Investigating Feasibility of Queuing Model to Eliminate Server Deadlock

Mudiana Mokhsin @ Misron\textsuperscript{1}, Rusyamimi Rosman\textsuperscript{2}, Nor Diana Ahmad\textsuperscript{3}, Anitawati Mohd Lokman\textsuperscript{4} and Hamizan Abdul Halim\textsuperscript{5}

\textsuperscript{1,2,3,4}Universiti Teknologi MARA, Malaysia, \{mudiana,nordiana,anita\}@tmsk.uitm.edu.my, syarosman@gmail.com

\textsuperscript{5}Open University Malaysia, hamizan.abdhalim@tm.com.my

ABSTRACT

In any object that represents a shared resource in a distributed system, it must be responsible in ensuring that it can operate correctly in a concurrent environment. Therefore, programmers should take concern in implementing any objects or processes which will not be intended to use in a distributed system so that it will stay safe in a concurrent environment. This is to avoid from any issue such as server deadlock to be happened. A deadlock occurs when two or more processes are waiting on the same resource and each of the process is waiting on the other process to complete before moving forward. Looking to this scenario, this study has two major purposes which are 1) to identify the occurrence of server deadlock in three different modules in SPP and 2) to propose feasibility of Queuing Model concept as the solution to reduce server deadlock occurrence. In conclusion, this study had showed that Queuing Model offers much more capability in handling deadlock issue. As for the recommendation the utilization of this model is hopefully can reduce the occurrence of server deadlock in SPP.

Keywords: Deadlock, Queuing Model, Transaction Wait for Graph, Two-Phase Locking, Queuing Model

1 INTRODUCTION

Today, Malaysia becomes one of the countries which provide the best healthcare system in region (Malaysian National Integrity Plan, 2005). These include all facilities which can be accessible by any people in Malaysia and the most import part is affordable. Due to the rapid growth of information technology, Ministry of Health (MOH) has taken a further step by introducing Information, Communication and Technology (ICT) in the healthcare system in Malaysia. Information systems have a great potential in reducing health care costs and improving the outcomes. ICT is making it possible for the communities to enhance the quality of healthcare and expand the range of healthcare services available to the population. In fact, the acceptance and implementation of ICT had given the big impact to the organization such as improved efficiency, lowered cost and improved customer satisfaction.

According to Vladimir Zwass (2014) information systems is an integrated set of components for collecting, storing and processing data and for delivering information, knowledge and digital product. Information system does offer its privileges or functions in managing health care costs and improving the quality of care. As most of the activity required in any health care institutions such as collecting patients’ data, storing the record, and as well as to retrieve the data, which all of these functions are provided in information system.

HeiTech Business Solution Group is one of the business outfits of HeiTech Padu Berhad, focusing on creating value for the public sector through implementation of e-Government solutions. HeiTech Business Solution Group is mainly in realizing and supporting the Government to be a customer-centric organization by providing integrated and perfect solutions as well as services to the citizen. Integrated Healthcare Solutions (IHS) focuses on the business of healthcare information management system on the automation and computerization of information system infrastructure in health care institutions. One of its innovations is the Sistem Pengurusan Pesakit (SPP), which is proposed to be implemented in general hospitals throughout Malaysia. Concurrent system or also known as distributed system in healthcare institution can be considered as critical process. From the research, there were a lot of processes involved since there were many departments reside in it, such as Pharmacy, Radiology, Wards and others. SPP or Sistem Pengurusan Pesakit (Patient Management System) has been established since 2008, which was introduced by Ministry of Health, and currently being used in few generals’ hospitals such as Hospital Kajang, Hospital Sungai Buloh, Hospital Seremban, Hospital Kuala Lumpur and few more
listed hospitals. Ministry of Health has planned to implement SPP in all general hospitals in Malaysia by end of this year.

II PROBLEM STATEMENT
As from an interview Fahmi M, 2013, the developer of the SPP, he mentioned that SPP currently is deployed in seven application domains and these domains are used by different departments and there were transactions that might be run concurrently or simultaneously by the SPP users.

There were few problem statements to be addressed in SPP, which is database systems in SPP cannot manage too much open connection in server which leads the application system become slow and server will down, which this always happen or at least will happen for once a month (Fahmi M, 2013). Plus, single or multiple jobs or processes which run simultaneously by many users during peak hour will cause the system performance to degrade, example in module Ward Management, for function release patient.

