A Novel Scheduling and Queue Management Scheme for Multi-band Mobile Routers

Md. Shohrab Hossain, Husnu S. Narman and Mohammed Atiquzzaman

School of Computer Science, University of Oklahoma, Norman, OK 73019

Email: {shohrab, husnu, atiq}@ou.edu

Abstract—Recent trends in Internet usage have seen large amount of multimedia data due to increasingly large number of mobile users. To facilitate higher bandwidth, modern mobile routers are capable of supporting simultaneous multi-band, leading to less interference, higher capacity. However, there exists no previous work that attempts to maximize utilization through sharing of traffic among different frequency bands. In this paper, we have proposed a novel scheduling algorithm for multi-band mobile routers which transmits different classes of traffic through different frequency bands to achieve improved performance. We have developed an analytical model to perform queuing analysis of the proposed multi-band system and derived various performance metrics, validated by extensive simulations. Our results show that the proposed architecture can ensure maximum possible utilization through sharing of capacities among the bands. Our proposed scheduling algorithm and related analytical model can help network engineers build next generation mobile routers with higher throughput and utilization ensuring least packet loss for different classes of traffic.

Index Terms—Analytical modeling, scheduling algorithm, queuing system, real-time traffic, next generation mobile routers.

I. Introduction

In recent years, there has been explosive growth of mobile users accessing large multimedia files (such as, high definition audio, video, images, etc.) over the Internet. To satisfy such a higher bandwidth requirement, today wireless routers are available commercially with simultaneous multi-band support of 2.4 and 5 GHz. Exploitation of rarely-used frequency bands in wireless networks reduces interference in heavily-used frequency band, thereby increasing total capacity of the wireless network.

Current simultaneous multi-band MRs make use of two different bands without trying to exploit the under-utilized frequency band. Moreover, there are different classes of traffic (such as, real-time) having strict delay constraints; some other signaling traffic (required for mobility management) are crucial for maintaining Internet connectivity of the mobile users. Therefore, it is essential to propose appropriate scheduling and queue management scheme for the multi-class traffic to ensure maximum possible utilization of the system resources in multi-band mobile routers [1].

There have been several research works [1]–[7] reported in the literature that attempt to extend current single band technology through the use of multiple frequency bands. Even though multi band usage has been widely investigated in cell networks [5], [6], it is relatively new concept in wireless networks. Verma and Lee [7] explained possible Wi-Fi architecture with multiple physical and link layers to support multiple frequency bands simultaneously. Singh et al. [2] proposed a method to assign end-devices different bands based on their distances from the access router. In [3], [4], authors proposed the use of 60 GHz frequency band (having low range) to attain faster data transfer. However, none of these works [1]–[7] propose any scheduling algorithm for multi-band system considering multi-class traffic, neither do they perform any queueing analysis to measure different performance metrics. This is a *novel work* that aims at attaining maximum possible band utilization and proposes a scheduling algorithm which exploits band-sharing.

The *objective* of this work is to analyze the performance of multi-band mobile routers while ensuring maximum possible utilization through band sharing.

The *contributions* of this work are: (i) proposing a band-sharing router architecture and a novel scheduling algorithm to ensure maximum possible utilization of the system, (ii) developing an analytical model for the performance evaluation of the proposed multi-band system, and (iii) comparing the proposed router architecture with the typical one.

Our proposed algorithm considers multi-class Internet traffic and schedules them through alternate under-utilized frequency bands, thereby reducing packet loss and delay. Results show the packet drop rate and throughput are significantly improved in proposed band-sharing architecture of the mobile router.

Our proposed scheme and related analysis will help network engineers build next generation mobile routers with higher throughput and utilization, ensuring least packet loss of different classes of traffic.

The rest of the paper is organized as follows. In Section II, we explain the typical architecture of multi-band mobile routers, followed by the proposed architecture in Section III. Section IV presents the analytical model to derive different performance metrics of the proposed architecture. In Section V, we present the simulation results that validates our analytical model; we also compare the proposed and typical architecture. Finally, Section VI has the concluding remarks.

