Markov Chain Based Priority Queueing Model for Packet Scheduling and Bandwidth Allocation

Reema Sharma[1], Navin Kumar[2](✉), and T. Srinivas[3]

[1] Department of ECE, The Oxford College of Engineering, Bangalore, India
sharma80reema@gmail.com
[2] Department of ECE, Amrita School of Engineering, Amrita Vishwa Vidyapeetham University, Bangalore, India
navinkumar@ieee.org
[3] Department of ECE, Indian Institute of Science, Bangalore, India
tsrinu@ece.iisc.ernet.in

Abstract. This paper considers classification of diverse traffic types in Internet of Things (IoT) based on importance of data rate, packet size and proposes a priority-based probabilistic packet scheduling strategy for efficient packet transmission. Reduction of peak resource usage, dynamic control of service rate corresponding to arrival rate and QoS buffer management are few main factors considered to develop this strategy. By calculating percentage of link bandwidth required for prioritized traffic in each cycle, we provide quality of service (QoS) to real time traffic in IoT and non-IoT applications. Different experiments including MPEG traffic traces and Poisson traffic are conducted to verify the proposed scheduler. Also, performance of scheduler for both IoT and Non-IoT applications is compared for different data rates. We observe that the proposed packet scheduler satisfies QoS requirements for both IoT and non-IoT traffic.

Keywords: Internet of Things · Average queue length · Quality of service
Packet scheduling algorithm · Delay sensitive applications
Service differentiation

1 Introduction

IoT devices offer numerous novel real world services which result in heterogeneous QoS constraints and brings forward the requirement of a scheduling scheme to achieve overall optimum performance. Current studies on classification and scheduling of IoT services [1–3] rarely considers size of packets, its data rate, type of packets and comparison with Non-IoT data. Also, stringent delay constraints and high bandwidth requirements of multi-user video transmission applications [4], and to provide adequate transmission opportunities to all video/image senders before their tolerable delay deadlines is a longstanding research problem in IoT. This paper considers some of these issues to investigate optimal approach for assigning scheduling priority levels and allocates required bandwidth.

N. Kumar and A. Thakre (Eds.): UBICNET 2017, LNICST 218, pp. 91–103, 2018.
https://doi.org/10.1007/978-3-319-73423-1_9

IoT has a broad research scope in several areas like healthcare, smart environments, structural health monitoring and transportation, etc. [1, 5]. Applications like structural health checking generally needs unfailing information release from every node to the destination node. Furthermore, the QoS necessity of traffic blocking is comparatively rigorous in terms of throughput and delay because of the association of critical continuous data. In each of these applications, lightweight smart objects are active participants which are capable of sensing different incidents and communicating it to various other devices. Current methods do not offer acceptable solutions for delay sensitive applications. Few slot allocation policies like rate adaptive round robin or round robin provide assured QoS but the allotment of the current slot is not based on the allotment of the previous slots and these policies are hence considered to be stationary.

In this paper, a policy is developed to consider immediate release of delay sensitive data in IoT applications. We are addressing the scheduling scheme for the packets to provide QoS services for different class of services. Policy is actively calculating the number of packets to be scheduled in current cycle from high priority queue based on increase or decrease in its average queue length in consecutive cycles. This results in assuring adequate allotment of bandwidth to each service class by avoiding excess allocation always to the high priority class. After assigning required bandwidth to high priority class, remaining bandwidth can be assigned to low priority class. The major contributions of this paper are:

- Classifying and scheduling packets based on its size, type and data rates to achieve less transmission delay for each priority class
- Analyzing theoretically with Markov Chain model and simulation experiment with MATLAB R2013 to explore and study the waiting time of packets for various priority-queuing schemes and discover out the optimal one
- Comparison of model for IoT and Non-IoT applications.

The rest of the paper is organized as follows. Few recent investigated works are discussed in Sect. 2. Probabilistic model and its assumptions are presented in Sect. 3. Section 4 discusses the detailed analysis. Simulation results are presented and discussed in Sect. 5. Finally, conclusion and future work is included in Sect. 6.

2 Related Work

Recently, IoT has attracted researchers from both industry and academia. Current research explores into various phases of IoT such as service oriented architecture based IoT [6], Web of Things [7], applications and clarifications related to IoT [8] and therefore various issues can be investigated. Many authors [9–13] presented surveys on IoT vision, IoT related projects, IoT enabling technologies, research issues like privacy, trust, energy consumption and resource insufficiency with certain application areas in IoT. Some of them are very important issues which provide useful discussion about QoS requirements but design of QoS models to provide priority to emergency applications are not extensively discussed.

