components can be found running on legacy hardware and operating systems. Fabro et al. [13] classify SCADA system technologies into categories that include legacy/proprietary technologies, which may have been deployed for more than 10 years, have moderate computing capabilities if compared to modern systems, and have limited or no vendor support. Moreover, the field devices (such as RTUs and PLCs) are generally resource-constrained devices. In such cases, a SCADA system provides limited system resources for data acquisition. This demands lightweight data acquisition tools for SCADA systems.

**Inadequate logging**: Collection of adequate records or logs of events that happened near incident time is crucial for successful investigation. Logging capabilities of SCADA systems are geared towards discovering and diagnosing process disturbances, not security incidents, and are thus often not adequate for forensic investigation [13]. Thus, in such cases, historical visibility needs to be improved in SCADA system components, which can cover information about incidents to the extent where forensic investigation can be performed effectively.

**Tremendous amount of data on lower layers**: (Recalling from Figure 2), SCADA system consists of multiple layers. Capturing and analyzing data from lower layers in SCADA systems is challenging due to the large amount of data generated by individual sensors. For example, in electricity grids, sensors may
generate up to 4,000 measurements per second and only condensed information is forwarded to higher layers [29].

**Suggestions for facilitating forensics in SCADA systems**

Forensic process in SCADA system can be improved through preparedness and the selection of appropriate tools. We discuss these issues in the remainder of this section.

**Data acquisition plan**

The data acquisition process should be fast and well targeted to acquiring the most relevant data related to an incident. This can be achieved through a data acquisition plan where the design and operation of the system is well documented, with unique features of the system, application data flow, and temporary and permanent data storage locations carefully enumerated. Furthermore, the plan should also outline what data should be acquired for which incidents.

Fabro *et al.* [13] proposed guidelines for creating a cyber forensic plan for control systems. It consists of three phases: First, identifying the system environment and its unique characteristics, including whether the system has modern computing capabilities, is still fully supported by vendors, uses contemporary operating systems, and has continuing support from open source community for any open source components. Second, defining environment-specific requirements such as the impact of vendor solutions on operating systems and third, identification and collection of data, such as activity and transaction logs.

**Data acquisition monitoring**

During forensic acquisition, no matter how careful a forensic investigator is when copying data, there is always a risk of upsetting the availability of SCADA services. However, the risk can be mitigated if the availability of SCADA system services can be monitored during the data acquisition so that the acquisition process can be stopped in case of any serious perturbation. A monitoring tool can facilitate this process by detecting the perturbation as soon as it occurs and automating the response, to avoid any serious damage to the system. EnCase Cybersecurity of Guidance software [14] is a relevant exemplar tool that can be used for monitoring data acquisition to some extent by integrating it with alert or event management systems and configuring it for auto response to alerts or events.

**Lightweight data acquisition**

The data acquisition tools should have a minimal impact on system resources so that adequate resources are available for SCADA system services to work properly during the data acquisition process. To get a preliminary idea of how resource intensive data acquisition tools are, we ran three well-known variants of the disk copy tool DD (i.e. WinDD [28], Garner’s DD [15], and DD on Linux-variants) to acquire the whole physical memory and hard disk data of a computer and recorded the resource consumption of the computer during acquisition for analysis. The data were acquired over a 100Mbps network (using the Netcat tool [15]) – a preferred way for forensic investigators.

To emulate a resource-constrained system, we used a PC with a Celeron 1.7GHz CPU, 384MB RAM and 40 GB Hard disk (having 7200 rpm speed). We used two different operating systems for our initial experiments i.e. Windows XP (Service Pack 2 (SP2)) and Centos 4 on the machine. While performing the experiments, we kept the machine idle. The idea was to leave all possible system resources available for data acquisition tools so that they could exploit the resources at their full capacity without any constraints. Moreover, for data acquisition over the network, we directly connected the PC through a crossover cable with the investigative machine (where the data were transferred) in order to avoid the
overhead of packet switching or routing. The investigative machine was a modern computer having a Core 2 Duo CPU, 4 GB RAM and 300 GB hard disk (running at 15000 rpm speed), which is unlikely to have caused any bottleneck in the performance of the data acquisition tools.

