Morkov Chain based Analytics of Requests Caching at Network Edge

Conference Paper · January 2017

Umer Majeed
Kyung Hee University

Morkov Chain based Analytics of Requests Caching at Network Edge

Umer Majeed and Choong Seon Hong
Department of Computer Engineering, Kyung Hee University, Yongin, 446-701 Korea
{umermajeed, cshong }@khu.ac.kr

Abstract

Edge computing is the solution to increasing computing demands by the ubiquitous proliferation of Internet of Things (IoT) services. However, for resource constraint IoT gateway and time sensitive applications, the data processing must be offloaded to cloud computing. In this paper, we modeled the number of request cached (for advanced processing on cloud) at IoT gateway as Morkov chain based queue. We found that the said Markov chain is ergodic and independent of initial state with certain steady state probabilities.

1. Introduction:

Edge computing also known as Fog computing is emerging technology to meet the computational demands of Internet of Things applications. However, if the IoT gateway or edge has limited computational or energy capabilities, the tasks are offloaded to cloud computing [1].

The computational and resource capabilities of IoT gateways can be as high as of a devoted IoT gateway such as an industrial router for a commercial IoT network or as low as of a smart mobile phone for wearables and personalized IoT devices [2].

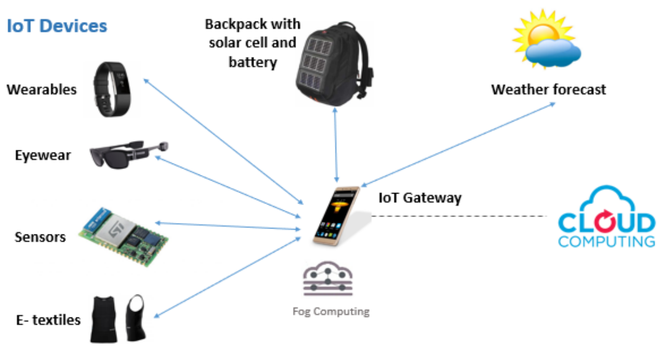

Figure 1: architecture of a personalized IoT system with a mobile IoT gateway powered by battery and renewable power

Figure 1 shows a typical personalized IoT architecture in which IoT devices sends request to resource constraint IoT gateway for advanced processing at cloud.

This paper models the number of requests (cached at network edge for further processing at cloud) as a Morkov chain based queue.

Rest of the document is formulated as follows. Section 2 will give system model followed by problem formulation in section 3. Section 4 shows Simulation Results and in section 5, we will conclude our research work.

2. System Model:

Our system model is IoT Architecture as shown in Figure 1. Different IoT devices send request to IoT gateway for processing at cloud. These requests are cached at IoT gateway also called “Edge”. The Edge has capacity of caching L requests. For simplicity Let suppose It caches $\bar{\lambda}$ requests per unit time and transfer $\bar{\mu}$ requests per unit time to the cloud on average. The time can be divided into time slots $\triangle$t where $\triangle$t is average time to either cache a request or transfer request to cloud. The edge is resource constraint. So, in a time slot, it can either cache a request which we call “arrival” of request in queue at edge or transfer it to cloud which we call “departure” of request from queue at edge. If the Edge is busy in arrival or departure of a request, it is responsibility of IoT devices to delay their requests until Edge is available.

3. Problem formulation:

The probability of arrival of requests in queue at edge in △t is $\lambda = \bar{\lambda}\triangle t$ and the probability of departure of requests in △t is $\mu = \bar{\mu}\triangle t$.

Since, there can be no simultaneous arrival or departure at queue because of resource constraint at edge. The number of requests i pending in queue at edge at time t can modeled as Morkov chain with L+1 states. The state 0 is when there are no request pending in queue and state L when queue is full up to its maximum capacity.

If queue is at state $i \in (0, L)$, then both arrival and departure is possible. But it can also be a scenario when no requests arrive at queue and there is no departure because of worse wireless network conditions. Transition probabilities of states for $i \in (0, L)$ can be given by

$$P_{i,i-1} = u \qquad (1)$$
$$P_{i,i} = 1 - \lambda - u \qquad (2)$$
$$P_{i,i+1} = \lambda \qquad (3)$$

However, $P_{0,0} = 1 - \lambda$ and $P_{0,1} = \lambda$ because at state 0 only arrival is possible and there is no request pending for departure.

