Implementation of virtual digital forensic class and laboratory for training, education and investigation

Damir Delija, Goran Sirovatka, Domagoj Tuličić, Marinko Žagar, Krešimir Hausknecht,Davorka Topolčić, Savina Gruičić

Tehničko veleučilište u Zagrebu /Informatičko računarski odjel, Zagreb, Republika Hrvatska

marinko.zagar@tvz.hr, damir.delija@tvz.hr, goran.sirovatka@tvz.hr,Domagoj.Tulicic@tvz.hr, Kresimir.Hausknecht@tvz.hr, Davorka.Topolcic@tvz.hr, Savina.Gruicic@tvz.hr

***Abstract* - In this paper, we present theoretical development and practical implementation of digital forensic laboratory computing infrastructure in a private cloud environment. The idea is to develop a scalable environment of forensic workstations and management facilities to support both the learning process and practical work in digital forensic education. Experience gathered in this process will be later used for setting up a production digital forensic laboratory in a cloud environment. To gauge student perception and experience in online learning and using virtual digital forensic classes and laboratory students filled the Student Perception Survey. The results are analyzed, and suggestions and conclusion used to improve future use.**

Keywords: Digital forensic laboratory, cloud infrastructure

I. INTRODUCTION

Digital forensics is defined differently and according to some authors, there are as many definitions as the digital forensic practitioners [2]. For this paper, it is important to choose definitions which address essential of the digital forensics and relations to training and education. Some of the definitions, like the following two, in our opinion, give the best definition of digital forensics:

- Digital Forensics is considered the application of science to the identification, collection, examination, and analysis of data while preserving the integrity of the information and maintaining a strict chain of custody for the data [3].
- The application of computer science and investigative procedures for a legal purpose involving the analysis of digital evidence after proper search authority, chain of custody, validation with mathematics, use of validated tools. Repeatability, reporting, and possible expert presentation [4].

For the understanding of digital forensics education and training, it is important to see that digital forensics is an engineering science, which is again part of computer science [6]. The profession of digital forensics requires continued education, training, and practices also it needs to smooth differences which practitioners from different fields have in their previous education and training [6]. At the moment there are issues with definitions, meanings, terms. Important concepts like case, evidence, etc. comes from law enforcement but lacks in technical implementations. All this issue must be addressed and solved in the practical implementation of digital forensic classes and related laboratory.

Since 2008 authors have been involved in digital forensics, digital forensic training, and education in various environments. Through this experience, various methods of deploying training infrastructure were tested. This experience is the core of the ideas presented in this article. Different approaches were used in different scenarios as various training and consulting contracts were done, from vendor defined professional training, to academic lectures and workshops.

II. DIGITAL FORENSIC LABORATORY AND DIGITAL FORENSIC EDUCATION

Digital forensic laboratory or DFL is hard to define, especially in the context of training and education. Based on the definitions of digital forensics and current practices digital forensic laboratory is an organization which provides digital forensic services. With such broad definitions, it is easy to implement different types of forensic services under a same legal and organizational umbrella, it is even possible to virtualize part or full laboratory processing in the legal limits and in the limits of best practices and standards [7]. Core services are processing digital forensic artifacts and digital evidence. If we want to provide this service, the laboratory must be organized and deployed in such a way that tools, procedures, infrastructure, and tasks can be easily executed.

Organization of the DFL is a complex issue where many factors interfere, the most modern approach is to define DFL as a business process based on the related business plan and metrics as it is suggested in [7], but problem is that this does not guarantee success since many historical, technical and legal issues can create huge problems.

Historically, the digital forensic lab is based on law enforcement practices, since 19th-century forensic science is introduced into the law enforcement process. Since than forensics is used as a tool to reconstruct events not only for legal reasons [6]. In practice, that means slow and incremental growth of law enforcement laboratory based on adding a new service into existing organizations. Such a model works well with traditional science. Introduction digital forensics creates a conceptual problem in using not optimal models of data processing. This practice has caused backlog crises in about early 2000 since the volume of digital artifacts needed to be processes raised exponentially while procedures, infrastructure, and tools were capable only on slow linear growth [8]. Introduction of the new concepts like digital forensic triage, parallelism in processing and laboratory management tools solved backlog situation, but the basic source of the problem is still there. As current trends of digitalization get the momentum we can soon expect again performance problems.

