Chapter

The Role of Penetration Testing in Forensic Multimedia Retrieval Process

Amr Adel and Brian Cusack

Abstract

Digital forensic investigators are faced with multimedia retrieval and discovery challenges that require innovation and application of evolving methodologies. This work is made more difficult in critical infra-structure environments where the acquired evidence is in many formats, types and presentations. Penetration testing is one of the techniques used to focus an investigation and to target the potential case information from the vulnerability identification phase, through to the media identification phase. In this chapter a review of these processes is made and a framework example developed to show how the investigator discovers relevant evidence. The problem for the digital investigator is the vast array of media in which evidence is stored or transmitted. Some work is from live retrieval and others static. A framework of methods that is flexible and adaptable to the context of investigation is proposed and the discovery methods for multimedia environments elaborated.

Keywords: penetration testing, digital forensics, critical infrastructures, evidence extraction, process framework

1. Introduction

Forensic Investigators conduct forensic examinations in order to identify evidence and to prevent future compromises of a system. The increasing volume of digital data to be managed and the diversity of media type is a contemporary challenge. The diversity of devices, operating systems, media and services present obstacles that require solution for efficient and effective professional practice. The variety of data sources, formats and styles poses a multimedia problem that requires working solutions for information access and content documentation. The acquired evidence can include different types of forensic data such as pictures, audios, videos, files, directories, and texts [1]. The systems for extraction are either live and functioning or static and stored. In either situation due processes, methods, standards and guidelines must be complied to achieve a repeatable practice for later auditing. In many instances copies are taken of the various media so analysis proceeds on identical images and not the original media. Investigation processes are segregated into phases to assure the best deployment of specialist skills and the preservation of the evidence [2]. Segregation is usually divided into preparation, acquisition, analysis, and reporting phases and sequenced towards a deliverable

that provides corrective actions [3]. In such a situation the system of work and the targeting of the work objectives are critical to the deliverable and the viability of an investigation. In this Chapter we derive a framework for investigation in an intensive multimedia environment and then demonstrate the targeting power of penetration testing techniques.

Critical infrastructures (CI) involve complex systems for the control and protection of assets, and the production and distribution of services to detect suspicious activities [4]. Any unplanned disturbance to these facilities seriously affects the quality of life and economic wellbeing of humans. Modern society depends on digital infrastructures to provide their management of services and the fair and timely distribution. For example, one day of disrupted power supply to a region of users stops work of all kinds and prevents the usual activities that support daily living [5]. Extended power failure causes long-term destruction of economic relationships and negatively affects the necessities for daily life. These systems require protection and one of the ways to do this is to use forensic investigation of events, and to do penetration testing before anything unplanned occurs [6]. In addition to other security provisions, forensic techniques are commonly implemented to document baseline configurations in order to detect abnormal activities, such as unauthorized access into network infrastructure. However, the challenge is to gain a fair estimation of the data provisions in the systems that are chaotically fill of large volumes of static and live data, and a full range of multimedia data types [7].

In this research we designed and tested an investigation framework for multimedia data types to address the challenges of evidence collection in CIs. The volume and complexity issues influence the evidence collection phase but also each environment has unique features from organizational cultures, administration designs, recovery tools, record structures, logging systems, and general usage patterns that all impact the scope and success of an investigation [8]. In addition, there are further challenges such as automation, volatility of data, and data mingling. Automation creates key information resources in order to handle the data and abstract data from its context. Volatility makes the process of collecting data difficult because the data within the collection process is removed, deleted, or overwritten [9]. Furthermore, Data Mingling is a serious problem of data mixing and the types being indistinguishable. Often, the sample of total data investigated in the forensic process comprises of both data related to the incident and data unrelated to the incident [10]. Forensic investigators require help to make sense of the complex multimedia contexts in which they have to work. An investigation framework that is responsive to CI complexities and has targeting features to make workloads manageable is required. The following sections describe how these requirements are designed and become functional in an investigation process.