This research has been proposed in order to solve or reducing the number of server deadlock by examining on the feasibility of the proposed model, which the server deadlock occurs in SPP (Patient Management System). The high number occurrence of server deadlock has caused many internal problems to the health care institutions where sometime it will delay the SPP users’ daily tasks.

The final output of this research project will hopefully give benefit to the different user in SPP either internal or external which will make their process or task in accessing SPP will run smoothly.

III LITERATURE REVIEW
A. Health Information System
Information technology application in healthcare has evolved with the technology today. As it can be considered as a crucial part, where medical instruments itself has develop in order to cater the latest technology. According to Nafees (2007), requirements on the registration and processing of medical services and bills that have been imposed by the government and insurance company have lead to the needs for the hospitals to use an integrated application system. Besides, all the systems in the hospitals should be integrated as it will be accessed by numbers of different users such as doctors, nurse, and physicians. But still, in a paper by Menon (2000), he stated that information technology implemented in healthcare industry is still relatively slow.

B. Distributed System
Distributed computing systems have been more popular these years and are widely used .Various definitions of distributed system has been made by varies of literature and authors. Meho V. et al. (2008) in their article, they defined distributed system as a system which consists of a serial independent processor unit and consist of concurrent processes accessing distributed system resources which are shared through message passing in the network environment. This shows that a distributed system may consist of many independent workstations at the same time. According Sklenar J. (n.d.), distributed systems have proven which they have gained several significance importance such as geographically distributed environment, speed up, resource sharing and fault tolerance. The implementation of distributed system can be done at any place which means it is accessible anywhere. Next, there is a need in a high speed for this kind of computation. Distributed system also can be considered as the best example of resource sharing where a large database or a system may be stored in several host machines, and being consistently updated or retrieved by number of users.

C. Performance Evaluation on Distributed System
Performance evaluation of computer and telecommunication systems has become an increasingly important issue. The process of evaluation for computer systems and networks are needed in every stage of computer system lifecycle. The needs of performance evaluation is needed when there is new technology in the market and the system designer want to compare or wants to do a benchmarking for alternative design, and also in considering the cost, to find the less or most cost effective technology. As nowadays there is a lot of latest and robust system and technology, there will be also tight competitions with other available systems in the market.

Real time database systems are identified as having timing constraint where it can be also found in health care systems applications where this timing constraints of real time database are typically specified in any deadline that need to be met or require a transaction that need to be completed by a specific time, where failure to do so might cause the results to lose its value or may cause delay on its.
D. Concurrency in distributed System

Distributed system currently has a lot of applications such as wireless communications, distributed transactions, office management and also high critical system such as aircraft navigation and also military system. It is important to improve the system operating efficiency as to enhance as well as the implementation of the concurrency transaction. The goal of this concurrency control is to allow several transactions to be executed simultaneously.

Concurrency control problem is the activity of coordinating concurrent process in database system. The main difficulty that has widely been addressed is to prevent the database updates being performed by one user without interfering with database retrievals or updates which also being done by other user. Concurrency control is the important consideration in designing a distributed system. This concurrency control is concern with deadlock where it should avoid from it to happen as it is important in maintaining the consistency of the database system. It is important in ensuring that during operation of any processes, each process should sees a consistent picture of the database, each process eventually terminates and finally the final database after all the processes terminate is consistent.

E. Server Deadlock

System availability, reliability and predictability are becomingly more important as system size and demand kept increased. This is happened to high end servers as example, web, database and data centers which are currently based on highly parallel computers. In these systems, the interconnected networks play important roles in achieving high system availability. Armoni (2000) said concurrency control and deadlock detection which is the most important problem which need more attention in sharing information in distributed system.

