A Digital Forensic Approach for Diverse Cloud Computing System

Zayyanu Umar¹, Francis S. Bakpo², Collins N. Udanor³ and Musa Ibrahim Kamba⁴

^{1,4}College of Science and Technology,
Federal Polytechnic, Birnin Kebbi

^{2,3}Department of Computer Science, University of Nigeria

Abstract:- Cloud services vendors might be restricted to specific resources, lacking some services to their customers' requests; this leads to the need for multiple cloud data centers to collaborate to share resources. With distinctions in various features and architectures, the cloud computing systems may be interconnected, and such networks may be exposed to instability or intrusion. Therefore, there is a need for efficient, resilient, versatile, reliable architecture and a model capable of detecting hazardous cybercrimes on joined diverse cloud service providers, and it can facilitate real-time digital forensic investigations. The state-chart diagram and network design methodologies were used in designing a forensic architectural framework to depict the behavioural nature of cyber-crime attacks in joined cloud platforms. In addition, brainstorming using statistical set theory was used to formulate a statistical model for detecting and identifying cyber-crime. Meanwhile, Visual Basic, version 6.0 and MySQL version 5.5.8, were used, respectively as front-end and back-end, in developing a digital forensic system that can enable the detection of cyber-crime in joined clouds platforms. Both two tools have all the required features and enable inter-connectivity between heterogeneous operating system platforms. Weka statistical tool, version 3.9.3. was deployed based on the dataset generated, for evaluating detection rate, false-positive rate, accuracy and precision of three classification models (Support Vector Machine and Random Forest and Naïve Bayes). The performance evaluation of generated dataset using WEKA's Three classifiers: Support Vector Machine, Random Forest and Naïve Bayes classifier in terms of detection rate were 93.87%, 99.74% and 99.89%, respectively, false-positive rate, were 3.01%, 2.45% and 0.026%, respectively, accuracy rate were, 95.53 %, 9 9.86 % and 99.92% respectively, while precision were, 96.45%, 99.16 % and 99.50% respectively.

Keywords: Cloud Framework, Cloud Computing, Cloud Centers Diversity, Digital Forensics

1.0 INTRODUCTION

Cloud services environment turns the purchasing of software and hardware by the academic sectors and industries valueless as cloud's data centers on the globe store sensitive data or information on cloud storage, not on institutions' storage devices anymore. To achieve high availability and access speed, businesses use cloud computing to move their data and processing to the cloud. Clients in the cloud are most concerned about security. As a result, many businesses are hesitant to move their IT needs to the cloud (Nasreldin et al., 2015). As a result, the industry regularly suffers from increased traditional assaults like Denial of Service (DOS) and Distributed Denial of Service (DDOS), tempting some cloud resources.

In an innovative technology of joining multiple clouds to ensure interoperability and reap other interconnection benefits, malicious users such as DOS and DDOS attackers gain access to specific resources on cloud storage servers with a negative ego, snatch service or gain access to some sensitive data (Alqahtany et al., 2016).

The security challenge is raised by keeping some sensitive data on the cloud exposed to the public machines, as joined multiple cloud services are a combination of public and private cloud environments by different service providers. Aside from that, various aspects of linked cloud services, like the lack of well-defined physical attributes, diverse service models, heterogeneous deployment methods, and other configurations, have given rise to a new set of cloud forensics dimensions.

Most cloud environments' data-centers are not consistent and cannot share services with others because everyone speaks a unique language. The biggest issue with integrating several cloud service providers with different configurations is that they cannot share resources with individuals (Garrison, 2010).

Samy et al. (2018) mentioned that "Every cloud service provider has their distinct cloud computing infrastructure that has its structure of the different cloud computing models. Each cloud computing service model (such as software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS)) has a unique structure, contents, and formatting of the log files. This variety of log files will be another challenge for digital forensic investigators. Therefore, the digital forensic investigators need to understand and able to extract the required evidence from it".

There are no service standards specific to joining two or more heterogeneous cloud data-centers. Some data-centers use Simple Object Access Protocol (SOAP), and others use Representational State Transfer (REST) as communication protocols. Every cloud platform has its particular features, such as security control and access control of logins (Elhozmari and Ettalbi, 2016).

