2018

Digital forensics investigative framework for control rooms in critical infrastructure

Brian Cusack

Amr Mahmoud

DIGITAL FORENSICS INVESTIGATIVE FRAMEWORK FOR CONTROL ROOMS IN CRITICAL INFRASTRUCTURE

Brian Cusack, Amr Mahmoud
Cyber Forensic Research Center, Auckland University of Technology, New Zealand
brian.cusack@aut.ac; amr.mahmoud@aut.ac.nz

Abstract

In this paper a cyber-forensic framework with a detailed guideline for protecting control systems is developed to improve the forensic capability for big data in critical infrastructures. The main objective of creating a cyber-forensic plan is to cover the essentials of monitoring, troubleshooting, data reconstruction, recovery, and the safety of classified information. The problem to be addressed in control rooms is the diversity and quantity of data, and for investigators, bringing together the different skill groups for managing data and device diversity. This research embraces establishing of a new digital forensic model for critical infrastructures that supports digital forensic investigators with the necessary information for conducting an advanced forensic investigation in Critical Infrastructures. The framework for investigation is presented here and elaborated. The extended work applies the framework to industry case studies and is not reported here.

Keywords
Digital Forensics, Control Room, Critical Infrastructure, Investigation

INTRODUCTION

The contribution of this paper comes from a substantial literature review and analysis of forensic methodologies that may be appropriate for control rooms in critical infrastructures. Figure 1 provides the context of study and the resultant framework is presented in Figure 2. Existing digital forensic models are designed and developed with specific characteristics for target areas. The target areas have traditionally been adequately covered by the models and investigator best practice guidelines (Reimer, 2013). For example, digital forensic models have been designed to perform network forensics, computer forensics, cloud network, and mobile forensics. These traditional techniques are no longer suitable to deal with the age of data. Large volumes of data "Big Data" is a new age that has the problem of dealing with large data sets that are analysed computationally in order to expose patterns. Non-traditional ways and design are required to deal with these large volumes of data and still maintain the integrity of investigations. A growth area for forensic investigation is in conducting forensic investigations in industrial control systems (Cherdantseva, 2016; Rodrigo and Morocho, 2017). These are a complex process not only because of the diversity of data, but also the variety of physical and logical partitions that are interconnected to the network including name nodes, data nodes and checkpoints. This type of investigation requires collecting all sources of information not only from a suspect computer, but also from the system itself. Most sensitive events and logs are recorded into the node controllers (Stouffer, Falco, and Scarfone, 2015). Therefore, this issue can be solved by applying Hadoop cluster to acquire the information from the HDFS file system metadata. Conducting a clustering reconnaissance phase will provides valuable information about data blocking, size, and replication factors based on a Block ID.

General investigation guidelines for control centres are fragmented across many documents and an assumption that what works in some areas of investigation will work everywhere (Casey, 2011; Reimer, 2013). The deliverable from this research has been a resolution of the current partitions of information in the literature into a working framework for large and diverse data sets in control rooms for critical infrastructures. The framework is called a Corrective Big Data Forensic Investigation Model for Critical Infrastructure. It was designed to control components interconnected to the network infrastructure such as data clusters, load balancers, servers, and engineering workstations. The proposed model was set to be implemented according to design science research methodology. This research methodology proved its capability of uncovering all vulnerabilities while lab testing was conducted. Throughout the testing, it was found that supplementary steps can be added to increase more acquisition of digital evidence and improve the quality of the credible sources of data. The model has been verified by pilot testing on case study scenarios and test data. The following sections of this paper elaborate the background and model components.

BACKGROUND CONTEXT

Figure 1 encapsulates the networks of a critical infrastructure organisation and layers each functionality against the requirements. This summary provides the general context of study.