Klepec and Kos [11] proposed a priority model with two queues and presented packet transit time behavior for a delay susceptible application for which bandwidth limit should be the least. The model is straightforward; the exploitation of higher priority was shown at the price of more packet losses for low priority data. Moreover, in order to analyze and monitor energy efficiency [14]; network topologies; issues related to performance of the network [15]; and the accessibility of bandwidth; a number of new methods have been devised. A buffer sharing scheme for resource distribution in wireless local area networks (WLAN's) under diverse traffic conditions is discussed in [16]. The results show that for heterogeneous traffic loads, transmission opportunities are not equally allocated by 802.11. They also showed that large buffer can help in providing this equality but at the expense of increased delay. The solution to delay sensitive applications is not efficient as discussed in the above studies. To guarantee instantaneous communication with no packet loss, queueing delay and specifically to address the delay critical applications, an efficient packet scheduling scheme with service priorities becomes necessary.

3 Probabilistic Model and Its Assumptions

The probabilistic representation and few hypotheses considered are discussed in this section. In the projected model, it is assumed that the data packets are categorized into prioritized (critical data) and non-prioritized (non critical data) traffic and is accumulated in two separate queues. The scheduler calculates the current departure packets using number of departure packets in the previous slot and the number of arrival packets in the current slot. It calculates a weighting coefficient for each queue which represents average number of packets that can be scheduled from the queue before moving to the next queue. The idea is based on weighted round robin scheduler. In this, the system serves each queue in a round robin manner and the calculated dynamic weights are assigned. Scheduling is executed at the beginning of each cycle. The scheduling mechanism differentiates the services based on the priority, by measuring the probability of traffic increased at the current time slot from the previous time slot, predicts the number of packets to be scheduled at the current time slot and the amount of bandwidth to be allocated to each service.

The configuration of the node and scheduling system is given in Fig. 1. Consider that the prioritized traffic queue has a maximum size of $B1$ and average buffer size of P_{avg} and non-prioritized queue has a maximum size of $B2$. It is assumed that packets arrive separately for all service classes follow a Poisson procedure with a mean rate of arriving packets per cycle as $\lambda = \sum_{i=0}^{i_{max}} i.P(i)$, the same is depicted in Fig. 1, where P(i) denotes the probability of i packets arriving in one round and i_{max} specifies the upper limit to packets arriving. Let $\lambda_1, \lambda_{HQ}, \lambda_{HL}$ are the packets arrived, accumulated in the buffer and dropped from the prioritized queue respectively, and $\lambda_2, \lambda_{LQ}, \lambda_{LL}$ are quantity of packets arrived, accumulated in the queue and dropped from non-prioritized queue respectively. The system consists of N cycles, each of which is further partitioned into different time slots. Here, every time slot carries a packet of variable size. Now, the system model formulation to compute the average quantity of packet serviced

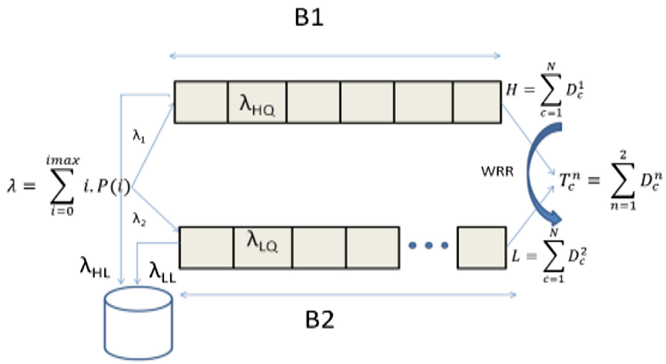

Fig. 1. System model

T_c^n in cycle $\boldsymbol{c}$ for nth class can be designed. The corresponding buffer queue size of the current cycle $\boldsymbol{c}$ is calculated as:

$$q_c^n = q_{c-1}^n - D_c^n + a_c^n \tag{1}$$

where a_c^n indicates quantity of packets which arrives in the nth queue during cycle c; D_c^n signifies the quantity of packets leaving from nth queue at cycle $\boldsymbol{c}$ and q_{c-1}^n represents quantity of packets stored in the queue during cycle (c − 1). The quantity of packets served in cycle $\boldsymbol{c}$ can be given as:

$$D_c^n = \min\left(q_c^n, b_c^n\right) \tag{2}$$

where, b_c^n is the quantity of packets which can be served according to the bandwidth availability and computed based on the anticipated method for n^{th} queue during cycle $\boldsymbol{c}$ (details given later). The respective queue is picked up in a round robin manner, thus n can be computed as:

$$n = (n' \bmod 2) + 1 \tag{3}$$

where, n' represents the previous class selected, n = 1 represents the prioritized service and n = 2 represents the non-prioritized service queue. Probability of average queue length of prioritized queue is calculated as:

$$P_{avg}(c) = 0.01 * P_{avg}(c-1) + (1-0.01) * P(q_c^n) \tag{4}$$

where $P_{avg}(c)$ is the probability of average queue size in current cycle, $P_{avg}(c-1)$ is the probability of average queue size in previous cycle and $P(q_c^n)$ is the probability of instantaneous queue size at current cycle. The sum of the packets serviced in one round can be computed as: $T_c^n = \sum_{n=1}^{2} D_c^n$. $H = \sum_{c=1}^{N} D_c^1$ and $L = \sum_{c=1}^{N} D_c^2$ are sum of prioritized and non-prioritized packets departed in N cycles where c = 1,2 …N.