According to Ali et al. (2018), "The footprints are found in the log files, which help to trace the malicious actor, but it is much complex to trace and correlate these files in the heterogeneous structure of the cloud. To handle the situation, the Cloud Service Provider (CSP) needs a systematic approach to collect evidence, which helps to trace the malicious activity".

Digital forensics was created as a discipline to assist law enforcement in dealing with the criminal usage of digital technology. The discipline is relatively new; thus, there is little consistency between industry and courts of law, which has led to a lack of standardized processes, training and tools (Lopez, Moon, and Park, 2016). The issue with these services is that there is no harmonized security interface to a heterogeneous security platform (Khan and Ullah, 2016).

Vol. 10 Issue 06, June-2021

ISSN: 2278-0181

V 01. 10 188ue 00, June-2021

In cloud computing, collecting log files from heterogeneous configured cloud servers and providing the log files securely to digital investigators have become the highly challenging base on a chain of custody of the evidence (Meyer and Stander, 2015; Umar et al., 2020). These logs are decentralized, as data stored in the cloud is replicated to multiple servers and data-centers. Multiple cloud users' log information may be stored together or can be spread over multiple servers.

Taylor et al. (2011) argued that current forensic tools and frameworks do not satisfy digital investigations in today's cloud computing environments and suggest that these tools be updated to a cloud service environment. Furthermore, as a result of the growth of severity of these attacks, there is a need for a design of digital forensic services that can meet with the new cloud systems due to incapability of the existing conventional digital forensic strategies and techniques and cannot be practical well with the new cloud systems (Ezz-El-din et al., 2016).

2.0 REVIEW OF RELATED WORKS

Several attempts were made by researchers to the development of a digital forensic framework for cloud computing environments; among them are;

Digambar, (2015): "A Novel Digital Forensic Framework for Cloud Computing Environment." The author developed a framework to speed up the investigation processes of the crime in a private cloud. The framework developed was meant for a single cloud computing environment.

Sibiya et al., (2012): "Digital Forensic Framework for a Cloud Environment". The author developed a proactive digital forensic framework titled Live Digital Forensic Framework for a Cloud (LDF2C). The framework was tested on a single cloud computing environment.

Kechadi & Nhien-An, (2015): "Digital Forensic Investigations in the Cloud a Proposed Approach for Irish Law Enforcement". The authors developed a forensic framework that gives obvious distinctions between traditional digital forensics systems and cloud-based forensic systems. Furthermore, it was designed and implemented strictly abiding by Irish Law Enforcement and their cloud computing services provision.

Umar, Eneh, et al., (2020): in an article titled "Joined Heterogeneous Clouds Resources Management: An Algorithm Design" built a system and algorithm that can handle the variability and complexities of the different clouds during the management of inter-cloud resources. The experiment result shows that the USER-BASE (UB1) can subscribe to Data Center1 (DC1) through Data Center 3 (DC3) that it initially subscribed with average time 301.05 with insignificant differences when utilizing resources from Data Center 3 (DC3).

Choi et al., (2014): "A Method of DDoS Attack-Detection using HTTP Packet Pattern and Rule Engine in Cloud Computing Environment". The authors devised an enhanced MapReduce method and then compared the proposed method's performance execution time and the Snort detection method. The experiment was taken in a single cloud computing environment.

Chen et al., (2015): "Cloud Computing Based Network Monitoring and Threat Detection System for Critical Infrastructures". The author developed a cloud computing-based network monitoring and threat detection system. The system was tested in a single cloud environment.

Mondal et al., (2017): "Enhancing a secure cloud computing environment by Detecting DDoS attack using fuzzy logic". The study used fuzzy rules for the detection of DDOS attacks in a single cloud computing environment.

The above-developed digital forensic tools and frameworks were not intended for the multiple joined cloud service. Also, the frameworks deployed were neither to solve the interoperability issues nor specifically in locating the cyber-crime such as DOS and DDOS in joined clouds to help the digital investigation.

3.0 METHODS AND MATERIALS USED

The state-chart diagram and network design methodologies were used in designing a forensic architectural framework to depict the behavioural nature of cyber-crime attacks in joined cloud platforms. In addition, brainstorming using statistical set theory was used to formulate a statistical model for detecting and identifying cyber-crime. Meanwhile, Visual Basic, version 6.0 and MySQL version 5.5.8, were used, respectively as front-end and back-end, in developing a digital forensic system that can enable the detection of cyber-crime in joined clouds platforms. Both two tools have all the required features and enable inter-connectivity between heterogeneous operating system platforms. Weka statistical tool, version 3.9.3. was deployed based on the dataset generated in evaluating detection rate, false-positive rate, accuracy and precision of three classification models (Support Vector Machine and Random Forest and Naïve Bayes).

