I.J. Modern Education and Computer Science, **2018, 3, 17-28**
Published Online March 2018 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2018.03.03

INSPECT- An Intelligent and Reliable Forensic Investigation through Virtual Machine Snapshots

K. Umamaheswari
Research Scholar, Research and Development Centre, Bharathiar University, Coimbatore, India.
Email: uma.tvr1981@gmail.com

S. Sujatha
Assistant Professor, Department of Computer Science, Bharathi Womens College, Chennai, India.
Email: sujaphd@gmail.com

Received: 13 November 2017; Accepted: 15 January 2018; Published: 08 March 2018

Abstract—Cloud computing is emerging as a popular paradigm that provides significant advances and utility-oriented services over shared virtualized resources. Despite the advantage of the cloud services, the majority of cloud users are reluctant to access the cloud due to unprecedented security threats in the cloud environment. The increasing cloud vulnerability incidences show the significance of cloud forensic techniques for the criminal investigation. It is challenging to gather the evidence from the abundant cloud data and identifying the source of the attack from the crime scene. Moreover, the Cloud Service Provider (CSP) confines the investigator to carry out the forensic investigation due to the prime concerns in the multi-tenant cloud infrastructure. To cope up with these constraints, this paper presents INSPECT, an investigation model that accomplishes adaptive evidence acquisition with adequate support for dynamic Chain of Custody presentation. By utilizing the VM log files, the INSPECT approach forensically acquires the corresponding evidence from the cloud data storage based on the location of malicious activity. It enhances the evidence acquisition and analysis process by optimally selecting and exploiting the required forensic fields alone instead of analyzing the entire log information. The INSPECT applies the Modified Fuzzy C-Means (M-FCM) clustering with contextual initialization method on the acquired evidence to recognize the source of the attack and improves the trustworthiness of the evidence through the submission of the chain of custody. By analyzing the Service Level Agreement (SLA) of the cloud users, it facilitates the source of attack identification from the clustered data. Furthermore, it isolates the evidence to avert deliberate modification by an adversary in the multi-tenant cloud. Eventually, INSPECT presents the evidence along with the chain of custody information regarding the crime scene. It enables the law enforcement authority to explore the evidence through the chain of custody information and to reconstruct the crime scene using the VM snapshots associated with timestamp data. The experimental results reveal that the INSPECT approach accomplishes a high level of accuracy in the investigation with the improved trustworthiness over the multi-tenant cloud infrastructure.

Index Terms—Forensic investigation, VM snapshots, SLA, FCM clustering, chain of custody, privacy, and multi-tenant.

I. Introduction

Cloud computing [1] has become a promising and popular technology that offers a variety of opportunities and significant economic benefits to the individuals and organizations. Though, the rapid proliferation of the Information Technology (IT) and an advantage of the low-cost as well as on-demand cloud services [2] tend to lead the cyber crimes. For instance, cyber criminals candidly utilize the cloud services with the intention of targeting the victim. In 2013, the Chinese cybercriminals effortlessly distributed the malware for the initial stage of DDoS attack through the cloud file hosting services in the Dropbox [3]. Also, the distributed nature of cloud computing infrastructure creates unique challenges to the forensic investigators. Accordingly, there is an essential need of an effective forensic investigation system to deal with the dynamic cloud environment while ensuring the data integrity [4]. In addition, manipulating the crime scene investigation is an arduous task in ever-changing large-scale distributed cloud data centers. To resolve those constraints above, the recent researches introduce the cloud forensics [5] that enables the post-attack log investigation to detect the evidence and the source of the attack in the cloud through accessing the Virtual Machines (VMs).

Owing to the massive amount of available cloud data storage, collecting the bit-level evidence from the byte-level data becomes a difficult process for the cloud forensic investigator [6]. VM Introspection (VMI) mechanism can indirectly analyze and manipulate the suspected VMs through snapshots [7, 8]. The standardization of Cloud Service Level Agreements (SLAs) [9] eases the forensic investigator to acquire the

evidence from the corresponding malicious incident under the control of CSP. Even though SLA provides the significant data, separately analyzing the SLA of every individual user creates time complexity during the investigation, which tends to mislead the context identification. The previous methods apply the various clustering methods like Ahmed Fahim's [10] to carry out the analytical tasks on the data collection. In contrast, these methods are relatively difficult to handle the forensic data due to its unique features such as dimensionality and vagueness. To support the large-scale forensic data, several research works employ the Fuzzy C-Means (FCM) clustering method that can deal with the uncertainty of events. However, it generates the clusters based on the random decision on initial centroid points, regardless of the data context. Moreover, the existing cloud forensic techniques [11] [12] focus on reducing the data investigation but do not concentrate on acquiring the evidence-based on the execution changes in the cloud infrastructure. Thus, this work introduces a unique endeavor to collect the evidence adaptively by mitigating the investigation data and maintaining the chain of custody in the multi-tenant cloud environment. It ensures the reliable investigation procedure and improves the sustainability of the proof pointed by any malicious activity detection system like DCGIDPS [13].

The major contributions of INSPECT methodology are formulated as follows.

- The main objective of the INSPECT approach is to determine and present the appropriate evidence with a chain of custody while ensuring the accuracy of investigation through the forensically sound investigation method in the cloud.
- The INSPECT approach exploits the VM log files to precisely gather the evidence and eases the malicious data acquisition through considering the forensic fields alone.
- It applies the M-FCM clustering method and utilizes the SLAs to adaptively identify the source of attacks, which promotes the investigation system to maintain a chain of custody for the corresponding evidence with the assistance of CSP.
- Instead of applying the random initialization method in the traditional FCM clustering method, the INSPECT approach conceptually determines the centroid points using the effective initialization method and clusters the malicious data, resulting in the accurate solutions and optimum global convergence.
- According to the clustered results, exploring the SLAs of the Cloud users diminishes the computational complexity and improves the accuracy of the source of attack identification.
- The INSPECT approach isolates the cloud instances to perform the uninterrupted as well as uncompromised investigation and to submit the evidential artifacts with the chain of custody information in the court using the Advance Forensic Format (AFF).
- The experimental results show that the INSPECT approach accomplishes a high level of forensic investigation accuracy and evidence trustworthiness.

II. Related Work

Most notably, the existing works in cloud computing have provided valuable contributions towards the improvement of the cloud forensic investigation presented in this paper. From the previous works, several prominent efforts are enumerated as follows:

A. Evidence Acquisition Methods

Cyber Crime Scene Investigation (C^2SI) [14] presents pay-as-you-go trace back model to explore the cloud environment over Tor, captures the real suspects by assigning Tor entry and Tor exit sentinels in the cloud. Although, it lacks to ensure the correctness of the extracted evidence, the VMI-based forensic examination model [15] identifies the malicious events by examining the running suspected VM using VMM, which leads the system to determine the volatile data. However, it causes the security risks to the cloud users if VMM is compromised, as VMM has the control of all resources. Also, it solely provides benefits for a particular application and requires an expert analyst. Later, to accurately reconstruct the crime scene and ensure reliable investigation in the cloud, the researchers [16] [17] focus on the temporal information associated with the evidence while acquiring the evidence for the investigation. However, it provides poor performance in a large-scale cloud environment. An approach [18] resolves the incongruity of the evidence by detecting the temporal inconsistencies in the VMs, which facilitates the sequence of event management and missing events detection in the VM. Though, it relies on the timestamp of the VM logs, which misleads the forensic investigation. The CloudVMI method [19] allows the introspection of the VM by virtualizing the VMI interface with the cloud environment to periodically monitor the allocated resources. However, it only considers the access history, fails to analyze the dependency and resource utilization. Forensic acquisition approach [20] explores two snapshots of the same VM to identify the disparity between the snapshots, including a post-attack snapshot of virtual disk and snapshot of the same virtual disk after determining that disk as the suspected disk. The snapshot-based approach promotes the in-depth analysis of the suspected VM and stores the VM snapshots in the persistent storage to regenerate the crime scene in the dynamic cloud environment rather than taking the snapshots of all the VMs [21]. However, it lacks focus on maintaining the trustworthiness of the evidence acquisition, according to the analysis of VM snapshots.

