High Performance Computing (HPC) Data Center for Information as a Service (IaaS) Security Checklist: Cloud Data Governance

Arun Kumar Singh
Assistant Professor, College of Computing and Informatics, Saudi Electronic University, Kingdom of Saudi Arabia-KSA. ORCID: 0000-0002-9974-4696. E-mail: a.singh@seu.edu.sa, arunsinghiiiita@gmail.com

Samidha Dwivedi Sharma
Assistant Professor, College of Computing and Informatics, Saudi Electronic University, Kingdom of Saudi Arabia-KSA. ORCID: 0000-0001-5526-709X. E-mail: ssharma@seu.edu.sa, samidhad2000@gmail.com

Received September 12, 2019; Accepted December 7, 2019

Abstract

This study focused on cloud Data Governance (DG) for High Performance Computing (HPC) Cloud data Centre focusing on IaaS cloud service. To ensure the service provided to users is secured, HPCC are required to be certified by International Organization for Standardization (ISO). Therefore, IaaS security checklist is needed to measure the IaaS service provided. In depth interview results shows that failure in implementing good DG not only will put HPC data center at risks, but also will leads to business failure and recovery process might take more effort than some organizations might have imagined. At the end of paper, a cloud DG security checklist for IaaS security is proposed based on the ISO/IEC 27012 and applied by identified HPC cloud data center to improve the security of IaaS services.

Keywords

Cloud computing; Security; Data governance; Advanced Persistent Threat (APTs); High Performance Computing (HPC); Information as a Service (IaaS)

Retracted for copyright issues
Introduction

Many organizations come to realize that cloud DG is critical after many cases of data breaches occurred. Due to many data breaches cases because of internal mistakes, it is necessary to have a person that own and control and monitor the data. According to (Khatri and Brown, 2010), governance is the decision for effective management and information technology that need to be made. While cloud DG is a guideline for the person that holds the responsibility for the data assets in the organizations. The increasing threats which came from internal and external factors require cooperation from all stakeholders to implement cloud DG properly. Since cloud service has been widely used, there are issues related to the quality of the service in terms of security from providers. To ensure security in the cloud, service provider must be certified with standards and compliance according to DG/IT governance (English, 2009). Therefore, there is a need for HPC data center needs to get certified by International Organization for Standardization (ISO) 27012 for cloud security. ISO is an information security standard that has been widely used as the reference in managing Information Security Management System (ISMS) such in proprietary, industrial and commercial areas.

ISO/IEC 27012 standard guides the information security aspects of cloud computing, recommending and assisting with the implementation of cloud-specific information security controls, supplementing the guidance in ISO/IEC 27002 and other ISO27000 standards. Data and information are critical and should be treated as an asset and need to be appropriately secured. At the end of this study, a security checklist for IaaS layer at HPC cloud DC is proposed based on DG implementation. Failure in implementing good DG not only will put risks on the data but also will incur more cost on the recovery process. 122 organizations reported by English (2009) have lost almost $1.2 trillion due to poor data quality management. He also estimated that many organizations wasted their resources on recovery phases from that data management failure. In (Gow et al., 2006) report, Gartner estimated that more than 50 percent data warehouse would face failure due to lack of awareness in DG. Therefore, lacking awareness in DG can lead to business failure and recovery process might take more effort than some organizations might have imagined. As mentioned above, even though most organizations have security measurement, without proper standard procedures, the implementation might not be efficient enough. Before designing and developing security measures, some steps and procedures need to be followed. The security measure developed will cover all levels of security in the organization. Research done by Borgschulte (2016) states that poor DG can lead to data breach. He continued to explain that one recent example which is Sony Pictures hack issue in December 2014. The incident happened when a group named Guardian of Peace (GOP) hacked into Sony Pictures servers and storage. They deleted, stole and exposed much information including unreleased films, actors' salaries, celebrities' aliases, Steve Jobs’ drama, Obama’s racism, snap chats, medical documents, and Christmas gifts (Betters, 2015). As reported by Elkind (2015), the group deleted everything stored on 3,262 of the organization’s 6,797 personal computers and 837 of its 1,555 servers. The
CEO of Sony Pictures quoted the attack as "the worst cyber-attack in U.S. history." The incident occurred was due to the lack of data security and poor DG policies implementation in the organizations (Borgschulte, 2016).