Table 1: Resource consumption of data acquisition tools

<table>
<thead>
<tr>
<th>Tools</th>
<th>Operating System</th>
<th>Device Data-acquired</th>
<th>CPU Idle Time (%age)</th>
<th>Free Physical Memory (%age)</th>
<th>Disk Queue Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>WinDD</td>
<td>Windows XP (SP2)</td>
<td>Physical memory</td>
<td>90.72</td>
<td>75.60</td>
<td>0.03</td>
</tr>
<tr>
<td>Garner’s DD</td>
<td>Windows XP (SP2)</td>
<td>Hard Disk</td>
<td>27.49</td>
<td>74.01</td>
<td>0.72</td>
</tr>
<tr>
<td>DD (on Linux-variants)</td>
<td>Centos 4</td>
<td>Physical Memory</td>
<td>51.98</td>
<td>79.69</td>
<td>0.0</td>
</tr>
<tr>
<td>DD (on Linux-variants)</td>
<td>Centos 4</td>
<td>Hard Disk</td>
<td>0.646</td>
<td>71.14</td>
<td>0.805</td>
</tr>
</tbody>
</table>

1CPU Idle Time: Average percentage of time during data acquisition that the CPU was idle
2Free Physical Memory: Average percentage of free physical memory during data acquisition
3Disk Queue Length: Average number of (read and write) requests outstanding on the hard disk during data acquisition

The experimental results (shown in Table 1) show that the tools did not exhaust the system resources for data acquisition per se and this may get better if SCADA systems use better hardware than used in the experiments. However the results do not guarantee that the tools are compatible with a particular SCADA environment and that no significant impact on the services would be generated during acquisition until they are run and tested on that or its equivalent environment (such as a testbed of a production environment). Moreover, the tools not included in the experiments may not necessarily show similar performance impact.

**Limitations of forensic analysis tools on SCADA systems**

To the best of authors’ knowledge, state of the art forensic analysis tools do not support the unique features of diverse SCADA environments, which include supporting SCADA protocols and numerous SCADA applications’ proprietary log formats etc. Thus plugins or modules for contemporary forensic tools need to be developed to augment the forensic analysis in SCADA systems.

**Research challenges**

There is an increasing interest in the security/forensic research community on SCADA systems. This is mostly due to the heightened focus of governments worldwide on protecting their critical infrastructures (including SCADA systems) by specifically allocating research funding for this cause. The critical nature of SCADA systems brings challenges to the research community including the following:

**Lack of appropriate testbeds**

While a security incident affecting a server in an office environment may lead to monetary loss, SCADA systems face threats with different consequences [24][25] and a security breach can lead to dangerous impacts on the environment and can have consequences for human life [26]. In addition, performance
requirements for protection systems used in SCADA systems have an impact on security features that are chosen. For example, in certain use cases, the overhead induced by asymmetric cryptography is not tolerable [27]. Thus, the research in this domain should be practical and conclusive, which cannot be achieved without having SCADA systems available for research purposes. Real SCADA systems are expensive to build and thus require significant research funding. To deal with this problem, the SCADA research community (especially in academia) mostly opt for the following approaches, each of which has its own merits and limitations.

- **Simulator:** A simulator provides a virtual environment to imitate the operations of a real-world process. There are simulators such as simSCADA[16] that specifically provide an environment for SCADA systems research. These are mostly used to imitate the network traffic of SCADA systems that occurs between the field devices and MTUs. The simulator is effective in reducing the purchase and installation cost of the field and communication devices. However, simulation errors may affect the results of experiments and therefore simulators typically do not provide the same level of confidence that a real system might.

- **Building small-scale SCADA systems:** Governments and academia use commercial hardware and software to build laboratory scale test beds of SCADA systems in order to control physical processes at small scale such as an industrial blower, gas pipeline, electric transmission, and petroleum storage tanks. For instance, the Mississippi State University [17] has developed a test bed mainly for teaching and research purposes, which has multiple industrial control systems to control different physical processes at laboratory scale. Furthermore, the Idaho National Lab (INL) has a test bed [18] that is a full scale electric power grid and is dedicated to control system cyber security assessment, standards improvements and training.

- **Industry collaboration:** The SCADA research community mostly tries to engage SCADA owners and operators as industrial partners when they apply for research project funding. The terms of agreement for the project usually involve technical assistance, SCADA facility access (at least to the test bed used by the operators for testing application patches from vendors) and financial support. Industrial collaboration provides close access to real-world SCADA systems and the technical personnel that experience the problems and understand the limitations of their SCADA system. However, industry collaboration is the hardest of all to achieve due to the critical nature of SCADA systems, which discourages SCADA owners and operators from cooperating with the research community.

**Tight-lipped SCADA system owners and operators**
Research in SCADA system could be more practical if it is done with the collaboration of SCADA owners and operators, who actually experience the problems. However, the critical nature of SCADA systems demands the owners and operators not share any information about their system with the SCADA research community in order to prevent any information leakage that can be exploited for evil intentions. This creates a gap between the efforts by the research community and the problems faced by the SCADA owners and operators. Governments in this situation are in a better position to play a mediator role and take initiative that can help in reducing this gap. For instance, the Australian government regularly organizes SCADA community of interest (CoI) meetings to provide a platform where SCADA owners and operators, SCADA vendors and researchers from academia are gathered together to make professional acquaintances and relationships and discuss current issues.
Existing Research Development
So far the research community has mainly focused on the security of SCADA systems, which is also evident from the number of publications on the topic. However, very limited work has been reported to date that deals with the forensic aspect of SCADA systems, which this section briefly describes.

