2010 12th IEEE International Conference on High Performance Computing and Communications

Scheduling and Queue Management for Multi-class Traffic in Access Router of Mobility Protocol

Md. Shohrab Hossain
Mohammed Atiquzzaman
University of Oklahoma, OK 73019
Email: {shohrab, atiq}@ou.edu

William Ivancic
NASA Glenn Research Center
Cleveland, OH 44135
Email: wivancic@grc.nasa.gov

***Abstract*—The traffic characteristics in recent years have changed considerably having more video and audio traffic with critical real-time constraints. This situation is compounded by the increased number of mobile nodes communicating over Internet. Due to such high data traffic, the access routers are often overloaded with packets with different priority. The access routers have to act quickly enough to avoid loss of connection of mobile nodes' ongoing communication. However, there is a lack of analytical models that focus on the internal queue management access router of IP-mobility protocols. In this paper, we have developed an analytical model using multi-class non-preemptive priority queues to measure average queuing delay, queue occupancy, and packet drop probability for different classes of data and signaling traffic in the access router. We present numerical results that reflects the impact of node density, service rate and traffic pattern on these measures. The analytical framework presented in this paper will be helpful to better manage access routers with different classes of data and signaling traffic causing least queuing delay and packet loss.**

***Index Terms*—Mobility protocols, analytical modeling, scheduling algorithm, queuing delay, real-time traffic, multi-class traffic**

I. Introduction

Increasing demand for mobility in wireless data networks has given rise to various mobility management schemes. IETF proposed Mobile IPv6 [1] and NEMO Basic Support Protocol (BSP) [2] to support host and network mobility, respectively that allow connections to remain alive while mobile nodes are on the move. In each IP-mobility protocol, the mobile nodes communicate with the Access Router (AR) through wireless channels for sending data packets and signaling packets, such as location updates, also known as binding updates.

Due to the widespread use of wireless and mobile devices that access the Internet, the number of mobile nodes under an access router is rising rapidly, resulting in increase of total packet arrival rates. Mobile nodes send signaling packets, such as binding update, refreshing binding update packets that need to be processed very quickly by the access routers to notify the mobility agent, such as location manager or home agent to keep the location database up-to-date. On the other hand, the traffic class having time critical deadline (e.g., real-time data), packets must be delivered to the destination before the deadlines. Otherwise, they are useless and can be dropped by the router.

There have been several research works on multi-class traffic that considered mobility. Nam et al. [3] present a two-layer downlink queuing model and a scheduling mechanism for providing lossless handoff and Quality of Service (QoS) in mobile networks. Iftikhar et al. [4] present a novel analytical model for multiple queue systems by considering two different classes of traffic that exhibit long-range dependence and self-similarity. Iftikhar et al. [5] also use a G/M/1 queuing system and construct analytical models on the basis of non-preemptive priority queuing, low-latency queuing and custom queuing. However, these papers lack the quantitative analysis of average queuing delay, queue occupancy and packet drop probability of data and signaling packets at the access router while managing IP-mobility.

In this paper, we have considered two classes of data traffic: Real Time (RT) and Non-Real Time (NRT) traffic. Another class of traffic are signaling traffic (binding update) sent by the mobile nodes to update mobility agents about its current location. We have proposed a scheduling algorithm that selects a packet based on its priority and the current location of the mobile node within a cell so that it produces the least loss to signaling traffic and handoff traffic, while allowing packet loss to real-time traffic as RT traffic are loss-tolerant. Based on the scheduling algorithm, we have derived expressions for average queuing delay, queue occupancy, and packet drop probability of each class of traffic. We have presented numerical results that reflect the impact of node density, service rate and traffic pattern on these measures.

Our *objective* of this paper is to analyze the performance of access router in processing real-time, non-real-time and signaling traffic while managing IP-mobility.

The *contributions* of this work are: (i) developing a mathematical model to estimate the average queuing delay, average queue occupancy and the packet drop probability at the AR of IP-mobility protocol and (ii) analyzing the impact of node density, service rate and traffic distribution on them.

The analytical model developed in this paper can be used to better manage the access routers and other network components for improved performance of future wireless and mobile networks.

