ISSN 2278-3091

Volume 8, No.1.3, 2019

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse7081.32019.pdf

https://doi.org/10.30534/ijatcse/2019/7081.32019

Data Leakage Detection in Cloud Computing Platform

Muhammad Azizi Mohd Ariffin[1], Khadijah Ab Rahman[2], Mohamed Yusof Darus[3] , Norkhushaini Awang[4] , Zolidah Kasiran[5]

[1]Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Shah Alam, Malaysia. mazizi@tmsk.uitm.edu.my
[2]Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Shah Alam, Malaysia. khadijah.ar@gmail.com
[3]Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Shah Alam, Malaysia. yusof@tmsk.uitm.edu.my
[4]Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Shah Alam, Malaysia. shaini@tmsk.uitm.edu.my
[5]Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Shah Alam, Malaysia. zolidah@tmsk.uitm.edu.my

ABSTRACT

The popularity of Cloud Computing technology has made it a norm for IT deployment in enterprise, education and government sectors. But technologies in the cloud such as hypervisor or web-based dashboard have vulnerabilities which can potential cause data leakage. The impact of data leakage is huge, data leak incident at a firm such as Exactis cause 340 million of customer record being exposed. Moreover, the incident lead to financial loss, reputational damage, loss of customer trust and compliance issue to the firm. Therefore, there is a need to address the threat of data leakage in cloud computing platform. This paper presents the topic of data leakage detection in cloud computing platform. First we will discuss about the threat of data leakage during VM migration process and web dashboard authentication in cloud computing platform. To detect the data leakage, this paper proposes a method which involves performing packet capture on the platform. To demonstrate the method, this paper will simulate the threats and verify the data leakage. The results show that data leakage can be detected and verified effectively using the method when cloud management traffic is not encrypted.

Key words: Data Leakage, Cloud Computing, Cloud Security, Information Security

1. INTRODUCTION

Nowadays, utilizing the cloud computing technology has become the norm in enterprise, education and government sectors [1]. This is because cloud computing offers more agile, flexible operation and has on-demand self-service features when compared with conventional server platform. Even though cloud computing offers many advantages, it has vulnerabilities which could lead to security issues. The very core technology in cloud computing is virtualization, because its liberate applications, servers, storage and desktops from becoming dependent to physical hardware layers, by abstracting resources in isolated virtual computing environments. In cloud platform, the host running on the platform is called as Virtual Machine (VM) and the common operations performed on VM are replication and migration. During those operations, the risk of data leakage may occur due to misconfiguration, software (e.g. hypervisor) bugs or poor management practice [2] [3] [4]. Moreover, the risk of data leakage could also occur during authentication in communication session when cloud users are accessing the cloud's self-care portal or dashboard [5]. Thus, the risk of data leakage could lead to security issues such as data breach on cloud platform which affect the confidentiality of data and legal compliance.

Several studies has been conducted on security threats and data leakage on the cloud [8] [9] [10] [11] [12] [19] [26] [27], but those papers did not investigate data leakage due to vulnerabilities of the hypervisor and dashboard of the cloud management software and demonstrate further a method to detect data leakage on cloud computing platform. The study of [12] [19] highlight the possibility data breaches or loss in cloud computing platform, but it did not demonstrate the mechanism that lead to the data breach and demonstrate a method to detect it. The work of [26] provides a survey on securing the cloud data under the condition where the secret key has been exposed. The work may able to secure the data on application layer but the work did not investigate and address data leakage due to vulnerabilities of hypervisor and cloud platform software. Moreover, the work of [27] regarding security protocol for data transmission in the cloud is more directed towards securing application while data is in motion, it does not fully address the data leakage issue originating due to cloud platform software. Thus there need a work to investigate and address the data leakage threat due to vulnerabilities of the hypervisor and the cloud platform software.

The security impact of data leakage on cloud computing platform is big, for example the data breach incidents which occur on Exactis [6] a cloud based data brokerage firm is exposing 340 million record of its customer to public internet. The personal data which has been leaked during the incident can be exploited for malicious intent and broke the trust of its customer. Data breaches incident also have lasting impact to the cloud computing security as it can cause massive financial and reputation loss to the organization [7]. Therefore, there is a need to address the risk or threat by detecting the data leakage on the cloud-computing platform.

This paper presents the topic of data leakage detection in cloud computing platform. First we will discuss about the threat of data leakage during VM migration process and web dashboard authentication in cloud computing platform. To detect the data leakage, this paper proposes a method which we perform packet capture during VM migration and dashboard authentication process. While the cloud process was running, different types of management and communication packet will be scrutinized and analyzed in order to detect data leakages. To demonstrate the method, this paper will perform experiment to simulate the threats and perform analysis to verify the data leakage. For the demonstration, this paper will develop a cloud testbed environment built on VMware VCenter and OpenStack software. The experiment result will verify the effectiveness of the method in detecting data leakage occurrence.

2. DATA LEAKAGE IN CLOUD COMPUTING PLATFORM

2.1 Cloud Computing Platform and Virtualization

Cloud computing has introduce new paradigm in computing and data centre deployment [13], and hosted services on the Internet are often associated with wide-ranging term of cloud computing [14] [27]. Generally cloud services has application programming interface (API) based on SOAP or REST standard for easy interaction with other web application [23]. The ability to converts computes hardware into standard abstract of software copies have caused datacentre technologies to evolve from physical infrastructure to virtualized infrastructure. Downtime of datacentre infrastructure has also greatly reduced due to virtualization technology features which employ virtual machine replication, cloning and automated compute resources control when compared to physically duplicating server hardware for redundancy. Moreover, workloads balancing between physical hardware could easily be done via live virtual machine migration.

