HEALTHCARE INTEGRATION BASED ON CLOUD COMPUTING

Roxana MARCU¹, Dan POPESCU², Iulian DANILA³

A high number of information systems are available on the market for each hospital department with high performance results but in healthcare we need access to data from everywhere at any time. This paper presents an e-health solution based on healthcare information systems integration and cloud computing concept with focus on medical imaging department. Hybrid cloud architecture (public – private) and a request management application has been proposed and analysed in terms of waiting time defined as Quality of Service criteria (QoS). The service has been modelled using queuing theory as two serially connected queues M/M/s and M/M/1.

Keywords: cloud computing, healthcare, information system, integration, quality of service

1. Introduction

Healthcare information systems (HISs) are considered as one of the most complex and challenging field. Currently are available a lot of information systems specialized on each health area: hospital management, pharmacies, insurance companies, resource management and the challenge is now to integrate these systems in order to access data when is needed having a minimum waiting time. Due to the high amount of data and the complexity of healthcare domain we will consider in this paper the radiology department as central point of analysis of clinical and administrative processes. Radiology in terms of information systems (RIS, PACS) has been described in [1] from the clinical and operational workflow perspective. In [2] has been presented a service oriented architecture (SOA) based integration of radiology information system that introduces an ERP system as central point of management. The aim of this paper is to extend the presented model in order to obtain complete integration with all technologies including mobile communications from the perspective of radiology data access and to analyse the results in terms of waiting time performance. The purpose of this

¹ PhD student, Faculty of Automatic Control and Computers, University POLITEHNICA Bucharest, Romania, e-mail: roxana.marcu@cti.pub.ro
² Professor, Faculty of Automatic Control and Computers, University POLITEHNICA Bucharest, Romania, e-mail: dan_popescu_2002@yahoo.com
³ PhD student, Faculty of Automatic Control and Computers, University POLITEHNICA Bucharest, Romania, e-mail: iulian.danila@gmail.com
approach is to convert an application into a number of components called services that can be shared and reused.

SOA Healthcare Information System integrates a large number of heterogeneous systems in a scalable solution (hardware and software) based on HL7 standards and web services delivering “reasonable response time under the different outpatient, inpatient and emergency system usage conditions” [3]. Available information at the right moment is the central requirement for healthcare therefore a better management of the resources and control of data is demanded.

Reports show 35% of healthcare organizations are either implementing or operating cloud computing in 2012 having as main concerns in implementing cloud healthcare: security of data/applications (51%), performance of cloud services (36%) and integrating cloud applications/infrastructure with legacy systems (31%). Currently this cloud technologies are mostly limited to conferencing and collaboration (29%), compute power (26%), office and productivity suites (22%) [4]. Cloud computing technology is still new, but actually research results describe low price and minimum resources architectures and low waiting time values for data access[5].

2. Background and related work

A medical images storage and retrieval solution based on a DICOM Server over cloud computing using Windows Azure Cloud and SQL Azure is described in [6]. Authors concluded that using Platform as a Service (PaaS) service model architecture reduces the management cost and increase security by the use of DICOM server.

A high level solution of sharing medical imaging over cloud is presented in [7]. Architecture integrates PACS system using a DICOM bridge router and cloud providers to connect medical imaging modalities and to provide access to images to the end-user over the Internet. The architecture demonstrates secure, maintainable and fast setup solutions for communication between hospitals as distinct/geographical dispersed entities.

Security and privacy as a requirement of a healthcare cloud platform is one of the most analysed and discussed topics among researches. In [8] the authors present a hybrid cloud infrastructure for Electronic Health Record (EHR) sharing and integration of medical data with focus on ownership of medical record. A mechanism of medical data protection is built up supporting both normal and emergency scenario. The hybrid cloud architecture is verified by the authors from a security perspective.

Increased storage capabilities supported by a cloud based platform confirm one of the main concerns that have been raised by most of the healthcare
organizations. As discussed in [9] a cloud platform that integrates a cloud storage layer covers local storage and remote storage using Internet resulting in theoretical infinite storage capacity.

As described in [1] the purpose of this research is to consider the healthcare organization as a business that fully integrates clinical processes. Cloud service and deployment model is determined based on the healthcare organization strategy expected to be accomplish. In order to validate this model we need to consider several services that are included in the model. As a scope of this paper we build the model of accessing radiology data in an integrated healthcare cloud platform.

In order to achieve the highest value of quality in the service that is to be designed it is required a strong analysis of the service system from dedicated resources and demands placed on the system perspective. Ideally in a healthcare information system a patient should never wait for a medical service but in this case we face a limited number of resources (medical staff, modalities, waiting times etc.) and an uncontrollable number of patients. Based on analyses of 141 papers presented in [10] queuing theory use in analysis of healthcare processes is feasible and recommended for QoS analysis for both physical queues that are created for clinical service access or for healthcare information system modelling processes. Queuing theory is used to predict the QoS measured in a variety of ways, as a function of both the demands on the system and the service capacity that has been allocated to the system [11].