The rest of the paper is organized as follows. In Section II,

The research work reported in this paper was supported by NASA Grant NNX06AE44G.

978-0-7695-4214-0/10 $26.00 © 2010 IEEE
DOI 10.1109/HPCC.2010.25

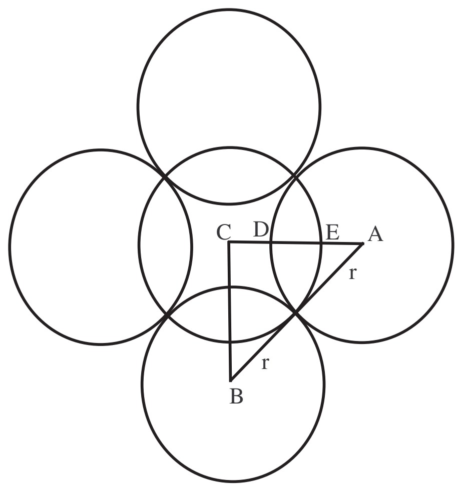

Fig. 1. Cell topology and overlapping among the cells.

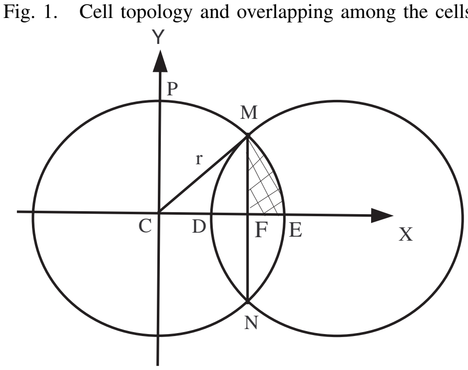

Fig. 2. Overlapping among the cells.

the cell topology is explained along with the compuation of the overlapped area of a cell. The analytical model is presented in Section III. Section IV analyzes the numerical results. Finally, Section V has the concluding remarks.

II. Cell topology and overlapping area

The cell topology is shown in Fig. 1 where each cell has overlapping regions with four neighboring cells. In this Section, we compute the overlapping area of a cell that will be used in queuing analysis (section III-E).

Let the radio coverage area of each cell be a circular region of radius r and two adjacent cells overlap at a maximum length of l along its diameter. In Fig. 1, let $AC = a = BC$, $DE = l$. Since the length of $AB = 2r$, we find that $a = \sqrt{2}r = r + r - l$. Therefore,

$$l = (2 - \sqrt{2})r \tag{1}$$

In order to find out the overlapping area, let us consider Fig. 2 where the center (C) of a cell is situated at the origin. The cord MN bisects the line segment DE at point F. Therefore, $DF = EF = l/2$. Since point C is the origin, then $CE = r$, $CF = CE - EF = r - l/2 = r/\sqrt{2}$. As point F is on the x-axis, the coordinate of point F is ($\frac{r}{\sqrt{2}}$, 0). Therefore the area of the region CFMP (let it be φ) is the area under the circle $x^2 + y^2 = r^2$ from the origin to point F and can be obtained by integrating between 0 to $\frac{r}{\sqrt{2}}$ as follows:

$$\varphi = \int_0^{\frac{r}{\sqrt{2}}} y dx = \int_0^{\frac{r}{\sqrt{2}}} \sqrt{(r^2 - x^2)} dx \tag{2}$$

Now, putting $x = r\sin\theta$, the limit becomes 0 to $\pi/4$. Thus

$$\varphi = r^2 \int_0^{\pi/4} \cos^2\theta d\theta = \frac{r^2}{4}\Big(\frac{\pi}{2} + 1\Big) \tag{3}$$

As the area of the cell is πr^2, the portion of the cell in the first quadrant (i.e., the region CPME) is $\pi r^2/4$. So the area of the shaded region (MFE) can be obtained as by subtracting Eqn. (3) from $\pi r^2/4$.

$$\text{Area of MFE} = \frac{\pi r^2}{4} - \frac{r^2}{4}\Big(\frac{\pi}{2} + 1\Big) = \frac{r^2}{8}\big(\pi - 2\big) \tag{4}$$