Virtualization is a disruptive technology which free up applications, servers, storage and desktops from dependent to physical hardware layers, by abstracting resources in isolated virtual computing environments. Different operating systems and application can run efficiently in parallel at the same host when the hypervisor layer is managing on top of hardware layer in the infrastructure layer. A complementing layer of cloud tools or software which provides automated management and self-service orchestration of the virtualization product were released later after years of virtualization technology has been rolled out. The cloud tools and software suite enhanced the virtualization platform offering by introducing scalability, elasticity, flexibility and pay-per-use charging mechanism which are not available before in IT and data centre deployment. To enhanced availability and protection in the cloud, methods to protect Virtual Machine (VM) such as VM replication was introduce, where a production VM is being copied to either same physical host or alternate host.

2.2 Data Protection in Cloud Computing

The computing resources such as CPU, memory and storage on the cloud are basically shared among multiple hosts or virtual machine [27]; therefore it is important to have a data protection scheme to protect data confidentiality, Table 1 below shows three types of data which need to be protected in the cloud deployment.

Table 1: Type of Data Need To Be Protected in the Cloud

Type	Risk Level	Description
Data in Transit	High	A condition where transferring data via wired or wireless connectivity through a network or Internet make it at risk
Data in Use	Medium	A condition where data is currently used, manipulated or being hold in memory or cache at endpoints of the network such as laptops, USB devices or mobile devices.
Data at Rest	Medium - Low	A condition where the data are stored in databases, storage devices or files systems for persistent storage.

Table 1 shows the type of data with its corresponding risk level of confidentiality breach. Data in transit has the highest risk level as it can be intercepted easily while it traverse on the network, thus encryption play a key role in securing the data [28]. The data in use has medium risk level as data stored in memory is volatile which make it harder to be intercepted. Lastly, the data at rest has medium to low risk level as hacker need to bypass security features which may present in the database system.

To protect virtual machines, there are several useful cases which show data protection is the best solution. As means of onsite backup for single site replication, we can replicate the virtual machine to different storage appliance within the same site. In the event where the original copy gets corrupted, this will ensure the easy recovery and availability of the virtual machine. Replicating the VM in the branch to HQ or any other site within the same hypervisor infra which has cluster configuration setup can also be another use case. The location of HQ can be in the same country of the branch site or even global or worldwide site. This will make a good case for offsite backup solution. And normally when data is stored (at rest) in networked environment it is crucial for enterprise or any organization to perform a threat assessment (e.g. MyRAM, HiLRA method) in order to determine the risk of storing the data in such environment [24].

2.3 Data Leakage Threat

Unintentional or accidental distribution of sensitive or confidential data to an unauthorized entity is the definition of data leakage [10]. Unauthorized transfer of classified information from an organization to an outside recipient was also associated with data leakage. By transmitting data processes, such as website forms, emails, instant messaging and file transfer which are unmonitored and unregulated to their destination could make data leakage issue become intensifies. According to the perspective of other researcher, undefined unsolicited revelations of information can also be defined as a data leakage [9].

According to Cloud Security Alliance [15], the rate of data leakage event being reported in cloud platforms are increasing. In the reported threats, hardware failures, insecure interfaces and APIs, data loss and leakage was shown as three most frequent events. More than half of the reported cloud incidents were contributed by those threats. While the threat of internal malicious activity, shared technology issues, abuse and criminal use, account or service hijacking and unknown risk profiles make up the rest of the threats in cloud computing. According to a survey in 2018 [16], there is increasing number of data breaches by industry which highlight the importance of addressing data leakage on the cloud. Figure 1 shows the result of the survey

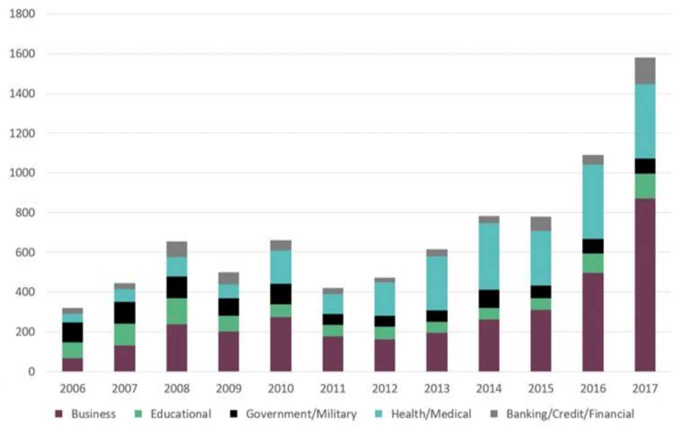


Figure 1: Number of Data Breaches by Industry [16]

Figure 1 shows the number of data breaches is increasing for every year according to industry. The peak data breaches are recorded in year 2017 which business or enterprise industry experiences the highest number of breaches which is alarming.

2.4 Data Leakage during VM Migration and Replication

As has been highlighted earlier, virtualization is one of the key components in cloud technology. In cloud computing, the host running on the platform is a virtual machine (VM) and the task of managing (e.g. resource allocation) it is handled by software known as hypervisor. VM replication serves as one of the methods of VM protection, where a production VM is being copied to either same physical host or alternate host. The VM replication could also being backed-up to alternate site. This replication process is a common practice done by assistance data protection software or hypervisor. By doing this, the VM is protected and can be recovered at any time without needing onsite backup mechanism. Virtual Machine replication can be done in at least two methods. Firstly via snapshot replication where VM current state is copied to other site and it is based on Changed Block Tracking technology which only changed blocks states is copied. The second method is via storage replication method where the whole original VM image file is copied on the storage later to the other site.