B. Forensic Investigation Methods

The Forensic analysis method [22] exploits the VM image migration concept to control the difficulty of managing the data of co-located users in the virtual

instance, which intends to isolate the data from the suspected VM. Cloud instance isolation method [23, 24] separates the suspected instances within the cloud infrastructure by exploiting the possible techniques that include server farming, instance relocation, failover, address relocation, sandboxing, and man-in-the-middle. It facilitates the forensic investigation of the evidence damage by the adversary and also preserves the privacy of other tenants in the cloud. Although, there is an essential need for isolation that isolates either data or suspected VM according to the severity or the source of the attack. An approach [25] employs the fuzzy clustering method to regenerate the crime event with the help of periodic VM snapshots defining attacks as clusters, which leads the investigator to restore the crime scene with proper sequence. However, it poses a complication when there is a global dispersion of multiple VMs under attack. An approach [26] presents an integrated conceptual framework to collect and preserve the forensically sound data from the cloud, eases the chain of custody management, and improves the evidence trustworthy. However, there is no hash image to validate the data integrity. FROST [27] presents the forensic tools for the OpenStack cloud platform that supports Infrastructure-as-a-service (IaaS) cloud, provides the opportunities for the forensic investigators to acquire the forensically sound evidence from the cloud regardless of CSP. However, it fails to ensure the data integrity from the malicious investigator and CSP. Adaptive evidence collection model [28] takes into account of the attack scenarios to precisely acquire the evidence in the cloud infrastructure. Because of that, the evidence collection strategy highly relies on the virtual and physical machine changes and jurisdiction changes.

However, the former cloud forensic techniques mislead the forensic investigator when there is a diverse attack scenario during evidence acquisition and evidence validation in the dynamic cloud. Also, lack of analyzing the contextual information in the cluster under attack leads the investigator to identify the inappropriate data as the evidence of the crime scene.

III. The Inspect Methodology

The rapid growth of the cloud computing technology exacerbates the digital forensic activities and consequently, creates a brand-new way for cyber crime investigation. To deal with the cloud forensic challenges and to cope up with the latest advancements, the digital forensic practitioners need to enrich their tools and technologies. The cloud forensics investigation model often struggles when handling the massive data, terminated data, and dynamic allocation and reallocation process during the collection of evidence of the crime event in the cloud infrastructure. Most of the existing cloud forensic techniques consider only on acquiring the evidence from the abundant cloud storage. However, these techniques fail to identify the entire sequential activities of the crime scene. Hence, this paper presents a novel cloud forensics model that focuses on candidly collecting the evidence from the VM log file storage of multiple tenants.

The proposed model targets to resolve the constraints related to multi-tenancy violations, a chain of custody, and malicious insiders. It employs an M-FCM clustering method to determine the evidence regarding the malicious activities. It also supports the completion of the investigation process in a reasonable time, even when there are changes in the cloud scenario. It involves two major phases such as identifying the malicious data and adaptively acquiring the source of the attack and isolating and presenting the evidence. The overall processing methodology of the proposed model is shown in Fig. 1.

Identifying the malicious data and adaptively acquiring the source of the attack: The INSPECT approach extracts the log files of the suspected VM that is identified by the malicious activity detection system. The INSPECT approach employs the M-FCM clustering method with the knowledge of SLA on the obtained malicious activities to firmly identify the cause for a specific set of malicious activities. According to the dynamic cloud scenario involving allocation, reallocation, and migration process, and its corresponding SLA, the proposed cloud forensic model, adaptively determine the convict through analyzing the possibility of the malicious activity. The changes in the cloud scenario tend to create the impact on determining the source of the attack.

Isolating and presenting the evidence: To evade the data breaches of other users, the INSPECT approach isolates the suspected evidence from the suspected VM. It tends to avert the cloud service interruption during a forensic investigation. It performs either task or VM isolation based on the suspected data location and leads the investigator to conduct evidence analysis process conveniently in their machine. Eventually, it submits the acquired evidence along with the source of an attack of a real suspect to the law enforcement system with the proof in terms of chain of custody. The VM snapshots also facilitate to assess the evidence acquisition by the law enforcement system in the cloud environment.

A. Identifying the malicious data and adaptively acquiring the source of the attack

The INSPECT approach targets on identifying the malicious data by extracting the VM logs alone. It leverages the forensic investigator after receiving the result of attack location from the malicious activity detection system. The malicious activity detector continuously monitors the cloud environment to determine the malicious activity. Initially, the malicious activity detection system such as Intrusion Detection System (IDS) determines the attack location concerning VM ID by exploring the fluctuations in the rules of a specific process in the cloud server. The forensic investigation process involves the malicious data collection. With the aim of achieving more accurate and rapid process of malicious data acquisition, the INSPECT approach develops a model for gathering the cloud data of the suspected VM to recover the data involving logs of VM instances. After determining the malicious data as

the evidence from the cloud, the INSPECT approach acquires the source of the attack with the assistance of SLA and a fuzzy clustering method. Determining the source of attack plays a crucial role in the forensic investigation to produce forensically sound evidence and also to maintain a chain of custody. To effectively identify the source of the attack, the INSPECT approach follows the two steps. Initially, it applies the M-FCM clustering method on the collected malicious data to aggregate the similar type of malicious data. Secondly, it analyzes the SLAs of each cluster that accelerates the source of attack identification process. Eventually, the INSPECT approach presents a set of the source of attacks with different scores for each malicious data cluster.

