A NOVEL SECURITY MODEL FOR DATA MANAGEMENT IN CLOUD COMPUTING

S. S El Ashry M. A. El-dosuky M. Z. Rashad H. S. Abdelkader

: Faculty of computer, Mansoura University, Egypt

Sobhank_raby@yahoo.com Dr_dos_ok@yahoo.com Magdi_z2011@yahoo.com Hatem6803@yahoo.com

Abstract: Data security is always the focus of huge possible cloud clients, also a big obstacle for its extensive applications. Till now there is no proficient mechanism for data security accustomed for the cloud environment, and various service types need different solutions for data protection. The objective of this research is to develop a new data security model for the cloud environment. The proposed model is decomposed of four modes of operations denoting the different ways for providing the service for better user convenience. The four modes are: Scheduling mode, Through-off mode, Virtual machine mode, and Batch mode. Analyzing the proposed model shows the advantages it has over previous models. Based on extensive experiments, the proposed model covers many security cases, and is robust in handling security threats while gaining user convenience.

Keywords: Cloud Computing, Data Management, Security Model, Simulation.

1. Introduction

Cloud Computing, abbreviated as CC, offers services to consumers according to the provided service level. There are three service categories as follows: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and software as a Service (SaaS) [1, 22]. For these three types, security and protecting personal privacy have been ranked as risk [2,23]. However using the conventional methods of cryptography become inappropriate to provide a proficient technique for the clouds environments [3]. In conventional software development, security issue has not a main concern, but user privacy protection in CC is a major issue. This is because the user data often in unencrypted pattern are located in some machines and this user data maybe contains several company operators, commercial sensitive data, as well as potential damage in materials privacy [4]. So, safety of client sensitive data that hosts in the virtual servers of the cloud is extremely significant. During the establishment of the cloud service system that achieve control of data security challenge, the cloud providers have to consider of all factors types and improves client trust level in the laws case, has to evaluate and obtain the system measure with principle of each phase design [5].

CC on World Wide Web is a significant improvement application, the cloud systems serve consumers without need to master control power hardware [6], the customer of the software service that the cloud offers, processing and storing data like individual profiles, credit cards, health and finance in addition to other data. The characteristics of cloud are for offering data storage services, processing, and platform
service outflow to universal consumers, however this material set result is data protection, leading to the popular reason for the enterprises otherwise individuals worried about using the clouds services solutions packages. Also, the data security in CC still needs more researches are done on this topic as it has several problems with no solution. Till now there is no proficient mechanism for CC data security system [2], and several CC service types need several solutions for data protection. This research objective is developing a new data security model for CC.

This paper is organized as the following: section 2 reviews the cloud working principle, CC management, and previous models. Section 3 shows the proposed model which is decomposed of four modes of operations denoting the different ways for providing the service for better user convenience. The four modes of operations are: Scheduling mode, Through-off mode, Virtual machine mode, and Batch mode. In section 4, the used tool and different configuration of the tested scenarios. In section 5, the results are described. Finally in section 6 the conclusions.

2. Related Work

2.1 The Clouds Working Principle.
The clouds systems consist of the network infrastructure provider (IaaS), and the platform service provider (PaaS) and software service provider (SaaS) [7,26]. From Google CC Trends analysis chart can be viewed that CC has well-respected degree grown year after year. Also, the database manufacturers’ products are joining the purpose of the clouds to support the database. For example: Oracle currently is operating a service of clouds computing platform (EC2) directly in Amazon [8]. So launching more and more data in the clouds service, causing that the data safety will still has a serious concern, as these data often include significant sensitive information for the organizations otherwise the individuals.

2.2 CC Management Approach.

Hardware devices of Clouds environment is offered via an external third party agency entrusted with clouds i.e. TTP Clouds otherwise an interior cloud. Clouds environments maybe limited in several organizations share the public cloud i.e. Public Clouds otherwise in private cloud enterprises i.e. Private Clouds [9, 10, 24].

