Preemptive and Non-preemptive Service of IoT Traffic Flows

Dimitar Atamian and Boris Tsankov

Abstract — **This paper investigates the Internet of Things traffic performance and Quality of Service guarantee strategies. Preemptive and non-preemptive services for different traffic flows are considered. Analytical and simulation approaches are used for Quality of Service parameters evaluation for both cases of exponential and constant service time distribution. Numerical results are presented and some conclusions are drawn.**

Keywords – **Internet of Things, Quality of Service, Priority queue discipline, Queue length and Waiting time.**

I. Introduction

The term Internet of Things (IoT) was first introduced a little bit more than a decade ago. Eight years ago ITU formally defined the term and the concept of IoT. In short it is the integration of information space and physical space.

IoT traffic sources generate heterogeneous traffic flow with messages of information with different importance and urgency that reflects to different Quality of Service (QoS) constrains. QoS guarantee for different IoT traffic flows is an important issue [1] - [3] and for healthcare monitoring in particular [4]. A usual way to guarantee QoS requirement is activation of different priorities. Any violation of the natural way of service, namely on a first-come, first-served basis might be considered as priority activation. In priority systems the customers with higher priority are chosen for service ahead of those with lower priority no mater of their arrival time. Priority queues are generally more difficult to model than non-priority ones. But it can be shown that as far as the selection for service from the queue is not related to service time size, one can apply tractable analytical models for QoS parameters evaluation [5, p. 323] as it is done in this paper. For some systems this is not possible and we engage simulation methods.

A priority is considered preemptive if the arriving message of higher priority interrupts (preempts) the service (if any) of the lower priority message. In case of non-preemptive priority (HOL, Head Of the Line priority) discipline the arriving message does not interrupt any service in progress but it is put to wait in the queue ahead of any waiting lower priority massages. The traffic flows of delay sensitive services are granted priorities and some like life critical warnings are granted preemptive priority. Preemptive and non-preemptive priority queuing systems for a long time ago are subject of many mathematical or telecommunication scientific papers or books to mention a few more resent books [5 – 9].

II. The teletraffic system

The corresponding teletraffic serving systems considered are M/M/1 and M/G/1. We restrict to only two offered traffics: one high priority, denoted with $A_h = \lambda_h/\mu_h$ and a low priority denoted with $A_l = \lambda_l/\mu_l$, where λ and μ are the arrival rate and the service rate accordingly. Notation L indicates the mean message number in the system, and L_q indicates the mean message number in the queue. Similarly W and W_q the mean waiting time in the system and the mean waiting time in the queue, respectively. The well-known and very useful Little's formula

$$L = \lambda W \tag{1}$$

and

$$L_q = \lambda W_q \tag{2}$$

apply for both high and low priority traffics substituting values with corresponding flow indexes. Other useful relations hold: $L_{q,h} = L_h - A_h$ and $L_{q,l} = L_l - A_l$.

No limitation on the queue length is imposed. This assumption is quite reasonable as IoT messages a relatively short and with modern technology is easy to build big enough buffers that are practically unlimited and almost never overflow. On the other hand for the important delay-sensitive applications the waiting time is the most restrictive among the QoS parameters. It is obvious that for unlimited queue

$$A_h + A_l = A < 1.$$

III. Performance analysis for M/M/1

Most wireless and Internet IoT data transmission can be modeled by means of a single server queue [3]. Interarrival times and service time are assumed to be exponentially distributed with this queuing system. That permits one to build a Continuous Time Markov Chain and a system of difference equations to be derived for the stationary queuing system

[1] Dimitar Atamian is with the Faculty of Telecommunications. at TU-Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, dka@tu-sofia.bg

[2] Boris Tsankov is with the Faculty of Telecommunications. at TU-Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. bpt@tu-sofia. bg

probabilities. Therefore one could apply analytical methods for performance investigations.