The overlapping region between two adjacent cells (MDNE), which is four times the shaded area (of MFE), equals $\frac{r^2}{2}\big(\pi - 2\big)$. For each cell, there are four neighboring cells with which it has overlapping regions (see Fig. 1). Hence, the total overlapping area of a cell (ϕ) is given by,

$$\phi = 4 \times \frac{r^2}{2}\big(\pi - 2\big) = 2r^2(\pi - 2) \tag{5}$$

The ratio of the overlapping area (Eqn. (5)) to the total area (πr^2) is given by

$$\gamma = \frac{2r^2(\pi - 2)}{\pi r^2} = \frac{2(\pi - 2)}{\pi} \tag{6}$$

From Eqn. (6), we find that $\gamma = 0.7268$ which means 72.68% area of the cell is in overlapping region whereas 27.32% area is non-overlapping.

III. Analytical model

First, the assumptions and notations of the model are listed in Sections III-A and III-B, Section III-C explains the node architecture of the AR, followed by the scheduling algorithm in Section III-D. Finally, Section III-E presents the queuing analysis.

A. Assumptions

To make the model analytically tractable, the following assumptions are made.

- Uniform density of the Mobile Nodes (MN) in the network.
- The number of MNs leaving the coverage area of an AR is equal to the MNs entering the area. Therefore, the net change in number of MNs under an AR is zero.
- The ratio of the overlapping area to the total coverage area is used as a measure of probability for being in handoff status.
- Packet arrival process is a Poisson process.

- Type of queue discipline used in the analysis is FIFO with non-preemptive priority among various traffic classes.
- One way (download) traffic is considered for simplicity.
- As Binding Update (BU) messages exchanged between MN and Home Agent are essential to track a mobile node, BU packets are assigned the highest priority.

B. Notations

The notations used in the analysis are listed below.

N_m Number of MNs in a cell (AR),
λ_{BU} Binding update packet rate at each MN,
λ_{RBU} Refreshing binding update packet rate at each MN,
λ_{RT} Real time packet arrival rate at each MN,
λ_{NRT} Non-real time packet arrival rate at each MN,
λ_i Packet arrival rate at class-i queue of AR,
λ Total packet arrival rate to the system,
μ Service rate of the system,
ρ System utilization factor,
r Radius of each cell,
γ Ratio of overlapping area to the total area of the cell,
$E(T_i)$ Average queuing delay of a class-i packet,
$E(n_i)$ Average queue occupancy of class-i packets,
$P_d(i)$ Packet drop probability of class-i packets.

C. Node architecture of AR

We have considered three classes of traffic: RT and NRT data traffic, and signaling traffic (BU and refreshing BU packets) exchanged between mobile nodes and mobility agents for location management. Depending on the location of the MNs, there can be two statuses of each packet: *handoff* (H) and *non-handoff* (NH). *Non-handoff* status is when the MN is inside the coverage area of the AR only, whereas *handoff* status when it is moving towards some neighboring cell and about to handoff.

While sending data packet in the overlapped coverage region, the MN may set some flag (*handoff flag*) to indicate that it is about to handoff to some new AR. Other (data) traffic and RBU packets are assumed to have a *non-handoff* status.

Fig. 3 shows the queuing architecture at the access router. Packets of all types arrive at the AR at a rate λ. The classifier categorizes the packets into six classes according to the traffic and handoff status. The packets are then queued appropriately. The scheduler follows the scheduling algorithm (Section III-D) to select the next packet to transmit. Thus, packets of higher priority can be considered to have been queued ahead of other lower priority packet as shown in Fig. 4.

D. Scheduling algorithm

The steps are as follows:

- If the arrived packet is a BU packet, it is scheduled immediately after the packets in the BU queues if there is any.
- When the data packets (RT or NRT) with handoff flag = 1 arrives at the AR, they are selected with a higher priority than data traffic with non-handoff status. In addition, RT traffic gets higher priority than NRT traffic.