2.3 The Classification of Security Requirements.
The of classification level for the security challenge must be considered as the following [7]: (1) security of server/s, (2) security of Internet (3) security of database (4) data security (5) security of applications.

2.4 Previous Models.
The basic data model of the distributed system is Client-server [11, 12, 13], There are many models are proposed for incorporating security in the distributed systems such as: Kerberos [14-17, 25] and GARS [18]. Among the problems facing these previous models are the delivery of service which is a significant aspect of quality of service (QOS), is postponed until verification process is performed. Another critique is that the service is not provided in a continues manner.

The [19] threat modeling process is as follows:

1. Identify the known threats to the system.
2. Rank the threats in order by decreasing risk.
3. Determine how you will respond to the threats.
4. Identify techniques that mitigate the threats.
5. Choose the appropriate technologies from the identified techniques.

2.4.1 STRIDE Model

A model you may find useful for data security in cloud environments is STRIDE by Microsoft, derived from an acronym for the following six threat categories: S stands for “Spoofing identity”, T stands for “Tampering with data”, R stands for “Repudiation”, I stands for “Information disclosure”, D stands for “Denial of service” and E stands for “Elevation of privilege” [20].

3. Proposed Model

Definition 1: Overall required security (ORS)

Overall required security (ORS) is the required security that managed by the provider.

Definition 2: Computing Capacity(C)

Computing Capacity(C) is the ability of the datacenters.

For heterogeneous

$$C = \sum_{k=1}^{d} v_k c_k$$  

(1)

Where d is number of the data centers, $v_k$ is number of the virtual machines in datacenter k, and $c_k$ is the computing capacity of virtual machine $v_k$.

For homogeneous

$$C = d \cdot v$$  

(2)

Where d is number of the data centers, and v is number of the virtual machines.

Notes: For simplicity, we will use equation (2). At design time C is varying, but once it is in operation is considered a constant threshold, because it does not make any difference if it is homogeneous or heterogeneous.

Definition 3: Total number of requests(R)

Total number of requests(R): the whole number of the users’ requests.

Similarity the total requests for heterogeneous

$$R = \sum_{i=1}^{u} \frac{(r_i t_i c_i)}{g_i}$$  

(3)

Where $u$ is number of the user bases, $r_i$ is rate of the requests per hour, $t_i$ is the duration of the simulation session, $c_i$ is computational size of request i, $g_i$ is grouping factor of user bases.

For homogeneous

$$R = \frac{u r t}{g}$$  

(4)
Where \( u \) is number of the user bases, \( r \) is rate of the requests per hour, \( t \) is the duration of the simulation session, \( g \) is grouping factor user bases.

For simplicity, we will use equation (4).

**Definition 4: Security Measure (M):**

Security Measure (M): is the measure of the total security threats, based on the following assumptions:

1) The ORS is proportional to size of user bases that assess the cloud services.
2) The ORS is proportional to the value of data.

We have two dimensions that are determine the required security:

1. Value of data (more value, more required security)
2. User base size (more requests, more required security) to decide the appropriate mode type for clients among the four components of the proposed model. This does not mean to ignore less-value data. These less-value data need to be secured but not as the security level of the more-value data.

### 3.1 The Model

After a deep study and analysis of the related existing systems, we designed our proposed model for clouds applications. This model is decomposed four modes of operations denoting the different ways for providing the service for better user convenience as in our suggested model we tried to avoid the mentioned problems in section 2.4 that faced implementing data security in distributed systems. This optimized model consists of four modes of operations as follows:

1. Scheduling Mode.
2. Through-off Mode.
4. Virtual machine Mode

Each mode of operation has specific characteristics.