Digital forensic education requires practice in a controlled environment as close as possible to current real situations. Many vendors who are providing such training recently developed a classroom configuration as close as possible to production labs, since procedures and tools are taught as soon as it is operational. The same approach can be seen in the academy, at least at some institutions where cybersecurity centers are closely linked with academic programs, Rochester Institute of Technology is one of the successful examples. It basically means infrastructure is duplicated and separated in such a way that security is not compromised, while everything else is close to the real environment. This approach has also hidden advantage of easy surging of both sides of infrastructure, training or production, since changing the role of part of the lab means reinitialization and connecting to another part of the infrastructure, with some extra care on licensing, since academic licenses strictly forbid commercial or real-world application.

In such an approach it is important to understand possible functional organization of lab or classroom, there are basic types or configuration of computers which can be implemented as building elements. It is possible to separate machines into:

- management and control services,
- storage and support service,
- forensic services.

Forensic services are provided through general purpose forensic workstations and purpose configured workstations which are built to specific tool or task requests. General purpose forensic workstation is primarily for analysis and reporting while task-oriented is streamlined for specific tasks which usually requires specialized hardware tools, like acquisition, data recovery from damaged media, etc. The usual problem with forensic workstations is which OS to choose since forensic tools are OS dependent. With the advent of Windows 10, it is possible to implement a reliable Linux subsystem on the Windows platform, providing us with the best of both forensic platforms.

Management and control services are provided through DFL management services and communication services, mostly it is a server-oriented task.

Storage and support services are provided as support for both forensic and management in the sense of providing cooperation infrastructure, printing, scanning, data storage, backup, data scrubbing, reconfiguration, etc.

Such configurations can be further virtualized and hosted on the cloud infrastructure, like Azure or private clouds.

III. DEVELOPMENT OF CLOUD-BASED DIGITAL FORENSIC EDUCATIONAL ENVIRONMENT

The current technology aims to virtualize and employ cloud infrastructure. The idea is to simplify the business process by providing only needed services to business while infrastructure being something to rent. It can be easily summarized as a „Cloud computing is the dynamic provisioning of IT capabilities (hardware, software or services) from third parties over a network.” [18].

In the context of digital forensics, there are two issues with such an approach, legal constraints and current capabilities and design of digital forensic tools. Technical issues are solvable, it depends on the technology which is already available.

Digital forensic workflow (detailed in [1] and [6]), which contains well-defined steps of acquisition of evidence, analyses, and reporting, has to be implemented in such environment. The acquisition step means extracting digital data from devices into the forensic environment. All other steps are also executed in a forensic environment where data are secured and preserved without changes. It is important to understand that environment where forensic work is done can be fully virtual or cloud-based, without any negative implication on the quality of the forensic process. There is some possible limitation on the first glance, but they are the same as for any other forensic environment [6].

The legal issue mostly addresses the ownership and confidentiality of digital evidence [18], both with the security of DFL which cannot be completely guaranteed if you are not the owner of all facilities and physical installations.

More important for education and training are limitations or issue based on the current state of digital forensic tools. Forensic tools and practices are not well suited for virtualization or automation. There are only a few tools, mostly open source, which is fully capable of automation and parallel work in the scalable environment, a good example of such tools is Sleuthkit, and its derivatives and partly bulkextractor. Other tools, especially commercial are designed as a single user single task model, which simply does not scale up for cloud environment or virtualization. Dell did some early work on cloud preparations [8], working extremely hard to achieve this with the current versions of tools like EnCase v6, even there was a significant capability in tools itself. The next generation of commercial tools like EnCase v7 [14] and FTK v3 [13] recognizes the need for parallelism and automation moving to two different approaches, database-

centric in FTK and direct cooperation like EnCase. Still, both tools kept their initial graphic user interface dependency, where it was hard to separate case processing from case analysis steps. The FTK database centralization led to an environment of central storage and work drones using built-in support for parallelism in modern databases. This approach led to a system with a heavy footprint and price not well suited for cloud environments. A less efficient approach is from EnCase, where dedicated back-end process was used for evidence processing, while the front end GUI is used for analysis and task control. It was easier to maintain on a set of almost identical nodes, creating much more cloud-friendly environment. The other forensic vendors still keep more single machine/server approach. Mobile vendors like Cellebrite [17] and Microsytemation are moving slowly into parallelism and virtualization while still being mostly single machine oriented. This is due to their primary orientation to data extraction from mobile devices, a type of data acquisition, which is almost impossible to do within the virtual environment.