4.0 PROPOSED FRAMEWORK

The researcher dwelled on developing a digital forensic framework to detect cyber-crimes in heterogeneous cloud computing platforms. A heterogeneous cloud computing platform refers to the interaction of cloud computing service environments of different protocols, standards, cloud service providers used in rendering services to clients. The lack of standards for managing and configuring heterogeneous infrastructures, as well as a digital forensic system for detecting cyber-crime, denies companies that need to use heterogeneous infrastructures in transaction models the benefits of infrastructure heterogeneity, such as resource management, regional advantages, pricing model, hypervisor style benefits, recovery, and so on. We presented a combined heterogeneous platform and developed an effective digital forensic system for cyber-crime detection in this paper. A

heterogeneous cloud computing environment is depicted in Figure 1.0 and the heterogeneity parameters amongst the joined cloud environments were displayed in Table 1:

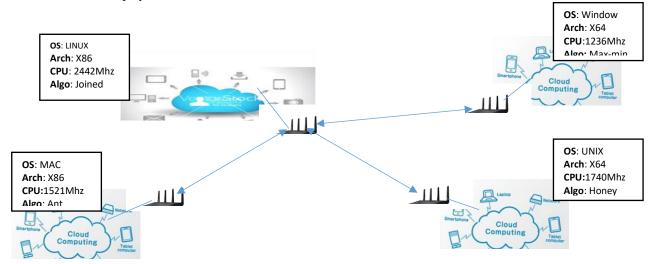


Figure 1.0: Heterogeneous Environment

Table 1: Cloud Service Provider (CSPs) Configurations

Provider	os	Arch	CPU (MHZ)	Apps (no)	Scheduling Algorithm	Pricing	Security Mechanism
CSP 1	KALI-LINUX	X86	4236	320	Joint	?	?
CSP2	WINDOW	X64	1232	20	Max-Min	?	?
CSP3	UBUNTU	X64	1740	15	Honey been foraging	?	?

The following developed statistical model is used in joining three different cloud setups for cyber-crime detection.

$$\mathbf{V} = ((\mathbf{X}^{\alpha} \ \mathbf{n} \ \mathbf{Y}^{\beta} \ \mathbf{n} \ \mathbf{Z}^{\theta}) - (\mathbf{X}^{\alpha} \ \mathbf{n} \ \mathbf{Y}^{\beta}) - (\mathbf{Y}^{\beta} \ \mathbf{n} \ \mathbf{Z}^{\theta}) - (\mathbf{X}^{\alpha} \ \mathbf{n} \ \mathbf{Z}^{\theta})) E \frac{Uj}{Si}$$

 X^{α} n Y^{β} n Z^{θ} stands for the Commonality of a joined set of Cloud service platforms, each with distinctions to others.

• Si Stands for a set of services that each Cloud service provider provides

$$S = \{S_1, S_2, S_3, S_4, \dots, S_M\}$$

• U_j Stands for a set of registered Users for each Cloud service provider U={U₁, U₂, U₃, U₄,.....U_M}

• E Stands for evidence generated while Cloud service providers interrelate with each other.

The following figure 2.0: Vein diagram represents the Statistical Formulae/Model generated:

ISSN: 2278-0181

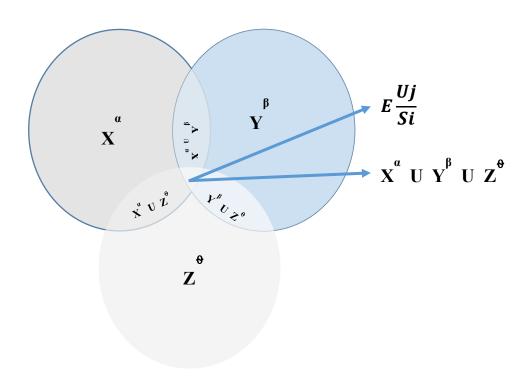


Figure 2.0: Vein diagram for interconnected Cloud Environments

The following figure 3.0 illustrates the digital forensic framework for cyber-crime detection