Data Leakage issue is possible to happen while transporting critical organizational data from one place to another. The sensitive data can leak from virtual machines, which are collections processes running on a virtualization platform that is known as hypervisor. The processes have its own designated memory spaces that can be accessed directly by the hypervisor. Hence all confidential information stored in the memory spaces on the virtual machine can be inspected. A compromised or improperly configured hypervisor can possibly impose risk to every virtual machine created within [8]. The hypervisor acts as a single point of access to the virtualization environment, and hence be a single point of failure. Virtual machine replication process in cloud environment can possibly trigger data leakage occurrences, if not meticulously managed [2]. Research studies done on virtualization and cloud computing area by researchers declares that there are possibilities of data leakage issue in the process of cloning virtual machines [9]. The vulnerability which triggered from the usage of cloning while delivering on-demand service is the potential data leakage resulted from virtual machine replication [4]. This paper will investigate whether there is any data leakage during VM replication process.

2.5 Data Leakage during Cloud Dashboard Authentication

Cloud computing platform provides web dashboard as an easy-to-use interfaces, which enable cloud administrators, users, and customers to remotely manage the cloud platform and the VMs. Before users able to access the dashboard, they need to authenticate via a login page. But many of web dashboard components provided by several cloud management softwares are vulnerable to web-based attacks, such as cross site scripting attacks (XSS) and a man-in-the-middle attack [5]. Moreover, there is a paper which reviews the security of OpenStack Horizon, and they identify that it has security weaknesses, such as not using SSL/TLS to encrypt the web traffic and having a weak password policy [17]. Thus, data leakage could potentially occur during the authentication process to access the dashboard and need to be investigated further.

2.6 Data Leakage Detection Method

To identify or verify whether data leakage occur in some of process or features of the cloud platform, we need a method to detect the data leakage. Few parameters need to be applied as data leakage detection measurement in comparing the data leakage model such as Traffic shape, regularity, distribution, data context and inter arrival time.

Data Leakage detection solutions depend on appropriate data classification [9]. If the data was not classified to different level and characteristics, the Data Leakage detection method will not be able to distinguish between confidential and normal traffic. Data classification is commonly used in military and government sector applications. The term used in military classification are such as ‘restricted’, ‘confidential’, ‘secret’ and ‘top secret’, which makes identification of confidential data easier. Academic data leakage detection/prevention solutions are presented as full study of particular concepts [9]. The academic DLP can be categorized into according to the application or the method. This paper will conduct a packet capture using a Wireshark as a method to detect data leakage based on packet parameters.

3. RESEARCH METHOD

This paper proposes a method to detect data leakage on cloud platform by performing packet capture during VM migration and dashboard authentication process. While the cloud process was running, different types of management and communication packet which being exchanged between cloud nodes and components will be scrutinized and analyzed in order to detect data leakages. To investigate and verify a data leakage on the cloud, this paper conduct two experiment which simulate VM replication and migration and Cloud user authentication processes. During Experiments, packet captured will be conducted using Wireshark tool while the experiment is running and packet analysis will be conducted on the packet dump. Data leakage can be detected and verified from the packet analysis. The following section will explain further regarding the experiments.

3.1 VM Migration Process Experiment

The first experiment is about simulating a VM replication and migration process in the cloud. The experiment will have two sites which both running VMware vCenter 6.5 as cloud management and vSphere ESXi 6.5 for hypervisor software. During the experiment the VM from site 1 will be replicated and migrated to site 2 by snapshot replication, Figure 2 shows the diagram of the sites. The VM is installed with Window Server 2008 operating system and will be running Active Directory (AD) services which will contain sensitive user’s information. Furthermore, the VM will also store sensitive information in a form of text file. The experiment will detect and verify whether the sensitive information from the AD and the text file will be leaked during the replication and migration process. The ESX host for site 1 is running on IBM x3100 M4 server with 3.10 GHz processor, 16GB DDR4 RAM and 500GB Hard disk. While, ESX host for site 2 is running on Acer Aspire E14 with 2.5 GHz processor, 16GB DDR4 RAM and 128GB Hard disk. The Wireshark host for performing a packet capture will be running on Asus A555L laptop with Window 10 OS. All of those hosts will be connected together via HP V1910-16G Ethernet switch, figure 3 show the network topology of the testbed.

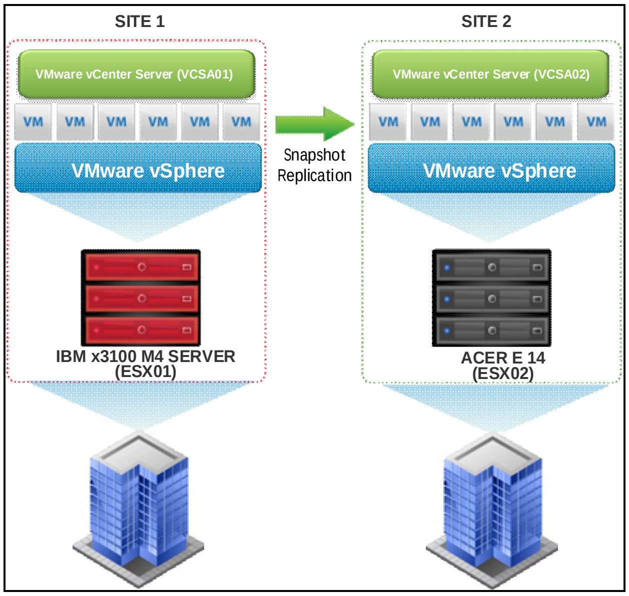


Figure 2: Diagram of Two Cloud Platform Sites

Figure 2 shows the logical diagram of the two cloud platform sites of the testbed. Site 1 physical host is IBM x3100 M4 server which running VMware vSphere software suite (hostname: ESX01), the site will host the VM running active directory (AD) application. Site 2 physical host is Acer E14 which will also running VMware software suite (hostname: ESX-02), the site will be accepting VM snapshot which will be migrated from site 1.