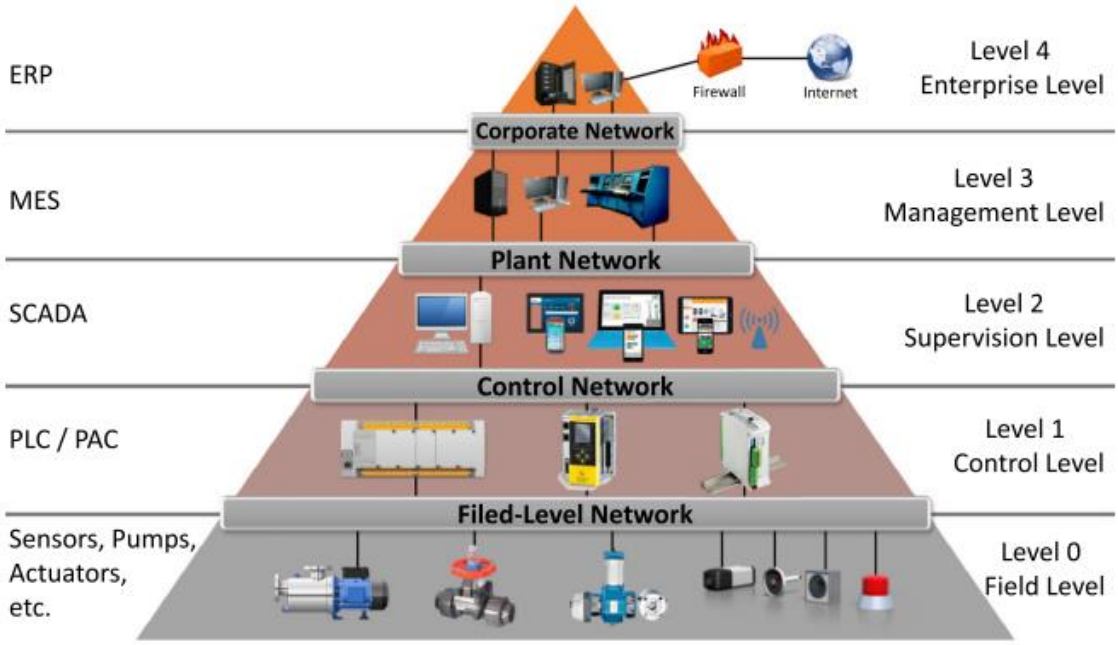

Figure 1. Critical infrastructure context

DIGITAL FORENSIC CHALLENGES

An effective forensic method in critical infrastructures must screen and record actions and user's events from the complex environment in order to enable overall access to the data (Carlini, 2016; Adrien, 2016). A microscopic view would support credible information about the user incoming and outgoing network packets within the network; for example, a time to live (TTL) connection, the volume of data packets transmitted, the size of each data packet, and identify the type of transmitted information related with each link. A macroscopic perspective, on the other hand, would support information of how many connections were successfully established, the average TTL of each request and the overall volume of data transmitted. Therefore, challenges in a number of critical phases have been identified as a part of the problem solving. The challenging areas are defined in collection, data analysis and reporting phases (Cornelius and Mark, 2008; Dhanunjoya, 2018).

Automation is critical for the collection of evidence. The challenges in the context are: Control system key information resources; data retention; volatility (this inviable because the data within the collection process is removed, deleted, or overwritten, and this can make it impossible to be detected in its original state (Horokuan, 2015); "Data Mingling" (Folkerth, 2015) (the data mixture and being indistinguishable); and data classification (Reimer, 2013). In the data analysis phase tool availability for the volume of data and the diversity of data require automation (Ghani, 2013). Sophisticated tools such as those that copy processes, examine evidence, and analyse programs for generating checksums in order to complete the verification, may not fit perfectly to some of the control systems technologies. Consequently, many digital forensic tools in different areas such as network forensics, database forensics, computer forensics, and mobile forensics will require testing before use for relevance to the environment (Grispos, Glisson, & Storer, 2015). Therefore, digital forensics vendors will have to apply new modifications to their software and frameworks in order to fill the gap and meet the challenge (Beebe, 2009; Watt and Slay, 2015).

The reporting phase must consider data volume, diversity and the usefulness of the report. The documentation must be presented in a usable format. This is usually in digital formats and it is reusable for system improvement and investigatory purposes. Documentation relevance and usefulness ensures the success of any forensic investigation in control systems environments but this requires preparation and design (Agarwal, & Kothari, 2015,

p.567). Asset owners have to take preparation steps in order to identify and detect any types of changes that could be done during operating system installation, configurations of devices, hardware, or any elements whose modified behaviour may affect the original equipment manufacturer specification (Weiss, 2010). Documentation allows reporting to have a factual basis but the challenges of timeliness, relevance and volumes still have to be addressed.