II. Typical multi-band router architecture

Commercial (simultaneous) multi-band mobile routers available today make use of two different bands (2.4 and
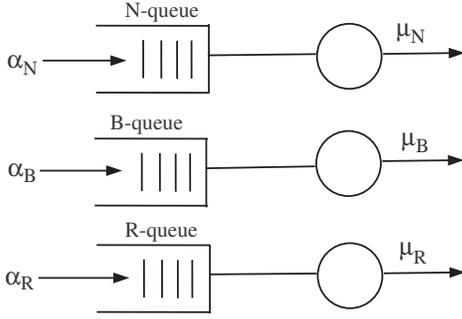
Fig. 1.   Architecture of a typical (simultaneous) multi-band mobile router.
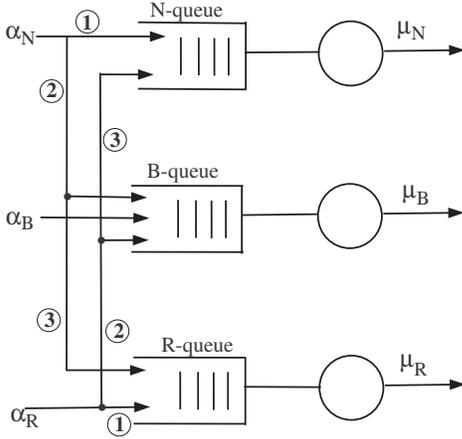
Fig. 2.   Proposed architecture of a simultaneous multi-band mobile router.

5 GHz) for different types of devices in a home network. The main principle of today's simultaneous multi-band MR is the non-sharing of bands among different flows of traffic. Moreover, some of the devices today (such as, IPTV) mostly deals with real-time traffic. Based on this fact, we have assumed that each of the band of typical simultaneous multi-band MR only deals with one type of traffic. This might be a slight deviation from the real MR used today. However, we have assumed this to compare our proposed architecture with typical simultaneous multi-band MR.

Fig. 1 shows the typical architecture of a simultaneous tri-band MR. Here, three bands are assumed to be used for three different classes of traffic: signaling traffic or Binding Update (BU), real-time (RT) and non-real time (NRT) traffic. Each class of traffic is solely assigned to each designated frequency band as shown in Fig. 1 and we name the corresponding queue as B-queue, R-queue and N-queue. There is absolutely no sharing of traffic among different bands even if one (or more) bands are under-utilized due to low traffic arrivals.

III. PROPOSED MULTI-BAND ROUTER ARCHITECTURE

In the proposed architecture, we have considered three different queues (Fig. 2), each of which corresponds to a frequency band of a simultaneous tri-band MR. As in Fig. 1, we consider three classes of traffic and each queue is des-
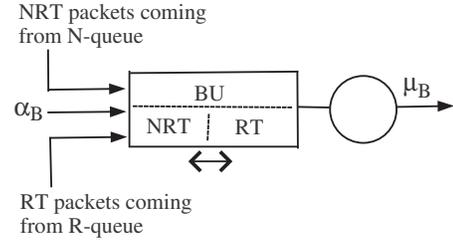
Fig. 3.   Queue corresponding to BU band.
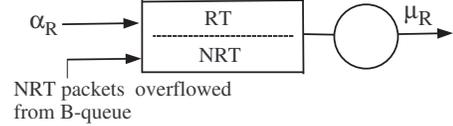
Fig. 4.   Queue corresponding to RT band.

ignated for each class of traffic. However, unlike the typical architecture, in the proposed architecture, traffic of one class can flow through other queues provided there are empty slots, thereby ensuring better utilization of buffer spaces. For example, if the B-queue has some empty spaces and a bursty RT traffic comes in, the overflowed RT traffic can be queued in the B-queue and subsequently served (or sent) through the B-server (transmitter). The numbers to the left of Fig. 2 designate the order in which the RT and NRT traffic can be queued to different queues, 1 meaning the first choice, and so on.

A. Time and space priority

The time and space priority for the three queues of the proposed architecture are explained in Figs. 3, 4 and 5. For B-queue, BU packets have the highest priority; RT and NRT packets have dynamic priority based on arrival rates (see Eqns.(4) and (5)). Regarding space priority, BU packets are queued in front of B-queue and if there are empty spaces, other types (RT and NRT) can be accommodated as shown in Fig. 3. R-queue can have only RT and NRT packets as shown in Fig. 4. RT traffic has higher priority over NRT traffic. Therefore, R-queue can have NRT packets only if RT packets cannot fill the R-queue at any instant and there are NRT packets overflowed from the B-queue. Finally, Fig. 5 shows the N-queue which is designated for NRT traffic. However, if there are empty spaces available in this queue, overflowed RT traffic out of B-queue can be enqueued in N-queue (see Fig. 2).