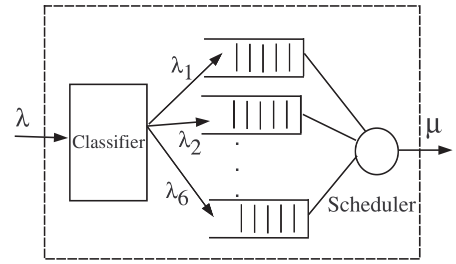

Fig. 3. Node architecture of the AR.

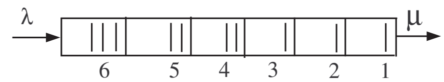

Fig. 4. Multi-class packets in the queue of the AR.

- While serving RT traffic, check the queue length of NRT traffic. If the NRT queue grows beyond *minThreshold* (usually three-fourth of queue size), allow dropping of RT traffic of different flows (if dropping is allowed).
- While serving RT traffic, if the NRT queue grows beyond *maxThreshold* (equals queue size), the transmission of RT traffic is suspended, and NRT traffic is transmitted until the NRT queue length comes below *minThreshold*.

As BU / RBU traffic has the highest priority, the queue associated with them will not grow much and with a reasonable size of the queue (in compliance with the node density) will ensure least loss of binding updates. As RT traffic (usually UDP packets) is loss-tolerant and useless if delivered after a long delay, the queue length of the RT traffic should be kept small. In case of overflow in RT queue, the (UDP) traffic will be lost which does not harm the RT communication. As NRT traffic (usually TCP traffic) is not loss-tolerant, NRT queue size is kept larger than RT queue. As NRT traffic is assigned the lowest priority, NRT packets (usually TCP) may be lost, resulting in the retransmission of the same packets.

E. Queueing analysis

Based on the proposed scheduling algorithm, we have computed the average queuing delay, queue occupancy and packet drop probability for BU packets, RT and NRT traffic through wireless media. We have used non-preemptive queuing which means lower priority packets in service will not be preempted by higher priority packets.

Let N_m be the number of mobile nodes under the radio coverage area of the AR. Among them, the fraction of MNs that are in the overlapped region are γN_m and these MNs will send data to the AR indicating this in the *handoff flag*. Without loss of generality, we can assume that γ fraction of the data packets will have handoff (H) status and rest will have non-handoff (NH) status. The probability that a MN will be in handoff mode is given by γ (see Eqn. (6) in Sec. II).

1) Packet arrival rates: In the overlapping region, binding updates are sent by MNs. As the BU packet rate at each MN is λ_{BU}, then BU packet arrival rate at the AR from the $N_m\gamma$

MNs will be given by

$$\lambda_1 = \lambda_{BU} \times N_m \gamma \tag{7}$$

Similarly, as the RBU packet rate at each MN is λ_{RBU}, RBU packet arrival rate at the AR from the $N_m(1-\gamma)$ MNs will be given by

$$\lambda_2 = \lambda_{RBU} \times N_m(1-\gamma) \tag{8}$$

For RT traffic, as the packet arrival rate to each MN is λ_{RT}, the packet arrival rate at AR with handoff status is

$$\lambda_3 = \lambda_{RT} \times N_m \gamma \tag{9}$$

Similarly, the RT packet arrival rate at AR with non-handoff status is given by

$$\lambda_4 = \lambda_{RT} \times N_m(1-\gamma) \tag{10}$$

For NRT traffic, as the packet arrival rate to each MN is λ_{NRT}, the packet arrival rate at AR with handoff status is

$$\lambda_5 = \lambda_{NRT} \times N_m \gamma \tag{11}$$

Finally, the NRT packet arrival rate at AR with non-handoff status is given by

$$\lambda_6 = \lambda_{NRT} \times N_m(1-\gamma) \tag{12}$$

Since the packet arrival rate to each queue is a Poisson process, then the total arrival rate of all classes of packets to the system will collectively be a Poisson process with rate $\lambda = \lambda_1 + \lambda_2$ +..+ λ_6. Let the service time of the system be exponentially distributed with mean $1/\mu$. Let $\rho_i = \lambda_i/\mu$. Therefore, the system utilization factor can be computed as

$$\rho = \frac{\lambda}{\mu} = \rho_1 + \rho_2 + .. + \rho_6 = \sum_{i=1}^{6} \rho_i \tag{13}$$

2) Mean packet delay and queue length: When a class-1 packet arrives, the system may be in the process of serving a packet of any other class. The probability that a class-1 packet finds a class-2 packet in service equals the fraction of time the server spends on class-2 packets which is $\lambda_2/\mu = \rho_2$. Thus class-1 packet finds any of the class-i packets in service is ρ_i for $2 \leq i \leq 6$.