PROPOSED SOLUTION

Conducting forensic investigations in industrial control systems have become a complex process not only because of the diversity of data, but also the variety of physical and logical partitions that are interconnected to the network including name nodes, data nodes and checkpoints (Ahmad, Hadgkiss and Suighaur, 2012). This type of investigation requires collecting all sources of information not only from the suspect computer, but also from the system itself. Most of sensitive events and logs are recorded into the nodes controllers (Quick and Choo, 2017). Therefore, this issue can be solved by leading a reconnaissance on Hadoop cluster to acquire the information of HDFS file system metadata. Conducting a clustering reconnaissance phase will provide valuable information about data blocking, size, and replication factors based on Block ID (Amadeo and Abdo, 2018). The Corrective Big Data Forensic Investigation Model for Critical Infrastructure (Figure 2) was designed to control components interconnected to the network infrastructure such as data clusters, load balancers, servers, and engineering workstation. It shows the link between big data clusters in big data rooms and engineering workstations in control rooms. One of the benefits gained from applying the Model was that forensic investigation could be made relevant across the variants of the complex system. The improved model shows the additional stages for conducting remote forensic investigation before performing the actual forensic investigation. These stages can assist in collecting and documenting digital evidence. The following sub-sections review the application to each scenario.

Engineering Workstation Forensic Investigation

A digital forensic investigation in engineering workstations or control rooms is a term that can be used in critical infrastructures to include all electronic devices that are interconnected with each other for sending/receiving messages or two-way communications, such as, mobile phones, laptops, computers, tablets, PDAs, programmable logic controllers, human machines interfaces, and supervisory control and data acquisition systems (Adelstein, 2006; Fehr, 2015). These systems and devices have their own storage systems. Either physical storage systems or virtual technologies such as cloud computing for logging all activities, incidents, and events (Martini and Choo, 2014). Conducting a forensic investigation on engineering workstations and applying physical and remote data acquisition will discover evidence that can be used for legal, employment, and other purposes (Jones and Leitha, 2016). In this type of investigation, physical and remote data acquisition are an advantage. The following paragraphs summarise each of the stages and the relevant relationships depicted in Figure 2.

Planning & Identification: at this stage, the incident has to be verified in order to collect fact sheets and plan for a handling strategy on the particular case. The major objective of this phase is to boost the productivity of gathering the necessary information about the incident and facilitate the process of data acquisition. Furthermore, obtaining authorisations and authentications are also compulsory, when the case needs an authorised access to the system for logging in to the log system at the initial stage. System settings are one of most important facts required to be obtained by official and authorised investigators for determining the device's system state when the incident occurred. System settings can include the system specifications of all machines that are under investigation, and the time or date. Moreover, conducting a network reconnaissance is the last step to obtain IP addresses of all machines and their mac addresses and any other information that could lead to personal ownership identification or related activities.

Search & Data Collection: at this stage, all information will be collected from the suspect machines as they have been put into the investigation process. This step will require more detailed information about the daily events for the user on the machine or device. All information that will be collected, will be taken into consideration and will be preserved for the next step. The collected data will go in to a complex process to determine whether the data acquired is considered as admissible or not admissible. If the data is admissible, it will go to further investigation for finding the relevant evidence for the case. If not, the data will be stored for a specific period of time and reserved for analyse later when the circumstances have changed. This stage aims to prepare all potential credible data to go through a parsing process, which is a more detailed analysis of the data. All necessary data will be available for an advanced level data acquisition from the control room.

Initial Assessment: at this stage, penetration testing will be conducted remotely for acquiring live data on the suspect machines when the users have not been formally informed that their machines are going through forensic investigation. This step will assist in preserving live data before the digital evidence gets damaged or corrupted. The aim of this step is to combat the anti-forensic tools used by advanced persistent threat (APT) attackers and professional hackers in critical infrastructures. Dead acquisition will be confirmed as the second step when evidence is found on the suspect machines. At this step, screenshots can be taken as a credible evidence of unauthorised access to the resources in the engineering workstation.

Data Examination: at this stage, the timeline will be analysed methodically. All data, fact sheets, system settings, parsed data, and data that came from the initial assessment will go to further processes of data analysis and examination. Timeline analysis will analyse the data from different perspectives. This is a vital stage and beneficial as it comprises evidence history such as what time the files have been accessed, modified, created and changed in a clear format that humans can understand. The data is collected using a diversity of applications and is released from the layer of metadata from the file system (recorded from Linux or Windows platforms) and then analysed. The timeline is fixed and application data reconstructed if required as a part of data analysis and examination. Furthermore, media and artefact analyses is addressed, for example, what applications have been executed, which archives have been opened downloaded, which documents have been clicked on, which records were checked, which files were deleted, where did the user browse to and many other properties. Another type of analysis, which is necessary for finding indirect paths of information is at the signature level. This analysis, when forensic investigators implement techniques and practices that will search for byte signatures of known folders, files and regular expressions that lead to the cookies. Furthermore, link analysis is employed to find the relationships and trusted links to other entities, servers, domains, email, people, and other relevant objects that can be traced to identify all possible communications.