To dynamically allocates weights with standard scale values, prioritized queue is assigned with two thresholds T_{min} = 0.083 and T_{max} = 0.3667 which act as indicators to achieve desired and acceptable QoS parameters. At these threshold values, the least blocking probability values for the considered simulation scenario are obtained. Assume *p* is the probability of serving prioritized packets and *q* is the probability of serving the non-prioritized packets. As, only two service queues are taken, the probability of serving the non prioritized packets can be given by q = 1 − p. Probability *p* can be further distinguished based on proposed scheduler into three different cases:

$$p = \begin{cases} \mathrm{p}_1 & \text{for } 0 \leq \mathrm{P}_{\mathrm{avg}} \leq \mathrm{T}_{\mathrm{min}} \\ \mathrm{p}_2 & \text{for } \mathrm{T}_{\mathrm{min}} < \mathrm{P}_{\mathrm{avg}} < \mathrm{T}_{\mathrm{max}} \\ \mathrm{p}_3 & \text{for } \mathrm{B1} > \mathrm{P}_{\mathrm{avg}} \geq \mathrm{T}_{\mathrm{max}} \end{cases} \tag{5}$$

where, p_1 = 0.3, is the probability of weight allocated to prioritized packets when the probability of average queue length P_{avg} is between 0 and T_{min}. This value of p_1 is chosen to provide minimum bandwidth to prioritized queue irrespective of the arrival rate. The p_2 is the probability of weight allocated to prioritized packets when average queue length increases and lies between T_{min} and T_{max} and is calculated by Eq. (6) and p_3 = 0.7 to 0.9 is the probability of weight allocated to prioritized packets when the average queue length increases beyond T_{max}. The value of p_3 are chosen to limit maximum bandwidth allotted to prioritized queue and to provide some processing of non-prioritized queue and reduce its blocking probability while providing guaranteed service to prioritized queue in each cycle.

$$\mathrm{p}_2 = \mathrm{p'} + (\mathrm{P}_{\mathrm{avg}}(\mathrm{c}) - \mathrm{P}_{\mathrm{avg}}(\mathrm{c}-1)).\frac{0.3}{(\mathrm{T}_{\mathrm{max}} - \mathrm{T}_{\mathrm{min}})} \tag{6}$$

where p' the probability of weight assigned to prioritized queue in previous cycle. $\mathrm{P}_{\mathrm{avg}}(\mathrm{c}) - \mathrm{P}_{\mathrm{avg}}(\mathrm{c}-1)$ is the change in probability of average queue (increase or decrease) in consecutive cycles. Eq. (5) shows a linear relationship between probability of weights allocated to priority service and probability of average queue size.

4 Model Analysis

The system is depicted by a probabilistic Markov Chain model. Since the investigational process of all transitional nodes is similar, a node is picked up arbitrarily to examine the algorithm because this scheduling algorithm can independently work in each router to schedule packets based on arrival rate. Knowing the scheduling time spent in one node and total nodes existing in the path chosen by the routing algorithm for a particular topology, the processing delay can be found. The blocking probability of prioritized and non-prioritized class can be calculated. The Markov chain model formulation to compute the average packets scheduled or departures T_c^n in cycle ***c*** for n^{th} service class is as follows.

Figures 2(a), (b) and (c) presents the state transition diagrams for state (x, y), where (x, y) represents a state in which *x* non-prioritized packets and *y* average prioritized packets are stored in their respective queues. P(x, y) is the probability of the system being in state (x, y). Four cases are discussed here: (i) when x < B2 and y < B1 (Fig. 2 (a)) (ii) when x $\geq$ B2 and y < B1 (Fig. 2(b)) (iii) when x < B2 & y $\geq$ B1. (Figure 2 (c)) (iv) when x $\geq$ B2 and y $\geq$ B1. Based on Figs. 2(a), (b) and (c), the balance equations for state (x, y) are computed as:

$$\begin{aligned}(\mathrm{p}\lambda_1 + (1-\mathrm{p})\lambda_2 + \mathrm{y}\mu_1 + \mathrm{x}\mu_2)\mathrm{P}_{\mathrm{x,y}} - \mathrm{p}\lambda_1\mathrm{P}_{\mathrm{x,y-1}}(1-\mathrm{p})\lambda_2\mathrm{P}_{\mathrm{x-1,y}} \\ - (\mathrm{y}+1)\mu_1\mathrm{P}_{\mathrm{x,y+1}} - (\mathrm{x}+1)\mu_2\mathrm{P}_{\mathrm{x+1,y}} = 0\end{aligned} \tag{7}$$