Whereas, $P_{L,L} = 1 - u$ and $P_{L,L-1} = u$ because queue is full at state L. no more requests can arrive in queue. Only departure is possible. The State representation diagram of Markov Chain based queue is shown in Figure 2.

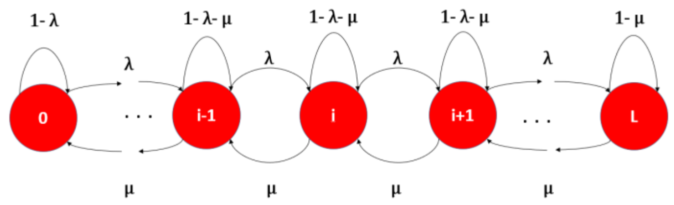

Figure 2: State Transition Diagram

We can see that Markov chain is clearly ergodic (irreducible, aperiodic and positive recurrent). Equivalent Transition Probability Matrix P of above Markov Chain is square matrix of order L+1. Most of the entries are zero.

It can be given as follows:

$$P = \begin{bmatrix} 1-\lambda & \lambda & & & & & \\ \mu & 1-\lambda-\mu & \lambda & & & & \\ & \mu & 1-\lambda-\mu & \lambda & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & \mu & 1-\lambda-\mu & \lambda \\ & & & & & \mu & 1-\mu \end{bmatrix}$$

The initial distribution **p**(0) is vector of probabilities of Markov chain states at initial time. Initial time can be a time when network condition has become good enough for departure and we have accumulated certain number of requests with probability **p**(0). The order of **p**(0) is L+1. We can now find probabilities of Markov chain based queue states at any time slot n as by below matrix product:

$$\boldsymbol{p}(n) = (P^n)^T \boldsymbol{p}(0) \qquad (4)$$

4. Simulation Results:

We have implemented the proposed queue model to observe the results. Markov Chain state distribution is used to find the probability of queue states at n time slot. We set cache capacity $L = 100$.

First, we set arrival probability as $\lambda = 0.3$ and departure probability as $u = 0.33$ and we did simulation for different initial conditions. we assume $\boldsymbol{p}(0) = [1, 0, 0, \dots\dots]^T$ that is queue is empty at start of process. Then, we assume $\boldsymbol{p}(0) = [\dots\dots, 0, 0, 1]^T$ that is queue is full at start of process.

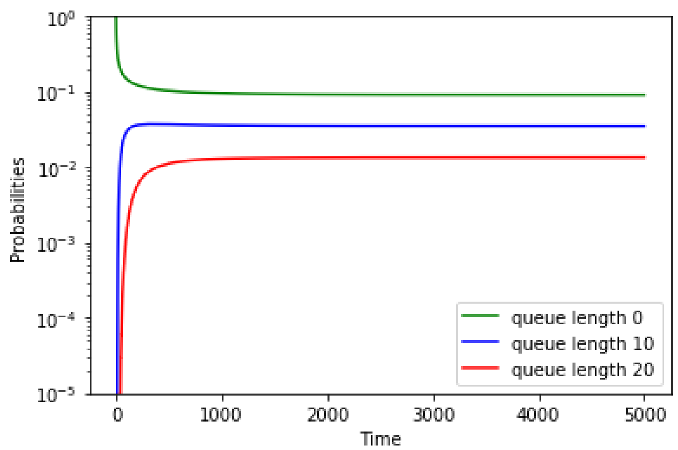

Figure 3: probability of queue states 0, 10, 20 at Time n and $\lambda = 0.3, u = 0.33$, $\boldsymbol{p}(0) = [1, 0, 0, \dots\dots]^T$

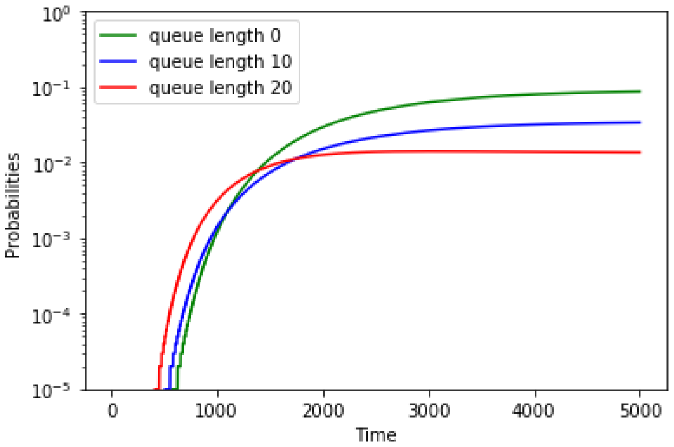

Figure 4: probability of queue states 0, 10, 20 at Time n and $\lambda = 0.3, u = 0.33$, $\boldsymbol{p}(0) = [\ \dots\dots, 0, 0\ , 1]^T$