2. Background literature

Industrial Control Systems in critical infrastructures support monitoring, administering, and controlling essential services. Therefore, by design architecture, components, and environments in CI, allow forensic capabilities to be implemented and to further mitigate the potential risk of security failure. Industrial Control System Architecture is deployed based on Service Oriented Architecture [11]. Hence, three different designs are found according to the architecture of the system. First, Supervisory Control and Data Acquisition (SCADA) systems apply central administration by using a central computer to communicate remotely through a Remote Terminal Unit (RTU). A Human Machine Interface (HCI) is linked to SCADA and facilitates the process of displaying, and monitoring processes. The typical uses

Level	Information System	Media Type
0	Sensor Networks, Internet of Things, and so on	Data, streams of text and digits
1	Programmable Logic Controller, Picture Archiving and Communication	Structured data, text, frames, objects
2	Supervisory Control and Data Acquisition	Ladder logics, objects, words, text and digits
3	Management Expert and Management Information Systems	Images, videos, text, files, directories
4	Enterprise Resource Planning	Files, directories, all manner of media type

Table 1.
CI Media types at organization levels.

for SCADA are in natural gas, electricity, and water distribution [12]. Second, Distributed Control Systems (DCS) distribute processes that have been controlled to devices for execution [13]. The typical uses of DCS are in manufacturing, chemical and electric power plants. Third, Non-Centralised System design allows for a number of control systems that do not require centralised administration. Accordingly, Programmable Logic Controllers (PLC) or any other control devices can be implemented and configured as a combination of Control System, Data Historian and Human Machine Interface [13]. This type of configuration is usually designed for manufacturing processes.

The distribution of media type is found spread evenly through the layers of an industrial control system for CI [14]. These layers are often described as starting at layer 0 where the sensors of the system and primitive data are found, through to layer 4 which is the enterprise level where the business applications and rich media reside [15]. In **Table 1** these layers and media types are described and elaborated to identify the data types and the diversity of media type a digital investigator must review in discovery processes. Discovery processes hence require extreme multimedia processing capabilities that can span the scope of data type and format found in a CI environment. This requires critical tool selection and the designing of staged and sequenced tool use for comprehensive discovery. The task is difficult and is challenged by the constant innovation and adoption of new data type and structures that come with new versions of software and new applications. The multimedia processing capability an investigator chooses reflects the design and scope of an investigation, and the professional capacity to adapt and acquire the necessary tools and techniques [16, 17].

3. Designing a framework

Primarily an investigator requires a systematized process framework to effectively guide an investigation through the known and unknown media types found in a CI investigation. The design proceeds through a phased approach outlined in **Figure 1**. A digital forensic investigation in engineering workstations or control rooms in CIs includes 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 machine interfaces, and supervisory control and data acquisition systems [16]. 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 [18–21]. Conducting a forensic investigation on engineering

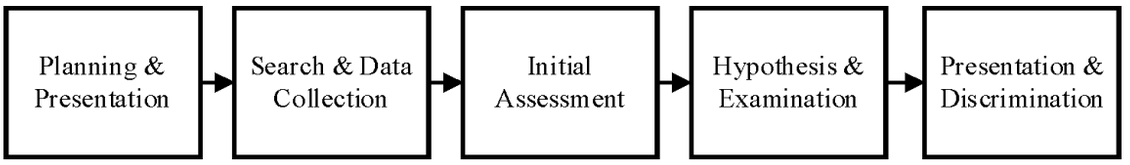

Figure 1.
The Five Phases.

workstations and applying physical and remote data acquisition will discover evidence in the different media that can be used for legal, employment, and other purposes [22]. In this type of investigation, physical and remote data acquisition are an advantage, and hence the investigation equipment must have the capacity to manage volumes, data complexities and multimedia types.