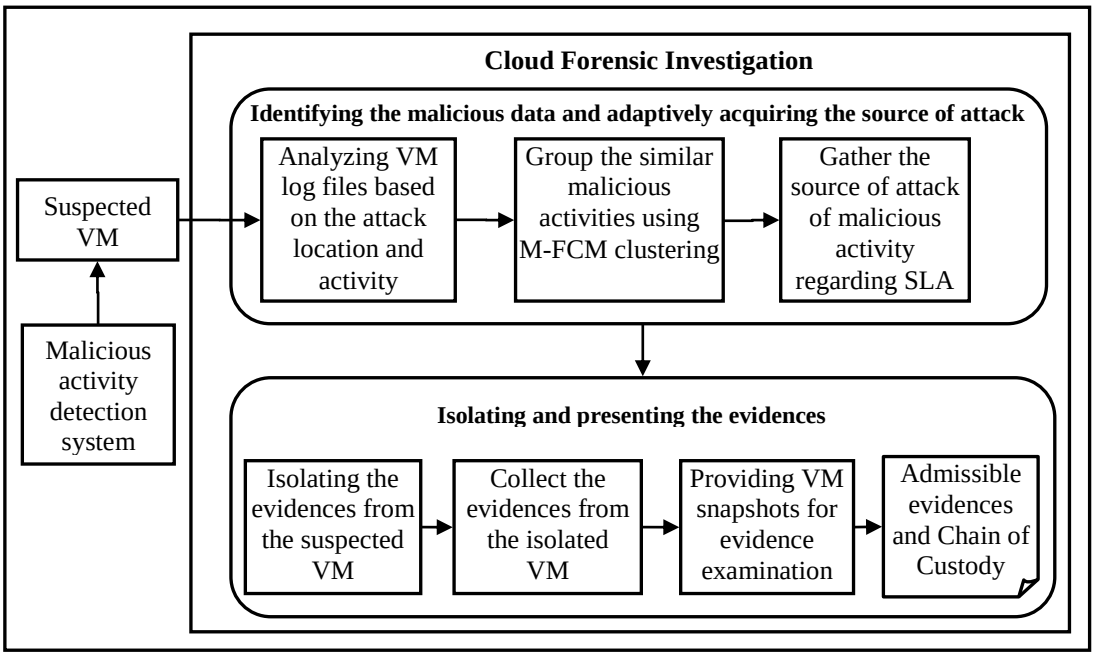

Fig.1 Intelligent and reliable forensic investigation using M-FCM clustering in the cloud

A.I Aggregating the identical malicious data by Modified Fuzzy C-means clustering

To precisely perform the adaptive forensic investigation, the INSPECT approach focuses on both the proactive and reactive analysis during evidence collection. It leads the system to preserve the most prominent evidence, which is more beneficial to the forensic investigation to take the immediate and strong suggestion regarding the crime event. Hence, the INSPECT approach employs the M-FCM clustering instead of applying the fuzzy logic function to determine the fuzziness of the malicious data adaptively. It examines the malicious activities on cloud infrastructure by analyzing the computational and storage data on the multiple VMs. The FCM clustering model observing the operations on the VMs and Fuzzy clustering process assigns the fuzzy membership functions and clusters the malicious activity of data to the corresponding cluster group. The fuzzification process outcomes the degree of fuzziness among the malicious data points. The M-FCM clustering algorithm segments a finite set of 'n' malicious data into the number of groups based on the fuzzy membership values. The INSPECT approach initially assigns the number of clusters 'C' based on the malicious data in the cloud. The traditional FCM algorithm randomly selects the initial centroid points, which creates the numerous iteration steps (r) and deprives to provide the accurate results. In contrast, the INSPECT approach selects the initial points in two levels: (i) selecting a set of random points ({P}) based on the input data, and (ii) finding the 'Min' and 'Max' as the two random points (rp_1, rp_2) from a set of random points ({P}). If the cloud data set comprises 'n' number of malicious data, the INSPECT approach randomly selects {P} by '$0.01*n^2/(n-1)$'. Consequently, it sorts the data points in ascending order and selects 'rp_1', and 'rp_2'. To avert the local convergence of data points in the cluster, the INSPECT approach determines the probability of each data point (x_i) in Representative Point (RP) selection using Equation (1).

$$\Pr(rp)_{x_i} = \frac{\mathrm{Dist}(x_i, rp^{(p)})}{\sum_{rp_i=1}^{k} \mathrm{Dist}(x_i, rp_i)} \quad (1)$$

Where, Dist(x_i, RP) refers to the distance between the data point (x_i) and representative $rp^{(p)}$ which is in proximity to x_i in RP. 'rp_i' is the representative point in RP, wherein 'i' varies from '1' to 'k', and 'k' denotes the total number of representative points in RP at that time, k Є C. Equation (1) reveals that the INSPECT approach adds the x_i to the RP list until k=C, when the distance between x_i and preceding rp_i is large. Accordingly, the INSPECT approach assigns the threshold value (α) based on the input data points to select x_i which has higher probability value than 'α' controlling the spreading factor. After finding the initial representative point, the INSPECT approach updates the membership matrix (a_{ij})

using equation (2).

$$a_{ij}=\frac{\left(\left|x_i-rp_j\right|\right)^{-2/(m-1)}}{\sum_{k=1}^{c}\left(\left|x_i-rp_k\right|\right)^{-2/(m-1)}} \quad (2)$$

Where, a_{ij} represents the membership of x_i in cluster j, xi refers the data affected by malicious activity, wherein $1\leq i\leq n$. $A = \{\mu_{ij}\}_{i,j=1}^{nk}$ is the matrix of membership degrees. By exploiting equation (2) and (3), the INSPECT approach identifies the secondary representative points and updates the matrix of membership degree respectively, until $|A^{r+1} - A^r|$ is less than the default value 1e-5, which is similar to the traditional FCM clustering method. This default value denotes the minimum improvement between the two values. If $|A^{r+1} - A^r|$ is greater than the default value, the INSPECT approach employs equation (1) to identify the representative points and sequentially updates the matrix of membership degree. In equation (3), 'μ' is the membership degree and m Є [1, ∞] is a weighting factor of each fuzzy membership function, and 'k' is the number of clusters, $1\leq k\leq C$.

$$rp_j=\frac{\sum_{i=1}^{n}\mu_{ij}^{m}x_i}{\sum_{i=1}^{n}\mu_{ij}^{m}} \quad (3)$$

Where, j=1,2,…,k

Finally, the INSPECT approach determines a set of representative points that are equal to the number of clusters. Accordingly, it aggregates the malicious data in different separate groups, wherein the aggregated cloud data in each cluster are affected by the identical malicious activity. It facilitates the process of identifying the source of attacks by the forensic investigator in the cloud infrastructure instead of separately examining each data residing in the cloud.

A.II. Determining the source of attacks using SLAs

With the aim of identifying the source of attacks, the INSPECT approach examines the SLAs negotiated between the CSP and the users, wherein the malicious activity suffers users. SLA information provides the notion of determining the malicious individual according to analyzing the possibility of launching the malicious activity. It includes security specifications, uptime statistics, and compliance. Moreover, it consists of the advance schedule for notification of network changes, data ownership, rights information, and system infrastructure. The INSPECT approach takes into the account of SLA concerns and leads to the identification of the causal relationships between a specific type of malicious activity, attack location, and attack launcher in that attack launcher is either a malicious insider, cloud user, VM, or third party.

The INSPECT approach explores the associated SLAs of each cloud user ID to determine the probability of attack launchers according to the scenario of a specific malicious cloud data and feasibility of inter-component at that location. Inter-component refers to the component that performs the intervention, wherein the component implies the Cloud User (CU), VM, CSP, and Third Party (TP). The intervention of the component varies according to the access policies in the cloud infrastructure. The INSPECT approach remarkably determines the component as a real suspect for a specific malicious data by sequentially analyzing the component behavior in the cloud environment. The hypervisor logs accomplish the sequential log matching of the components.

For example, each Cluster (CL) comprises a set of malicious data with a unique session ID and user ID, CL = {A, B, C}. According to the requirement of each task, the system assigns the SLAs for each user, reflecting each request associated with an inter-component. Consider, the SLAs of malicious data, CL = {A(CU, VM3, CSP), B(CU, VM2), C(CU, VM3, CSP, TP)}. After clustering the malicious data, the INSPECT approach measures the correlation among the data in each cluster based on the CU, VM, CSP, and TP. By utilizing the SLA information, the INSPECT approach obtains the coarse-grained results of the suspects such as CU and VM. To further identify the fine-grained results, the INSPECT approach focuses on the timestamp information of each attack and the type of user of A, B, and C. As a consequence, it facilitates the determination of the intervention of components in all the malicious data in each cluster by sequentially matching this information with the hypervisor logs. Finally, the INSPECT approach determines the feasibility of suspects such as VM for a set of malicious data A, B, and C.