A) *Preemptive Priority*

The queuing system in this case is easier as the presence of messages of low priority class has no effect on the service of high priority messages. Therefore mean number of high priority messages in the queue is

$$L_{q,h} = \frac{(A_h)^2}{1-A_h} \tag{3}$$

as it is well known from the M/M/1 queuing system without priorities. For the low priority messages there is

$$L_{q,l} = \frac{A_l - A_h A_l - A_h A_l \frac{\mu_l}{\mu_h}}{(1-A_h)(1-A_h-A_l)} - A_l \tag{4}$$

The mean waiting time is obtained applying the Little's formula (2). For traffics with equal service rate (that is quite possible case in practice) $\mu = \mu_h = \mu_l$ and Eq. (4) is simplified.

B) *Non-Preemptive Priority*

The expressions here are a bit more complex (see for example [7]):

$$L_{q,h} = \frac{\lambda_h\left(\frac{A_h}{\mu_h}+\frac{A_l}{\mu_l}\right)}{1-A_h} \tag{5}$$

$$L_{q,l} = \frac{\lambda_l\left(\frac{A_h}{\mu_h}+\frac{A_l}{\mu_l}\right)}{(1-A_h)(1-A)} \tag{6}$$

IV. Performance analysis for M/D/1

Many applications of data transmissions (and IoT, in particular) the service times are not exponentially distributed. Often the service time is constant [8, ch.13.5]. Unfortunately for the M/G/1 queuing systems with priority the mathematical expressions are cumbersome [5, ch.6], [8, ch.13.4], [9, ch.5.4]. We conduct simulations for preemptive end no-preemptive systems and constant service time (queuing system M/D/1).

For M/D/1without priorities one can apply the well known Pollaczek-Khintchine's formula. (See for example [8, ch. 13.5]). The system parameters of M/D/1like *L* and *W* are exactly two times less than corresponding M/M/1parameters. Therefore we make the heuristic proposal to apply (3)... (6) for constant holding time after a division of two:

Preemptive priority:

$$L_{q,h} = \frac{(A_h)^2}{2(1-A_h)} \tag{7}$$

Non-Preemptive priority:

$$L_{q,h} = \frac{\lambda_h\left(\frac{A_h}{\mu_h}+\frac{A_l}{\mu_l}\right)}{2(1-A_h)} \tag{8}$$

$$L_{q,l} = \frac{\lambda_l\left(\frac{A_h}{\mu_h}+\frac{A_l}{\mu_l}\right)}{2(1-A_h)(1-A)} \tag{9}$$

and compare the non-preemptive priority results with the simulation for the same system parameters. The obvious aim is to evaluate how reasonable is to apply the Eqs. (8) and (9) for M/D/1 system in IoT applications.

V. Numerical results

An index *s* on the following figures denotes the results from a real-time simulation.

Non-preemptive priority simulation is compared to calculations based on Eqs. (8) and (9). The simulation is performed for two streams with equal traffic – equal packet rate from 100 to 800 packets per second and fixed and equal packet size of 150 bytes. Results are shown on Fig. 1.

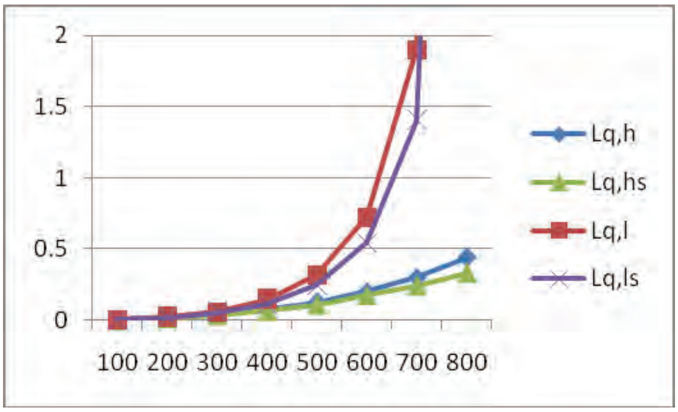

Fig. 1. Non-preemptive priority queue length

Additional simulations are performed with a variety of parameters for both classes of traffic flows and non-preemptive priority. Different service time (packet length) – 1:5 for high:low priority streams with equal packet rates is shown on Fig. 2 as an example compared with analytic results from Eqs. (8) and (9):