Each investigation requires phases that develop the focus for evidence collection and then pass the findings to the next phase for further refinement. Planning and Identification is a starting point for a structured investigation. At this stage, the incident has to be verified in order to collect fact sheets and plan for a capability handling strategy for 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 [23]. Critically the acceptance of multimedia types by the acquisition tools allows credibility to be established and documented against the brief scope. If media types cannot be collected then the performance and adequacy of the investigation are brought into question. Furthermore, obtaining authorizations and authentications are also compulsory, when the case needs an authorized access to the system for media acquisition. System settings are one of most important facts required to be obtained by 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 [23]. At each of the context levels in **Table 1** different evidence is located and each data and media type must be accommodated in the framework design.

The search and data collection stage employs discovery techniques that allow all information in the multiplicity of media types present to be collected. The investigation process requires detailed information about the daily events for the users in the systems and machines or devices. All information that is collected, will be taken into consideration and preserved for relevancy determination. The collected data goes in to a complex process to determine whether the data acquired is compliant to evidentiary standards and the acquisition process and the deliverable are reproducible by others. If the data is admissible, then it will go to further analysis for case relevancy and positioning in the data log. If not, the data will be stored for a specific period of time and reserved for analysis later when the circumstances may have changed. This stage aims to prepare all potential credible data to go through a parsing process, which is a more detailed analysis and sieving of the data. All necessary data is available to construct and to reconstruct a walkthrough of the control room.

A penetration testing phase is useful to target and to identify weaknesses in the system under investigation [24]. It is conducted remotely for acquiring live data on the system often when the users have not been formally informed that their machines are going through forensic investigation [25]. 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 [26]. Dead or static acquisition will be confirmed as the second step when relevant evidence is found. At this step, screenshots can be taken as a credible evidence of weaknesses and potential vulnerabilities to the work system.

The data examination stage features methodical assessment of all data, fact sheets, system settings, parsed data, data that came from the initial assessment, and media. Further processes of data analysis and examination also assure each media type is correctly processed and tools are found to process any irregular types. Timeline analysis and other perspectives allow systematic categorization and documentation of the relevant elements of information for the case [27]. 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 regardless of the operating system or format, and then analyzed. The timeline is fixed and application data reconstructed if required as a part of the data analysis and examination. Media and artefact analyses is addressed by, for example, what applications have been executed, which archives have been opened or downloaded, which documents have been clicked on, which records were checked, which files were deleted, where did the user browse, and many other properties. Another type of analysis, which is necessary for finding indirect paths of information is at the signature level. This analysis is where forensic investigators implement techniques and practices that will search for byte signatures of known folders, files and regular expressions that lead to the cookies. Link analysis is employed to find the relationships and trusted links to other entities, servers, domains, email, images, audio, people, and other relevant objects that can be traced to identify all possible communications [28].

Finally reporting and presentation is the stage that 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 are required for completion; also suggestions for enhancing procedures, guidelines, policies, applications, and other aspects of the forensic process investigations required in the target infrastructure [29]. This step is essential as it is important for the stakeholders in order to determine what strategies they must think about for future preparation. It includes a capability statement with respect to the investigation ability to process all multimedia formats or otherwise. The report has to be formulated in a form that is acceptable to the court or for any legal, employment or administrative purpose.

4. The framework

Digital forensic investigation frameworks have typically been developed for specialist areas of investigation by selecting standardized and repeatable process steps. In the former section we have described such phased steps for the generation of an investigation guideline for CI. However, what has yet to be addressed is the unique system and architectures of CI designs. A CI divides into work stations and control rooms. These are the two areas in which evidence must be collected by an investigator. The workstations interface at each of the CI levels described in Section 2 and **Table 1** and carry live data and stored data that can include volatile components such as RAMs and Flash memory. The digital investigator has to strategically plan

for the full range of devices and media types, and to tactically deploy capability to act effectively and efficiently in these environments. The digital investigator is also faced with enormous volumes of data and not just the variability of formats. To cope with volumes our modelling proposes deployment of Hadoop architectures to manage the big data volumes, and the selection of relevant evidences. **Figure 2** is designed to include these features and to deliver sufficient guidance to a digital investigator that they can manage the challenges of a CI environment. The framework provides control of the investigation from the five central phases where each phase appropriately connects to the big data issues on the right, and the workstation and control room issues on the left.