B. Isolating and presenting the evidence

After determining the evidence as well as the source of attacks, the INSPECT approach targets to isolate the evidence from the cloud storage to ease the forensic investigation. The evidence isolation is an integral part of the forensic investigation process and prevents the evidence from further tampering. If there is any loss in continuity and tampering with the evidence, the gathered evidence becomes ineffective in the admissibility level. Hence, the INSPECT employs the isolation method to protect the admissibility of the evidence. The advantage of cloud instances, such as migration in the cloud environment facilitates the evidence isolation during a forensic investigation.

The consideration of SLAs is essential while isolating the cloud instances that comprise the malicious data. In SLAs, the cloud service provider has the responsibility to maintain the Confidentiality, Integrity, and Availability (CIA) of the cloud instances, and the cloud user has the responsibility to protect the CIA in terms of the content of files. To obtain CIA of the cloud instances, the INSPECT approach employs the sandbox techniques that segregate the evidence without hindering other tenants in the multi-tenant cloud environment.

B.I. Evidence Isolation, Preservation, and Presentation

The INSPECT approach employs the Secure SHell (SSH) to isolate and preserve the evidence during a forensic investigation in the cloud environment. It restricts the actions of the malicious process during isolation based on the security policy and ensures the realistic forensic investigation in an isolated environment. It facilitates the evidence isolation and leads the proactive measures against the future malicious attacks on the collected evidence. The forensic investigator needs to present the evidence that is gathered in the acquisition phase to the court of law in the form of a chain of custody. The presentation is a process of submitting the organized report regarding the conducted actions of the criminal investigation, such as evidence collection, analysis, and preservation. To effectively maintain the chain of custody log, the INSPECT approach employs the Advance Forensic Format (AFF). AFF is an on-disk format that stores the forensic information in which files are to be digitally signed to ensure the long-term integrity and provide the chain of custody. The INSPECT approach converts the disk image into an AFF file in which disk image stores the entire investigation process. The converted AFF file comprises the data in the form of encrypted data from entire disk, wherein encryption is performed by the Secured Hash Algorithm 1 (SHA-1). The INSPECT approach retains and preserves the data regarding the user, evidence, access, and history changes by exploiting the AFF, which builds a better chain of custody during cloud forensic investigation. The INSPECT approach submits the chain of custody information with the support of AFF to the law enforcement system. To validate the acquired evidence by the INSPECT approach, the jury employs the corresponding VM snapshots of the crime scene through event reconstruction and also reconstructing the AFF. AFF facilitates the evidence validation process by the jury, wherein the reconstructed AFF reveals that the entire cloud forensic investigation process performed by the INSPECT. Finally, the jury justifies the evidence with the source of attacks with the knowledge of CSP and consequently, adjudges the real suspect of the corresponding crime scene.

IV. Experimental Evaluation

This section describes the experimental evaluation of M-FCM and evaluation of a prototype implementation of the proposed INSPECT model in which the evaluation is done in the OpenNebula cloud management architecture.

A. Experimental setup

Initially, the experiment exploits the WEKA tool to show the performance improvement of the M-FCM clustering algorithm that is utilized by the INSPECT. The INSPECT approach runs its prototype model in OpenNebula that is an open source software offering the elastic and the most feature-rich solution for the comprehensive management of virtualized data centers. OpenNebula provides an Infrastructure-as-a-Service (IaaS) and manages the heterogeneously distributed data centers. It can able to work with the Kernel-based Virtual Machine (KVM), Xen, and VMware virtualization facilitating the accommodation of multiple hardware and software. It supports the different VMMs and provides the dynamic services to the end-users by enabling faster delivery, easy scale-up, and scale-down services in IaaS. The interaction between the cloud user and OpenNebula is based on the OpenNebula Command Line Interface (CLI). To perform the further process of evidence acquisition and source of attack identification, the INSPECT approach applies the M-FCM clustering algorithm.

A.I. Forensic Data

The evaluation of the INSPECT approach employs the cloud data that include the data affected by two different types of attacks such as Date spoofing and Sender spoofing attacks on Email application.

Consider the attack launched VM runs the task as the email application in the cloud environment. The sample Email data is gathered from the EDRM Micro dataset [29]. Email applications often deal with various vulnerabilities such as junk mail, viruses, offensive text or pictures, unauthorized software, legal liabilities, forged messages, date or time spoofing, unauthorized disclosure of sensitive information, and so on. Most notably, email date and address spoofing are two prime forms of Email spoofing. Accordingly, the INSPECT experiment validates the forensic investigation process on the Email header data. Hence, this experiment extracts the email application log files such as Email header information. The data type in the cloud email log instances is divided into normal data and forensic data that are based on the fields that are required to perform the investigation. From the email header information, all the fields are taken as the input for clustering method to cluster the similar data instances. In the case of a forensic investigation, several specific fields are enough to cluster the malicious data instances. Several specific forensics fields include Recipient mail ID, Recipient IP address, Receiving data, Return path, Sender IP address, Received-Sender Policy Framework (SPF) domain, Sender mail ID, and Sending data. Then, the date spoofing and sender spoofing attacks are created on the Email header data.

A.II. Evaluation metrics - Clustering algorithm

Precision: It is the ratio between the number of accurately clustered malicious data and the total number of clustered malicious data.

Recall: It is the ratio between the number of accurately clustered malicious data and the total number of relevant malicious data.

F-measure: It is the harmonic mean of precision and recall.

Clustering accuracy: It is the percentage of accurately clustered data, including normal as well as malicious data.

A.III. Evaluation metrics - Investigation method

Investigation time: It is the total time spent by the investigator to collect the evidential artifacts from the cloud.

Investigation Accuracy: It is the percentage of evidence collection accuracy on a malicious user at a specific time, which relies on the mapping of crime event with the acquired evidence.

B. Analysis of clustering algorithms

To assess the performance of the Clustering algorithm for cloud forensic data, the INSPECT approach implements its M-FCM clustering algorithm with the baseline clustering algorithms on the cloud data. By exploiting WEKA tool, the performance of the M-FCM algorithm is compared with several conventional clustering algorithms such as FCM, K-means, Canopy, and Expectation Maximization (EM) clustering algorithms. The significant performance improvement of the proposed clustering algorithm is illustrated through the following metrics.

B.I. Data type Vs. Precision

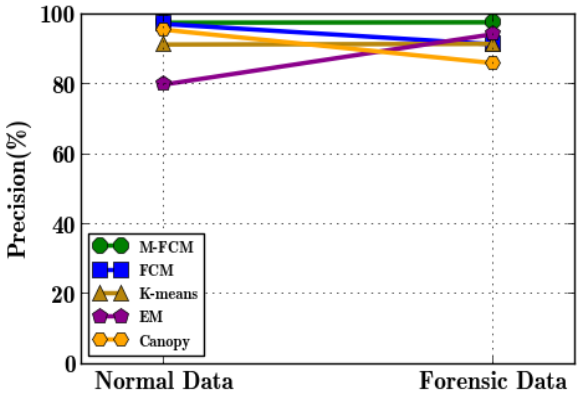

Fig.2. Data type Vs. Precision

Fig. 2 illustrates the comparison of precision of the M-FCM clustering algorithm with the baseline clustering algorithms such as FCM, K-means, EM, and Canopy clustering algorithms. It presents the percentage of precision value for the variation of data types such as normal data and forensic data, wherein the normal and forensic data differs from the selected fields referring attributes. The M-FCM clustering algorithm yields fair performance for forensic data than the normal data by 0.1% and also than the traditional FCM clustering algorithm by 6.5% respectively. By clustering the forensic data with the forensic fields alone and contextually initializing the random points, the performance of the M-FCM gets the higher precision value of forensic data. For the forensic data, the precision of K-means and EM algorithms suddenly escalate the precision than FCM clustering algorithm due to the evidence convergence in the random point selection of FCM clustering.