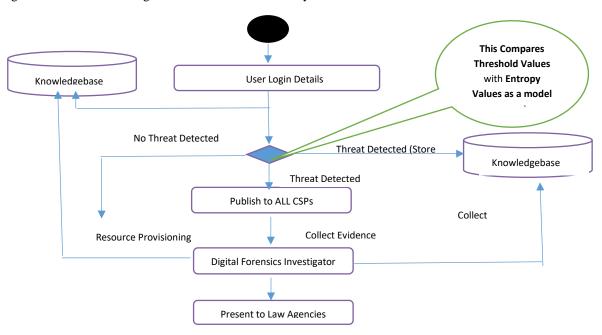


Figure 3.0: Digital Forensic Framework for DDOS Detection

Once user packets come in, the service request details such as Source IP, Target IP, Protocol Used, Source Port, Destination Port, Request Packet Length, Transaction time, Transaction Classification will be stored in Knowledge-Base A, then an algorithm will conduct a comparison between threshold value (Mean and the standard deviation) and Entropy Value (Which calculates the average amount of information covered per message); if a threshold value is not less than the entropy value, then it is considered as an intruder found in the environment (Threat Detected), at this juncture, the intruder's service request details (Source IP, Target IP, Protocol Used, Source Port, Destination Port, Request Packet Length, Transaction time, Transaction Classification) will be stored in Knowledge-Base B. Subsequently, that evidence will be shared or published onto storage media of all other joined cloud environments storage media and other cloud environments' local storages for onward to law enforcement agencies (such as

ISSN: 2278-0181

courts and police.). If the threshold value is less than the threshold value, the flow is normal; the cloud environment is not attacked, **Service Provisioning** will be allowed to the service request.

5.0 SIMULATION OF THE PROPOSED DIGITAL FORENSIC APPROACH

The system was developed using visual basic 6.0 as a from-end and MySQL server version 5.5.8 as a back-end to capture and store transaction packets among the number of joined cloud computing platforms. The following figure 4.0 is the first screen to display when the software is run:

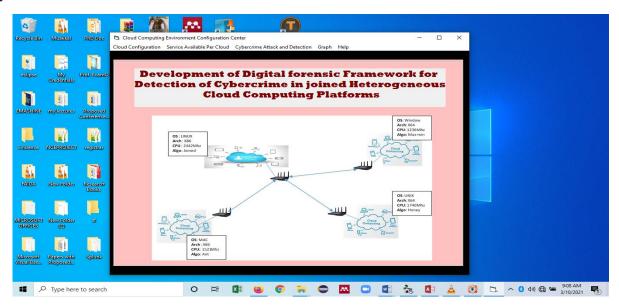


Figure 4.0: First screen with menu

The above diagram contains a menu for cloud computing environment configurations, services available for each environment, Cloud attacks and Detection, help that comprises documentation, and the author. The appendix titled as codes displays the system developed codes and figure 5.0 is for cloud configuration:

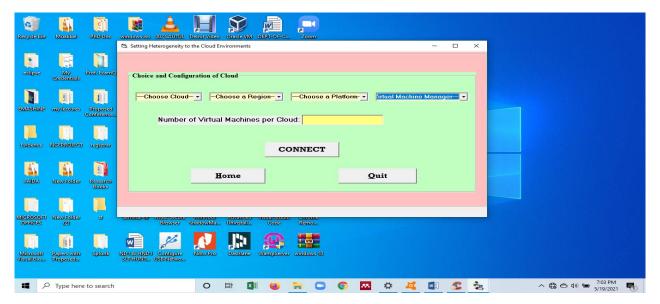


Figure 5.0: Cloud Configuration

The above figure is for setting some joined cloud environments, the world region of each, the governing platform (Operating system) and Virtual Machine Manager for each cloud environment (Hypervisor). Once the configuration is done and the button CONNECT is clicked, the cloud environment set will be connected. The following figure 6.0 shows how attacks can be launched