Kilpatrick et al. [19] [20] [21] proposed a network-forensic architecture for capturing and subsequently analyzing sensor data and control actions in SCADA network. The architecture comprises of two main entities: agents and a data warehouse. The agents are placed at strategic locations within the SCADA network to capture the network traffic in its local network segment and forward a relevant portion of packets (synopsis) to the data warehouse. The data warehouse after analyzing a synopsis creates its digital signature and stores it with the synopsis to the agent’s designated storage area. The storage design uses a relational database and query mechanisms are employed to support forensic investigations. Modular design of agents with configurable synopsis engines is used to deal with diverse SCADA protocols, some of their deployment implementation variations, subsets of standard protocols or proprietary protocols. A prototype of the architecture based on the Modbus TCP protocol was developed using two control devices and one HMI station.

Valli [22] proposed a framework that produces forensically verified signatures for the Snort intrusion detection system (IDS) for known and published vulnerabilities of SCADA and control systems, enabling forensic investigators to find the traces of exploits during analysis. He looked for vulnerability announcements or traces at several relevant sites such as Blackhat, hacker, vendor, and CERT sites and reproduced the vulnerability scenarios. He examined the vulnerabilities of SCADA communication protocols i.e. Modbus and DNP3. His experiments involved an attacker, victim/target machine, Snort IDS and network sniffer. The attacker executes attacks to target a victim machine that is running SCADA software. The network sniffer captures all network traffic in a tcpdump binary capture file for analysis to generate snort rules. The modus operandi of exploit is then analyzed and used to create a ruleset to reduce or stop the attack. The ruleset is later included in the Snort configuration to test its resilience when under stained attack.

Slay and Sitnikova [23] raise a discussion that in order to develop a generic approach for forensics in SCADA and control systems, a big picture approach needs to be taken where a good understanding of forensic computing process and a range of technical and procedural issues within the process needs to be understood at the government, industry and academic levels.

Conclusion
We have pointed out that performing a forensic investigation on a SCADA system is different from other networks such as corporate networks, home networks etc. due to the underlying industrial processes that are controlled. The critical nature of SCADA systems demands that forensic investigators are well trained and have a good understanding of the issues required to handle such systems. The early engagement of forensic investigators to get accustomed to a particular environment is highly encouraging. Moreover, investigators should also encourage SCADA owners and operators to initiate the steps that can facilitate investigation when it is needed. For instance, maintaining a data acquisition plan and regular testing of data acquisition tools to assure that the tools would not affect the availability of the SCADA system services during forensic investigation are very important.
SCADA-focused forensic research is essential to address the unique issues and challenges in order to facilitate forensic investigations. Thus, it is necessary to engage the forensic research community with SCADA owners/operators and investigators (actively working on SCADA systems) so that the research problems in this domain can be highlighted and efforts toward their solutions can be made. Moreover, SCADA owners and operators and governments (in particular) need to organize these efforts and provide resources, funds and suitable access of researchers to SCADA systems (by encouraging industry to make collaborations with the research community).

Acknowledgment
This work was supported in part by NSF grant CNS #1016807.

References


Authors’ short biographies

Irfan Ahmed is a postdoctoral research associate in the department of computer science at the University of New Orleans. His research interests include industrial control system security, digital forensics and malware detection and analysis. He received his PhD from Ajou University, South Korea. Contact him at irfan.ahmed@uno.edu.

Sebastian Obermeier is a Principal Scientist at ABB Corporate Research, Switzerland. His research interests include IT security for Industrial Control Systems and database technology. He received his doctorate degree in computer science from the University of Paderborn, Germany. Contact him at sebastian.obermeier@ch.abb.com.

Martin Naedele is the R&D Program Manager for Industrial Software Systems at ABB Corporate Research. His research interests include software engineering and IT security. He received a PhD in computer engineering from ETH Zurich. He is GIAC certified security auditor (GSNA), and a member of IEEE, ACM, and INCOSE. Contact him at martin.naedele@ch.abb.com.

Golden G. Richard III is Professor of Computer Science and University Research Professor at the University of New Orleans. His research interests include digital forensics, reverse engineering, and operating systems internals. He received his Ph.D. from The Ohio State University in 1994 and has been employed by the University of New Orleans since that time. He is a member of USENIX, ACM, IEEE, and the American Academy of Forensic Sciences (AAFS). Contact him at golden@cs.uno.edu.