As a computer, the digital forensic workstation has a key role in digital forensics. It is an implementation of the computing platform on which various forensic tasks are done. Such a machine needs to have performance and versatilities to keep with tools and task. Traditionally from user rated reasons, a forensic workstation is a Microsoft Windows based for the law enforcement forensic tasks, while other types of forensic tasks often use Linux based machines. Since the virtual environment usually means cloning and machine duplication such workstation should be based on the configurations with the simple uniform maintenance and administration. From a training and education perspective, there is another requirement to have a scalable machine able to perform various tasks with different tools, with simple failback/restore mechanism in case of catastrophic incidents. The key element in digital forensics is a volume of evidence data to be processed, in the training/education environment evidence data volume is usually much smaller than real case evidence data, but the same principles for workstation configuration apply as in the real work.

IV. IMPLEMENTATION IN TVZ PRIVATE CLOUD

TVZ cloud infrastructure is implemented in Microsoft technology as a private cloud with resources for education and research. TVZ cloud DFL at the moment consists of a set of resources:

•20 virtual forensic workstations;

•1 virtual Foreman lab management server;

•Onedrive infrastructure for data storage;

•SharePoint infrastructure for documentation and cooperation storage (knowledge base, communication groups, etc.)

•LMS (TVZ Moodle) infrastructure for class materials

The TVZ Digital Forensics Laboratory is planned to support education (student work), research (work of TVZ staff) and forensic investigation (authorized external investigators). Up to now, the DFL was organized as part of the ICT infrastructure through the use of clouds of TVZ and the virtualization mechanisms present. The laboratory will primarily be used in teaching, performing practical exercises and theoretical part of the course in cybersecurity subjects, in accordance with the study description [15]. Laboratory organization must be such that new methods and cases can be easily implemented, and that simple transfer of knowledge method is also provided

The laboratory consists of a set of computers of various configurations, mostly virtualized on the TVZ cloud. The basics of the lab are the machines in the cloud. Workstations can be virtual and physical, depending on the requirements and capabilities, access to the lab servers is over the TVZ web portal, in the same way as access to the TVZ cloud infrastructure is provided and secured. Security, data backup, maintenance is implemented through the cloud service of TVZ and appropriate support. The lab is trying to be maximally virtual to achieve maximum independence from physical machines, and thus also achieve mobility and error resilience. All server and workstation configurations are streamlined to their tasks.

In the case of special needs such as licensing (i.e. access to licenses), the specific procedure and access mechanism are either permanent or temporary implemented. Licensees are provided from license servers inside of TVZ or out of TVZ. At the moment example is EnCase, in the future, some other tools will get licensed externally for DFL use.

The basic configuration of the system is a Microsoft Windows machine with a Linux subsystem. Each of the particular server and workstation configurations is designed for a specific task or a specific class lecture, while all are derived from the basic configuration. Detailed descriptions of how each configuration is done are kept in the knowledge base and workbooks.

Access to the lab is provided to students, teachers and to the staff who have the approval and the valid user account on TVZ.

The DFL consists of functional units:

- Workstations – for performing digital forensic tasks
- Server - provision of services and support
- Communication Infrastructure - Data Transmission, Security, etc.,

Servers are defined according to the services provided by the lab. Accordingly, they may be part of a laboratory, or external, providing a service or even special servers completely out of the TVZ infrastructure.