B.II. Data type Vs. Recall

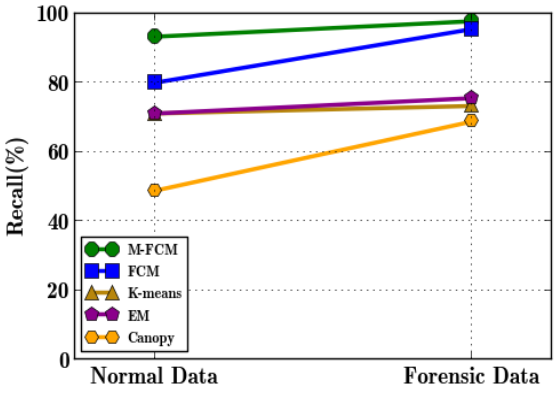

Fig.3. Data type Vs. Recall

The performance improvement of the M-FCM algorithm is shown in Fig. 3 for different data types such as normal data and forensic data. The recall of the M-FCM algorithm substantially increases by 4.47% with the variation of data type from normal data to forensic data. It reveals that both M-FCM and FCM clustering algorithm effectively supports the uncertain forensic data even when selecting the restricted attributes for clustering. Though, the modified FCM obtains higher recall value by 2.2% than the traditional FCM clustering algorithm. It is because of the input data based initial clustering centroid point selection and membership value updating in the M-FCM. Moreover, other baseline clustering algorithms such as K-means, EM, and Canopy accomplishes the recall value around 73%. It is because, even though K-means and EM algorithms provide significant results, these algorithms lack to ensure the quality of the clustering results.

B.III. Data type Vs. F-measure

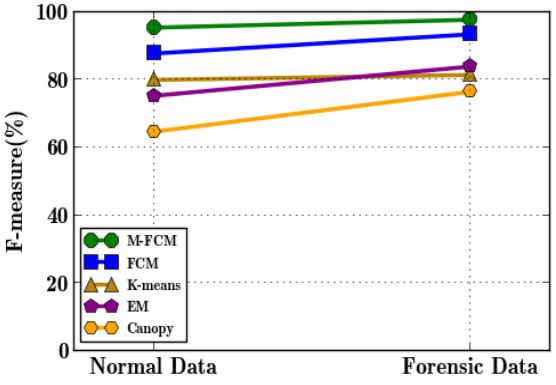

Fig.4. Data type Vs. F-measure

Fig. 4 demonstrates the F-measure value of the proposed M-FCM and several existing clustering algorithms. The results presented in Fig. 4 shows that the M-FCM algorithm is evaluated with the four different clustering algorithms when testing the data type belongs to the same set of normal and forensic fields. The FCM clustering algorithm achieves 11.99% higher performance than the K-means clustering algorithm due to the improvement in the data partition. Accordingly, the M-FCM clustering algorithm obtains 4.3% and 16.29%

higher F-measure value than the FCM and K-means clustering algorithms respectively. The proposed initial centroid point selection tends to improve the effectiveness as well as the efficiency of the forensic data clustering. In essence, it leverages the M-FCM clustering algorithm with high accuracy and less number of iterations during clustering. Moreover, the INSPECT approach analyzes the SLA of the cloud users based on the cluster results, which tends to improve the investigation accuracy significantly.

B.IV. Data type Vs. Clustering accuracy

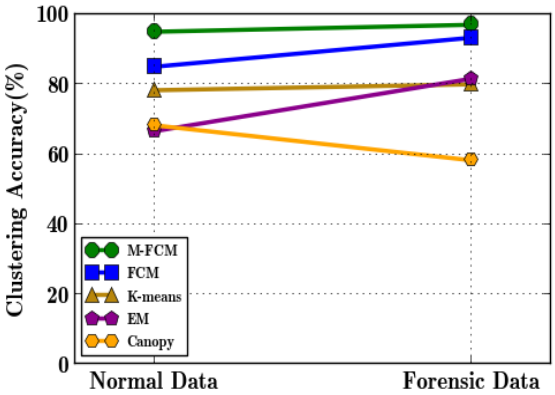

Fig.5. Data type Vs. Clustering accuracy

The comparative results of the proposed M-FCM clustering and the previous clustering algorithms are shown in Fig. 5. It illustrates the percentage of the clustering accuracy while varying two different data types such as normal and forensic data with the various attributes. Even though the normal data have a high number of attributes than the forensic data, it tends to inaccurate data clustering due to the deviation from the context and also its faster convergence. Whereas in forensic data, the optimal fields are taken as the forensic fields that have a high level of uncertainties and thus, the dynamic partition and membership value updating in the FCM clustering leads to improve the clustering accuracy by 8.3%. Moreover, M-FCM clustering algorithm further improves the clustering accuracy at 97% by averting faster data convergence, increases the distance between the cluster centroid points. Due to the hard partitioning of the data in K-means clustering, the FCM clustering achieves high clustering accuracy for both the normal and forensic data.

C. Discussion of results

In the real cloud environment, evaluating the performance of the INSPECT approach is paramount through analyzing the benefits of the cloud provider, forensic investigator, and cloud users. To validate the significant performance improvement of the INSPECT approach, the evaluation system discusses the investigation time, data recovery efficiency, and the investigation accuracy. The INSPECT approach improves the performance concerning scalability because the forensic investigation does not rely on the number of resources and requests on the remote server. It considers only the bit-level information about VM logs for forensic investigation rather than exploring the entire cloud storage. Thus, the INSPECT approach does not add up the burden to the cloud service provider regarding the scalability.