Reporting & Presentation: The last stage contains reporting the results of the analysis and then presenting it to requested recipients. This step includes stating potential risks, clarifying the actions taken, specifying what other arrangements that are required. Also suggesting enhancements to procedures, guidelines, policies, applications, and other aspects of the forensic process investigations required in the target infrastructure. This step is essential as it is important for the stakeholders in order to determine what strategies they must think about for future preparation. The report has to be formulated in a form that is acceptable to the court or for any legal, employment or administrative purpose.

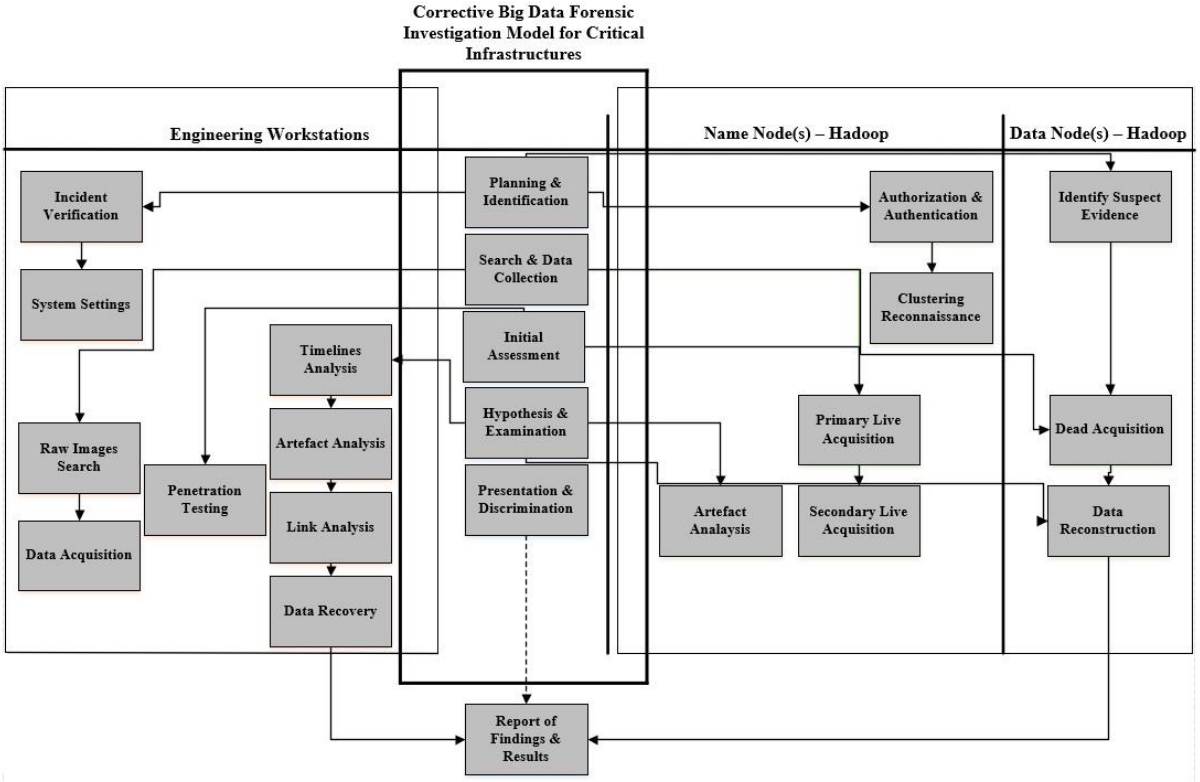

Fig. 2. Corrective Big Data Forensic Investigation Model for Critical Infrastructures

Hadoop HDFS Forensic Investigation

A digital forensic investigation in big data rooms using big data platforms, such as Hadoop HDFS, is not a common occurrence. A big data platform is used in critical infrastructures to include all logical nodes that are interconnected with each other for sending/receiving messages or two-way communications, such as, primary nodes, secondary nodes or checking-out nodes, and data nodes. These nodes have their own storage systems and distributed file systems technology such as Hadoop HDFS for logging all activities, incidents, and events. Conducting a forensic investigation on Hadoop HDFS and applying live and dead data acquisition will give evidence that can be used for all legal, employment, and other purposes. In this type of investigation, live and dead data acquisition are an advantage. The following paragraphs give a detailed description of each stage of investigation.