4.1 The five phases of investigation

The framework design centers the five phases of digital investigation between the two challenges in the CI environment – the media complexity and the data volumes. An investigator proceeds through the five phases described in Section 3 to assure completion and compliance with standardized procedures. The systematic and sequenced approach allows concentration on the system in focus and the completion of the professional activities associated. The investigator has the deliverable and the budget in mind at all times. Different types of evidence require different treatment and handling while data format and media type determine adequate access for imaging. By staging the investigation phases in the center of the framework the work system is established and the challenges of the environment are managed, phase by phase. On the left hand side the complexities of the CI workstation context and on the right hand side the strategy for managing large data quantities, are specified. The investigator can hence branch left and right to effectively acquire evidence, while maintaining the phased requirements for due processes.

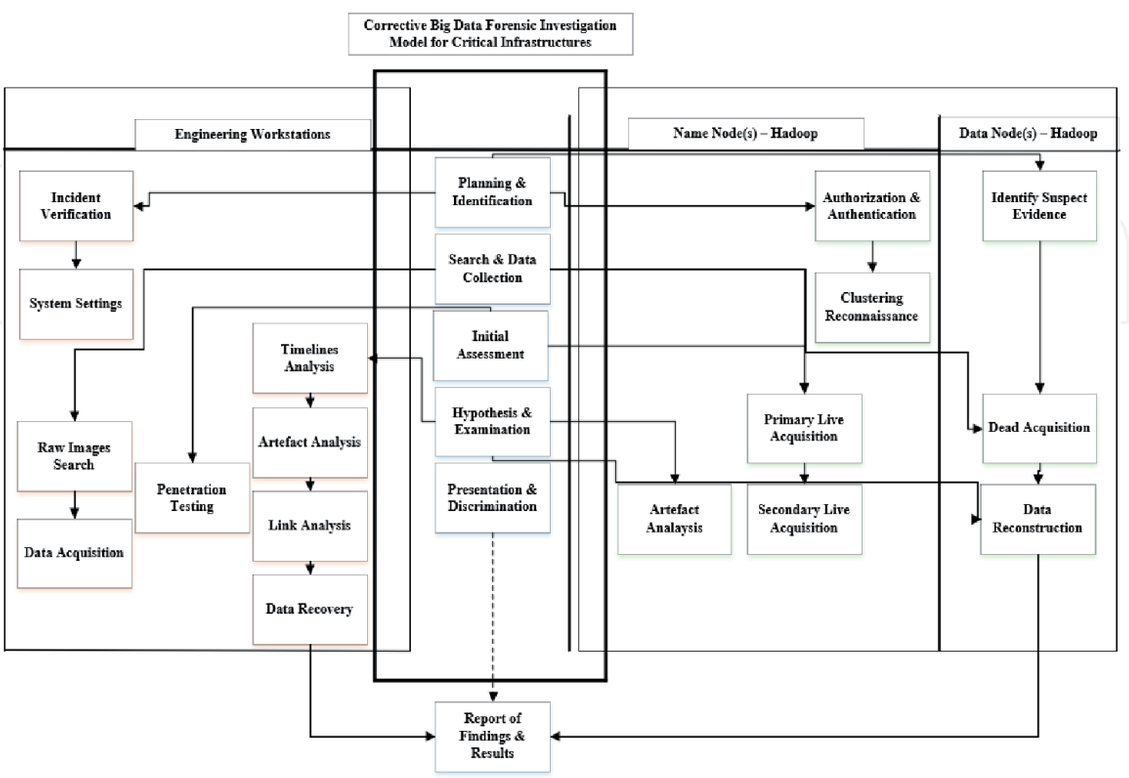

Figure 2.
A CI Investigation Framework.