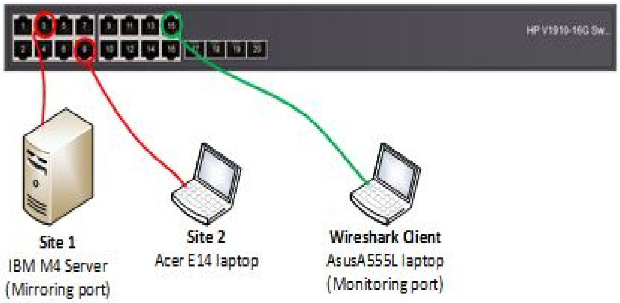


Figure 3: Testbed Network Topology

Figure 3 shows the testbed network topology. All physical host in the testbed will be connected together by HP V1910-16G Ethernet switch. One port will connected to site 1 and another port will be connected to site 2 host. Then another port which will mirror the port of site 1 and 2 will be connected to Asus 555L which is a Wireshark client host which the packet capture will be performed.

The first experiment will be conducted as in flowchart shown in figure 4 below. According to the method flowchart of figure 4, firstly we the setup Vmware vCenter server appliance and VSphere replication appliance to prepare the cloud infrastructure. Then we create and run the VM on the platform with AD services running. After that, we configure the port on the switch to mirror all traffic passing through ports which are connected to the ESX host. Then we configure the replication parameters such as destination site IP and start the packet capture on the Wireshark client. When everything is set, the VM replication will be initiated and its progress can be tracked on the cloud management console. When the replication and migration process has completed, we can stop the packet capture and perform packet analysis to detect for any data leakage.

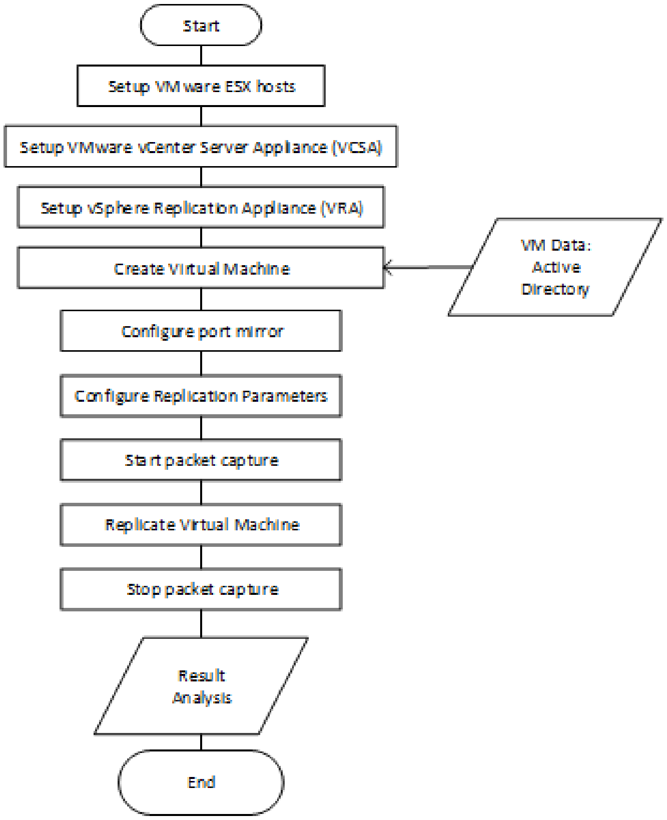

Start
Setup VMware ESX hosts
Setup VMware vCenter Server Appliance (VCSA)
Setup vSphere Replication Appliance (VRA)
Create Virtual Machine
VM Data: Active Directory
Configure port mirror
Configure Replication Parameters
Start packet capture
Replicate Virtual Machine
Stop packet capture
Result Analysis
End

Figure 4: Testbed Network Topology

3.2 Cloud Dashboard Authentication Experiment

The second experiment is about simulating an authentication process between cloud user and the web dashboard login page. In this experiment, the cloud platform is based on OpenStack (Mitaka Release) which is Open-Source and widely deployed cloud management software [20]. The OpenStack controller node has a web based dashboard which provides cloud users and admin an easy-to-use interface to manage virtual machine, monitor performance and to execute other management task. The web dashboard uses HTTP protocol to facilitate the communication between client and server. In this experiment, the OpenStack controller host is running on Dell Optiplex 990 workstation which has 3.40GHz processor, 16GB RAM and 1TB hard disk. Moreover, there are two hosts which will be used as a client and packet capture machine. All of the host in this experiment, will be connected together using an unmanaged Ethernet switch, figure 5 shows the testbed network topology.

Before user authentication process on the web dashboard is initiated, we will begin the packet capture. To start the authentication process, user will use web browser on client host and access the dashboard by entering controller node IP on the URL. Then user will need to enter their username and password in order to access the control panel, figure 6 shows the login page of OpenStack Web Dashboard. When the authentication process is complete, we will stop the packet capture and begin the analysis on the packet dump.

