Exploring Queue Behavior in Local Cloud Buffer Using Markov Chain

†Shashi Raj Pandey and *Choong Seon Hong
Department of Computer Engineering, Kyung Hee University, Yongin, 446-701 Korea
Email: {†shashiraj,*cshong}@khu.ac.kr

Abstract
While the excessive growth in mobile applications, Mobile Cloud Computing (MCC) has been proposed to maximize the computational performance of mobile nodes. For further enhancement cooperative local and remote (server) cloud concepts are evolving where the two computing and storage resource sharing partners interact with each other to improve the performance of MCC. However, the computational resources of local cloud are usually limited and therefore efficient queuing model design becomes critical during such cooperation. This paper analyzes the queue behavior in the local cloud with limited computational resource applying Markov Chain techniques and provides the convergence in probability for each state.

1. Introduction
In recent years the amount of mobile applications has increased rapidly. To facilitate this growth in use of mobile application Mobile Cloud Computing (MCC) has been proposed. MCC helps mobile device to save energy and improve their computational performance by providing abundant computational resource available cloud platform to offload application tasks. However, the real problem about dealing with latency sensitive task is crucial which directly affects the quality of service (QoS). Therefore it is desired to bring both computational and storage resources as near to mobile users as possible. This is the primary objective of local cloud.

To address the delay requirement in MCC [1] cloudlet is proposed to deploy some local cloud servers such that the delay requirement is met. But still cloudlet has some computational limitation. Taking advantages of cloudlet in [2] authors proposed an architecture to associate cloudlets which in fact helped to overcome computational resource limitation of cloudlet.

Figure 1 A cellular network with local and remote cloud
Also, the evolution of Internet of Things (IoT) has emphasized upon the efficient modeling of local cloud computing platform such as Fog Computing. In these scenario authors [3] has proposed a cooperative scheduling scheme that improves the QoS by managing proper execution of delay sensitive applications in between local and cloud servers. Generally, the role of
scheduler is exploited in most of the models to handle queues in the buffer so as to meet the provided objectives effectively.

In this paper we have modeled a queuing system in a local cloud and analyze its behavior upon different dependent parameters related to local cloud computational resource and arrival of incoming data packets for a specific task. For this purpose we have used Markov Chains to formulate our queuing system.

At first we find the stationary distribution probability function for our model with necessary condition for convergence and evaluate the system’s limiting probability behavior with queue lengths upon different arrival rates [4]. We also simulate the system to realize the convergence for further analysis.

2. System Model

Consider a local cloud sub-system from Fig (1). Here, we assume that local cloud has S number of homogenous virtual servers. We have tasks of type i (=1, 2, ..., N) arriving towards the cloud with the rate defined by poison processes as \( \lambda_i \). Let \( \mu \) be the service rate of the system defined as an exponential function which is same for each type of tasks. Considering for concurrent arrival and departure, when data arrives at the local cloud the scheduler passes it in the queue with its own kind. For this purpose we have buffer for each task which is the size of queue length pre-defined runs from 0 to \( L \). Here [4] we can easily prove our system as a Ergodic Markov chain (i.e, irreducible, aperiodic and positive recurrent). Therefore, we can get the stationary distribution \( (\pi_j) \) by solving the balanced equation with limit distribution equations. In reference to fig [2], we have

\[
\mu(i-\lambda) \pi_i = \lambda(1-\mu) \pi_0 \quad (1)
\]

\[
\mu(i-\lambda) \pi_i = \lambda(1-\mu) \pi_{i-1} \quad (2)
\]

and

\[
\sum_j \pi_j = 1 \quad (3)
\]

Solving recursively eq (1) and using (2) and (3) we can have;

\[
\pi_j \approx (1-\alpha) \alpha^j \quad (4)
\]

where,

\[
\alpha = \lambda(1-\mu)/\mu(1-\lambda)
\]

The general model of queue evolution is implemented to realize the convergence of Markov Chain for different initialization and prove the ergodicity.

4. Simulation Results

This section explains results obtained from simulation for the formulated problem.

For showing the convergence we have chosen arrival rate for a particular task \( \lambda \) as 0.3 and service rate as 0.33. The buffer size of system for each task is considered to be same (\( L = 10 \)).

![Figure 2 Markov Chain System for a single task](image)

![Figure 3 Limiting probabilities for different \( \lambda \) (L=60)](image)
Fig.4 and Fig.5 shows the convergence of the state probabilities with different initialization. Here we observe the final results are independent with initialization.

5. Conclusion and Future Work

The limitation of local cloud computing resources and the proliferation of mobile applications have challenged MCC to seek coordinated queue design in the local cloud. For this purpose it is important to model local buffer size such that it addresses the latency, time-delay constraint and computational issues in an effective way. This paper at first models the local cloud system as a markov chain, and then proves its convergence for different initialization.

We found out that for a fixed service rate, the size of a buffer plays an important role for tasks with different arrival rates. It means the design of buffer should be done such that it restricts queue overflow situation while ensuring computational and delay constraints for the execution of an arriving task. To manage this, an improved and efficient cooperation model between Local and Remote cloud relative to local buffer’s design is considered as a future work.

Acknowledgment

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (B0190-16-2017, Resilient/Fault-Tolerant Autonomic Networking Based on Physicality, Relationship and Service Semantic of IoT Devices) *Dr. CS Hong is the corresponding author

REFERENCES