According to Little's theorem [6], average number of packets waiting in the system is equal to the average delay times average arrival rate of the system. This theorem can be used in some part of the system. Let us apply this theorem for queue of class-i. As the average packet delay for class-1 packets is $E(T_1)$ and average queue occupancy is $E(n_1)$ with arrival rate λ_1, we have

$$E(n_1) = \lambda_1 E(T_1) \tag{14}$$

Again, when a packet of class-1 arrives at the system, there are, on the average, $E(n_1)$ packets on class-1 queue. Since class-1 packets have the highest priority, the mean time delay of the packet that has just arrived depends on $E(n_1)$ packets which are already buffered in the class-1 queue plus the packet in service. As $E(n_1)$ includes the class-1 packet in service (if any), the mean time delay of class-1 packet in the system is the queuing delay and the service time ($1/\mu$). Therefore,

$$\begin{aligned} E(T_1) &= \frac{E(n_1)}{\mu} + \frac{1}{\mu} + \frac{1}{\mu}(\rho_2 + \rho_3 + ... + \rho_6) \\ &= \frac{E(n_1)}{\mu} + \frac{1}{\mu} + \frac{1}{\mu}\sum_{i=2}^{6} \rho_i \end{aligned} \tag{15}$$

Now using Eqn. (14) in the above Eqn. (15) and after simplification, we get

$$E(T_1) = \frac{\left(1 + \sum_{i=2}^{6} \rho_i\right)}{(1-\rho_1)\mu} = \frac{1}{\mu} + \frac{\rho/\mu}{1-\rho_1} \tag{16}$$

Hence the Eqn. (14) can be rewritten as

$$E(n_1) = \frac{\left(1 + \sum_{i=2}^{6} \rho_i\right)\rho_1}{1-\rho_1} = \rho_1 + \frac{\rho\rho_1}{1-\rho_1} \tag{17}$$

While considering the delay of class-2 packets, we have to consider average number of class-1 packets ($E(n_1)$) in the system as they have the higher priority of service. So the class-2 packet which arrives in the system have to wait for all the class-1 packets buffered in the class-1 queue and class-2 queue.

$$\begin{aligned} E(T_2) &= \frac{E(n_1)}{\mu} + \frac{E(n_2)}{\mu} + \frac{1}{\mu} + \frac{1}{\mu}(\rho_3 + \rho_4 + ... + \rho_6) \\ &= \frac{E(n_1)}{\mu} + \frac{E(n_2)}{\mu} + \frac{1}{\mu} + \frac{1}{\mu}(\rho - \rho_1 - \rho_2) \end{aligned} \tag{18}$$

Using Little's theorem for class-2 queue, that is, $E(n_2) = \lambda_2 E(T_2)$ in Eqn. (18) and after simplification, we get

$$E(T_2) = \frac{1}{(1-\rho_2)\mu}\left[1 + \frac{\rho}{1-\rho_1} - \rho_2\right] \tag{19}$$

Hence, average queue length of class-2 queue is given by,

$$E(n_2) = \frac{\rho_2}{(1-\rho_2)}\left[1 + \frac{\rho}{1-\rho_1} - \rho_2\right] \tag{20}$$

Therefore, the general expression for packets of class-i can be derived as,

$$\begin{aligned} E(T_i) &= \frac{E(n_1)}{\mu} + \frac{E(n_2)}{\mu} + .. + \frac{E(n_i)}{\mu} + \frac{1}{\mu} \\ &+ \frac{1}{\mu}(\rho_{i+1} + \rho_{i+2} + ..\rho_k) \\ &= \frac{1}{\mu}\left(\sum_{j=1}^{i} E(n_j) + \sum_{j=i+1}^{k} \rho_j + 1\right) \end{aligned} \tag{21}$$