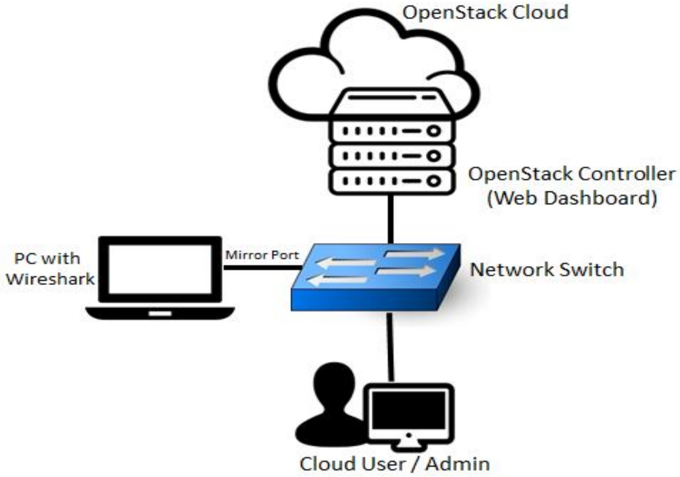
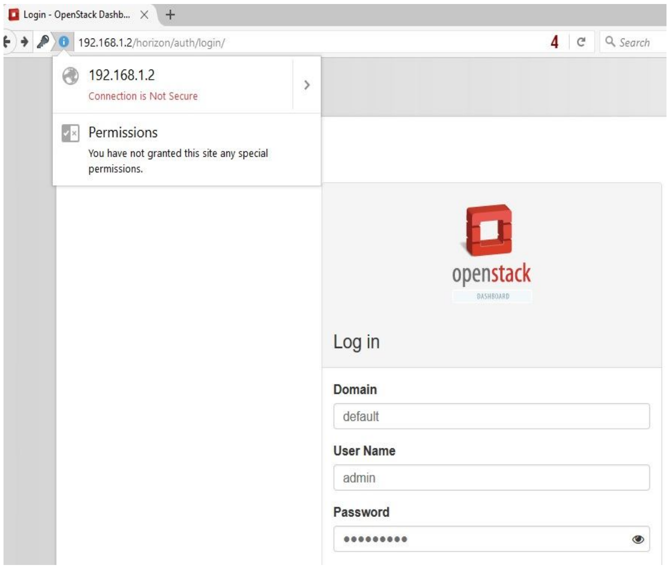


Figure 5: Second Experiment Testbed Network Topology

Figure 5 shows the testbed network topology used in the second experiment, all hosts will be connected via unmanaged Ethernet switch. One port will be connected to the physical machine which host the OpenStack controller which running the cloud dashboard while another port will be connected to a laptop which represent the cloud admin workstation. Another port which will mirror all port will be connected to a pc with Wireshark which we will perform a packet captured.

Figure 6 below shows the OpenStack web dashboard interface which used in the experiment. The web interface will ask users to enter their credentials before able to proceed to the cloud dashboard and control panel.



192.168.1.2/horizon/auth/login/

192.168.1.2
Connection is Not Secure

Permissions
You have not granted this site any special permissions.

Domain: default
User Name: admin
Password: •••••••••

Figure 6: OpenStack Web Dashboard

4. RESULTS AND ANALYSIS

In this section, we will explain the result of the experiment research and provide an analysis regarding data leakage detection.

4.1 VM Migration Experiment Result

The analysis on the packet capture is conducted based on detection parameters which are applied as display filters in Wireshark tool. The detection parameters are any username, password, or Active Directory data and other data strings. These parameters are monitored on packets sent out from VRA01 (Site 1) vSphere Replication Appliance IP address (192.168.0.21) using port 80, as VM replication blocks are sent to Site02 (192.168.0.198) from this appliance IP. The definition of sensitive data which will be flag in this research:

- Active Directory data such as username, password, Full Name and email address.
- Other possible strings of AD data

To perform effective analysis on the packet dump, four display filters were applied in Wireshark:

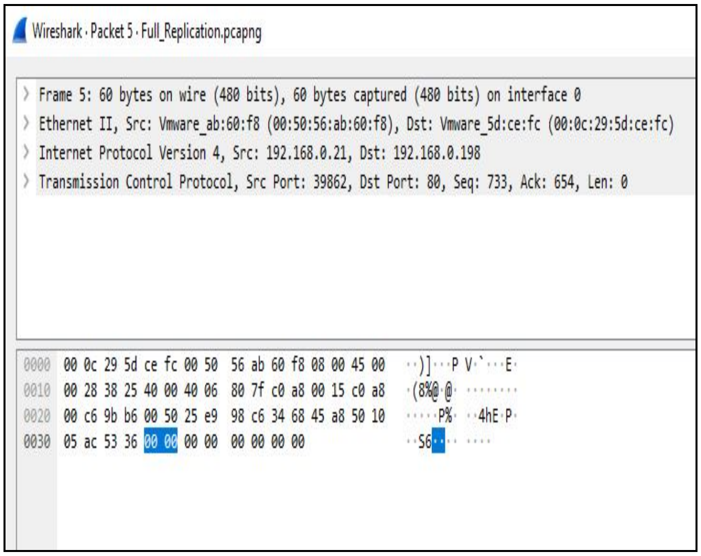
1. Display Filter 1: ip.addr == 192.168.0.21 and ip.dst == 192.168.0.198
2. Display Filter 2: ldap
3. Display Filter 3: Search String “password”
4. Display Filter 4: tcp contains user

After the filter has been applied, the result of the analysis is shown in Table 2. Figure 7 shows the Wireshark interface showing the packet capture after the filter has been applied.