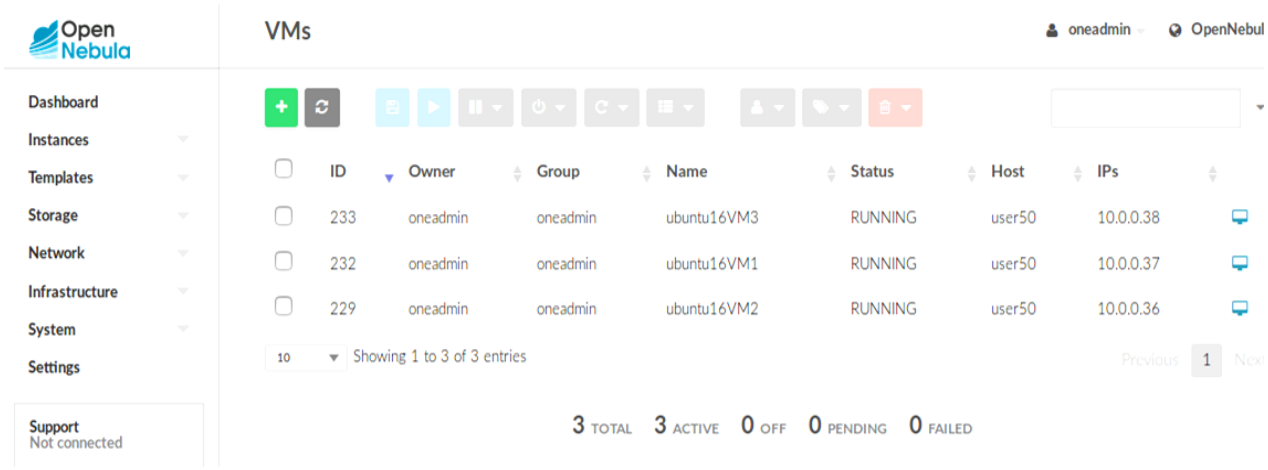

Fig.6. OpenNebula Cloud Infrastructure

The proposed INSPECT methodology eases the forensic investigation by exploiting the bit-level information about the crime scene. It tends to lead the investigator to complete the investigation in a reasonable time. It improves the investigation accuracy due to the consideration of the M-FCM clustering and SLAs while acquiring the evidence. The analysis of clustering algorithms in Section IV.B reveals that the M-FCM clustering algorithm provides a fair performance than the other algorithms, hence, the INSPECT employs the M-FCM clustering algorithm to perform the cloud forensic investigation effectively. In essence, the INSPECT approach applies the M-FCM clustering algorithm on the gathered malicious data points to identify the source of the attack and to determine a specific evidence. Thus, an inaccurate investigation is averted.

The INSPECT approach provides specifically a couple of benefits to the cloud user. First, it improves the

feasibility to perform the cloud forensic investigation in its services even when the cloud user runs the service on a remote cloud. Second, the cloud user can allocate the resources to perform the forensic investigation. Thus, the INSPECT approach ensures the benefits to both the user and the provider while ensuring the data privacy.

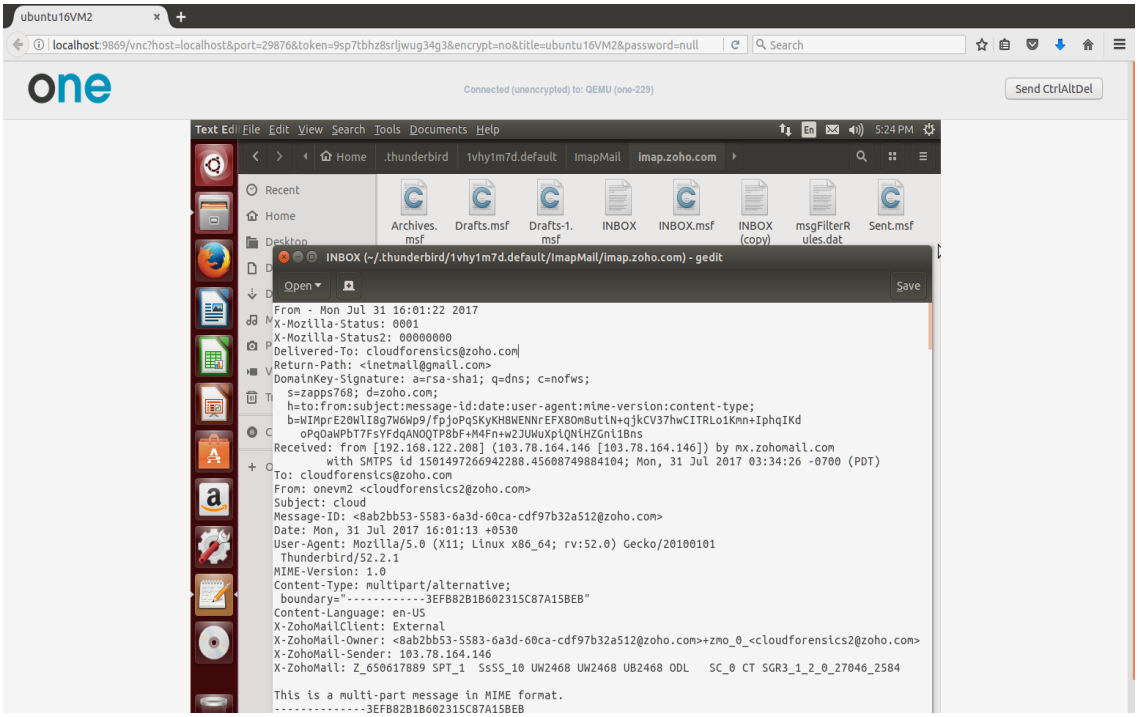

Fig.7. Email Log File Acquisition from Cloud

V. Case Study

The INSPECT approach selects and analyzes the IaaS scenario, and it runs the proposed algorithm with the prime objective of forensic investigation in the OpenNebula cloud middleware. Consider, a user rent a VM to run the email application. Assume, the intention of the attacker is to spoof the email through the cloud access. The cloud VM enables the users to perform multiple services such as to test new software, run a website, and run the applications. This case study assumes that there are two types of attacks such as sender spoofing and date spoofing. The overall process of the INSPECT approach is illustrated in the following steps.

Step 1: Initializing the INSPECT Forensic process

After determining the malicious activity in the host by IDS, the INSPECT approach initializes the forensic method in the cloud environment. In OpenNebula, with the help of the hypervisor, the investigation model receives the unique information of attack location from IDS and maps the received IP address or machine ID with the cloud storage to obtain the cloud log information on that particular machine. From the Figure 6, the OpenNebula has three suspected VMs such as VM ID: 229, VM ID: 232, and VM ID: 233 under the host of User-50 among the multiple VMs. Each VM runs multiple tasks by the multiple users in the cloud.

Step 2: Gathering the data regarding malicious activity

In the cloud infrastructure, if an attack is known, IDS can determine the malicious activity and consequently take the snapshots. Otherwise, the system utilizes the stored snapshots of the email application execution on the suspected VM to recognize the malicious activity residing location and to perform the post forensic investigation after receiving the command of initiating the forensic process on a specific location. The OpenNebula enables the dynamic snapshot creation. This information facilitates the INSPECT approach to determine the evidential artifacts related to the malicious activity.

Step 3: Acquiring the log files of the email application

The INSPECT approach intends to capture the evidential artifacts from the suspected VMs that are located in the cloud data storage. It attempts to capture the cloud log files from the cloud with the authorization of the CSP. The suspected VM handles the email application by multiple cloud users. Consider the malicious activity in email application involves the Sender spoofing and Date spoofing. The forensic investigator obtains the corresponding email logs from the cloud storage, wherein email logs in the form of header information which is used to perform header forensic investigations. Fig. 7 shows the email header logs in the cloud storage.

Step 4: Forensic field selection and attack scenario analysis

To perform the header forensics on the email application logs, the INSPECT approach selects the forensic fields rather than exploiting the entire fields of the header data. Consequently, it facilitates the forensic investigation in terms of restricting the investigation within the significant information of malicious data and precisely determining the source of the attack through clustering method. To aggregate the similar log indexes or malicious data, the investigator employs the source information regarding the attack scenario and attack pattern. The email log indexes with a set of forensic fields are illustrated in Fig. 8.