Therefore, it is essential for all organizations to implement good DG to avoid threats that not only will put data at risks, but also result in business failure.

**Cloud Data Governance (DG) Security at IaaS**

**Cloud DG**

The cloud computing enables users to think beyond HPC on premises infrastructure. Three services such as public cloud, private cloud, and hybrid cloud available, while the layers are Software as a Service (SaaS), Platform as a Service (PaaS) and IaaS. Public cloud is when the cloud service providers provide the service over the internet for all users, and some services are free, or the users only pay for the storage or bandwidth that they subscribe. Some examples of public cloud are Google Drive from Google, Dropbox, and OneDrive from Microsoft. While a private cloud is a cloud service that serves to specific users only. Users also are pay according to the services that they subscribed. According to Azure (2019) two cloud models can service private cloud which is IaaS and PaaS. Hybrid cloud is the combination of public cloud and private cloud. Organizations can manage their data for public access or keep confidential information to private storage flexibly.

Example of a hybrid cloud service is Microsoft Azure. SaaS is a software solution where organizations rent an app from provider and users use it via a web browser over the internet. All services are managed by cloud service providers. PaaS is when the organization developed its app and rent the infrastructure from cloud service providers. Unlike SaaS, the organization has more flexibility to the functions of the apps, but the backbone of the services such as infrastructure, middleware, and app software is managed by the cloud service providers. IaaS is when an organization is free to manage their installation, configuration and software requirement such as operating system and applications while cloud service providers manage the infrastructure only, same as PaaS. Even though some companies had implemented DG, however, it is different from cloud governance. Therefore, it is crucial to have cloud DG within the DG itself to isolate the roles and responsibilities that specifically for cloud use and functions. The details about cloud DG is explained further in the below section.

Cloud DG has been received attention due to the concerns raised in cloud computing technology. Some organizations may have implemented DG or IT governance. Some organizations also may have heard corporate governance and information governance. However, these governance roles and functions are varied as explain in Ruithe, Benkhelifa and Hameed (2018), which discussed the differences between the mentioned governance above. They explained that cloud governance
is a new term in the information technology area where there is no clear definition for it. However, according to Saidah and Abdelbaki (2014), cloud governance is a set of guidelines on managing the cloud services which includes security, availability, privacy, compliance, and location of data in real-time. Therefore, to get the bright idea where the cloud governance lies, Al-Ruith, Benkelifa and Hameed (2018) describe the position for cloud governance in governance domains as can be seen in Figure 1. There are challenges in implementing cloud DG such as poor communication plan among stakeholders (Khatri & Brown, 2010; Grob & Schill, 2012; Benkhelifa & Al-Ruith, 2017) and lack of awareness DG among big organizations. Holt, Ramage, Kear and Heap (2015) stated that 45 percent of their respondents do not have DG policies.

Saed, Aziz, Ramadhani and Hassan (2018) stated that according to a research conducted by UBM, none of their respondents know how to define DG. Therefore, it leads to many security issues mainly related to data and cloud itself. Many security issues such as unauthorized access from other users in the same cloud, unknown location of data and lost data ownership to cloud provider, can be the effect from the improper of implementation of cloud DG. Governance can be in many forms of documentation which are either in guideline, assessment, checklist or policy. In Saed, Aziz, Ramadhani and Hassan (2018) few samples of previous works that focus on security checklist for IaaS security were reviewed. The further explanation of the security checklist for this study is explained below section.