Again using Little's theorem for class-i queue and simplifying, we get

$$E(T_i) = \frac{1}{\mu(1-\rho_i)}\left(\sum_{j=1}^{i-1} E(n_j) + 1 + \rho - \sum_{j=1}^{i} \rho_j\right) \tag{22}$$

Therefore, average queue length of class-i queue is given by,

$$E(n_i) = \frac{\rho_i}{(1-\rho_i)}\left(\sum_{j=1}^{i-1} E(n_j) + 1 + \rho - \sum_{j=1}^{i} \rho_j\right) \tag{23}$$

TABLE I
VALUES OF PARAMETERS USED IN THE NUMERICAL ANALYSIS.

Parameter	Value	Parameter	Value
γ	0.7268	r	500 m
T_r	60 sec	T_{lf}	30 sec
T_{ovr}	10.9 sec	μ	30000 pkts/sec
λ_{RT}	128 pkts/sec	λ_{NRT}	32 pkts/sec
λ_{BU}	0.275 pkts/sec	λ_{RBU}	0.122 pkts/sec

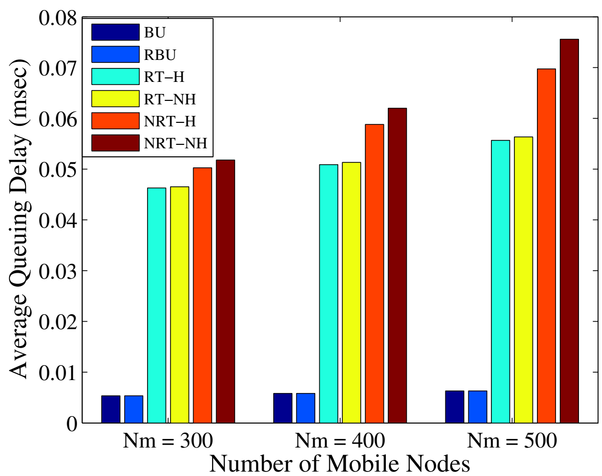

Fig. 5. Average queuing delay at the six queues for different number of MNs.

3) Packet drop Probability: Let us assume that N_i denotes the size of the ith queue. Therefore, the packet drop probability at each queue is the probability of the queue being full. Thus the packet drop probability at ith queue can be obtained as follows:

$$P_d(i) = \frac{\rho_i^{N_i}(1-\rho_i)}{1-\rho_i^{N_i+1}} \tag{24}$$

IV. NUMERICAL RESULTS

In this section, the numerical results are presented to show the effect of node density, service rate and traffic distribution on the average queuing delay, queue occupancy and packet drop probability at the AR. We have denoted the class-1 through class-6 as BU, RBU, RT-H, RT-NH, NRT-H and NRT-NH for clear understanding of the reader.

Table I lists the parameter values used for producing results. As the radius of the cell is 500 m, the total cell area is 785398.16 sq. m. If the subnet residence time is T_r sec, during γT_r sec MN will reside in one of the four overlapping regions of the cell (see Fig. 1). So during $\gamma T_r/4 = T_{ovr} = 15$ sec an MH will reside in one overlapping region, and three binding update messages (add new IP address, set primary, remove old IP address) are sent by an MN during T_{ovr} sec. Hence, $\lambda_{BU} = 3/T_{ovr}$ = 3/ 10.9 = 0.275 packet/sec As on the average during $(1-\gamma)T_r$ sec an MN resides inside the cell with no overlapping, $\lambda_{RBU} = (T_r/T_{lf})/(1-\gamma)T_r$ = 0.122 packet/sec. Size of BU/RBU packets are assumed to be 60 bytes. For RT traffic, we have assumed the average data rate is 512 kbps rate with 512 bytes of packets, while for NRT traffic, we assumed 128 kbps data rate with same packet size. Percentage of users accessing RT and NRT traffic are assumed to be 5% and 95%, respectively.