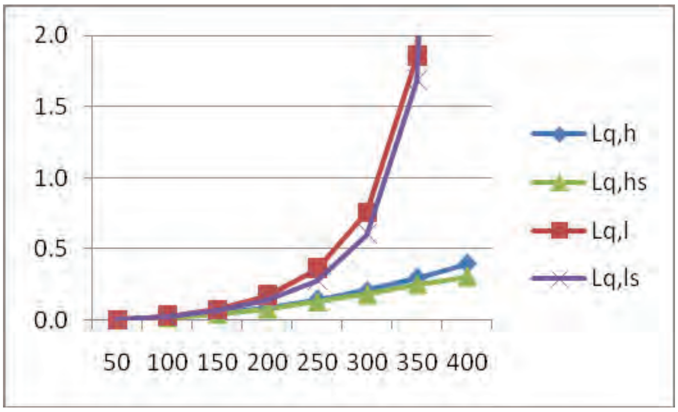

Fig. 2. Non-preemptive priority with 1:5 service time

Non-preemptive priority with equal service time but with different packet rate – 1:5 for high:low priority streams (low priority packet rate shown on x- axis) is simulated and compared with analytic results from Eqs. (8) and (9). Results are shown on Fig. 3.:

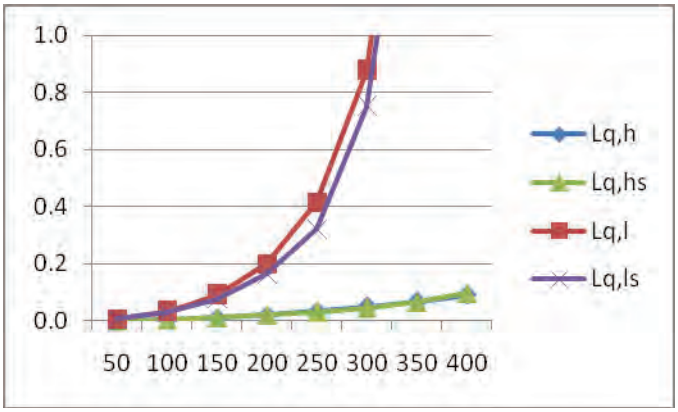

Fig. 3. Non-preemptive priority with 1:5 packet arrival rate

VI. Conclusion and Future Work

Based on the above study and results we can conclude that the formulas - Eqs. (8) and (9) are applicable in wide range of cases for the non-preemptive priority and M/D/1 system in IoT.

For future research work, the authors plan to develop to extend his researches over system combining more priority flows and preemptive and no-preemptive disciplines as well.

Acknowledgement

This work was supported in part by the Grant DDVU02/13/17.12.2010 "Public and Private Multimedia Network Throughput Increase by Creating Methods for Assessment, Control and Traffic Optimization" of the Bulgarian Science Fund.

References

[1] Q. Zhang and D. Peng, Intelligent decision-making service framework based on QoS models in the Internet of Things, *11th Int. Simp. on DCABES*, 2012, pp.103-107.

[2] J. Jiong, J. Gubbi, L. Tie and M. Palaniswami, Network architecture and QoS issues in the Internet of Things for a smart city, *2012 Int. Simp. on ISCIT*, pp. 956-961.

[3] W.-X. Li, J. Xu and H. Jiang, Queuing state analysis on a Hybrid Scheduling strategies for heterogeneous traffics in IoT, *2012 Int. Conf. on Computer Science and Service Systems.*, pp.1007-1008.

[4] V. Rohokale, N. Prasad and R. Prasad, A cooperative internet of things (IoT) for rural healthcare monitoring and control, *2nd Int. Conf. Wireless VITAE, 2011*, pp. 1-6.

[5] G. Giambene, *Queuing Theory and Telecommu-nication Networks and Applications*, Springer, 2005.

[6] J. N. Daigle. *Queuing Theory with Applications to Packet Telecommunication*, Springer, 2005.

[7] D. Gross et al, *Fundamentals of Queuing Theory*. Wiley, 2008.

[8] V. B. Iversen, *Teletraffc Engineering and Network Planning.* Technical University of Denmark, 2006, (http://www.com.dtu.dk/education/34340/)

[9] N. Chee-Hock and S. Boon-Hee, *Queueing Modeling Fundamentals with Application to Communication Networks.* Wiley, 2008.