Deadlock control technologies can be divided to three parts which is deadlock avoidance, deadlock detection and deadlock prevention. Deadlock avoidance is where it only accepts request which will lead to safe states, deadlock detection is where it tracks resource allocation and process states and deadlock prevention removes the “hold and wait” conditions. According Mokdad L. et al. (2012), deadlock is a persistent and circular wait condition which means that each process in involved in deadlock need to waits for resources held by other processes while holding the resources needed by others as well. Deadlocks also are a common type of faults in distributed system. Armoni (2000) said it also has been identified as the most serious problems in concurrent programming systems. The authors added that to detect and optimize the deadlock issue is very difficult in a distributed database system. Therefore, Ling Y. et al. (2006) said deadlock occurs in transactions in distributed system where concurrent access of data resources by transactions might affect the consistency of data resource. Deadlock has to be prevented as it can lower the runtime cost and also better response time. Hold and wait results in deadlock when it leads to circular waits plus it is more tedious to check the circular wait condition in distributed system compared to centralized system. In reducing the wait times, all the transactions should get all the resources they need first instead of doing the rollback of transactions on detection of deadlock which this will cause additional overhead.

F. Queuing Model

As the technology has evolved in today’s information age, where all activities are highly interdependent in sharing resources, thus, waiting lines or processes in queues is a common phenomenon. Queuing theory has been introduced by a Danish engineer, A K Erlang, where he applied this theory to study on the behavior of telephone. His theory also has developed some queuing results which still remain as the backbone of evaluation in queuing performance till today. According Meho V. et al. (2008), queuing theory is one of the fields in stochastic modeling, which defines as an application of probability theory and analysis of real world problems.

Queuing theory is the mathematical study of waiting in lines or queue. Breuer & Baum (2005) said this queuing theory enables mathematical analysis of the related processes such as arriving at the queue, waiting in the queue and being served at the queue. In addition, he also claimed that this theory can be as benchmark in evaluating performance of a system such as measure the average waiting time in the queue or waiting time in the system and any other processes that might take place while waiting in the queue or system. Queuing model is helpful in determining how to operate a queuing system in most effective way, as it helps on how to perform the service.
### III METHODOLOGY

In accomplishing this research, the methodology that has been applied consists of five main phases which are problem identification and planning, data collection, knowledge acquisition, queuing model concept analysis and implementation and lastly result and finding. All five phases has been summarized in Table 1 as follows:

<table>
<thead>
<tr>
<th>Phase</th>
<th>Activity</th>
<th>Deliverable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Identification and Planning</td>
<td>identifying research background, problem statement, research objective, research scope and research significance</td>
<td>Research background, problem statement, research objective, research scope and research significance are clearly defined.</td>
</tr>
<tr>
<td>Data Collection</td>
<td>1) interview with stakeholders involved - software engineers, implementation analysts, nurse engineers, SPP users, nurses and doctors</td>
<td>1) interview with stakeholders involved - software engineers, implementation analysts, nurse engineers, SPP users, nurses and doctors</td>
</tr>
<tr>
<td>2) interview with stakeholders involved - software engineers, implementation analysts, nurse engineers, SPP users, nurses and doctors</td>
<td>2) interview with stakeholders involved - software engineers, implementation analysts, nurse engineers, SPP users, nurses and doctors</td>
<td></td>
</tr>
<tr>
<td>Knowledge Acquisition (Data Collection)</td>
<td>Analyzing interview findings related to server deadlock as well as its impact</td>
<td>Knowledge acquisition (Data Collection) Analyzing interview findings related to server deadlock as well as its impact</td>
</tr>
<tr>
<td>Queuing Model Concept Analysis and</td>
<td>3) interview with stakeholders involved - software engineers, implementation analysts, nurse engineers, SPP users, nurses and doctors</td>
<td>3) interview with stakeholders involved - software engineers, implementation analysts, nurse engineers, SPP users, nurses and doctors</td>
</tr>
<tr>
<td>Implementation</td>
<td>3) interview with stakeholders involved - software engineers, implementation analysts, nurse engineers, SPP users, nurses and doctors</td>
<td>3) interview with stakeholders involved - software engineers, implementation analysts, nurse engineers, SPP users, nurses and doctors</td>
</tr>
</tbody>
</table>

Table 1. Five Main Phases in the Methodology.

### IV RESULT AND ANALYSIS

Each objective will be explained and discussed on how data has been gathered and analysis on the result. All data collected from stakeholders which is mostly from the interviews will be analyzed based on the understandings and knowledge gained from journal reviews as well as details from the stakeholders. For first objective, researcher will explain on how deadlock occurrence were detected and for second objective, researcher will need to dig into the current state of SPP server and database architecture and how the process of the transactions being done, and finally how Queuing Model concept can be implemented there in order the occurrence of server deadlock can be reduced.