4.2 Workstation and control room investigation

The workstations and control rooms context requires structured and planned entry. The control for investigation comes from the central digital investigation phases and the management constraints. At any step acquired evidence can include different types of forensic data such as pictures, audios, videos, text, files, directories, and so on. The multi layered challenge of the environment has to be addressed by strategy and tools that have proven effectiveness for data identification, time matching, multi-tenancy acquisition, data ownership differentiation, live forensic acquisition, privacy and privilege compliance, operating systems variation, media variation, format variations, and cloud compatibility. Sophisticated tools such as those that copy processes, examine evidence, analyze programs for generating checksums in order to complete the verification may not fit perfectly to some of control systems technologies. Control system technologies are also time stamped by the history of the system emergence and some data formats and operating systems may not be current. Consequently, many digital forensic tools demonstrate limited scope and require careful matching and mapping to the CI contexts to assure compatibility and effectiveness.

Importantly penetration tests are featured for the workstations and control rooms between the 'Search & Data Collection', and the 'Initial Assessment' phases. The penetration tests can confirm and limited the scope of further investigation. They can also provide vulnerability clues that redeploy of the 'Search & Data Collection' is done again for efficient targeting of areas for further investigation. This is a core component any CI forensic investigation. The major function of each one of these core components is to make sure that environments have correctly disclosed all the media for collection, and assurance is gained that complete analysis may proceed. The overall performance of an investigation will be limited unless the CI environmental and context variations can are fully addressed.

4.3 Big data investigation

The Hadoop context requires structured and planned entry for execution. The control for investigation comes from the central digital investigation phases and the management constraints. At any step acquired evidence can include different types of forensic data but the strategy is to organize the data into category and class nodes, and also data nodes. This organization and technical capability structures the data fields to optimize access at each phase of the central investigation plan. Live and dead nodes are discovered in a Hadoop architecture. They both contribute the necessary information needed to complete the digital forensic investigation on big data volumes. Nodes information is identified based on the different levels described in **Table 1**, such as node name with port number and IP address, last contact, admin state and additional information related to the data management and storage time and structure features. The scope includes all the logs created and stored on the cluster which contain the log files of data nodes, name nodes, secondary name nodes, the history server, user logs, the node manager, and the resource manager for all nodes. These files are vital for the process of hypotheses examination. To examine the Hadoop cluster, multimedia data acquisition techniques are used for the search and data collection. Data acquisition comes as a bit-by-bit copy of the content such as journal status, storage, log files, images, directories and logical database objects. The forensic examination is conducted through extracting system and nodes information using a range of proprietary and open source tools that are all selected and customized for the media type and performance. In this way the investigation phases can be executed in the big data context.

5. Penetration testing targeting

Cost efficiency is a critical factor in any digital investigation. Many elements, such as complexity and data volumes, hinder the efficient completion of investigation in CI environments. Consequently, strategies have to be employed to speed the completion without compromising the integrity of an investigation. Penetration testing is one such strategy. It is usually controlled and handled by penetration testers or qualified auditors and security specialists who are contracted in to scope the system and to identify useful investigation targets before the formal investigation proceeds. A penetration test seeks out the vulnerabilities of the system that an attacker could exploit, and where the system weaknesses are located. Such tests are performed from inside and outside the CI network infrastructure in order to test the overall performance of the network. The tests also determine the security level by categorizing the potential risks from high to low on the different interfaces. CI systems are a combination of applications interconnected to the control plane by network, hosts or branch networks. Penetration testing is a simulation process where real world attacks are made on potential targets to simulate the scope of hackers, attackers and other intruders. Penetration testing is also a valuable step towards developing a secure system that has assessed and mitigated potential vulnerabilities.

A basic penetration test may involve scanning for hosts' IP addresses in the network in order to check whether they are offering services with known vulnerabilities or hidden vulnerabilities that may be used in exploitation processes. The process would then extend to scanning ports for each host in the network and identifying unwanted opened ports that could be used as a gateway to the system. After following the penetration test plan the findings are reviewed and documented to be sent to stakeholders and investigators for action.

The objectives for penetration testing are [29, 30]:

- Preparing for the most effective starting test targets;
- Identification of security risks;
- Improving the performance of security systems;
- Prepare before an event occurs to prevent it; and,
- Reduce critical situations and potential crisis.