3. System architecture based on cloud

In this section, we propose as a solution to fully integrate healthcare clinical and administrative processes focusing on radiology department a cloud computing platform that is to be used by several units (hospitals) and entities (departments). As a deployment model we consider a hybrid cloud for the healthcare environment, composed by several private clouds that are representing each specific department (Radiology Private Cloud, Emergency Private Cloud) and a public cloud used to integrate all departments/hospitals and the third parties. The purpose of the public cloud that is used for unit integration is to avoid the mesh networks within the private clouds connections. As described in [5] the communications between the two public clouds will be held through HL7 CDA messages after the link between the applications will be establish using an application defined in the public cloud. DICOM and HL7 are successfully used to integrate a large numbers of applications from various medical domains [12]. The radiology data will be stored in the radiology private cloud and all departments/hospital can access medical patient data when is needed, the connection being established by the private cloud application. From the perspective of a network of
entities, defining the private clouds, with the scope of using information stored in diverse locations using a central management application defined in the public cloud we propose the architecture from Fig.1.

In a healthcare environment all patients are equal but they are split in different priority classes and this should be the base of modelling a healthcare service. Requests that are linked to critical patients need to be handled first even though there are other requests for the same type of data but the overall waiting time for each request should be kept to minimum.

Queues and queuing theory in healthcare for patient/resource scheduling and data access represents a challenge for healthcare IT research domain due to the need of understanding clinical and information processes from service modelling and management perspective. In this paper we focus on data access due to radiology images that represent a special type of data. Cloud architecture proposed in Fig.1 considers a PACS database integrated in a private cloud and a management application integrated in a public cloud in order to obtain full access to data for any point. Requests for medical images are coming in the presented hybrid cloud architecture from at least 4 points: mobile infrastructure (mobile phones, tablets, laptops), hospital management system (ERP), radiology department (intern specific system or workstations) and from emergency department (workstations). Another point of request entry in the system can be considered the medical staff that requests data from home based on authentication data provided by the cloud provider.
In order to collect all requests we propose a request management application to build up the entry queue in the public cloud system. Each request that is entering in the cloud is linked to a patient with a certain priority class assigned from 1 to r. Based on the priority class the application creates j queues after first come first served (FCFS) principle and offer as input to the cloud infrastructure a FCFS queue composed by serial concatenation of all j queues in priority order: FCFS₁; FCFS₂; … FCFSᵢ. The algorithm is as follows:

while TRUE do
  check request_type = request_data;
  case priority
    when 1: add r to FCFS₁;
    when 2: add r to FCFS₂;
    when j: add r to FCFSᵢ;
  end case.
end while.

Using Kendall notation for queues we consider a model of a cloud system composed by two queues M/M/s entry queue and M/M/1 database precede queue. Based on Burke’s theorem and Jackson’s theorem the system can be modeled as an open Jackson network in which the queues can be analyzed independently [13-15].

Considering probability that a request from M/M/s will enter in M/M/1 before database p or access directly the database (1-p) we propose as mathematical model for the cloud system two serially connected queues with arrival rate λ for M/M/s and λp for M/M/1 (Fig.2.).

\[
\begin{align*}
\text{FCFS} & \xrightarrow{\lambda} \text{M/M/s} & \xrightarrow{\lambda p} \text{M/M/1} & \xrightarrow{\lambda(1-p)} \text{DB}
\end{align*}
\]

Fig.2. Cloud service mathematical model: two serially connected M/M/s and M/M/1 queues

M/M/s queue is to be analyzed as a priority service with Poisson arrivals λ, multiple servers s and identically distributed service time μ for all r priority classes [9]:

\[
\mu = \mu₁ = \ldots = \muᵣ
\] (1)
Utilization ratio for a priority class \( j \) is defined as \( \rho_r = \frac{\lambda_r}{\mu} \) and the cumulative utilization ratio for classes 1 through \( r \) as:

\[
\sigma_r = \sum_{k=1}^{r} \rho_k \quad \text{for} \quad \sigma_0 = 0
\]

(2)

In this case we consider the expected waiting time in the first queue [9]:

\[
W'_q = \frac{W_0}{(1-\sigma_{r-1})(1-\sigma_r)} \quad \text{for} \quad \sigma_0 = 0
\]

(3)

where \( W_0 \) is the expected time until one of the \( s \) servers is available considering probability that the system is busy:

\[
W_0 = \left( \frac{\lambda}{\mu} \right)^s \frac{s\mu}{s\mu - \lambda} P_0
\]