*Planning & Identification***:** at this stage, basic properties of Hadoop HDFS has to be identified by a qualified forensic investigator in order to plan for the best strategy to initiate the forensic investigation procedure. Identifying those properties requires obtaining necessary authorisations for gaining access to the highest credentials on the system, identifying the name node address and its jurisdictions. The purpose of establishing these credentials is to acquire metadata system specifications and files, which provides the forensic investigator with the information for the process to progress. Examples of this information can include: block ID, block size, and replication factors with all nodes installed on the system. RAM memory acquisition should come first, as any delay of this process can risk losing potential forensic evidence. Inbuilt commands of the Hadoop system are required to be implemented as an initial step of acquiring cluster administration. This step can be done in a number of ways, for example, "Hadoop fsck", and "dfsadmin -report", as well as offline image viewer. It is highly recommended to access the system remotely from a virtual forensic workstation and execute these commands for reducing the risk of evidence validity, and minimise the interaction with the name node cluster. The Planning and identification operation will assist in the next phase, which is the search and data collection phase to collect only admissible evidence.

Search & Data Collection: at this stage, the data sources have been identified for further investigation. The Checkpointing operation is established for performing the task of including admissible evidence and excluding inadmissible evidence. This operation is meant to be prior to File System Image acquisition and analysis. To lighten the risk on the live cluster of data corruption in this stage, and to go in parallel with the concept of reducing cluster interaction during the forensic process, the checkpointing operation is carried outside the system in a virtual environment. The Hadoop is configured in pseudo distributed mode set specifically for forensic workstation investigations. In the stage of search and data collection, all copies of file system images and edits logs are collected, placed in the forensic workstation by the inbuilt command "checkpoint -force". This command will update the name node with the latest operations done on the system to give descriptive information about each transaction and event. This operation will assist in validating the credibility of the digital artefact in the next stage, which is the initial assessment.

Initial Assessment & Data Examination: these stages are working in parallel with the Hadoop HDFS architecture, as live and dead acquisition are linked with each other and required to be confirmed for analysing the data collected from the previous stages. The data that has been received from the different data sources such as RAM and clusters of name nodes, secondary nodes, and data nodes. It will go through analysis and assessment processes. Live artefact acquisition on the name node is performed to target the HDFS directory and system administration and to allocate the data storage of all partitions and nodes installed on the Hadoop. Data blocks will be matched with all block IDs in order to get the final outcome of the live analysis of assigning each operation to each user. These requirements are substantial because they relate to the internal block ID, which is specified to the HDFS data block. It is itemized, and the physical start offset address that the block is located within has the storage of data node. Moreover, the differential live analysis report versions of the file system image files, in terms of pre-checkpointing and post-checkpointing operations can support forensic investigators with beneficial information that clarifies the importance to identify any obvious inconsistencies. Dead acquisition of artefacts on the data nodes is performed as a vital part of forensic investigations to specify the suspect nodes in the workstation, so it can be investigated thoroughly. It is anticipated that the forensic investigator is now residing physically on the system

and doing dead acquisition on the basis of affected block IDs with suspected nodes. This process allows forensic investigators to select and target only required data nodes for initial imaging and assessing processes. Data reconstruction is one of critical processes in Hadoop HDFS, due to the complexity of its data structuring. This part involves data carving for reconstructing the deleted block IDs found on the HDFS. Reconstructing the deleted block IDs will enable examiners to validate the type of action made to delete the particular block ID.

Reporting & Presentation: The last stage contains reporting the results of the analysis and then present it to requesting recipients and stakeholders. All results found are documented in this phase to state the plan of action for all potential risks as well as recommendations for a safeguard plan for protecting the privacy of the information included in the report. Then the report will be presented in a formal format.

CONCLUSION

This paper has reported a digital forensic investigation framework for control rooms in critical infrastructure contexts. This is a big step forward towards addressing the data diversity and volume problems that have defeated previous attempts to design a comprehensive solution. Previous attempts to generalise such frameworks have stopped at particular installations, equipment or tool selections. The literature reviewed was comprehensive and the Corrective Big Data Forensic Investigation Model for Critical Infrastructures is an advancement on previous attempts. The further research is to apply it to an industry context and to implement further design science quality improvement design cycles.