$$(\mathrm{p}\lambda_1 + \mathrm{y}\mu_1 + \mathrm{x}\mu_2)\mathrm{P}_{\mathrm{x,y}} - \mathrm{p}\lambda_1\mathrm{P}_{\mathrm{x,y-1}} - (1-\mathrm{p})\lambda_2\mathrm{P}_{\mathrm{x-1,y}} - (\mathrm{y}+1)\mu_1\mathrm{P}_{\mathrm{x,y+1}} = 0 \tag{8}$$

$$((1-\mathrm{p})\lambda_2 + \mathrm{y}\mu_1 + \mathrm{x}\mu_2)\mathrm{P}_{\mathrm{x,y}} - \mathrm{p}\lambda_1\mathrm{P}_{\mathrm{x,y-1}} - (1-\mathrm{p})\lambda_2\mathrm{P}_{\mathrm{x-1,y}} - (\mathrm{x}+1)\mu_2\mathrm{P}_{\mathrm{x+1,y}} = 0 \tag{9}$$

In Fig. 2(a), when non-prioritized queue is filled, the new coming packets will be dropped which is shown by returning back to the same state. Similarly, as soon as the prioritized queue is filled then those incoming packets are dropped as shown in Fig. 2 (b). To obtain the blocking probabilities of service classes, the above equations need to be solved to obtain state probabilities P(x, y). So, consider a non-complex system to solve blocking probability. The corresponding state transition diagram is shown in Fig. 2(d). To make the computation easier, consider $\mu 1 = \mu 2 = \mu$ where the service rate μ is taken as the average service rate of two traffic. In the state diagram, if the prioritized queue is full, p(1, 1) state is not considered instead it is shown as loss of λ_2. If non- prioritized queue is full, then p(0, 1) state is considered as p(1, 1) state. The balance equations for the structure are re-written based on Eqs. (7), (8) and (9). The blocking probability for a non-complex system [17] is derived and then the result is extended for a complex system.

$$\mathrm{p}_1\lambda_1\mathrm{P}_{(0,0)} + \mathrm{p}_3\lambda_1\mathrm{P}_{(1,0)} = \mu_1\mathrm{P}_{(0,1)} \tag{10}$$

$$\mathrm{q}_1\lambda_2\mathrm{P}_{(0,0)} = (\mu_2 + \mathrm{p}_3\lambda_1)\mathrm{P}_{(1,0)} \tag{11}$$

$$\mathrm{P}_{(0,0)} + \mathrm{P}_{(1,0)} + \mathrm{P}_{(0,1)} = 1 \tag{12}$$

From Eq. (10)

$$\mathrm{p}_3\lambda_1\mathrm{P}_{(1,0)} = \mu_1\mathrm{P}_{(0,1)} - \mathrm{p}_1\lambda_1\mathrm{P}_{(0,0)}$$

Substituting in Eq. (11)

$$\mathrm{q}_1\lambda_2\mathrm{P}_{(0,0)} = \mu(\mathrm{P}_{(0,1)} + \mathrm{P}_{(1,0)}) - \mathrm{p}_1\lambda_1\mathrm{P}_{(0,0)}$$

From Eq. (12)