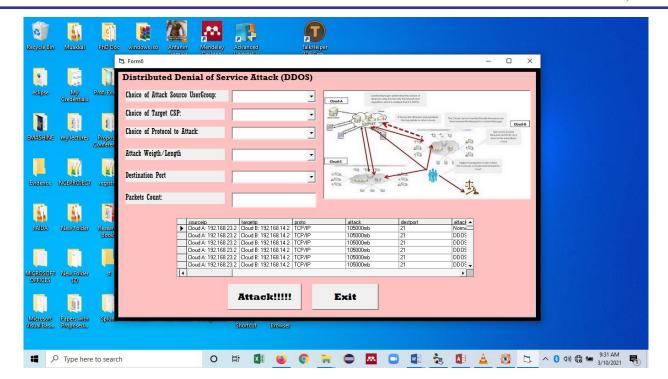


Figure 6.0: DDOS attack on cloud Environment

Testing inputs are Attack Source User-group, Target CSP, Protocol used in an attack, Attack request length, Destination Port and Packet Count. Once the user inputs the required data, then click on **Attack!!!!** Button, the attack is launched to the target cloud environment specified.

The following figure 7.0 shows how the proposed model was deployed to detect the DDOS cyber-crime:

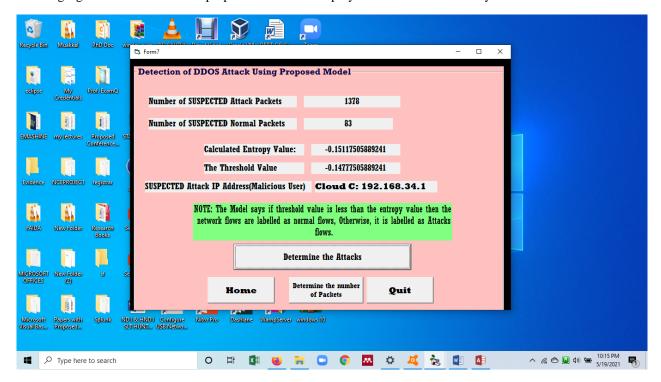


Figure 7.0: DDOS Attack Detection using the proposed model

The button labeled as **Determined the number of Packets** has to click, first to display the number of suspected attack packets and the number of suspected normal packets. Then the user clicks on the button **Determine the Attacks**. The calculated

Entropy and Threshold values will display; the system compares the two; if a threshold is less than the entropy value, the flow is Normal; otherwise, the flow is Attacked flow. As stated in the designed model, a suspected Attack IP will also display.

The attacked flow data is shared among the joined cloud environments. Therefore, a digital investigator will consult each cloud environment for compilations as figure 8.0 as below shows:

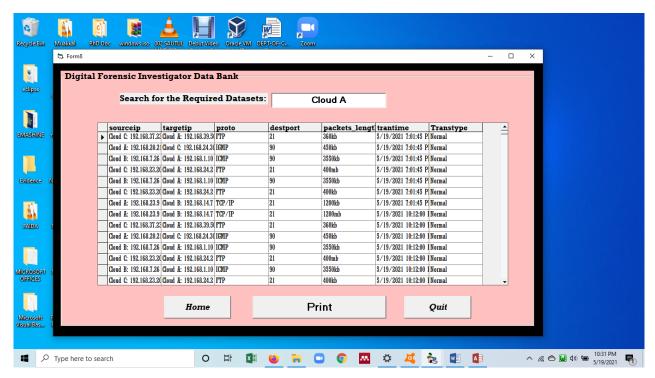


Figure 8.0: All Transactions from Cloud_A

With the limitations of the existing generated datasets, the author deemed it appropriate to use generated datasets after implementing and testing the developed system. As stated above, MySQL server version 5.5.8 was used in capturing the traffic packets and entire transaction datasets.

Each record in the generated dataset, as shown in figure 9.0, consists of six (6) features, and each record can either be a normal or attack traffic packet request. These features include Source IP, Target IP, Protocol, Destination Port, Request/Attack weight, and Status of attack/attack type.

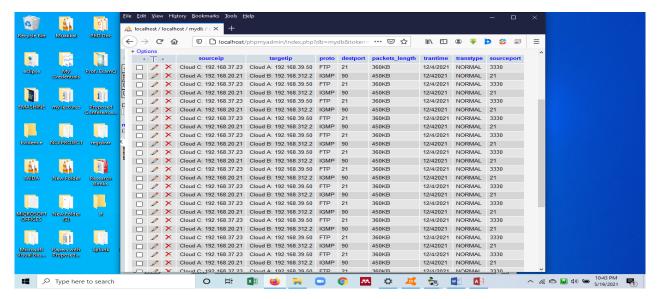


Figure 9.0: Generated traffic Packets Request

The paper now, focuses on analyzing the generated dataset of Distributed Denial of Service (DDOS) Attacks.