<table>
<thead>
<tr>
<th>Mode Name</th>
<th>User base size</th>
<th>Value of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scheduling Mode</td>
<td>( &lt; C )</td>
<td>( &lt; \theta )</td>
</tr>
<tr>
<td>Through-off Mode</td>
<td>( &gt; C )</td>
<td>( &gt; \theta )</td>
</tr>
<tr>
<td>Batch Mode</td>
<td>( &lt; C )</td>
<td>( &gt; \theta )</td>
</tr>
<tr>
<td>Virtual Machine Mode</td>
<td>( &gt; C )</td>
<td>( &lt; \theta )</td>
</tr>
</tbody>
</table>

Table 1 shows the four modes of operation with their features. In scheduling mode the user base size is less than computing capacity and the value of data is less than a specific threshold. In through-off mode the user base size is greater than computing capacity and the value of data is greater than a specific threshold. In batch mode the user base size is less than computing capacity and the value of data is greater than a specific threshold. In virtual machine mode the user base size is greater than computing capacity and the value of data is less than a specific threshold.
3.2 The Algorithm

We have two dimensions that are determine the required security:
1-Value of data (more value, more required security)
2-User base volume (more requests, more required security) to decide the appropriate mode type for clients among the four components of the proposed model. Figure 1 shows the main algorithm.

4. Evaluation

4.1 Tool

4.1.1 CloudAnalyst Simulator

It is a simulator which supports simulating large-scale Clouds applications to study these applications behavior with a variety of configuration. This tool can helping developers to understand applications distribution ways across clouds infrastructures as well as evaluate the services like the providers incoming by using Services Brokers and application performance optimization [21].

4.2 Configuration

We used CloudAnalyst software in our simulation with the following configuration: D=1, v=10, so C=1*10=10, u=1, scheduling techniques is RR, t = 1 hour.

<table>
<thead>
<tr>
<th>Algorithm 1 main ( )</th>
</tr>
</thead>
<tbody>
<tr>
<td>d is number of the data centers</td>
</tr>
<tr>
<td>v is number of the virtual machines</td>
</tr>
<tr>
<td>C is the computing capacity</td>
</tr>
<tr>
<td>u is number of the user bases</td>
</tr>
<tr>
<td>r is rate of the requests per hour</td>
</tr>
<tr>
<td>g_u is the users grouping factor</td>
</tr>
<tr>
<td>g_v datacenter grouping factor</td>
</tr>
<tr>
<td>t is duration of the simulation session</td>
</tr>
<tr>
<td>θ is a constant that represents a threshold for security measure</td>
</tr>
<tr>
<td>M is the security measure</td>
</tr>
<tr>
<td>C=d.v</td>
</tr>
<tr>
<td>R = \frac{u.r.t}{g}</td>
</tr>
<tr>
<td>If R-C&lt;0 &amp; M&lt;θ &amp; g_v&gt;1 Then Scheduling mode</td>
</tr>
<tr>
<td>Else If R&gt;C &amp; M&gt;θ &amp; Then Through-off mode ( )</td>
</tr>
<tr>
<td>Else If R-C&lt;0 &amp; M&lt;θ &amp; g_v&gt;1 Then Batch mode</td>
</tr>
<tr>
<td>Else If R-C&lt;0 &amp; M&lt;θ &amp; g_v&gt;1 Then Virtual Machine mode ( )</td>
</tr>
<tr>
<td>Else</td>
</tr>
<tr>
<td>Error</td>
</tr>
<tr>
<td>End If</td>
</tr>
</tbody>
</table>

Figure 1: The Main Algorithm.
4.3 Test Scenarios:

The first scenario consists of one Data center with 10 virtual machines, \( R=C=10 \) & \( gu=1, gv=1 \), we notice that the overall response time is 300.24ms and Data Center processing time: is 0.26 ms. The second scenario consists of one Data center with 10 virtual machines, \( R=15 \) & \( gu=1, gv=1 \), we notice that the overall response time is 300.92 ms and Data Center processing time: is 0.25 ms. The third scenario consists of one Data center with 10 virtual machines, \( R=5 \) & \( gu=1, gv=1 \), we notice that the overall response time is 300.14 ms and Data Center processing time: is 0.26 ms. In the next scenario there is one Data center with 10 VMs, \( R=10, \) & \( gu=10, gv=10 \), we notice that the overall response time is 300.48 ms and Data Center processing time: is 0.36 ms. Whereas in the fifth scenario consists of one Data center with 10 virtual machines, \( R=15 \) & \( gu=10, gv=10 \), we notice that the overall response time is 299.69 ms and Data Center processing time: is 0.37 ms. The sixth scenario consists of one Data center with 10 virtual machines, \( R=5 \) & \( gu=10, gv=10 \), we notice that the overall response time is 300.09 ms and Data Center processing time: is 0.26 ms. In the next scenario, there is one Data center with 10 virtual machines, \( R=10 \) & \( gu=10, gv=1 \), we notice that the overall response time is 300.38 ms and Data Center processing time: is 0.26 ms. The eighth scenario consists of one Data center with 10 virtual machines, \( R=15 \) & \( gu=10, gv=1 \), we notice that the overall response time is 299.58 ms and Data Center processing time: is 0.26 ms. Finally, in the last scenario one Data center with 10 virtual machines, \( R=5 \) & \( gu=10, gv=1 \), we notice that the overall response time is 299.98 ms and Data Center processing time: is 0.26 ms. Table 2 shows the settings of tested scenarios.

**NOTE:** As shown in table 2, the difference between the response times (ms) of the previous tested scenarios is so little. Clearly, the difference is in decimal percentages. Thus the settings maps of the previous scenarios look like the same and also the figures of response times of tested scenarios seem the same. So, presenting the map of only one scenario (to show the GUI of the used simulator and implementing for this scenario) will be sufficient. Figure 2 shows CloudAnaylst GUI & settings map of test scenario #1. And there is no need to show the maps of the rest scenarios. Figure 3 shows the response time of test scenario #1, we notice that the simulation time is concentrated in the duration from 0-2 hours (the first two hours) as the simulation duration is \( t = 1 \) hour.

![Simulation Complete](image)

Figure 2: CloudAnaylst GUI & settings map of test scenario #1.
Table 2: Simulation settings & results of experiments.

<table>
<thead>
<tr>
<th>Simulation settings</th>
<th>Overall average response time (ms)</th>
<th>Data Center processing time for a request (ms)</th>
<th>VM Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 One Data center with 10 Virtual Machines, R=10 &amp; ( g_u=1, g_v=1 )</td>
<td>300.24</td>
<td>0.26</td>
<td>$1.11</td>
</tr>
<tr>
<td>2 One Data center with 10 Virtual Machines, R=15 &amp; ( g_u=1, g_v=1 )</td>
<td>300.92</td>
<td>0.25</td>
<td>$1.02</td>
</tr>
<tr>
<td>3 One Data center with 10 Virtual Machines, R=5 &amp; ( g_u=1, g_v=1 )</td>
<td>300.14</td>
<td>0.26</td>
<td>$1.18</td>
</tr>
<tr>
<td>4 One Data center with 10 Virtual Machines &amp; R=C &amp; ( g_u=10, g_v=10 )</td>
<td>300.48</td>
<td>0.36</td>
<td>$1.11</td>
</tr>
<tr>
<td>5 One Data center with 10 Virtual Machines &amp; R&gt;C &amp; ( g_u=10, g_v=10 )</td>
<td>299.69</td>
<td>0.37</td>
<td>$1.02</td>
</tr>
<tr>
<td>6 One Data center with 10 Virtual Machines &amp; R&lt;C &amp; ( g_u=10, g_v=10 )</td>
<td>300.09</td>
<td>0.37</td>
<td>$1.18</td>
</tr>
<tr>
<td>7 One Data center with 10 Virtual Machines &amp; R=C &amp; ( g_u=10, g_v=1 )</td>
<td>300.38</td>
<td>0.26</td>
<td>$1.11</td>
</tr>
<tr>
<td>8 One Data center with 10 Virtual Machines &amp; R&gt;C &amp; ( g_u=10, g_v=1 )</td>
<td>299.58</td>
<td>0.26</td>
<td>$1.02</td>
</tr>
<tr>
<td>9 One Data center with 10 Virtual Machines &amp; R&lt;C &amp; ( g_u=10, g_v=1 )</td>
<td>299.98</td>
<td>0.26</td>
<td>$1.18</td>
</tr>
</tbody>
</table>

Figure 3: The response time of test scenario #1.
5. Results and Discussion

Table 3 summarized the results.