Important matters that come into consideration for planning CI penetration testing are aspects such as the scope, the intensity, the approach, the implementation techniques, and where to start. Each of these considerations will now be reviewed. The scope of the penetration test considers which systems and the degree to which each system will be tested. The cost may be reduced and complexity of the solution by limiting the extent of the testing in three categories:

- By performing Full penetration testing, the test will examine the overall performance and system safety policies of the target system.
- By performing Limited penetration testing, the access will include specific parts of the systems such as systems that are suspected hosts instead of testing the whole system.

- By performing Focused testing, where either one part of the system is tested or one service of the systems. The approach will provide only information about the test part not general information about the overall system security status.

The intensity of penetration testing is determined by the urgency of the situation. The urgency is measured by risk and is categorized into four metrics:

- Aggressive, is the highest level of penetration testing which generates a vast amount of network traffic about the infrastructure. The penetration tester tries to exploit all possible vulnerabilities in the system to identify whether the system is infected or secured. Some examples of aggressive attacks could be used in penetration testing such as Denial of Service attacks and buffer overflows. Calculated, cautious, and passive techniques are employed to get the best results. Covert and overt approaches are also used to sequence information gathering, and to achieve a comprehensive overview of a system. Different implementation techniques are also applied that differentiate characteristics of penetration tests and customize for the CI environment. The best approach, the motivation, and the important considerations when developing the optimal methodology and plan require sensitizing to the CI challenges. The implementation of an effective penetration testing plan can make an investigation cost efficient and deliver the best results earlier.

6. Conclusion

Conducting forensic investigations in industrial control systems is a complex process, not only because of the diversity of data and media, but also the variety of physical and logical partitions that are interconnected to the network including name nodes, data nodes and checkpoints. The research has delivered a framework for systematizing the process steps of investigation, and assuring the key issues of volume, format diversity, and management of data, are addressed. The innovation of featuring penetration testing into the investigation processes provides cost efficiencies and targeting towards completeness in an investigation. It steps beyond dependence on tool extraction of evidences, and justifies following the trail of evidence from the point(s) of greatest weakness and to the evidential media within the scope of a case. Such innovation improves assurance of completeness in an investigation and rigor for the methodologies. Digital forensic investigators are challenged by multimedia retrieval and data diversity. The proposed framework of methods is flexible and adaptable to multimedia environments, and assures control over the discovery processes.

Author details

Amr Adel[1*] and Brian Cusack[2]

1 Whitecliffe College of Technology and Innovation, Auckland, New Zealand

2 AUT University, Auckland, New Zealand

*Address all correspondence to: amra@whitecliffe.ac.nz

IntechOpen

References

[1] Phillips DM, Mazzuchi TA, Sarkani S. An architecture, system engineering, and acquisition approach for space system software resiliency. Information and Software Technology. 2018;**94**:150-164

[2] Khader M, Hadi A, Al-Naymat G. HDFS file operation fingerprints for forensic investigations. Digital Investigation. 2018;**24**:50-61

[3] Choure C, Patil LH. A Literature Survey on Intrusion Detection and Protection System using Data Mining. International Journal of Advance Research, Ideas and Innovations in Technology 2018;4:1.

[4] Regulation P. Regulation (EU) 2016/679 of the European Parliament and of the Council. REGULATION (EU). 2016;p.679.

[5] Qiu S, Liu J, Shi Y, Li M, Wang W. Identity-based private matching over outsourced encrypted datasets. IEEE Transactions on cloud Computing. 2015;23;6(3):747-59.

[6] Ahmad I, Abbas H, Raza A, Choo KK, Sajid A, Pasha M, Khan FA. Electronic crime investigations in a virtualised environment: a forensic process and prototype for evidence collection and analysis. Australian Journal of Forensic Sciences. 2018;4;50(2):183-208.