(4)

and

\[
P_0 = \left[ \sum_{j=0}^{s-1} \left( \frac{\lambda}{\mu} \right)^j \frac{\lambda^j}{j!} \frac{s\mu}{s\mu - \lambda} \right]^{-1}
\]

(5)

M/M/1 is defined as a non-preemptive queue is to be analysed as a priority service with a single server, Poisson arrivals in each class and identically distributed service time. The expected waiting time in the second queue [9]:

\[
\sum_{k=1}^{r} \frac{\rho_k}{\mu} \frac{\mu}{(1-\sigma_{r-1})(1-\sigma_r)}, \quad \text{for} \quad \sigma_0 = 0
\]

(6)

The total waiting time in the system is to be calculated as:

\[
W_{\text{total}}(\lambda_r, s) = W'_q(\lambda_r, s) + W'_q(\rho\lambda_r, 1)
\]

(7)

4. Performance analysis

In order to analyze the proposed healthcare architecture we focus on having the required data at the right moment of time defining this as QoS criteria. In terms of validating the cloud infrastructure we consider the variance of the waiting time of requests in the system using numerical calculations and
simulations of the mathematical model inputs. Therefore we consider three priority classes 1 (high), 2 (medium) and 3 (low) with a specific arrival rate of requests in the system $\lambda_1=5$ requests/s, $\lambda_2=10$ requests/s and $\lambda_3=15$ requests/s and queue specific service time $\mu_{Q1}=20$ requests/s and $\mu_{Q2}=100$ requests/s. Further we will consider $W_1 =$ waiting time for class1 (high) patients, $W_2 =$ waiting time for class2 (medium) patients and $W_3 =$ waiting time for class3 (low) patients.

For the first queue that is created by the public cloud application (Request management application described in section 3) after FCFS principle defined as M/M/s the waiting time in the queue is dependent on the number of servers that are allocated for the first queue. Assuming the variable defined upper ($r$, $\lambda_r$, $\mu_{Q1}$), waiting time ($W_r$) for each priority class is decreasing considerable if the number of servers, $s$, is increasing (Table 1).

<table>
<thead>
<tr>
<th>No. of servers</th>
<th>Class 1 ($\lambda_1 = 5$)</th>
<th>Class 2 ($\lambda_2 = 10$)</th>
<th>Class 3 ($\lambda_3 = 15$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1666.66</td>
<td>5000</td>
<td>15000</td>
</tr>
<tr>
<td>2</td>
<td>79.36</td>
<td>333.33</td>
<td>818.18</td>
</tr>
<tr>
<td>3</td>
<td>4.02</td>
<td>30.30</td>
<td>98.03</td>
</tr>
<tr>
<td>4</td>
<td>0.18</td>
<td>2.57</td>
<td>11.78</td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
<td>0.19</td>
<td>1.29</td>
</tr>
</tbody>
</table>

Graph plotting (Fig.3) shows that increasing the number of servers significantly decreases the waiting time in the first queue for all priority classes.

A request that is processed by the first queue has the probability $p$ to be placed in the second queue. Considering the structure of the first queue (requests
are processed depended on priority class) we are interested how is influenced the waiting time of the lowest priority classes by the arrival rate of the highest priority class. For these analyses the nominal arrival rate for priority class 1 was multiplied by 1.5. Assuming the variable defined in the beginning of the section \((r, \lambda_r, \mu_Q)\) and multiplying \(\lambda_1\) by 1.5, waiting time \((W_i)\) for each priority class is increasing (Table 2).

### Table 2

<table>
<thead>
<tr>
<th>(\rho \times 10^3)</th>
<th>Class 1 ((\lambda_2 = 5))</th>
<th>Class 2 ((\lambda_2 = 10))</th>
<th>Class 3 ((\lambda_3 = 15)) (base)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.52</td>
<td>1.85</td>
<td>5.04</td>
</tr>
<tr>
<td>0.07</td>
<td>0.81</td>
<td>2.29</td>
<td>5.83</td>
</tr>
<tr>
<td>0.11</td>
<td>1.26</td>
<td>3.04</td>
<td>7.22</td>
</tr>
<tr>
<td>0.16</td>
<td>2.03</td>
<td>4.42</td>
<td>9.85</td>
</tr>
<tr>
<td>0.25</td>
<td>3.38</td>
<td>7.30</td>
<td>15.65</td>
</tr>
</tbody>
</table>

In case of significantly increase of the arrival rate for class 1, \(W_2\) and \(W_3\) are exponentially increased (Fig. 4) but it can be seen that the influence on \(W_2\) is less than \(W_3\) therefore the quality of the system is not dramatically affected since we are focusing on highest priority classes.