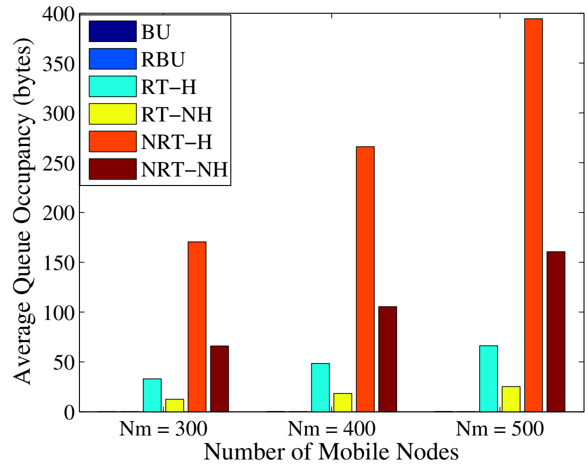

Fig. 6. Average queuing occupancy at the six queues for different number of MNs.

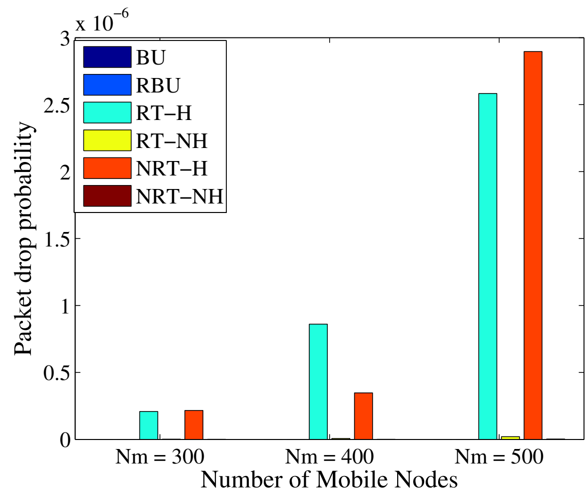

Fig. 7. Packet drop probability at the six queues for different number of MNs.

A. Impact of node density

In Fig. 5, average queuing delay at the six queues are shown for varying number of MNs in the cell. Results show that queuing delay of NRT traffic rises for higher node density whereas the queuing delay for signaling traffic remains unchanged ensuring faster delivery of signaling traffic.

In Fig. 6, average queue occupancy at the six queues are shown for different node density at the cell region. It is found that average queue occupancy is almost zero for signaling traffic as they are served very quickly by the AR. The NRT traffic in handoff status has the highest queue occupancy as the arrival rate is the highest unlike RT traffic which is served with higher priority to meet the deadline.

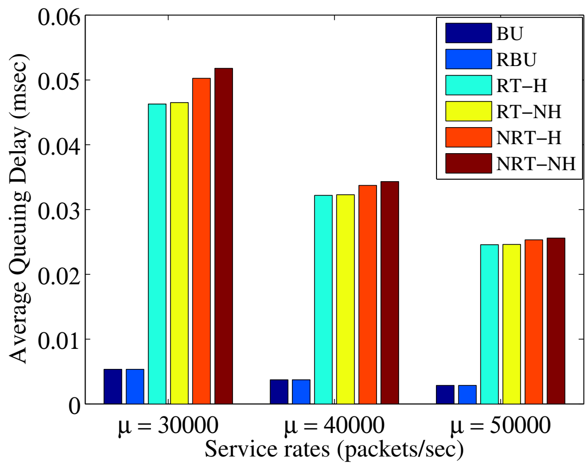

Fig. 8. Average queuing delay at the six queues for different service rates.

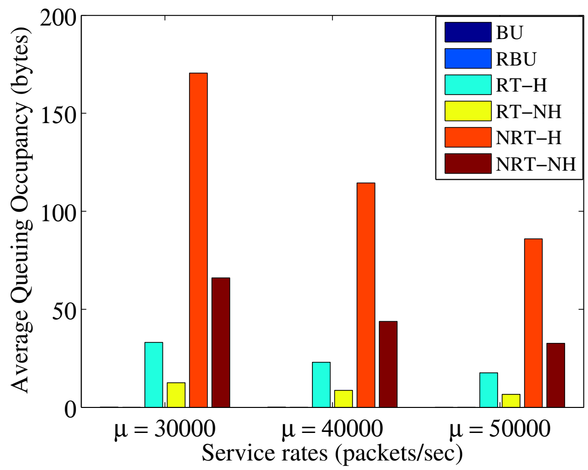

Fig. 9. Average queuing occupancy at the six queues for different service rates.