Vol. 10 Issue 06, June-2021

6.0 RESULT DISCUSSIONS

After deploying the Weka Environment, the following results show the detailed analysis of captured dataset/pcaps by MySQL server version 5.5.8 using selected machine learning algorithms. Table 2 and Table 3, Table 4 and Table 5, Table 6 and Table 7 show the analysed results of Support Vector Machine, Random Forest and Naïve Bayes Algorithms, respectively.

i. Support Vector Machine Result

Table 2: Detailed Accuracy by Class

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.811	0.011	0.993	0.811	0.893	0.754	0.991	0.986	NORMAL
0.989	0.189	0.718	0.989	0.832	0.754	0.991	0.993	ANOMALY
0.869	0.069	0.903	0.869	0.873	0.754	0.991	0.988	W. Avg.

Table 3: Confusion Matrix

A	В	classified as
13449	11743	ANORMAL
13044	11024	BANOMALY

ii. Random Forest Result

Table 4: Detailed Accuracy by Class

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.992	0.000	1.000	0.992	0.996	0.987	0.999	1.000	NORMAL
0.989	0.008	0.983	1.000	0.991	0.987	0.999	0.998	ANOMALY
0.994	0.003	0.994	0.994	0.994	0.987	0.999	0.999	W. Avg.

Table 5: Confusion Matrix

A	В	Classified As
13449	1833	ANORMAL
11713	22265	BANOMALY

iii. Naïve Bayes Result:

Table 6: Detailed Accuracy by Class

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.999	0.000	1.000	0.999	1.000	0.999	1.000	1.000	NORMAL
1.000	0.001	0.999	1.000	0.999	0.999	1.000	1.000	ANOMALY
1.000	0.000	1.000	1.000	1.000	0.999	1.000	1.000	W. Avg.

Table 7: Confusion Matrix

Tuble 7: Comusion Mulia					
A	В	Classified As			
25180	12	ANORMAL			
2	24066	RANOMALY			

Table 8 and Table 9 compared SVM, Random Forest and Naïve Bayes classification modesl based on the performance measure.

Table 8: Algorithms performance comparisons

Algorithm	True Positive Rate	False Positive Rate	True Negative Rate	False Negative Rate
SVM	93.87%	3.01%	96.98%	6.12%
Naive Bayes	99.89%	2.45%	99.99%	1.34%
Random Forest	99.74%	0.026%	99.97%	0.24%

Table 9: Accuracy and Precision

ruble y. Heedraey and Heelbion					
Algorithm	Accuracy	Precision			
SVM	95.53 %	96.45%			
Naïve Bayes	99.92%	99.50%			
Random Forest	99.86 %	99.16 %			

Vol. 10 Issue 06, June-2021

ISSN: 2278-0181

7.0 CONCLUSION

The new digital forensic approach for the heterogeneous cloud computing environment was developed and simulated, the dataset was generated. The analysis was conducted on the generated dataset in Weka environment deploying some preferred supervised learning algorithms. After analysis, the results show that the Naïve Bayes resulted in the highest accuracy and detection rates and the lowest false rates. The results specify that the Naïve Bayes model's classification capability is inherently superior to other two classification models; the Detection rate of the three (3) classifiers: support vector machine, random forest and Naïve Bayes) were 93.87%, 99.74% and 99.89%, respectively. False-positive rate, were 3.01%, 2.45% and 0.026%, respectively. Accuracy were 95.53%, 99.86% and 99.92%, respectively, while precision were 96.45%, 99.16% and 99.50% respectively.