![Fig. 4. Impact of increasing arrival rate for class 1 on class 3 and class 2 waiting times](image)

Assuming the variable defined in the beginning of the section \((r, \lambda_r, \mu_Q)\) and the arrival rate for class 2 is increased the total waiting time \((W_1)\) for priority
class 1 is not affected and the waiting time for priority class 3 ($W_3$) is increasing exponentially (Table 3).

Table 3

<table>
<thead>
<tr>
<th>$\rho \times 10^3$</th>
<th>Class 1 ($\lambda_1 = 5$)</th>
<th>Class 2 ($\lambda_2 = 10$)</th>
<th>Class 3 base ($\lambda_3 = 15$)</th>
<th>Class 3 base ($\lambda_3 = 15$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.52</td>
<td>1.85</td>
<td>5.04</td>
<td>126.05</td>
</tr>
<tr>
<td>0.15</td>
<td>0.52</td>
<td>2.63</td>
<td>6.73</td>
<td>126.05</td>
</tr>
<tr>
<td>0.22</td>
<td>0.52</td>
<td>3.99</td>
<td>10.19</td>
<td>126.05</td>
</tr>
<tr>
<td>0.33</td>
<td>0.52</td>
<td>6.65</td>
<td>18.97</td>
<td>126.05</td>
</tr>
<tr>
<td>0.50</td>
<td>0.52</td>
<td>13.19</td>
<td>54.18</td>
<td>126.05</td>
</tr>
</tbody>
</table>

Fig. 5 shows that the impact of increasing the arrival rate of class 2 is not affecting $W_1$ but is dramatically affecting $W_3$ in case this is increased uncontrollable.

Assuming the variable defined in the beginning of the section ($r$, $\lambda_r$, $\mu_Q$) the total waiting time ($W_i$) for each priority class having the probability $p$ [0,1] for reaching the second queue, is not significant affected by the increase of priority class 1 arrival rate (Table 4). Increase of waiting time $W_3$ could affect system performance therefore the total waiting time needs to be analyzed in order to decide if the requests need to be handled depending also on the execution time.
Table 4

<table>
<thead>
<tr>
<th>No. of servers</th>
<th>Class 1 (λ₁ = 5)</th>
<th>Class 2 (λ₂ = 10)</th>
<th>Class 3 (λ₃ = 15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16.90</td>
<td>51.21</td>
<td>153.55</td>
</tr>
<tr>
<td>2</td>
<td>1.02</td>
<td>4.54</td>
<td>11.73</td>
</tr>
<tr>
<td>3</td>
<td>0.27</td>
<td>1.51</td>
<td>4.53</td>
</tr>
<tr>
<td>4</td>
<td>0.23</td>
<td>1.23</td>
<td>3.66</td>
</tr>
<tr>
<td>5</td>
<td>0.23</td>
<td>1.21</td>
<td>3.56</td>
</tr>
<tr>
<td>6</td>
<td>0.23</td>
<td>1.21</td>
<td>3.55</td>
</tr>
</tbody>
</table>

Having the probability $p$ for each request to hit the second queue the performance of the system in terms of total waiting time is important to be analyzed (Fig.6).

Fig.6. Impact of increasing number of servers in $Q₁$ on total waiting time of the requests in the system $W_r(s) = W_{M/M/1}(s) + W_{M/M/1}$

We would like to emphasize that even though increasing the arrival rates for the highest priority classes exponentially increases the waiting time of the lowest priority classes the total waiting time in the system is not affected since there is a probability $p$ that the requests will be placed in the second queue.

5. Conclusions

In this paper we proposed a cloud computing based infrastructure for healthcare environments with focus on accessing radiology data. Using the cloud computing technology healthcare is improving considerably in terms of data
access. This has been analysed using queuing theory providing a mathematical model composed by two systems (M/M/s and M/M/1) based on which a numerical simulation has been done in order to validate the model in terms of quality of service. The performance analysis in terms of waiting time has shown that the proposed architecture significantly improves the quality of the healthcare services. Having an integrated healthcare infrastructure will save physical space and improve the efficiency of the medical staff. Costs of the proposed infrastructure are not higher than another infrastructure since the medical units will only rent the infrastructure to store medical data as they need. Also, the proposed architecture performance is incomparable in terms of quality with a traditional non-integrated informational environment. Even though the medical unit runs an information system, if this is not integrated in an overall healthcare infrastructure the waiting times for data access are very high. The hybrid cloud (public and private) ensures the security of data and communications between departments due to private cloud technology and messages that are used in the public cloud.

Future work will include design and analysis of a resource scheduling service in cloud.

REFERENCES