Table 2: Analysis Result of First Experiment

No	Display filter applied	Filter result	Follow TCP stream result
1	ip.addr == 192.168.0.21 and ip.dst == 192.168.0.198	Lists applicable packets, but no sensitive data found	No sensitive data found
2	ldap	No packet applicable to this display filter	Not applicable
3	Search String “password”	No packet applicable to this display filter	Not applicable
4	tcp contains user	No sensitive data found	No sensitive data found

Table 2 shows the analysis result of the first experiment. There are 4 results when filter is being applied to the Wireshark. When filter based on IP address of VRA01 site 1 and site 2, no sensitive data was found on the packet. When filter based on LDAP (used by AD) protocol was applied, no packet was found. When filter based on “password” string and TCP packet was applied, we also did not found any sensitive data. Figure 7 below shows the Wireshark interface did not show any sensitive data when the filter has been applied.



```
> Frame 5: 60 bytes on wire (480 bits), 60 bytes captured (480 bits) on interface 0
> Ethernet II, Src: Vmware_ab:60:f8 (00:50:56:ab:60:f8), Dst: Vmware_5d:ce:fc (00:0c:29:5d:ce:fc)
> Internet Protocol Version 4, Src: 192.168.0.21, Dst: 192.168.0.198
> Transmission Control Protocol, Src Port: 39862, Dst Port: 80, Seq: 733, Ack: 654, Len: 0

0000  00 0c 29 5d ce fc 00 50  56 ab 60 f8 08 00 45 00   ··)]···P V·`···E·
0010  00 28 38 25 40 00 40 06  80 7f c0 a8 00 15 c0 a8   ·(8%@·@· ········
0020  00 c6 9b b6 00 50 25 e9  98 c6 34 68 45 a8 50 10   ·····P%· ··4hE·P·
0030  05 ac 53 36 00 00 00 00  00 00 00 00               ··S6···· ····
```

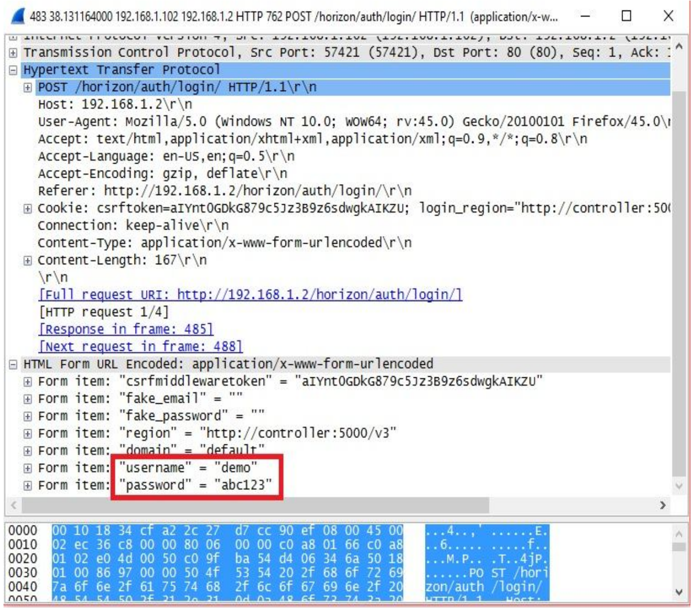
Figure 7: Wireshark Interface after filter has been applied

From the analysis conducted, we found out that there is no data leakage during the VM replication and migration processes using VMware vSphere ESXi hypervisor. This is because we did not find any sensitive information such as user’s information or password been exposed during the packet dump analysis. After going through documentation from Vmware [21], we found out that the migration process is encrypted and that may explain why we did not find any exposed sensitive data.

4.2 Cloud Dashboard Authentication Experiment Result

For this experiment, the analysis is also conducted based on detection parameters which are applied as display filters in Wireshark. The parameters are monitored on packets sent out from user web client (IP address: 192.168.1.102) to OpenStack controller (IP address 192.168.1.2) using port 80. The detection parameters included username, password, token string and other sensitive information, figure 8 below shows the packet capture result of when display filter is applied in Wireshark tool. To perform effective analysis on the packet dump, three display filters were applied:

1. Display Filter 1: ip.addr == 192.168.1.102 and ip.dst == 192.168.1.2
2. Display Filter 2: http
3. Display Filter 3: Search String “password”



```
483 38.131164000 192.168.1.102 192.168.1.2 HTTP 762 POST /horizon/auth/login/ HTTP/1.1  (application/x-w...
⊞ Internet Protocol Version 4, Src: 192.168.1.102 (192.168.1.102), Dst: 192.168.1.2 (192.1
⊞ Transmission Control Protocol, Src Port: 57421 (57421), Dst Port: 80 (80), Seq: 1, Ack: 1
⊟ Hypertext Transfer Protocol
  ⊞ POST /horizon/auth/login/ HTTP/1.1\r\n
    Host: 192.168.1.2\r\n
    User-Agent: Mozilla/5.0 (Windows NT 10.0; WOW64; rv:45.0) Gecko/20100101 Firefox/45.0\r
    Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8\r\n
    Accept-Language: en-US,en;q=0.5\r\n
    Accept-Encoding: gzip, deflate\r\n
    Referer: http://192.168.1.2/horizon/auth/login/\r\n
  ⊞ Cookie: csrftoken=aIYnt0GDkG879c5Jz3B9z6sdwgkAIKZU; login_region="http://controller:50
    Connection: keep-alive\r\n
    Content-Type: application/x-www-form-urlencoded\r\n
  ⊞ Content-Length: 167\r\n
    \r\n
    [Full request URI: http://192.168.1.2/horizon/auth/login/]
    [HTTP request 1/4]
    [Response in frame: 485]
    [Next request in frame: 488]
⊟ HTML Form URL Encoded: application/x-www-form-urlencoded
  ⊞ Form item: "csrfmiddlewaretoken" = "aIYnt0GDkG879c5Jz3B9z6sdwgkAIKZU"
  ⊞ Form item: "fake_email" = ""
  ⊞ Form item: "fake_password" = ""
  ⊞ Form item: "region" = "http://controller:5000/v3"
  ⊞ Form item: "domain" = "default"
  ⊞ Form item: "username" = "demo"
  ⊞ Form item: "password" = "abc123"

0000  00 10 18 34 cf a2 2c 27  d7 cc 90 ef 08 00 45 00   ...4..,' ......E.
0010  02 ec 36 c8 00 00 80 06  00 00 c0 a8 01 66 c0 a8   ..6..... .....f..
0020  01 02 e0 4d 00 50 c0 9f  ba 54 d4 06 34 6a 50 18   ...M.P.. .T..4jP.
0030  01 00 86 97 00 00 50 4f  53 54 20 2f 68 6f 72 69   ......PO ST /hori
0040  7a 6f 6e 2f 61 75 74 68  2f 6c 6f 67 69 6e 2f 20   zon/auth /login/
0050  48 54 54 50 2f 31 2e 31  0d 0a 48 6f 73 74 3a 20   HTTP/1.1 ..Host:
```

Figure 8: Wireshark Interface after filter has been applied

Figure 8 shows the Wireshark interface after the filter has been applied. From the figure we can clearly saw the username and password been captured by the Wireshark. It was capture from the HTTP packet captured during the communication session.

From the analysis of this experiment, there is a data leakage occurring during the authentication process on the web dashboard of OpenStack Mitaka Release. This is because, during our packet analysis we able to expose the username and password of cloud user in plaintext, figure 8 shows cloud user’s password is captured in plaintext. The sensitive information is exposed during the authentication process as the communication session between client and the OpenStack controller is using HTTP protocol which is not encrypted [22]. Even though the dashboard may have input filtering feature which may prevent a web based attack such as SQL injection [25], based on the experiment result it does not protect user credential data from eavesdropping.

This shows that method use in this paper can effectively detect data leakage which occurs on the cloud platform and its infrastructure. After data leakage has been detected, we can start to address the vulnerabilities of the cloud components (e.g. hypervisor, dashboard) which cause the data leakage.

5. CONCLUSION

This paper presents the topic of data leakage detection in Cloud Computing Platform. We propose a method which involves performing packet capture to detect data leakage which can potentially occur during VM replication and migration and also during web dashboard authentication process. The result of the experiment to detect data leakage during VM replication indicated that there is no data leakage occurring during the process as Vmware vSphere hypervisor encrypts the VM file before migrating it to other site. Moreover, the result of the second experiment to detect data leakage during web dashboard authentication shows that there is data leakage as the communication session was not encrypted in OpenStack software. These shows, the method can effectively detect data leakage on cloud platform. When compared with related work, other methods more focus on data leakage occurring on the application level. In the future, we want to expand the investigation by detecting potential data leakage on other cloud component such as block storage, APIs request and telemetry. We also want to investigate data leakage on other cloud platform or software such as Citrix, CloudStack and public cloud solution such as Amazon AWS and Google cloud.

ACKNOWLEDGEMENT

This research was supported by Industry, Community & Alumni Network Office of Fakulti Sains Komputer dan Matematik which provide us with a grant for the research. We also want to extend our gratitude to fellow colleague from Computer Technology and Networking department for comment and support needed for this research.

REFERENCES

1. Weins, K. (2016). Cloud Computing Trends: 2016 State of the Cloud Survey. [online] Rightscale.com. Available at: http://www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2016-state-cloud-survey https://doi.org/10.1109/MSP.2010.187
2. Harnik, D., Pinkas, B., & Shulman-Peleg, A. (2016). Side Channels in Cloud Services: Deduplication in Cloud Storage. In IEEE Security & Privacy, vol. 8, no. 6, pp. 40-47.