$$P_{(1,0)} + P_{(0,1)} = 1 - P_{(0,0)}$$

Therefore,

$$q_1\lambda_2 P_{(0,0)} = \mu(1 - P_{(0,0)}) - p_1\lambda_1 P_{(0,0)}$$

$$\mu = P_{(0,0)}[q_1\lambda_2 + \mu + p_1\lambda_1]$$

$$P_{(0,0)} = \frac{\mu}{[p_1\lambda_1 + q_1\lambda_2 + \mu]}$$

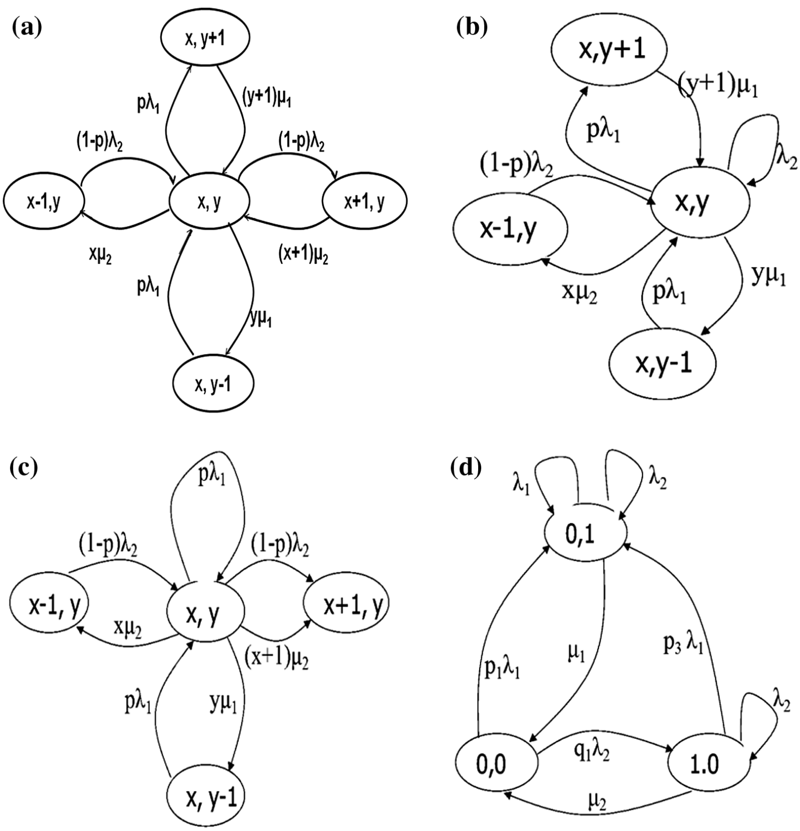

Fig. 2. (a) when x < B2 & y < B1 (b) when x $\geq$ B2 & y < B1 (c) when x < B2 & y $\geq$ B1 (d) State transition diagram of a single-channel system for proposed scheme

Substituting $P_{(0,0)}$ in Eq. (11), we get:

$$P_{(1,0)} = \frac{q_1 \lambda_{2*} \mu}{[(p_1\lambda_1 + q_1\lambda_2 + \mu)(\mu + p_3\lambda_1)]}$$

Substituting $P_{(0,0)}$ & $P_{(1,0)}$ in Eq. (10), we have:

$$\mu P_{(0,1)} = \frac{p_1\lambda_1 * \mu}{(p_1\lambda_1 + q_1\lambda_2 + \mu)} + \frac{q_1\lambda_2\mu * p_3\lambda_1}{[(p_1\lambda_1 + q_1\lambda_2 + \mu)(\mu + p_3\lambda_1)]}$$

$$P_{(0,1)} = \frac{\lambda_1[\mu p_1 + p_1p_3\lambda_1 + q_1p_3\lambda_2]}{(p_1\lambda_1 + q_1\lambda_2 + \mu)(\mu + p_3\lambda_1)} \tag{13}$$

The blocking probabilities can be calculated as:

1. For the Prioritized queue:

$$\text{Block}_{\text{prob1}} = P_{(0,1)} + (1 - p)P_{(1,0)} \tag{14}$$

which evidently involves two parts: (i) the probability that prioritized packet reaches state (0, 1) and is lost. (ii) the probability that a prioritized packet reaches state (1, 0) but due to probability $(1 - p)$, it is lost.
2. For Non-Prioritized queue:

$$\text{Block}_{\text{prob2}} = \left(P_{(0,1)} + P_{(1,0)}\right) + \frac{\lambda_1}{\lambda_2} pP_{(1,0)} \tag{15}$$

This probability can also be considered as consisting of two parts: (i) the probability that a low priority packet arrives either at state (0, 1) or (1, 0) and is getting dropped; (ii) The probability that a non-prioritized packet arrives at state (1, 0) and is getting lost due to probability p.

5 Simulation Results and Analysis

In this section, we present the simulation results to validate the efficiency of the proposed scheme and to prove that it can support service differentiation. The simulation results are plotted using MatLab R2013b. We have conducted three experiments to investigate the scheduler efficacy. In the first experiment, high priority packets are taken as MPEG traces and low priority packets as Poisson traffic with variable size and under different system loads. In the second experiment, for both high and low priority, Poisson traffic with variable size packets is considered. Third experiment compares IoT and Non-IoT cases and tests the scheduler working when very large packet sizes with high data rate for high priority class arrives as Non-IoT data. Extensive simulation has been conducted to test the scheduler effectiveness for IoT traffic under different data rates. Buffer sizes for both high priority and low priority queues are taken as 10. Details of simulation scenario are given in Table 1.

Table 1. Traffic adopted for conducting experiment

Parameters	IoT traffic (Exp1)		IoT Traffic (Exp2)		Non-IoT traffic (Exp3)	
Service type	High priority	Low priority	High priority	Low priority	High priority	Low priority
Traffic type	MPEG-4	Poisson	Poisson	Poisson	MPEG-4	Poisson
Packet size in bytes	136 to 1000	50 to 702	50 to 702	50 to 702	136 to 424536	50 to 702
Number of flows	2 flows at same time		2 flows at same time		2 flows at the same time	
First datarate	36.8 Kbps to 112.3 Kbps	73.36 Kbps to 203.3 Kbps	0.64 Kbps to 84.3 Kbps	66.8 Kbps to 107.52 Kbps	32.6 Kbps to 50.5 Mbps	358.8 Kbps to 475 Kbps
Second datarate	21.7 Kbps to 89.2 Kbps	64 Kbps to 175 Kbps	0.32 Kbps to 82.96 Kbps	55.3 Kbps to 88.7 Kbps	21.7 Kbps to 40 Mbps	277.5 Kbps to 300 Kbps
Third datarate	16 Kbps to 64 Kbps	45 Kbps to 164 Kbps	0.20 Kbps to 65 Kbps	45.4 Kbps to 64 Kbps	21.7 Kbps to 32 Mbps	160.6 Kbps to 250 Kbps