By functions we can define servers for:

- Laboratory management:
    - Foreman machine [12] - conducting the forensic procedure, part of the laboratory
    - Licensing – can be external
- Laboratory services:
    - Authentication of the user-part of the infrastructure

- Messaging communication/mail - part of the infrastructure
- Groupware for collaboration not yet defined - part of the infrastructure

- Data storage

The special type of servers is a license server providing license for products used. It is a not part of a lab usually on a separate machine, there are special requirements for network traffic due to licensing protocols.

Other special servers are forensic product drones, for FTK [13], EnCase [14], FidelisCybersecuirity [16] or GRR [15], machines that allow parallel work, at the moment such machines are only planned but not yet tested.

There are also network simulators - servers that enable the creation of a virtual network for traffic simulation, this type of machine is planned for advanced network forensic tasks like FidelisCybersecuirty, it is yet unclear how such functionality can be implemented in the current private cloud with available resources.

Forensic workstations are of two types:

- Physical workstations used as a portal entry point into the private TVZ cloud and for forensic data acquisition,
- The virtual - predefined virtual machine with configuration for a specific product, part of teaching class.

V. FUTURE DEVELOPMENT

Future development of the idea will follow the forensic as a service approach which was proposed by some forensic tool vendors recently. To achieve this goal some radical changes must be made in commercial forensic tool concepts with much better standardization. Open source tools like Sleuthkit and Autopsy are already capable of such implementation, but it lacks sophistication and support of the commercially available tools. The key steps are already done by forensic vendors like Access Data and Open text/Guidance software in the sense of providing their tools with parallelism in processing and some basic automation. Other vendors like NUIX are much less common and much more expensive but have a more elaborate approach to parallel processing and automation. Current development of the Python language as a widely used forensic tool also provides valuable automation/parallelism capabilities for digital forensics [9]. As it was presented in recent work about Python EnCase cooperation it is possible to [9] interact with commercial forensic tools from Python.

The key concept is embedded automation support, scripting, which must be available in the original commercial tool. It is important to notice that EnCase has for a long time internal cooperation capabilities well beyond other forensic tools on the scene. This feature provides an important step up in performance and reliability, but also a possibility to move to EnCase to cloud infrastructure. More or less similar features are available for other commercial digital forensic tools, as it is described in [9]. In the future, this will provide full "forensic as service" implementations. From training/education perspective, it will be important to keep the current state of a single workplace as a configuration, or as a type of workplace where students can learn the basics of the digital forensics.

TVZ DFL class will be updated and kept up with new tools, as the resources will be available. Results will be used in production forensic laboratory infrastructure. Lesson learned from current implementations clearly shows that such a simple structure as it is now can be used in forensic work. Even if it is not possible to go to a cloud infrastructure for DFL it is possible to clone configurations and use it as starting points. Currently, configuration developed for TVZ private cloud is used as a basic setup for purpose-built digital forensic laboratory, which is under development, and expected to be in production soon.

From the more theoretical analyses viewpoint is necessary to address a lack of control in the current setup, which requires better integration of workstations with cloud infrastructure. This is, unfortunately, a vendor-specific issue with many compatibility problems. From current experience, it is also clear that private cloud infrastructure in academia lacks reliability and support required for real heavy duty implementations.

Current DFL will be expanded with new resources and new products as new study requires, mostly for the network forensics and for the mobile forensics and incident response. It will be also expanded for other subjects in the curriculum like for the cryptography, math, etc.

VI. ANALYSIS OF QUESTIONNAIRE "PERCEPTIONS OF THE LEARNING EXPERIENCE OF ONLINE LEARNING - LEARNING IN A CLOUD - WORKING IN A VIRTUAL LAB "

Student opinion is the most important feedback on how the concept of DFL and concept of the virtual class are working and how well are accepted. The end term questionnaire both with term papers shows a positive image among students, data were collected for all enrolled 43 students. Students in Digital Forensics class on average acquired 88.33% points what gives them excellent grade. This is a subject which requires students to do 8 practical exercises in the lab trough solving tasks with EnCase and Autopsy forensic tools for checking skills, and, they had two partial exams for checking their knowledge.

Since both real lab classroom and virtual one was available for students it is possible to value usefulness of both. During winter semester of 2018/2019 in the first year of the new digital forensic study program [19], students are pleased with virtual lab concepts, mostly because of ability to work on it, in any moment during the semester. It allows them to better use their time and other resources needed for the study. Unfortunately, the lack of virtual presence for the lectures prevents additional benefits for students.