Figure 1. Governance Domains

**Threat in IaaS cloud**

A threat is any harm impact that can bring severe effect to the system such as hacking, hijacking, system vulnerability, and any attacks or unintentionally such as natural disaster. Some effect of
threats may lead to business disaster. A vulnerability is a weakness in the system that may lead to threats. Example of system vulnerability is bugs in a patch update, error in system development and using an old patch. Mistakes from human error also risky enough to put the system in danger. Lately, many attacks were targeting cloud storage and (Kim, 2010) claimed that attacks on cloud had increased up to 300 percent and the attacks were succeeded due to an easy password, phishing attacks and third-party security breaches. They also reported that sophisticated attacks were increasing due to the latest security requirement which most companies do not have yet. Another report from (IBM Corporation, 2007) also stated that human error contributes 424 percent increasing attacks in the cloud. The human error mentioned in the report is a misconfiguration in cloud infrastructure. This error leads to one severe attack which is ransomware. Ransomware is known as a malware that threatens victims by locking their files and folders and ask for some ransom payment in return to unlocking it. Some attackers also threaten to publish the files or folders if the owner did not want to pay any ransom money. Examples of ransomware that hit major industries are WannaCry, NotPetya, and Bad Rabbit. Table 1 shows that list of research that had been done focusing on threats in the cloud. Most of the research was done by professional bodies such as Symantec, CSA, and Kaspersky. However, some researchers also contribute their knowledge to identify and propose suggestions to improve and solve this issue.

Table 1. Research of threats in cloud

<table>
<thead>
<tr>
<th>Method</th>
<th>Conclusion</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey (Harada, 2011)</td>
<td>Japanese and European Union (EU) have similar thought about threats in cloud computing. However, Japanese are more concern about the quality of cloud service rather than attacks from threats.</td>
<td>This survey was conducted to compare the Japanese perception of Cloud Security with ENISA Cloud Computing Risk Assessment.</td>
</tr>
<tr>
<td>Survey – 1,900 IT Professionals in US-based companies (Cyber Security Survey, 2017)</td>
<td>All companies concern about cloud security threats such as loss of data visibility and manageability.</td>
<td>This objective of this survey is to identify the awareness and concern about current security trends such as cloud and mobile threats.</td>
</tr>
<tr>
<td>Survey, Questionnaire (Cloud Service Allowance, 2017)</td>
<td>Twelve threats had been identified by CSA based on two stages of research conducted. The respondents are from the various background of positions and companies.</td>
<td>This report provides insight to users about the current trend of cloud computing threats. It is also providing real case examples to show how severe the impact of these threats.</td>
</tr>
<tr>
<td>Report from Symantec Website Security (Davis and Rankin, 2016)</td>
<td>Due to the increase of threats attack from various resources, it is essential that system owner to understand their system so that both parties can do their responsibility to do &quot;digital hygiene.&quot;</td>
<td>This report is to inform that the increasing threats in the various platform such as computers and mobile phones that use cloud and internet.</td>
</tr>
</tbody>
</table>
Based on Table 1, it shows that threats in cloud computing are increasing and it raises concerns among users. Researchers also urge all parties involved in using the cloud to do their responsibility by together maintaining and managing the cloud so that threat attack can be prevented more effectively. Basically, threats in cloud computing can be more dangerous than the on-premise infrastructure even though cloud service providers can ensure they had provided the best protection to their services, but there are no specific measurement or guidelines in users' side that cloud service providers do their work well. With other threats that come from malicious insiders as had been discussed above, to ensure the security in the cloud, it requires cooperation from every stakeholder. Another threats that had been discussed by researchers specifically in IaaS layer includes data loss or leakage threats, data breaches, insecure application program interface, malicious insiders, Denial of Service (DoS)/DDoS, account or service hijacking/identity theft insufficient due diligent, shared technology vulnerabilities, weak identity and access management, system and application vulnerabilities, Advanced Persistent Threats, abuse of cloud service, threats in virtual machine, data theft, data security and privacy issue, trust issue, Service Legal Agreement (SLA)/Legal issue, threats with cloud service providers, availability and reliability issues, authentication and authorization, man-in-the-middle attack, spoofing, injection attack, loss of control, malware, phishing, backdoor, and social engineering.