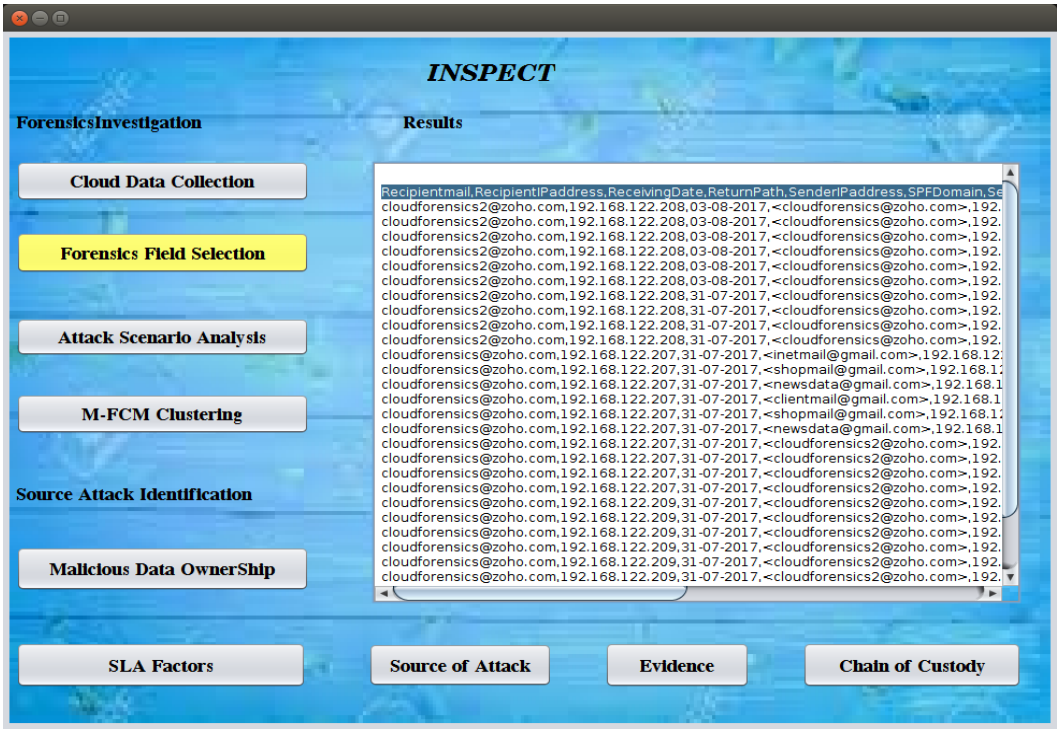

Fig.8. Forensic fields of the e-mail application

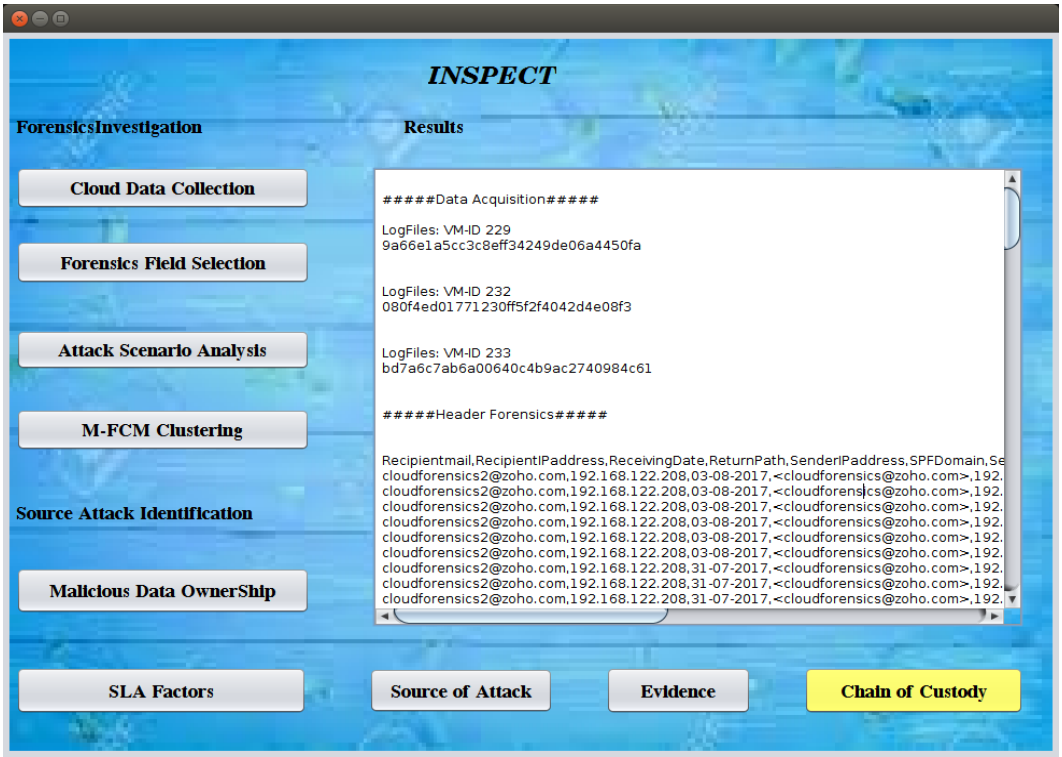

Fig.9. Chain of Custody of the INSPECT Process

Step 5: Adaptive forensic data acquisition using M-FCM

The INSPECT approach aggregates the similar malicious data in terms of numerical values by M-FCM clustering. The forensic data acquisition process includes a source of attack identification to maintain the chain of custody and determine a real suspect of the corresponding crime scene in the cloud. By utilizing the M-FCM clustering and SLA information with the assistance of the CSP, the INSPECT approach performs the forensic data acquisition from the suspected VMs that contain the evidence of the crime scene. The INSPECT employs the gathered email log files of the suspected VMs and generates 3 clusters based on the attack scenario using the M-FCM clustering. From the cluster results and SLA of each cloud user, the INSPECT approach explores the intervention components for each cluster. According to the probability of intervention components for the tasks involved in the clusters, several components have the high probability score to launch such a malicious activity. Thus, the source of the attack is determined for the specific attack types of data spoofing and sender spoofing.

Step 6: Isolation and Forensic investigation analysis

To tackle with the multi-tenancy, the INSPECT approach isolates the evidence and further investigates the evidential artifacts through the forensic investigator. It submits the evidence with the chain of custody information with the dynamic updating of the investigation process to improve the trustworthiness of the forensic investigation. Finally, the INSPECT approach leverages the investigator to present the evidence and source of the attack along with the chain of custody. The Chain of custody information of the INSPECT approach is presented in Fig. 9.

VI. Salient Features

- The INSPECT approach exploits the forensic fields alone from the cloud data, which reduces the computational complexity and averts the misclassification in the suspect identification.
- Moreover, its M-FCM clustering method avoids the random generation of centroid points by clustering the data points in a contextual manner, which improves the investigation accuracy.
- By analyzing the SLA of the Cloud user who belongs to the clustered data, the INSPECT approach improves the correctness of the source of attack identification and reduces the investigation time.

AFF based chain of custody submission ensures the trustworthiness of the forensic investigation in the cloud environment.

VII. Conclusion

This paper has presented an intelligent and reliable forensic investigation model for the multi-tenant cloud environment. In the proposed model, the INSPECT approach employs the VM logs and M-FCM clustering method to acquire the evidence from the large-scale cloud forensically. It assists to adaptively find the source of attacks from the clustered data generated by FCM with contextual initialization. The INSPECT approach accurately and quickly determines the appropriate evidence and source of attack by analyzing the SLA information of the clustered users, in which the CSP gives SLA. Then, it isolates the evidence to perform the uninterrupted and uncompromised forensic investigations in the cloud infrastructure. Finally, it preserves and presents the acquired evidence with the help of AFF comprising the information related to the forensic investigation, which maintains the chain of custody and facilitates the presentation process. The experimental evaluation demonstrates that the proposed INSPECT approach yields a fair and better investigation accuracy with great concern in the multi-tenant cloud infrastructure.