#### A. Objective 1 – To identify the occurrence of server deadlock in three different modules in SPP

In this objective, numbers of interviews have been done with few stakeholders such as software engineers, onsite engineer, as well as SPP users’, doctor and nurses. Occurrence of server deadlock has been collected based on server log provided by software engineers. Besides, onsite engineer also explained on the database and server architecture which also relates with the server deadlock occurrence. While for the SPP users, they have shared their user experience in using SPP system. Further analysis on the interview and data gathered are as per below and for this part researcher only includes summary of interview from SPP’s software engineers and SPP’s users.

In achieving the first objective, occurrence of server deadlock has been detected in three identified modules which are Create Order, Transfer Patient and Ward Management. Create Order modules consist of three parts in it which is in Laboratory, Medication and Radiology. This module is used for creating any new orders in these three areas. Figure below shows the example of deadlock captured in server log.

**Figure 1. Sample Deadlock captured in SPP’s Server Log for the Three Selected Modules.**

#### B. Objective 2 – To propose feasibility of Queuing Model concept as the solution to reduce server deadlock occurrence

In achieving this objective, detailed interview has been done thoroughly with SPP onsite engineer. Based on the interview, he explained that the server and database architecture which can be considered as a part of process of the concurrent transactions in SPP distributed system. As the main technical difficulty in achieving the goal of concurrency control is to prevent any updates from being interfered with other processes such as database retrieval or updates performed by other users. This queuing model or queuing theory has been chosen as it can act as a tool which helps to compute the size of the queues and also the time for each jobs spend in it.

He also stated that currently SPP is implementing the concept of database connection pool, which this concept is being used to solve the problems in accessing database. This connection pooling concept enables administrator to establish a pool of database connections which any applications can share on the available application server. Based on
As described in previous section, the concept of database connection pool being implemented in order to control the multi execution or concurrency process in SPP but still it produce deadlock to uncontrolled processes coming in from the users. In Queueing Model concept, by knowing the number of resources which is in this case is the number of server, it is able to calculate on the optimal time of a job should spend in a queue, with the equation derived from it.

The selection of Queueing Model is due to its queuing theory which based on a number of assumptions about the behavior of users and also the nature of distributions in modeling their arrival rates since in many applications such as SPP, even when the arrivals rate are continuous and unpredictable so the arrivals exhibits an exponential or "memory less "distribution (Elleithy & Komaralingam, 2011). Then number 1 denotes to the number of server where here in SPP only one main database server will be used in processing the transactions.

As being explained in previous section, the concept of database connection pool are being implemented in order to control the multi execution or concurrency process in SPP but still it produce deadlock to uncontrolled processes coming in from the users. In Queueing Model concept, by knowing the number of resources which is in this case is the number of server, it is able to calculate on the optimal time of a job should spend in a queue, with the equation derived from it.

The main concern here is on the time that a job or the transaction spent and the response time. As being suggested by the onsite engineer itself, he stated that in avoiding this issue, programmer or software engineers need to fine tune on the coding or scripts to give a better performance. Besides, connection pool size also can be altered by adding more size to it.

D. Analysis on Queuing Model Concept – M/M/1 Queue Model

According to Maysam Hedayati (2010), basic queuing theory stated that every service or resources may consist of a single queue feeding one or more servers. As in every queue there will be an arrival rate (A), which is the average rate at which new jobs arrive at the queue. Then time of the average amount that a server takes to process a job (Ts), while (Tq) is where the average amount of time for each job spends in queue. According to the queuing theory, if the arrival rate is less than the service rate (1/ (Ts)), the queue is stable. While for a Markov Model queue concept, which is will be proposed for this research project according to the current state of SPP’s system, M/M/1 queue will be used since the amount of time between job arrivals is random and unpredictable so the arrivals exhibits an exponential or “memory less “distribution (Elleithy & Komarlingam, 2011). Then number 1 denotes to the number of server where here in SPP only one main database server will be used in processing the transactions.
the Queuing Model itself did not limit on the time of job spent in a server as how database connection pooling concept did, where in Queuing Model concept, if a customer or transaction arrives at a time the server is unavailable to provide service to it, it is forced to wait in the queue temporarily until it can start receiving service. These authors also added that if there is more than one transaction in the waiting queue, where at the time the server is available, the concept of queue discipline will be used, such as either FIFO or LIFO or any queue discipline that has been implemented by the system.