For threat at IaaS, this study decided to focus on the APTs at HPC IaaS.

**Advanced persistent threat (APTs) at IaaS**

According to (Kamal, 2014), APTs attack is a powerful advance attack. The word advance shows that someone expert in hacking can do the method of the attack in advance and. While persistent shows that the ability of this attack to be in the system for quite long time without getting tracked by organization's anti-virus and threats shows that the attack had been planned very well and the stated that APTs are combination of advance and malware attack that is targeting specific organization with objectives to steal information or for threats. APTs is not just a sudden attack like hacking, it involved a very well-planned step by step attack that will destroy the whole system in just a few seconds (Siddiqi & Shackleford, 2016).

The attack might be in the system for a very long time even before IT specialists know they existed in there. The attackers possess very advanced knowledge and skills in cyber threats making them very hard to detect and even defend the system. Their targets are more towards companies or organizations such as government agencies, hospitals, financial institutions, banks and military which holds valuable and secretive information so that when the information was stolen or misused, it will lead to severe impact. APTs is targeting their victims to steal their information like the attack on Google and disclosing private and confidential information like the attack on Sony Pictures. These attacks are very systematic and were planned well before the incident. The attacker has the intention to attack with some purposes. Whether the size of the organizations is small or big, as long as the data is valuable to be attacked, then any
organizations may be the victim. Table 2 provides example of an attack that targeting
government organizations to steal high-value sensitive data.

Table 2. List of famous APTs attacks

<table>
<thead>
<tr>
<th>Name of the attack</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Titan Rain:2003</td>
<td>Based on China, where hackers targeted the US Government to steal highly sensitive military data. The cybersecurity attack also included APT attacks where the attack focuses on the high-end organization such as NASA and FBI.</td>
</tr>
<tr>
<td>Sykipot Attacks:2006</td>
<td>The actors use vulnerabilities in Adobe Acrobat and Adobe Reader to target the victims which consist of telecommunication companies, defense contractors and government department in the UK and US government. The actors use the most commonly used APTs tactic which is spear-phishing where they send malware contained email to corporate and government system.</td>
</tr>
<tr>
<td>GhostNet:2009</td>
<td>Another attack based in China, it is a large scale of cyber espionage attack which had infiltrated and compromising computer system to more than 100 countries. Researchers concluded that China wants to be seen as leaders in “information war.&quot; The severity of this attack is where all infected machines can be turned into spying and listening devices, and it can be controlled remotely.</td>
</tr>
<tr>
<td>Stuxnet Worm:2010</td>
<td>This is one of the most sophisticated APTS attacks where usually, APTS attack was initiated via malware-attached email. However, this attack was initiated using USB keys and infecting all devices that are not connected to the internet for security reasons. Stuxnet was used against Iranian industrial infrastructure.</td>
</tr>
<tr>
<td>Deep Panda:2015</td>
<td>Another cyberwar between US and China government where this attack had compromised more than 4 million US personal records and some secret service staff was afraid might have stolen.</td>
</tr>
</tbody>
</table>

Recent research by Brook, Field and Shackleford (2016) indicated that the way in order to prevent APTs from to get chance to start the attack is by spreading awareness among employees to not just click emails or attachment and open any suspicious downloaded folder or files. APTs are challenging to detect and eliminated as it can move between networks and normal traffic a few days or maybe months before the attack. It also requires advanced security controls and IT staff training in order to eliminate and do prevention measures to stop another attack. Min, Xiao, Xie, Mohammad and Mandayam (2018) mentioned that APTs cannot be detected in the system because when the attackers got accessed into the system, they will change the command in the security system so that it will think that they are authorized users. The attackers also will disable some security protection to allow them planting more malware so that they can continue to be in the system undetected. Some researchers suggested that security awareness is crucial among employees (Chen, Desmet and Huygenes, 2014; Daly, 2016; Lee, 2017; Patil, 2015). This is because, since APTs started from the human error itself, providing the best security protection will not be worked when the leading cause is not prevented. Therefore, all employees should be
trained and informed to not just click any links in the email sent by the unknown sender or plug in any USB in CPU without knowing what the contents in it.