After the lectures during the first semester of specialized professional study of information security and digital forensics, the questionnaire was conducted. In total 40 respondents answered the questions, some respondents with only partial answers.

The first group of questions was related to the effectiveness of teaching in the "cloud". The response gets

TABLE I ANSWERS RELATED TO THE EFFECTIVENESS OF TEACHING IN THE CLOUD

Online learning in the online cloud - the laboratory is or can be effective	Scale 1-5 Results:
Because it offers comfort	4,38
Because it allows meeting the individual learning needs	3,79
Because it contributes to the effectiveness of communication in a group	3,15
Because it increases the sense of community with teachers and fellow students	2,82
Because it encourages greater participation and interaction of students	3,38

a numerical grade form very efficient (5) to not efficient (1), as it is presented in Table I.

The effectiveness of learning in the online cloud has been recognized in "offers comfort," and meet the individual learning needs." In contrast to this, it is possible to see the low efficiency in the "community feel with the teacher and fellow students," as well as "effective communication in the group." An assessment of "participation and interaction of students" not proved to be a measurable because the respondents differently interpreted and uniformly appraised as "something more", "same" and "slightly less" effective.

The second group of questions was related to the preference of teaching in the cloud, results are summarized in Table II.

TABLE II ANSWERS RELATED TO THE PREFERENCE OF TEACHING IN THE CLOUD

Second Group questions related to the preferred form of:	1 most prefer, 4 least preferred
a) Lectures and exercises in the classroom	2,40
b) Online courses with online laboratory	2,77
c) A combination of lectures and computer laboratory	2,03
d) A combination of lectures and online tasks with online laboratory	2,20
a) Lectures and exercises in the classroom	2,40

Respondents most favored form of teaching "a combination of lectures and computer laboratory" - although as it is visible online learning offers "comfort", which follows "a combination of lectures and online tasks with online laboratory "and the least favored" an online course with online laboratory „even less than" lectures and exercises in the classroom". Respondents recognize their need for communication with the teacher in the form of "lectures" and put it in front of "convenience".

The third group of questions allowed respondents free form responses. There were four questions, results are summarized here according to a question.

a) What do you think is the biggest benefit of online learning - working in a lab in the cloud?

The participants in unison say the biggest benefit is the ability to plan the times and places of work, there is no need to install software, flexibility, alignment of learning with other obligations.

b) What do you think is or may be the biggest drawback - in the laboratory in a cloud?

Participants identified lack of mutual communication, and especially the lack of communication and feedback from the teacher, as an additional source of information and the person who can solve the problem or issue immediately without the need for the independent search. Participants identified the lack of communication in the group as well as the occurrence of loss of motivation, commitment, as well as the monotony in an online environment. Technical problems have been recognized also, the slowness of response and time limit of the connection duration. Participants emphasized the need for precise and detailed operating instructions as well as the need for increased intensity (frequency) of the exercises.

c) How would you describe your own experiences with online learning - work in the lab in the cloud?

Participants emphasize a sense of satisfaction with the ability to work at a time when it suits them, as well as the possibility of access to all the material at any time and any place. Although satisfied with the access they point out the need for the teaching of the "live teacher", "classical education", because they lack the teacher.

d) Your comments or suggestions for improving the work in the online lab where you work this semester.

Participants indicate and emphasize the need: to simplification of the access procedure, increased computing resources for more comfortable work. Participants suggest intense exercises (in blocks), with more facilitation of the teacher and the variety of forensic tools. They ask to be given tasks on a daily basis, and the possibility for the arranged contact with the teacher during the online work. It is recognized the requirement for additional content, and for materials for those who want to know more.

VII. CONCLUSION

The work with virtual laboratory and class shows some interesting findings, student questionnaires shows that students are more comfortable with virtual lab environment but the learning outcomes were not satisfied. This situation is more related to uncertainty with related curriculums as it is noted in a recent report "The cybersecurity Workforce gap" [10] which shows that appropriate skills are not taught in most of the cybersecurity curriculums, the same situation is actually presented in the questionnaire findings.