In Second phase, we set $\lambda = 0.33$ and $u = 0.3$. first, we assume $\boldsymbol{p}(0) = [1, 0\ , 0, \dots\dots]^T$ and then $\boldsymbol{p}(0) = [\ \dots\dots, 0, 0\ , 1]^T$

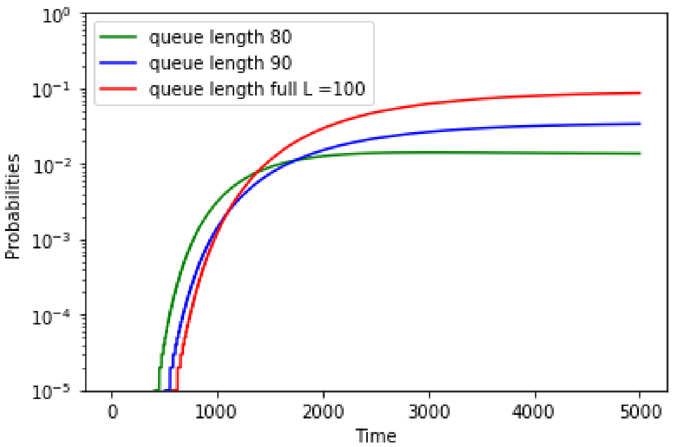

Figure 5: probability of queue states 80, 90, 100 at Time n and $\lambda = 0.33, u = 0.3$, $\boldsymbol{p}(0) = [1, 0\ , 0, \dots\dots]^T$

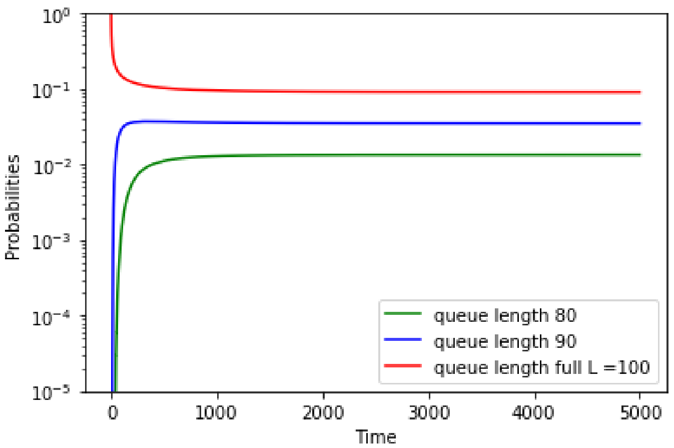

Figure 6: probability of queue states 80, 90, 100 at Time n and $\lambda = 0.33, u = 0.3$, $\boldsymbol{p}(0) = [\ \dots\dots, 0, 0\ , 1]^T$

We find that queue states probabilities become independent of initial conditions after significant number of time slots. However, the number of time slots needed for convergence depend upon initial conditions.

We also found symmetricity in probabilities of states for $\lambda = 0.3, u = 0.33$, $p_i(n)$ for $i = 0, 10, 20$ and $\lambda = 0.33, u = 0.3$, $p_i(n)$ for $i = 100, 90, 80$ which can be seen in Figure 3-6.

We also found that if the probability of arrival is less than probability of departure the probability of empty queue will be high.

5. Conclusion and Future Work:

In this paper, we modeled and analyzed the number of requests cached at edge as Morkov chain based queue. We found that state probabilities become independent of initial conditions after substantial time slots. The analysis can be useful for further insights and statics.

Acknowledgement:

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF- 2017R1A2A2A05000995). This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No.2015-0-00557, Resilient/Fault-Tolerant Autonomic Networking Based on Physicality, Relationship and Service Semantic of IoT Devices). *Dr. CS Hong is the corresponding author.

[References]

[1] Y. Mao, C. You, J. Zhang, K. Huang, K. B. Letaief, A survey on mobile edge computing: the communication perspective, (2017).

[2] Jalali, F., Smith, O. J., Lynar, T., & Suits, F. (2017, August). Cognitive IoT Gateways: Automatic Task Sharing and Switching between Cloud and Edge/Fog Computing. In *Proceedings of the SIGCOMM Posters and Demos* (pp. 121-123). ACM.