[7] Mouhtaropoulos A, Li CT, Grobler M. Digital forensic readiness: are we there yet. Journal International Computers, Law & Technology. 2014;**9**:173

[8] Jones J, Etzkorn L. Analysis of digital forensics live system acquisition methods to achieve optimal evidence preservation. IEEE SoutheastCon. 2016:1-6

[9] Shrivastava G, Kumar P, Gupta BB, Bala S, Dey N, editors. Handbook of Research on Network Forensics and Analysis Techniques. IGI Global; 2018

[10] Genge B, Graur F, Haller P. Experimental assessment of network design approaches for protecting industrial control systems. International Journal of Critical Infrastructure Protection. 2015;**11**:24-38

[11] Cherdantseva Y, Burnap P, Blyth A, Eden P, Jones K, Soulsby H, et al. A review of cyber security risk assessment methods for SCADA systems. Computers & security. 2016;**56**:1-27

[12] Stouffer K, Falco J, Scarfone K. Guide to industrial control systems (ICS) security. National Institute of Standards and Technology. 2008

[13] LeSaint J, Reed M, Popick P. System security engineering vulnerability assessments for mission-critical systems and functions. In: Proceedings of the Annual IEEE Systems Conference (SysCon). 13 Apr 2005. pp. 608-613

[14] Obregon L. Secure architecture for industrial control systems. SANS Institute InfoSec Reading Room. 2015 Sep.

[15] D'Orazio CJ, Choo KK. Circumventing iOS security mechanisms for APT forensic investigations: A security taxonomy for cloud apps. Future Generation Computer Systems. 2018;**79**:247-261

[16] Lutui R. A multidisciplinary digital forensic investigation process model. Business Horizons. 2016;**59**(6):593-604

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

[18] Martini B, Choo KK. Cloud forensic technical challenges and solutions: A snapshot. IEEE Cloud Computing. 2014;**1**(4):20-25

[19] Liu A, Fu H, Li Y. Secure and Trustworthy Forensic Data Acquisition and Transmission in a Cloud Infrastructure. World Scientific Book Chapters. 2018:167-191

[20] Qiu S, Liu J, Shi Y, Li M, Wang W. Identity-based private matching over outsourced encrypted datasets. IEEE Transactions on cloud Computing. 2015; 23;6(3):747-59.

[21] Awodele O, Onuiri EE, Okolie SO. Vulnerabilities in network infrastructures and prevention/ containment measures. In: Proceedings of Informing Science & IT Education Conference (InSITE) 2012.

[22] Broad J, Binder A. Hacking with kali, Practical Penetration Techniques. Waltham: Syngress; 2014

[23] Baloch R. Ethical hacking and penetration testing guide. Vol. 29. CRC Press; 2017

[24] Green J. Staying ahead of cyber-attacks. Network Security. 2015;1;(2):13-6.

[25] Singh S, Sharma PK, Moon SY, Moon D, Park JH. A comprehensive study on APT attacks and countermeasures for future networks and communications: challenges and solutions. The Journal of Supercomputing. 2019;**75**(8):4543-4574

[26] Northcutt S, Shenk J, Shackleford D, Rosenberg T, Siles R, Mancini S. Penetration testing: Assessing your overall security before attackers do. Sponsored by Core Impact, SANS Analyst Program 2006;3(6):22.

[27] Bradbury D. Point to own: the problem with hacking tools. Computer Fraud & Security. 2011;**2011**(11):12-14

[28] Yeo J. Using penetration testing to enhance your company's security. Computer Fraud & Security. 2013;**2013**(4):17-20

[29] Ficco M, Choraś M, Kozik R. Simulation platform for cyber-security and vulnerability analysis of critical infrastructures. Journal of computational science. 2017 Sep 1;**22**:179-186

[30] Sabillon R, Serra-Ruiz J, Cavaller V. An effective cybersecurity training model to support an organizational awareness program: The Cybersecurity Awareness TRAining Model (CATRAM). A Case Study in Canada. Journal of Cases on Information Technology (JCIT). 2019;21(3):26-39.