From a practical viewpoint it is simpler to maintain virtual laboratory/class than the traditional one, even in a situation with inadequate technical support and skills. The most troublesome situation is an implementation of a commercial product which requires reliable infrastructure and adaptive configurations what private academic cloud is unable to provide reliably.

The introduction of a class using the lab in the private cloud increases comfort and students satisfaction but requires greater preparation of materials for independent students work. Although students are satisfied because they can work when it suits them best, questionnaire presents students need to communicate with each other and with the teacher what during lectures and laboratory work. This is an important issue which should be addressed in further improvements, providing kind of live interaction among participants.

LITERATURE

[1] F.Kovacs, "Windows 10 as a Forensic PlatformSTI Graduate Student Research", June 15, 2018, Available: https://www.sans.org/reading-room/whitepapers/forensics/paper/38475

[2] "Division of Undergraduate Education/Office of Information Technology, May 19, 2014, Online Learning-Student Perception Survey" Available: http://sites.uci.edu/eee/files/2015/09/onlinelearningsurvey20092014.pdf

[3] J. R. Kiper, "Towards a Digital Forensics Instructional Framework," Sans Inst. InfoSec Read. Room, 2017.

[4] K. Kent, S. Chevalier, T. Grance, and H. Dang, "Guide to integrating forensic techniques into incident response," NIST Spec. Publ. 800-86, p. 121, 2006.

[5] K. Zatyko, "Commentary: Defining Digital Forensics," Forensic Magazine, 2007. [Online]. Available: https://www.forensicmag.com/article/2007/01/commentary-defining-digital-forensics.

[6] G.C.Kessler, "Is Mobile Device Forensics Really 'Forensics'?", NIST Mobile Forensics Workshop, Gaithersburg, June 2014, Available: https://www.nist.gov/sites/default/files/documents/forensics/4-Kessler-201406_NIST_Kessler_v2.pdf

[7] C.Valli, A.Jones, "Building a Digital Forensic Laboratory", Syngress, 2011, ISBN: 9780080949536

[8] B.Garbade, "Dell's Approach to Digital Forensics", CEIC,2011. now EnFuse digital forensic conference

[9] C.Hosmer, „Python Forensics",Syngress, 2014 ISBN: 9780124186835

[10] W.Crumpler, J.A.Lewis, "The Cybersecurity Workforce Gap", CSIS,2019, Available: https://www.csis.org/analysis/cybersecurity-workforce-gap

[11] M.Žagar,D.Delija,G.Sirovatka, "Setting Up Digital Forensics Laboratory:Experience of Zagreb University of Applied Sciences", Available: https://www.researchgate.net/publication/326112183_Setting_up_digital_forensics_laboratory_Experience_of_Zagreb_University_of_Applied_Sciences

[12] S.Holmes, "Foreman: forensic case management system" Available: https://bitbucket.org/lowmanio/foreman/

[13] F.Carbone, "Computer Forensics with FTK", Packt Publishing, 2014, ISBN 9781783559022

[14] S.Widup, "Computer Forensics and Digital Investigation with EnCase Forensic v7", McGraw-Hill Osborne Media, 2014, ISBN 9780071807920

[15] D.Delija,M.Žagar,G.Sirovatka, "Analiza pripreme Google Rapid Response (GRR) sustava za potrebe nastave digitalne forenzike i kibernetičke sigurnosti", MIPRO 2018, Available: http://docs.mipro-proceedings.com/ce/ce_80_4774.pdf

[16] FidelisCybersecurity product, Available : https://www.fidelissecurity.com/

[17] "Cellebrite: mobile forensic", Available: https://www.researchgate.net/publication/326112183_Setting_up_digital_forensics_laboratory_Experience_of_Zagreb_University_of_Applied_Sciences

[18] K.Paslin, "Demystifying the Cloud: A Lawyer's Guide to Cloud Computing", CEIC 2011 now EnFuse digital forensic conference

[19] "ELABORATE: Diplomski specijalistički stručni studij informatike (smjer računarstvo)", TVZ, 2010 (in the archive of TVZ and Agency for Science and Higher Education)