3. Annal Ezhil Selvi, S., & Anbuselvi, R. (2015). An Analysis of Data Replication Issues and Strategies on Cloud Storage System.
4. Grobauer, B., Walloschek, T., & Stocker, E. (2011). Understanding Cloud Computing Vulnerabilities. in IEEE Security & Privacy, vol. 9, no. 2, pp. 50-57, March-April. https://doi.org/10.1109/MSP.2010.115
5. Somorovsky, J., Heiderich, M., Jensen, M., Schwenk, J., Gruschka, N. and Lo Iacono, L. (2011). All your clouds are belong to us. Proceedings of the 3rd ACM workshop on Cloud computing security workshop - CCSW '11. https://doi.org/10.1145/2046660.2046664
6. Greenberg, A., Newman, L., Dreyfuss, E. and Palmer, A. (2018). Marketing Firm Leaked Database With 340 Million Records. [online] WIRED. Available at: https://www.wired.com/story/exactis-database-leak-340-million-records/ [Accessed 27 Mar. 2019].
7. Saed, K., Aziz, N., Ramadhani, A. and Hafizah Hassan, N. (2018). Data Governance Cloud Security Assessment at Data Center. 2018 4th International Conference on Computer and Information Sciences (ICCOINS). https://doi.org/10.1109/ICCOINS.2018.8510612
8. Kanoongo, B., Jagani, P., Mehta, P., & Kurup, L. (2014). Exposition of Solutions to Hypervisor Vulnerabilities. International Journal of Current Engineering Technology. 324444.
9. Alneyadi, S., Sithirasenan, E., & Muthukkumarasamy, V. (2016). A survey on data leakage prevention systems. Journal of Network and Computer Applications, vol. 62, issue C, pp. 137-152. https://doi.org/10.1016/j.jnca.2016.01.008
10. Shabtai, A., Elovici, Y., & Rokach, L. (2012). A Survey of Data Leakage Detection and Prevention Solutions. https://doi.org/10.1007/978-1-4614-2053-8
11. Yasmin R., Memarian, M.R., Hosseinzadeh, S., Conti, M., & Leppänen, V. (2018). Investigating the Possibility of Data Leakage in Time of Live VM Migration. In: Dehghantanha A., Conti M., Dargahi T. (eds) Cyber Threat Intelligence. Advances in Information Security, vol 70. Springer, Cham. https://doi.org/10.1007/978-3-319-73951-9_13
12. Satya Suryateja, P. (2018). Threats and Vulnerabilities of Cloud Computing: A Review. INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING, 6(3). https://doi.org/10.26438/ijcse/v6i3.297302
13. Zhiguo W., Jun’e, L., & H. Deng, R. (2012). HASBE: A Hierarchical Attribute Based Solution for Flexible and Scalable Access Control in Cloud Computing. IEEE Transactions on Information Forensics and Security, vol. 7, no. 2, april 2012, pp: 743 – 754. https://doi.org/10.1109/TIFS.2011.2172209
14. Sen, J. (2013). Security and privacy issues in cloud computing. Architectures and Protocols for Secure Information Technology Infrastructures, pp.1- 45, 2013. https://doi.org/10.4018/978-1-4666-4514-1.ch001
15. Cloud Security Alliance. (2017). The Treacherous 12 - Top Threats to Cloud Computing + Industry Insights [Ebook] (p. 25). Retrieved from https://downloads.cloudsecurityalliance.org/assets/research/top-threats/treacherous-12-top-threats.pdf.
16. Data breaches increasing - latest statistics | Leonovus inc. (2019). Retrieved from https://www.leonovus.com/blog/network-breaches-increasing-latest-statistics/.
17. Albaroodi, H., Manickam, S. and Singh, P. (2014). CRITICAL REVIEW OF OPENSTACK SECURITY: ISSUES AND WEAKNESSES. Journal of Computer Science, 10(1), pp.23-33. https://doi.org/10.3844/jcssp.2014.23.33
18. Alneyadi, S., Sithirasenan, E., & Muthukkumarasamy, V. (2015). Detecting Data Semantic: A Data Leakage Prevention Approach. In the Proceedings of the 2015 IEEE Trustcom/BigDataSE/ISPA, August 20 - 22, IEEE Computer Society Washington DC, USA, vol. 1, pp. 910-917. https://doi.org/10.1109/Trustcom.2015.464
19. Kirar, A., Yadav, A. K., & Maheswari, S. (2016). An efficient architecture and algorithm to prevent data leakage in Cloud Computing using multi-tier security approach. 2016 International Conference System Modeling & Advancement in Research Trends (SMART), Moradabad, pp. 271-279. https://doi.org/10.1109/SYSMART.2016.7894534
20. Rosado, T. and Bernardino, J. (2014). An overview of openstack architecture. Proceedings of the 18th International Database Engineering & Applications Symposium on - IDEAS '14. https://doi.org/10.1145/2628194.2628195
21. Setty, S. (2017). vSphere 6.5 Encrypted vMotion Architecture and Performance - VMware VROOM! Blog. Retrieved from https://blogs.vmware.com/performance/2017/01/vsphere-6-5-encrypted-vmotion-architecture-performance.html.
22. Ristov, S., Gusev, M., & Donevski, A. (2014). Security Vulnerability Assessment of OpenStack Cloud. 2014 Sixth International Conference On Computational Intelligence, Communication Systems And Networks. doi: 10.1109/cicsyn.2014.32.
23. Ali, M., Zolkipli, M., Zain, J., & Anwar, S. (2018). Mobile Cloud Computing with SOAP and REST Web Services. Journal Of Physics: Conference Series, 1018, 012005. doi: 10.1088/1742-6596/1018/1/012005
24. Mohd Ali, F., & Hadzril Wan Ismail, W. (2011). Network security threat assessment model based on fuzzy algorithm. 2011 IEEE International Conference On Computer Science And Automation Engineering. doi: 10.1109/csae.2011.5952688
25. Abu Othman, N., Mohd Ali, F., & Mohd Noh, M. (2014). Secured web application using combination of Query Tokenization and Adaptive Method in preventing SQL Injection Attacks. 2014 International Conference On

Computer, Communications, And Control Technology (I4CT). doi: 10.1109/i4ct.2014.6914229

26. Amrulla, G., Mourya, M., Sanikommu, R., & Afroz, A. (2018). A Survey of : Securing Cloud Data under Key Exposure. International Journal Of Advanced Trends In Computer Science And Engineering, 7(3), 30-33. doi: 10.30534/ijatcse/2018/01732018
27. ElSharif Karrar, D., & Idris Fadl, M. (2018). Security Protocol for Data Transmission in Cloud Computing. International Journal Of Advanced Trends In Computer Science And Engineering, 7(1), 1-5. doi: 10.30534/ijatcse/2018/01712018
28. Lord, N. (2019). Data Protection: Data In transit vs. Data At Rest. Retrieved 24 July 2019, from https://digitalguardian.com/blog/data-protection-data-in-transit-vs-data-at-rest