Deadlock also can be categorized as a blocking in any concurrent systems with finite capacity resources, as example a single server. When a queue reach its maximum capacity and the flow of process to the servers are stopped, and in this case if another process from other domain also coming in at the time the servers stopped, deadlock may happen. In modeling this Queuing Model concept, researcher consider a queuing model formed by M servers, which we will define on the number of servers, service time distribution, queue capacity and also queue discipline. We define the arrival process at each node $i$, $1 \leq i \leq M$. Let $\lambda_i$ denotes the total arrival rate and $P0_i$, $1 \leq i \leq M$, is the probability of the arrival tries to enter node $i$. Then the Poisson arrival process at node $i$ has parameter $\lambda P0_i$. Let $P = \{ P_{ij} \}$ ($1 \leq i, j \leq M$) where $P_{ij}$ is the probability that a transaction or a job may leave node $i$ and try to enter node $j$. Then, let $Si$ denote the state of node $i$, where it comprise of numbers of processes or transactions in node $i$, which is denoted by $n_i$, and other components such as service or queue discipline and service time distribution. Next, let $\mu_i$ denote the node’s service rate, which is $1/\mu_i$, the average service time. Let $B_i$ denotes the maximum number or processes entered in node $i$ than in the queue and the servers ($B_i = c + s$), $1 \leq i \leq M$. thus, the total number of jobs or processes in node $i$ will satisfies the constraint $n_i \leq B_i$ (Balsamo et al., 2003). So here, when the queue reaches to its finite capacity or loads of processes in it ($n_i = B_i$), the node is considered as full and deadlock may happens. So in order to avoid or reduce the deadlock from occurs, this concept can be applied by maintaining the number of jobs in a queue as being proposed in this concept. This also can be proved where by using the concept of queuing theory, it able to help in computing the size of the queues and also the time that jobs spend in them (Elleithy & Komaralingam, 2011).

V CONCLUSION

Based on the overall analysis and arguments that has been discussed in previous section, here the researchers can conclude that by using Queuing Model concept, which offers much more capability in handling deadlock issue, are hopefully can reduce the occurrence of server deadlock in SPP. This can be further be seen when experimental testing or simulation are being done by using this concept.

The feasibility of concept offered by Queuing Model and M/M/1 Model which can suits best on current SPP architecture, as current implementation, database connection pool still not able to cater on the problems of server deadlock. Since database connection pool has limited capabilities, it only can serve on the maximum time of each pool. It needs to be determined by its administrator based on the use of each application domains. In other words, it has to be seen on the usage of the applications domains whether it contains a lot of processes and users involved in it or vice versa.

ACKNOWLEDGMENT

This research is supported by Research Management Institute, Universiti Teknologi MARA, Malaysia, and the Ministry of Education Malaysia under the RAGS Grant Scheme [Project Code: 600-RMI/RAGS 5/3 (79/2012)]. Gratitude and appreciation also sent to Heitech Business Solution team who gave valuable contribution towards the completion of this project.

REFERENCES


Nafees, A. (2007). Queuing Theory and Its Application Analysis of the Current SPP architecture. As current implementation, database connection pool still not able to cater on the problems of server deadlock. Since database connection pool has limited capabilities, it only can serve on the maximum time of each pool. It needs to be determined by its administrator based on the use of each application domains. In other words, it has to be seen on the usage of the applications domains whether it contains a lot of processes and users involved in it or vice versa.

ACKNOWLEDGMENT

This research is supported by Research Management Institute, Universiti Teknologi MARA, Malaysia, and the Ministry of Education Malaysia under the RAGS Grant Scheme [Project Code: 600-RMI/RAGS 5/3 (79/2012)]. Gratitude and appreciation also sent to Heitech Business Solution team who gave valuable contribution towards the completion of this project.

REFERENCES


Nafees, A. (2007). Queuing Theory and Its Application Analysis of the Current SPP architecture. As current implementation, database connection pool still not able to cater on the problems of server deadlock. Since database connection pool has limited capabilities, it only can serve on the maximum time of each pool. It needs to be determined by its administrator based on the use of each application domains. In other words, it has to be seen on the usage of the applications domains whether it contains a lot of processes and users involved in it or vice versa.