Even though the INSPECT approach reduces the computational complexity and improves the investigation accuracy, it does not focus on the distributed cloud environment during evidence acquisition, which may create multi-jurisdictional issues. In future, the INSPECT intends to explore the new ways of ensuring the privacy of data into the cloud forensic investigation model over the distributed cloud environment.

References

[1] Zhang, Q., Cheng, L., and Boutaba R, "Cloud computing: state-of-the-art and research challenges", Journal of internet services and applications, Vol.1, No.1, pp.7-18, 2010

[2] Mohit Agarwal, Gur Mauj Saran Srivastava, "Cloud Computing: A Paradigm shift in the way of Computing", I.J. Modern Education and Computer Science, Vol. 12, No.1, pp.38-48, 2017

[3] K. Higgings, "Dropbox, WordPress Used As Cloud Cover In New APT Attacks", Dark reading, 2013[Online]Available: http://www.darkreading.com/attacksbreaches/dropbox-wordpress-used-as-cloud-cover-innew-apt-attacks/d/d-id/1140098

[4] Taylor, M., Haggerty, J., Gresty, D., and Lamb, D, "Forensic investigation of cloud computing systems", Network Security, Vol.2011, No.3, pp.4-10, 2011

[5] Zawoad, S., and Hasan R, "Cloud forensics: a meta-study of challenges, approaches, and open problems", arXiv preprint arXiv:1302.6312, 2013

[6] Dykstra, J., and Sherman A. T, "Acquiring forensic evidence from infrastructure-as-a-service cloud computing: Exploring and evaluating tools, trust, and techniques", Digital Investigation, Vol.99, pp.S90-S98, 2012

[7] Suneja, S., Isci, C., de Lara, E., and Bala V, "Exploring vm introspection: Techniques and trade-offs", In ACM SIGPLAN Notices, Vol.50, No.7, pp.133-146, 2015
[8] Ruan, K., Carthy, J., and Kechadi, T, "Survey on cloud forensics and critical criteria for cloud forensic capability: A preliminary analysis", In Proceedings of the Conference on Digital Forensics, Security and Law, Association of Digital Forensics, Security and Law, p.55, 2011
[9] Ruan, K., James, J., Carthy, J., and Kechadi T, "Key terms for service level agreements to support cloud forensics", Springer, In IFIP International Conference on Digital Forensics, pp.201-212, 2012
[10] Ahmed Fahim, "A Clustering algorithm based on local density of points", I.J. Modern Education and Computer Science, Vol.12, pp.9-16, 2017
[11] Pichan, A., Lazarescu, M., and Soh S. T, "Cloud forensics: Technical challenges, solutions and comparative analysis", Digital Investigation, Vol.13, pp.38-57, 2015
[12] Alqahtany, S., Clarke, N., Furnell, S., and Reich C, "Cloud forensics: a review of challenges, solutions and open problems", IEEE International Conference on Cloud Computing (ICCC), pp.1-9, 2015.
[13] Umamaheswari, K. and Sujatha, S. "Impregnable defence architecture using dynamic correlation-based graded intrusion detection system for cloud", Defence Science Journal, Vol.67, No.6, pp.645-653, 2017.
[14] F. Xinwen, L. Zhen, Y. Wei, and L. Junzhou, "Cyber Crime Scene Investigations (C2SI) through Cloud Computing", IEEE 30th International Conference on Distributed Computing Systems Workshops (ICDCSW), pp.26-31, 2010
[15] Hay, B., and Nance K, "Forensics examination of volatile system data using virtual introspection", ACM SIGOPS Operating Systems Review, Vol.42, No.3, pp.74-82, 2008
[16] Thorpe, S, and Ray I, "File timestamps for digital cloud investigations", Journal of Information Assurance and Security, Vol.6, No.6, 2011
[17] Zawoad, S., and Hasan, R, "Chronos: Towards Securing System Time in the Cloud for Reliable Forensics Investigation", IEEE 40th Annual In Computer Software and Applications Conference (COMPSAC), Vol.1, pp.423-432, 2016
[18] Thorpe, S., and Ray I, "Detecting temporal inconsistency in virtual machine activity timelines", Proceedings of Journal of Information Assurance and Security (JIAS), Vol.7, No.1, 2012
[19] Wook Baek H, Srivastava A, and Van der Merwe J, "Cloudvmi: Virtual machine introspection as a cloud service", IEEE International Conference on Cloud Engineering (IC2E), pp.153-158, 2014
[20] Hirwani M, Pan Y, Stackpole B, and Johnson D, "Forensic acquisition and analysis of vmware virtual hard disks", Proceedings of the International Conference on Security and Management (SAM), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2012
[21] Rani DR, and Geethakumari G, "An efficient approach to forensic investigation in cloud using VM snapshots", IEEE International Conference on Pervasive Computing (ICPC), pp.1-5, 2015
[22] Zhou G, Cao Q, and Mai Y, "Forensic analysis using migration in cloud computing environment", Information and Management Engineering, pp.417-423, 2011
[23] Delport W, Köhn M, and Olivier MS, "Isolating a cloud instance for a digital forensic investigation", In ISSA, 2011
[24] Delport, W., and Olivier, M, "Isolating instances in cloud forensics", Springer, In IFIP International Conference on Digital Forensics, pp.187-200, 2012
[25] Belorkar, A., and Geethakumari G, "Regeneration of events using system snapshots for cloud forensic analysis", In India Conference (INDICON), Annual IEEE, pp.1-4, 2011
[26] Martini, B., and Choo, K. K. R, "An integrated conceptual digital forensic framework for cloud computing", Digital Investigation, Vol.9, No.2, pp.71-80, 2012
[27] Dykstra, J., and Sherman A. T, "Design and implementation of FROST: Digital forensic tools for the OpenStack cloud computing platform", Digital Investigation, Vol.10, pp.S87-S95, 2013
[28] Pasquale, L., Hanvey, S., Mcgloin, M., and Nuseibeh B, "Adaptive evidence collection in the cloud using attack scenarios", Computers and Security, Vol.59, pp.236-254, 2016
[29] http://www.edrm.net/resources/data-sets/edrm-micro-datasets/

Authors' Profiles

Mrs K. Umamaheswari obtained her Master's degree from Bharathidasan University, Tiruchirappalli, in 2004. She is currently pursuing her PhD at Research and Development Centre, Bharathiar Univerisity, Coimbatore, India. Her areas of research include: Cloud security, virtualisation, machine learning, and data mining.

Dr S. Sujatha received her MSc (Computer Science) from Anna University, Chennai, in 2002. She obtained her PhD from Department of Mathematics, Anna University, Chennai, in 2009. Currently working as an Assistant Professor in Bharathi Women's College(A), Chennai, Tamil Nadu, India. Her current area of interest includes: Information and network security, cryptography, MANETs, soft computing and cloud computing.

How to cite this paper: K. Umamaheswari, S. Sujatha, " INSPECT- An Intelligent and Reliable Forensic Investigation through Virtual Machine Snapshots", International Journal of Modern Education and Computer Science(IJMECS), Vol.10, No.3, pp. 17-28, 2018.DOI: 10.5815/ijmecs.2018.03.03