**Materials and Methods**

Preliminary study using extensive LR was used to identify threats in the IaaS cloud and the mitigation plans that were suggested by researchers. From there, a set of security checklist is proposed, and the in-depth interview was conducted to investigate the security implementation cloud data security and verification with experts also were evaluated to ensure the matters that were investigated are correct and reliable for proposing the final security checklist.

To develop the final security checklist, a DG cloud security model is developed as based on the checklist. It is also based on the in-depth interview that was conducted with two respondents from cloud DCs and two respondents from cloud experts. From the in-depth interview session, the data is analyzed using a content analysis method. Below is the explanation on the content analysis method from two respondents of cloud DCs. It is to show how the content analysis was conducted and how the initial model for DG cloud security for IaaS layer is produced as shown in Table 3.

<table>
<thead>
<tr>
<th>Phase 1</th>
<th>Extensive Literature Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>List of Security issues in IaaS Layer</td>
</tr>
<tr>
<td>2</td>
<td>Threats mitigation plans</td>
</tr>
<tr>
<td>3</td>
<td>Initial HPC DG Security checklist for IaaS</td>
</tr>
</tbody>
</table>

**Table 3. Research Framework**

<table>
<thead>
<tr>
<th>Phase 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. In-depth Interview</td>
<td>1</td>
<td>Components in Security Measurement</td>
</tr>
<tr>
<td>2. Content Analysis</td>
<td>2</td>
<td>Initial HPC DG Security Model for IaaS</td>
</tr>
<tr>
<td>Final HPC DG Security Model for IaaS</td>
<td>4</td>
<td>HPC DG Security checklist for IaaS</td>
</tr>
<tr>
<td>HPC DG Security checklist for IaaS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the preliminary study, extensive LR activities conducted and at the end of this phase, two outputs are gathered. The outputs are a list of threats in IaaS and APTs and initial security checklist. In phase two, the initial security checklist is used in an in-depth interview for data collection.

As shown in Figure 2, there are two respondents identified for in depth interview and they are from higher educational institution that has IaaS in HPC data center. The respondents for the validation phase are experts in industry security cloud. Interview session with experts seeks the validation whether the proposed checklist is reliable and can be used to measure cloud security. The experts must be a person who is certified in network security or cloud practitioner.
Content analysis is a method widely used in the qualitative research study. Content analysis is a method that is identifying the main points in the text scripts, and the content is summarized until main themes and categories are identified. There are few steps to conduct content analysis as can be seen in Figure 3.

For step one is finding meaning units in the script. Meaning units are only related statement in the interview scripts which required for analysis. The unrelated conversations are taken out to assist the researcher in focusing the meaning input that also known as script from respondents.
Script from respondents answering the statement from the checklist and include reasons and opinion of any security measures implemented in the cloud DC. Therefore, to proceed with the content analysis, the only script from the respondents is used. The transformation steps from meaning units/scripts to theme shown in Figure 3.