<table>
<thead>
<tr>
<th></th>
<th>without grouping</th>
<th>with grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>balance</td>
<td>300.24</td>
<td>300.48</td>
</tr>
<tr>
<td>R&lt;(C)</td>
<td>300.14</td>
<td>300.09</td>
</tr>
<tr>
<td>R&gt;(C)</td>
<td>300.92</td>
<td>299.69</td>
</tr>
</tbody>
</table>

Figure 4 shows the summary of experiments based on whether there is grouping or not.

• Three cases are tried without grouping for the users & virtual machines, three cases with grouping for both the users & virtual machines, three cases with grouping for the users & without grouping for the virtual machines. CloudAnalyst doesn’t accept that the grouping factor for the virtual machines greater than grouping factor of the users.

• In the first three scenarios, we notice that overall average response time (ms) increase in scenario #2 when we increase the total requests of users (R) compared with scenario #1 and the overall average response time (ms) decrease when the total requests of users (R) decrease as in scenario #3.
• When \( R=C=10 \) (balance mode), the overall average response time (ms) increase in case of grouping for the users & virtual machines (i.e. grouping factor of the users & Datacenter =10) and also increase in case of grouping for the users & without grouping for the virtual machines (i.e. grouping factor of the users=10 & grouping factor of the Datacenter =1) when it is compared with the case (scenario#1) of without grouping for the users & virtual machines (i.e. grouping factor of the users & Datacenter =1).

• When \( R=5 \) (less than balance mode), the overall average response time (ms) decrease in case of grouping for the users & virtual machines (i.e. grouping factor of the users & Datacenter =10) and also decrease in case of grouping for the users & without grouping for the virtual machines (i.e. grouping factor of the users=10 & grouping factor of the Datacenter =1) when it is compared with the case (scenario#3) of without grouping for the users & virtual machines (i.e. grouping factor of the users & Datacenter =1).

• When \( R=15 \) (greater than balance mode), the overall average response time (ms) decrease in case of grouping for the users & virtual machines (i.e. grouping factor of the users & Datacenter =10) and also decrease in case of grouping for the users & without grouping for the virtual machines (i.e. grouping factor of the users=10 & grouping factor of the Datacenter =1) when it is compared with the case (scenario#2) of without grouping for the users & virtual machines (i.e. grouping factor of the users & Datacenter =1). This may be counter-intuitive, but there is some logic behind that. If there is a grouping, the response time decreases regardless that the total user requests is less than or greater than computing capacity.

6. Conclusions

In this research we discuss the data security challenge to cloud and propose an optimized secure model for data management in cloud computing environment. We can summarize the main contributions as follows:

• The model hypothetical foundation is the overall required security is determined by two factors which are the total number of requests (\( R \)) with focusing in measuring it and the security measure (\( M \)). Determine mode of operation is based on \( M \) and \( R \).

• This hypothesis leads to dividing the model into four modes of operation namely: Scheduling mode, Through-off mode, Virtual machine mode, and Batch mode.

• Based on extensive experiments, the proposed model covers many security cases, and is robust in handling security threats while gaining user convenience.

• Among interesting observations is the effect of the grouping on the overall required security. If there is a grouping, the response time decreases regardless that the total user requests is less than or greater than computing capacity.

A possible future work direction is implementing the proposed model by incorporating different combinations of symmetric encryption and hashing algorithms. A clustering algorithm such as K-mean [27] can be used in calculating value of data (\( M \)). Then, Hidden Markov Model (HMM) [22] may be used in determining the appropriate mode for users dynamically.
References