5.1 Impact of p on Blocking Probability

Figure 3(a) shows blocking probability between two classes against increasing value of probability p from p_1 to p_2 and then to p_3, in Exp1 and Exp2 with third data rate. We can observe that if value of p increases, blocking probability of the high priority packets decreases and for low priority it increases. So, the required service differentiation can be achieved by adjusting p according to the tolerable blocking probability.

It can be observed that due to this scheduling scheme there is a continuous decrease in the high priority blocking probability and simultaneous increase in blocking probability of low priority traffic. This is under the condition where prioritized and non-prioritized packet load is continuously increasing. Prioritized packets priority (in terms of bandwidth allocation) keeps on increasing as probability p increases from p_1 to p_2 and then to p_3. From the analysis, for Exp1 and Exp2, it can be easily verified that if bandwidth provided to high priority is greater than or equal to 90%, then in both experiments (Fig. 3(a)), blocking probability is similar for high and low priority services. For bandwidth less than 90% for high priority traffic, blocking probability increases to 1%. Although the packet sizes for high priority packets in Exp1 is of bigger size than Exp2 for IoT but it gets compensated with the higher bandwidth provided by the scheduler.

Figure 3(b) shows impact of probability p on blocking probability for Exp3 for non-IoT traffic. We have considered very high data rate and packet sizes in Exp3. We verified our scheduler to test its efficacy under the condition when variable and big packet of sizes of 136 bytes to 424536 bytes is transferred as high priority packets. It can be verified that for high and low priority traffic, blocking probability for non-IoT is slightly increased as compared to IoT traffic due to its high data rate and packet size. As compared to IoT cases, in non-IoT the blocking probability is 0.05% more for the bandwidth greater than 90% for high priority traffic. However, for the bandwidth less than 90% for high priority traffic, the blocking probability increases to 1% even in non-IoT traffic.

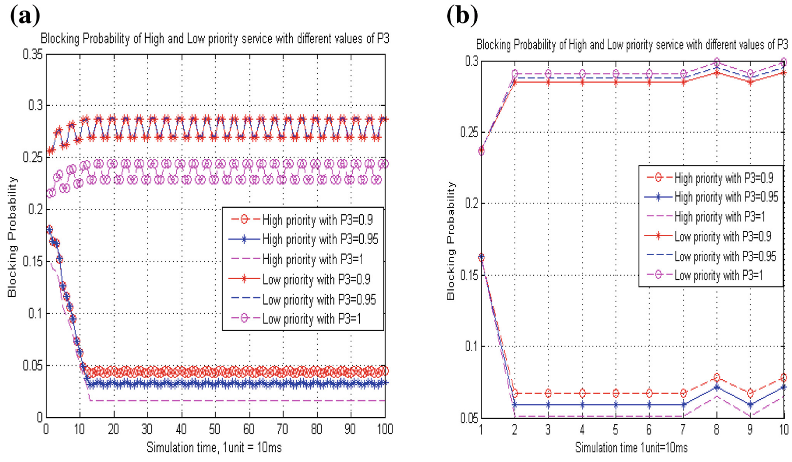

Fig. 3. Blocking probability of high and low priority traffic with different p_3 values in (a) Exp1 & Exp2 (b) Exp3

5.2 Impact of Different Data Rates on Average Blocking Probability

Figures 4(a), (b) and (c) describes average blocking probability of high and low priority traffic with different data rates in Exp1, Exp2 and Exp3 respectively. We observe that with Exp1 blocking probability decreases gradually for high priority traffic as compared to Exp2. In Exp2 blocking probability decreases at a faster rate in the beginning for prioritized packets and then becomes constant. This is due to the reason of comparatively bigger packet sizes and more data rate taken in Exp1 as compared to Exp2. In both experiments, when data rate is less, then average blocking probability is also less. Improvement for non-prioritized packets can be seen for less data rate. Here value of p_3 is 0.9. In Exp3 (Fig. 4(c)) also when data rate for non-IoT traffic is high (first data rate) its blocking probability is more. If data rate is reduced, blocking probability is also reduced for high priority traffic.