In Fig. 7, the packet drop probability for the six classes of packets is shown for different node density at the cell region. The probability increases for higher node density for RT and NRT traffic. However, the packet drop probability for signaling traffic is zero. The RT class has the smaller queue size resulting in more packet drop. In addition, data traffic (RT and NRT) has the higher arrival rate with handoff status causing more loss.

B. Impact of service rate

In Fig. 8, average queuing delays are shown for different service rates of the AR. The queuing delay decreases for higher service rate. Again, the delay for signaling traffic is the least whereas that of NRT traffic is the highest.

In Fig. 9, average queue occupancy for the six classes of traffic are shown for different service rates of the AR. Average occupancy at the queue decreases for higher service rate. For RT traffic, it is much less than the NRT traffic due to higher priority and smaller queue size.

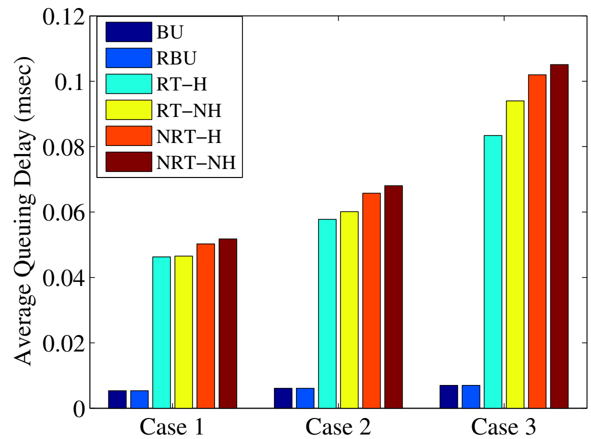

Fig. 10. Average queuing delay at the queues for different traffic.

C. Impact of traffic distribution

Fig. 10 shows the queuing delay for different classes of traffic varying the percentage of users accessing RT and NRT traffic. In case 1, MNs accessing RT and NRT traffic are assumed to be 5% and 95%, respectively. For case 2 and case 3, the distributions are 25%-75% and 50%-50%, respectively. It is found that as the RT traffic increases the queuing delay for data traffic increases.

V. Conclusion

In this paper, a scheduling algorithm has been proposed that selects packet based on the priority and current location of the mobile nodes, producing least loss to signaling traffic, while allowing packet loss to real-time traffic as RT traffic are loss-tolerant. Based on the scheduling algorithm, we have derived expressions for average queuing delay, queue occupancy, and packet drop probability of each class of traffic. Results showing the impact of node density, service rate and traffic distribution on those measures have been explained. The analytical framework presented in this paper will be helpful to better manage access routers with different classes of data and signaling traffic causing least queuing delay and packet loss.

References

[1] D. Johnson, C. E. Perkins, and J. Arkko, "Mobility support in IPv6," IETF RFC 3775, June 2004.
[2] V. Devarapalli, R. Wakikawa, A. Petrescu, and P. Thubert, "NEtwork MObility (NEMO) basic support protocol," RFC 3963, Jan 2005.
[3] H. N. Nguyen and I. Sasase, "Downlink queuing model and packet scheduling for providing lossless handoff and QoS in 4G mobile network," *IEEE Transactions on Mobile Computing*, vol. 5, no. 5, pp. 452–462, May 2006.
[4] M. Iftikhar, T. Singh, B. Landfeldt, and M. Caglar, "Multiclass G/M/1 queueing system with self-similar input and non-preemptive priority," *Computer Communications*, vol. 31, pp. 1012–1027, 2008.
[5] M. Iftikhar, B. Landfeldt, and M. Caglar, "Towards the formation of comprehensive SLAs between heterogeneous wireless DiffServ domains," *Telecommunication Systems*, vol. 42, pp. 179–199, 2009.
[6] G. Bolch, S. Greiner, H. de Meer, and K. S. Trivedi, *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*. New York, NY: Wiley-Interscience, 1998.