Assigning codes must take thorough consideration and justification to avoid misinterpretation of codes. The condensations might have a double meaning which can be assigned to one or more codes. However, to resolves this issue, the sentences need to be referred from the script. The codes should be assigned based on the earlier intended meaning from the original sentences. From this step also, the condensations that have same sentences and in the same codes can be taken out. Usually, codes are no more than four words. Some codes were taken from the condensation it referred, but some codes did not. Codes can be amended according to the closest match meaning from the condensed sentence. The content analysis steps conducted as below:

- First, the codes are identified based on what the condensed sentence means. Then, they were grouped to the close meaning with other codes. After that, a suitable word is used to represent all the grouped codes before. The last step in the content analysis is the identification of categories and themes.
- After the codes are identified, they were grouped, and suitable categories are assigned to it. The categories were selected based on the components in cloud data security and privacy that had been identified above. The assignment of codes to category must be referred to the definition of each component, and it must be suited with each category. After that, themes are assigned to each category as example in Table 4.
- The themes identified based on the fundamental of security and privacy that had been assigned to the respective codes. The condensed sentences and codes had been arranged and grouped according to its similarity properly. The categories and themes also had been identified based on the components in data security privacy and CIA triad respectively. By referring to Table 4, the analysis can be finalized, and from here, cloud DG can be developed and proposed.

Table 4. Categories and Themes for DC1

<table>
<thead>
<tr>
<th>Security Measurement in HPC Data Center 1</th>
<th>Codes</th>
<th>Categories</th>
<th>Themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>- We have multi factor authentication in the cloud</td>
<td>Access Control</td>
<td>Data Security</td>
<td>Availability in Cloud DC Security</td>
</tr>
<tr>
<td>- All researchers can access server anywhere</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- We control access from other users in the same cloud</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- We follow process and procedure when migrating data, but not our own process</td>
<td>Data Center Management</td>
<td>Compliance Management</td>
<td>Integrity in Cloud DC Security</td>
</tr>
<tr>
<td>- We do not do data duplication because our data is not big and incur the cost to implement it</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- We review access permission</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Need to make sure authorized user can access the files</td>
<td>Data Privacy</td>
<td>Privacy Protection</td>
<td>Confidentiality in Cloud DC Security</td>
</tr>
</tbody>
</table>
Final HPC DG cloud security model for IaaS

The final security checklist model had been produced based on the comments and suggestions from the experts as shown in Figure 4. Ten measurement parameters identified. The based from the content analysis of expert validation which are; Access Control, Data Management, Risk Management, Cloud Risk, Configuration Management, Threat Management, Cloud Service Provider, Data Center Management, Standard Guideline, Data Privacy, and Data Protection. Based on these 10 codes, there are new three codes that had been suggested by the expert which are Cloud Risk, Risk Assessment, and Standard Guideline. Besides the codes that were suggested by cloud DC, other components in cloud security needs to be focused on when developing a security assessment for cloud DC.

Figure 4. Cloud Security Model – Expert Validation
Conclusion

Final security checklist will be in the form of policy where it will properly document. It is to show that this checklist will be used as official security measurement HPC3. In this documentation, there will be an explanation for the checklist, the target users, responsibilities, and scope of functions. This checklist is divided into three components which are Availability in Cloud Security, Integrity in Cloud Security and Confidentiality in Cloud Security which taken from the categories in the content analysis. While under each part, there will be subcomponents. Under Availability in Cloud Security, the subcomponents that will be listed as Data Security which has Access Control and Data Management. In the meantime, subcomponents that fall under Integrity in Cloud Security are Risk Management which has Cloud Risk, Risk Management, Configuration Management, and Threat Management. One more sub-component that falls under this component is Compliance Management which has Standard Guideline, Cloud Service Provider, Data Center Management and Security Policy. Based on the result and findings from data analysis performed above, one suggestion can be suggested which is to study the effectiveness of security checklist implementation after applying for ISO standardization. After this security checklist is implemented in the HPC data center, the effectiveness of the implementation needs to be studied or monitored. This is to ensure the security of IaaS cloud service is continued properly even after it has been certified by ISO standardization. It also will ensure the HPC data center give priority to the most important component in security of IaaS cloud service.

References


Gow, B. W. et al. (2006). Effective data governance in banking - Banking failure is not an option. IBM, NY.


---

**Bibliographic information of this paper for citing:**


Copyright © 2019, Arun Kumar Singh and Samidha Dwivedi Sharma.