REFERENCES

- [1] Ali, S. A., Shahzad, M., & Farhan, S. (2018). Challenges in Cloud Forensics. International Conference on Cloud and Big Data Computing, 6–10.
- [2] Alqahtany, S., Clarke, N., Furnell, S., & Reich, C. (2016). A forensic acquisition and analysis system for IaaS: Architectural model and experiment. Proceedings - 2016 11th International Conference on Availability, Reliability and Security, ARES 2016. https://doi.org/10.1109/ARES.2016.58
- [3] Chen, Z., Xu, G., Mahalingam, V., Ge, L., Nguyen, J., Yu, W., & Lu, C. (2015). A Cloud Computing Based Network Monitoring and Threat Detection System for Critical Infrastructures ★ Big Data Research, 1, 1–14. https://doi.org/10.1016/j.bdr.2015.11.002
- [4] Choi, J., Choi, C., Ko, B., & Kim, P. (2014). A method of DDoS attack detection using HTTP packet pattern and rule engine in cloud computing environment. https://doi.org/10.1007/s00500-014-1250-8
- [5] Digambar, P. (2015). A Novel Digital Forensic Framework for Cloud Computing Environment (Issue 2015). BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI.
- [6] Elhozmari, M., & Ettalbi, A. (2016). Towards a Cloud Service Standardization to ensure interoperability in heterogeneous Cloud based environment. 16(7), 60–70.
- [7] Ezz El-din, H., Manjaiah, D. H., Hemdan, E. E. D., & Manjaiah, D. H. (2016). A Cloud Forensic Strategy for Investigation of Cybercrime. Proceedings of IEEE International Conference on Emerging Technological Trends in Computing, Communications and Electrical Engineering, ICETT 2016. https://doi.org/10.1109/ICETT.2016.7873667
- [8] Garrison, C. p. (2010). Digital forensics for network, internet and cloud computing. Elsevier Inc.
- [9] Kechadi, T., & Nhien-An, L.-K. (2015). Digital Forensic Investigations in the Cloud A Proposed Approach for Irish Law Enforcement. January 2016.
- [10] Khan, S. U., & Ullah, N. (2016). Challenges in the adoption of hybrid cloud: an exploratory study using systematic literature review. *The Journal of Engineering*, 2016(5), 107–118. https://doi.org/10.1049/joe.2016.0089
- [11] Lopez, E. M., Moon, S. Y., & Park, J. H. (2016). Scenario-based digital forensics challenges in cloud computing. Symmetry, 8(10). https://doi.org/10.3390/sym8100107
- [12] Meyer, G., & Stander, A. (2015). Cloud Computing: The Digital Forensics Challenge. 285–299.
- [13] Mondal, H. S., Hasan, T., Hossain, B., Rahaman, E., & Hasan, R. (2017). Enhancing Secure Cloud Computing Environment by Detecting DDoS Attack Using Fuzzy Logic. 3rd International Conference on Electrical Information and Communication Technology (EICT), December, 7–9.
- [14] Nasreldin, M. M., El-hennawy, M., Aslan, H. K., & El-hennawy, A. (2015). Digital Forensics Evidence Acquisition and Chain of Custody in Cloud Computing. *International Journal of Computer Science Issues*, 12(1), 153–160.
- [15] Samy, G. N., Maarop, N., Abdullah, M. S., Perumal, S., Albakri, S. H., Shanmugam, B., Jeremiah, P., & Hasan, S. (2018). Digital Forensic Investigation Challenges based on Cloud Computing Characteristics. *International Journal of Engineering & Technology*, 7(4), 7–11. https://doi.org/10.14419/ijet.v7i4.15.21361
- [16] Sibiya, G., Venter, H. S., & Fogwill, T. (2012). Digital Forensic Framework for a Cloud Environment. IST-Africa 2012 Conference Proceedings, 1-8.
- [17] Taylor, M., Haggerty, J., Gresty, D., & Lamb, D. (2011). Forensic investigation of cloud computing systems. *Network Security*, 11(3), 4–10. https://doi.org/10.1016/S1353-4858(11)70024-1
- [18] Umar, Z., Bakpo, F. S., Alkali, M., Tanko, A., Kamba, M. I., & Ezema, E. (2020). Statistical Model for Cybercrime Detection in Joined Heterogeneous Cloud Computing Environments. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(5), 7844–7851. https://doi.org/10.30534/ijatcse/2020/134952020
- [19] Umar, Z., Eneh, A., & E, O. G. (2020). Joined Heterogeneous Clouds Resources Management: An Algorithm Design. *Journal of Mechanics of Continua and Mathematical Sciences*, 15(8), 39–52. https://www.journalimcms.org/journal/joined-heterogeneous-clouds-resources-management-analgorithmdesign/