5.3 Comparison of Average Blocking Probability in Exp1&2 and Exp1&3

Figures 5(a) and (b) shows the average blocking probability of high and low priority traffic for both Exp1 with Exp2 and Exp1 with Exp3 respectively. For only IoT traffic with different data rates, as shown in Fig. 5(a); we observe that if packet size of prioritized traffic is reduced (Exp2) or if data rate is reduced; the average blocking probability of both prioritized and non-prioritized traffic is reduced. This implies that the length of the packet size can also play an important role in analyzing the performance of any model. Therefore, a smaller size of packet would be considered for better performance. Figure 5(b) compares IoT and non-IoT cases. It is observed that for the non-IoT traffic, the average blocking probability for both high and low priority traffic is more than IoT traffic because of high data rate of non-IoT applications. Table 2 clearly explains impact of increasing probability p on blocking probability of high and low

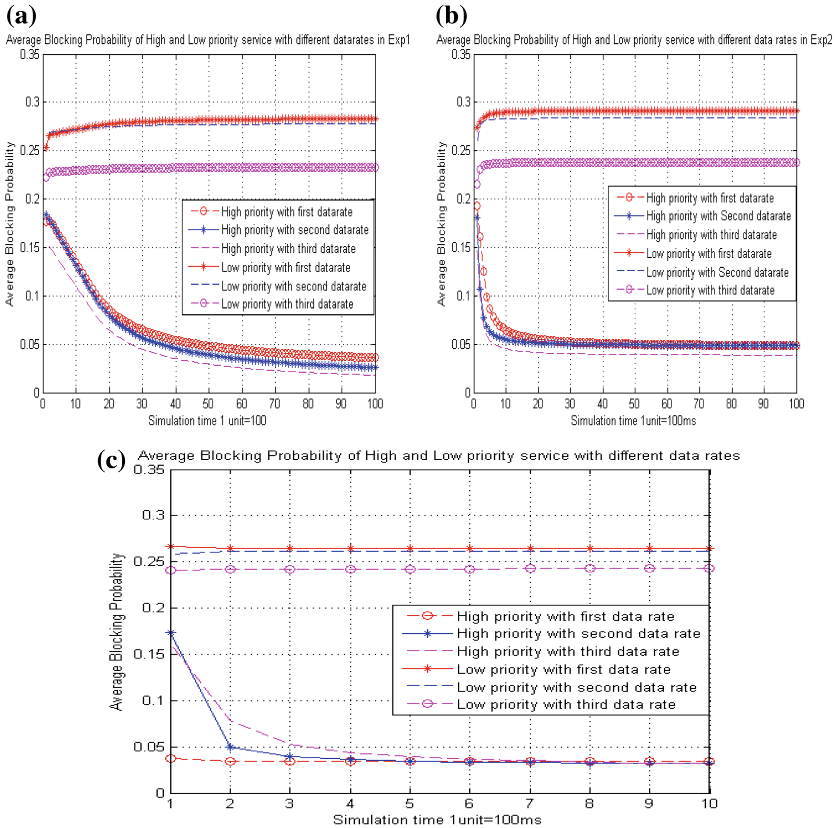

Fig. 4. Average blocking probability of high and low priority traffic for different data rates in (a) Exp1 (b) Exp2 (c) average blocking probability of high and low priority traffic with different data rates in Exp3

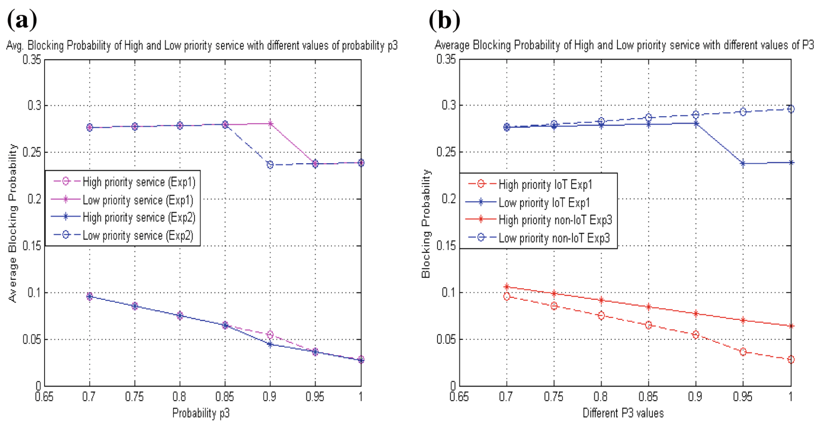

Fig. 5. Average blocking probability of high and low priority traffic with different p_3 values in (a) Exp1 and Exp2 (b) IoT (Exp1) and Non-IoT (Exp3)

Table 2. Impact of increase of probability p_3 on average blocking probability

S.No.	p_3	Blocking probability for IoT traffic (Exp1)		Blocking probability for IoT traffic (Exp 2)		Blocking probability for Non-IoT traffic (Exp 3)	
		High priority	Low priority	High priority	Low priority	High priority	Low priority
1	0.7	0.0956	0.2764	0.0956	0.2765	0.1061	0.2767
2	0.75	0.0852	0.2775	0.0851	0.2777	0.0987	0.2800
3	0.8	0.0750	0.2787	0.0748	0.2789	0.0915	0.2834
4	0.85	0.0649	0.2798	0.0646	0.2800	0.0844	0.2866
5	0.9	0.0551	0.2809	0.0444	0.2374	0.0775	0.2899
6	0.95	0.0366	0.2881	0.0359	0.2384	0.0706	0.2930
7	1.00	0.0284	0.2890	0.0275	0.2393	0.0639	0.2961

priority traffic. In all three cases with decrease in blocking probability of high priority emergency traffic, there is simultaneous increase in low priority blocking probability values which proves that the dynamic scheduling scheme is effective in achieving adjustable service differentiation in IoT and non-IoT applications.