REFERENCES

Adelstein, F. (2006). "Live forensics: diagnosing your system without killing it first." *Communications of the ACM* 49(2), 63-66.

Adrien, B. (2016). "Security of Industrial Control Systems and Cyber Physical Systems". First Workshop, CyberICS 2015 and First Workshop, WOS-CPS 2015 Vienna, Austria, September 21–22, 2015 Revised Selected Papers. Vol. 9588. Springer.

Agarwal, R. and Kothari, S. (2015). "Review of digital forensic investigation frameworks." *Information Science and Applications.* Springer, Berlin, Heidelberg, 561-571.

Ahmad, A., Hadgkiss, J. and Ruighaver, A. (2012). "Incident response teams–Challenges in supporting the organisational security function." *Computers & Security* 31(5), 643-652.

Amadeo, M. and Abdo, H. (2018). "A safety/security risk analysis approach of Industrial Control Systems: A cyber bowtie–combining new version of attack tree with bowtie analysis." *Computers & Security* 72, 175-195.

Beebe, N. (2009). "Digital forensic research: The good, the bad and the unaddressed." IFIP International Conference on Digital Forensics. Springer, Berlin, Heidelberge.

Carlini, E., (2016). "A decentralized and proactive architecture based on the cyber physical system paradigm for smart transmission grids modelling, monitoring and control." *Technology and Economics of Smart Grids and Sustainable Energy* 1(1), 5.

Casey, E. (2011). *Digital evidence and computer crime: Forensic science, computers, and the internet*. Academic press, New York.

Cherdantseva, Y. (2016). "A review of cyber security risk assessment methods for SCADA systems." *Computers & Security* 56, 1-27.

Cornelius, E. and Mark, F. (2008). "Recommended practice: Creating cyber forensics plans for control systems". No. INL/EXT-08-14231. Idaho National Laboratory (INL).

Dhanunjaya, V. (2016). "Collecting Volatile and Non-Volatile Data." LinkedIn, 26 Mar. 2016, www.linkedin.com/pulse/collecting-volatile-non-volatile-data-vuppala-dhanunjaya.

Fehr, R. (2012). “The Basics of Ladder Logic.” *Electrical Construction & Maintenance (EC&M) Magazine*, 5 Apr. 2012.

Folkerth, L. (2015). "Forensic Analysis of Industrial Control Systems." SANS Institute InfoSec Reading Room.

Ghani, M. (2013). "A Review of Communication Protocols for Intelligent Remote Terminal Unit Development." *Telkomnika* 11(4), 81-90.

Grispos, G., Glisson, W. and Storer, T. (2015) "Recovering residual forensic data from smartphone interactions with cloud storage providers." arXiv preprint arXiv:1506.02268.

Horkan, M. (2015). "Challenges for IDS/IPS deployment in industrial control systems." SANS Institute reading room (2015).

Jones, J. and Letha E. (2016). "Analysis of digital forensics live system acquisition methods to achieve optimal evidence preservation." IEEE SoutheastConference.

Martini, B. and Choo, R. (2014) "Cloud forensic technical challenges and solutions: a snapshot." *IEEE Cloud Computing* 1(4), 20-25.

Quick, D., and Choo, R. (2017). "Google drive: forensic analysis of data remnants." *Journal of Network and Computer Applications*, 40, 179-193.

Reimer, H. (2013). “Securing Electronic Business Processes Highlights of the Information Security Solutions”. Europe 2013 Conference. Springer Fachmedien Wiesbaden.

Reith, M., Carr, C. and Gunsch, G. (2016) "An examination of digital forensic models." *International Journal of Digital Evidence* 1(3), 1-12.

Rodrigo, R. and Morocho, F. (2017). "Digital Forensics Tools." *International Journal of Applied Engineering Research* 11(19), 9754-9762.

Stouffer, K, Falco, J. and Scarfone, K. (2015). "Guide to industrial control systems (ICS) security." NIST special publication 800.82, 16-16.

Watt, A. and Slay, J. (2015). "First responders actions to cope with volatile digital evidence." *International Journal of Electronic Security and Digital Forensics* 7(4), 381-399.

Weiss, J. (2010). *Protecting industrial control systems from electronic threats*. Momentum Press, Michigan.