6 Conclusion

A simple and flexible probabilistic scheme has been proposed to offer service differentiation and to provide QoS to emergency applications in IoT. Analytical and simulation results showed that the dynamic scheduling scheme is effective in achieving adjustable service differentiation in IoT and non-IoT applications where large amount of data needs to be transferred continuously at low rate and high rate respectively for long period of time. Also, if any emergency traffic needs to be given priority, this scheduler reduces blocking probability even in the case of congested network. The proposed scheme is tested for variable size packets with different data rates and the expected results are obtained. We also verified that the scheduler satisfies QoS requirements in both IoT and Non-IoT applications.

References

1. Monares, A., Ochoa, S.F., Santos, R., Orozco, J., Meseguer, R.: Modeling IoT-based solutions using human-centric wireless sensor networks. J. Sens. **14**(9), 15687–15713 (2014)
2. Sun, Y., Ma, H., Liu, L.: Traffic scheduling based on queue model with priority for audio/video sensor networks. In: IEEE International Conference on Pervasive Computing and Applications, pp. 709–714 (2007)
3. Dainotti, A., Pescape, A., Claffy, K.C.: Issues and future directions in traffic classification. IEEE Netw. **26**(1), 35–40 (2012)
4. Khalek, A.A., Caramanis, C., Heath, R.W.: Delay-constrained video transmission: quality-driven resource allocation and scheduling. IEEE J. Sel. Top. Sig. Process. **9**(1), 60–75 (2015)

5. Jin, J., Gubbi, J., Luo, T., Palaniswami, M.: Network architecture and QoS issues in the internet of things for a smart city. In: IEEE International Symposium on Communications and Information Technologies (ISCIT), pp. 956–961 (2012)
6. Spiess, P., Karnouskos, S., Guinard, D., Savio, D.: SOA-based integration of the internet of things in enterprise services. In: IEEE International Conference on Web Services, pp. 968–975, July 2009
7. Gupta, V., Poursohi, A., Udupi, P.: Sensor network: an open data exchange for the web of things. In: IEEE Intenational Conference on Pervasive Computing and Communications Workshops, pp. 753–755, April 2010
8. Thoma, M., Meyer, S., Sperner, K., Meissner, S., Braun, T.: On IoT-services: survey, classification and enterprise integration. In: IEEE International Conference on Green Computing and Communications, pp. 257–260, November 2012
9. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. Elsevier J. Future Gener. Comput. Syst. **29**, 1645–1660 (2013)
10. Mashal, I., Alsaryrah, O., Chung, T.-Y.: Performance evaluation of recommendation algorithms on Internet of Things services. Elsevier J. Phys. A: Stat. Mech. Appl. **451**, 646–656 (2016)
11. Klepec, B., Kos, A,: Performance of VoIP applications in a simple differentiated services network architecture. In: IEEE International Conference Eurocon 2001 at Bratislava, Slovakia, vol. 1, pp. 214–217, July 2001
12. Said, O., Masud, M.: Towards internet of things: survey and future vision. Int. J. Comput. Netw. **5**(1), 86–94 (2013)
13. Al-Fagih, A.E.: A framework for data delivery in integrated Internet of Things architectures. Ph.D. thesis, Queen's University Kingston, Ontario, Canada, April 2013
14. Liang, J.M., Chen, J.J., Cheng, H.H., Tseng, Y.C.: An energy-efficient sleep scheduling with QoS consideration in 3GPP LTE advanced networks for internet of things. IEEE J. Emerg. Sel. Top. Circ. Syst. **3**(1), 13–22 (2013)
15. Awan, I., Younas, M.: Towards QoS in internet of things for delay sensitive information. In: Matera, M., Rossi, G. (eds.) MobiWIS 2013. CCIS, vol. 183, pp. 86–94. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-03737-0_10
16. Duffy, K., Ganesh, A.J.: Modeling the impact of buffering on 802.11. IEEE Commun. Lett. **11**(2), 219–221 (2007)
17. Yang, L., Jiang, Y., Jiang, S.: A probabilistic preemptive scheme for providing service differentiation in OBS networks. In: IEEE Globecom, vol. 5, pp. 2689–2693 (2003)