B. Scheduling algorithm

The proposed scheduling policies are explained as follows:
- Attempts are first made to queue different class of traffic in their corresponding buffer.
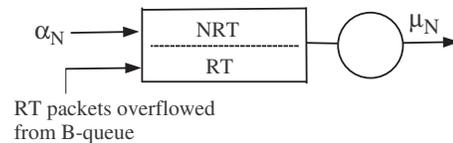
Fig. 5.   Queue corresponding to NRT band.

- If N-queue (or R-queue) overflows, corresponding traffic is forwarded to B-queue.
- If B-queue does not have enough empty slots to handle the overflowed NRT and RT packets, they race for slots in B-queue based on priority (see Eqns. (4) and (5)).
- If overflowed RT packets cannot be accommodated in B-queue, they are queued in N-queue, if space is available in the queue.
- If the R-type packets cannot even be accommodated in N-queue, they are dropped from the system.
- Similar policy is enforced when dealing with NRT packets in the B-queue and then in the R-queue.

IV. ANALYTICAL MODEL

In this section, we present our analytical model to derive various performance metrics of the proposed multi-band MR architecture.

A. Assumptions

To make the model analytically tractable, the following assumptions have been made:
- Packet arrival follows Poisson distribution.
- Type of queue discipline used in the analysis is FIFO with non-preemptive priority among various traffic classes [8].

B. Notations

The notations used in the analysis are listed below. To simplify our notation, we use T as the common notation for different traffic class types, and $T \in \{$ B, N, R $\}$.

N_T Queue size of T-queue in the MR,
α_T Total packet arrival rate at T-queue of i-th MN,
μ_T Service rate at T-queue of i-th MN,
$\sigma_{T_{BQ}}$ Priority of class-T traffic in B-queue,
$P_{dT_{XQ}}$ T-type packet drop probability in X-queue, where $X \in \{B, N, R\}$,
$E(D_T)$ Average queuing delay of class T packets,
$E(n_T)$ Average queue occupancy of class T packets,
P_{dT}^{sys} Final packet drop probability of class T packets.

C. Total arrival rates in each queue

For queuing analysis of the proposed system, we need to determine the total arrival rate of all class of traffic in each queue. In general, overflow in a queue can happen when the arrival rate is larger than the buffer space or when the service rate is smaller than the arrival rate and there is no buffer space left, as specified by the following conditions:

$$\alpha_T > N_T \quad or \quad \mu_T < \alpha_T \tag{1}$$

1) B-queue: For B-queue, number of packets overflowed from N-queue and R-queue goes to B-queue. Thus, the arrival rates of N-type and R-type packets to the B-queue (denoted by α'_N and α'_R) can be obtained as follows:

$$\alpha'_N = \alpha_N P_{dN_{NQ}} \tag{2}$$

$$\alpha'_R = \alpha_R P_{dR_{RQ}} \tag{3}$$

where $P_{dN_{NQ}}$ and $P_{dR_{RQ}}$ are packet drop probabilities of N-type packets in N-queue and R-type packets in R-queue. Priorities of different classes are taken into account while allowing traffic into B-queue. Priority of B packets in B-queue is $\sigma_{B_{BQ}}$ = 1. Priorities of other classes of traffic in B-queue are measured as follows:

$$\sigma_{N_{BQ}} = \frac{\alpha_R}{\alpha_B + \alpha_R + \alpha_N} \tag{4}$$

$$\sigma_{R_{BQ}} = \frac{\alpha_N}{\alpha_B + \alpha_R + \alpha_N} \tag{5}$$

Now, the total (effective) arrival rate of all class of traffic in B-queue can be obtained as follows:

$$\alpha_{B(Total)} = \alpha_B + \alpha'_N + \alpha'_R \tag{6}$$

2) N-Queue: The N-queue is designated for NRT traffic. However, if there are empty buffer spaces available in the N-queue (due to low NRT arrival rate), this queue can be used to transmit RT traffic that are overflowed from the B-queue. Let α''_R denotes the arrival rate of RT packets in N-queue. Therefore, the total arrival rate (of both N and R-type packets) in the N-queue is as follows:

$$\alpha_{N(Total)} = \alpha_N + \alpha''_R \tag{7}$$

3) R-Queue: The R-queue is designated for RT traffic. However, if there is empty buffer space available in the R-queue (due to low RT arrival rate), this queue can be used to transmit NRT traffic which has been overflowed from the B-queue. Let α''_N denotes the arrival rate of N-type packet going to R-queue. Therefore, the total arrival rate (of both R and N-type packets) in the R-queue is as follows:

$$\alpha_{R(Total)} = \alpha_R + \alpha''_N \tag{8}$$

D. Computing drop probability

The packet drop probability of R-type packets in R-queue can be obtained using standard M/M/1/N formula as follows [9]:

$$P_{dR_{RQ}} = \frac{\rho_R^{N_R}(1 - \rho_R)}{1 - \rho_R^{N_R+1}} \tag{9}$$

where $\rho_R = \frac{\alpha_R}{\mu_R}$. Similarly, the packet drop probability of N-type packets in N-queue can be obtained as follows:

$$P_{dN_{NQ}} = \frac{\rho_N^{N_N}(1 - \rho_N)}{1 - \rho_N^{N_N+1}} \tag{10}$$

where $\rho_N = \frac{\alpha_N}{\mu_N}$. Let us assume that the priority of R-type packets is higher than that of N-type packets in B-queue. Therefore, while computing the R-type packet drop probability in B-queue, we can safely consider only B-type and R-type packets in B-queue. Let us define utilization in B-queue considering only B-type and R-type packets be $\rho_{BR} = \frac{\alpha_B + \alpha'_R}{\mu_B}$. Thus, the packet drop probability of B-type packet in B-queue, denoted by $P_{dB_{BQ}}$, can be obtained as follows [10]:

$$P_{dB_{BQ}} = \frac{\rho_{BR}\rho_1^{N_B}(1 - \rho_1)(1 - \rho_{BR}^{N_B+1})}{(1 - \rho_1^{N_B+1})(1 - \rho_{BR}^{N_B+2})} \tag{11}$$

where $\rho_1 = \alpha_B/\mu_B$ and $\rho_2 = \alpha'_R/\mu_B$. Using Eqn. (11), the packet drop probability of R-packets in B-queue can obtained as follows [10]:

$$P_{dR_{BQ}} = \frac{(1 - \rho_{BR})}{\left(1 - \rho_{BR}^{N_B+2}\right)} \rho_{BR}^{N_B+1}$$
$$+ \frac{\alpha_B}{\alpha_R'} \left(\frac{(1 - \rho_{BR})}{\left(1 - \rho_{BR}^{N_B+2}\right)} \rho_{BR}^{N_B+1} - P_{dB_{BQ}} \right) \quad (12)$$

Hence, the RT packet arrival in N-queue can be obtained as follows:
$$\alpha_R'' = \alpha_R' P_{dR_{BQ}} \quad (13)$$

The total arrival of N-queue is the sum of two arrival rates α_N and α_R'' (see Eqn. (7)); the former has the higher priority than the latter. Therefore, following a similar approach as in Eqn. (11), we can compute $P_{dN_{NQ}}$. Then we can follow similar approach in Eqn. (12) to compute $P_{dR_{NQ}}$ which is the final drop of R-type packets from the system. That is, $P_{dR}^{sys} = P_{dR_{NQ}}$. The computation of N-type packet drop probability follows similar steps as followed for R-type packets. Therefore, $P_{dN}^{sys} = P_{dN_{RQ}}$.

E. Average queue length

Each queues behaves as M/M/1/N queue. Therefore estimated queue length can be obtained as follows:

$$E(n_T) = \begin{cases} \dfrac{\rho_T - (N_T+1)\rho_T^{N_T+1} + N_T \rho_T^{(N_T+2)}}{\left(1 - \rho_T\right)\left(1 - \rho_T^{N_T+1}\right)} & , \text{if } \rho_T \neq 1 \\ \dfrac{N_T}{2} & , \text{if } \rho_T = 1 \end{cases}$$
$$(14)$$

The average queue occupancy of R-type depends on the queue occupancy of R-packets in R-queue, B-queue and N-queue. This is computed as follows:

$$E(n_R^{sys}) = E(n_R^{RQ}) + E(n_R^{BQ}) + E(n_R^{NQ})$$
$$= E(n_R^{RQ}) + \left(E(n_{B+R}^{BQ}) - E(n_B^{BQ})\right) + \left(E(n_{N+R}^{NQ}) - E(n_N^{NQ})\right) \quad (15)$$

For computing $E(n_R^{RQ})$, we need to put $N_T = N_R$, $\rho_T = \rho_R = \alpha_R/\mu_R$ in Eqn. (14). Similar approach can be used for the rest of the terms in Eqn. (15).
Since B-type packets are only queued in B-queue, therefore,

$$E(n_B^{sys}) = E(n_B^{BQ}) \quad (16)$$

To compute average queue occupancy of N-type packets in the system, a similar approach as in Eqn. (15) can be used:

$$E(n_N^{sys}) = E(n_N^{NQ}) + E(n_N^{BQ}) + E(n_N^{RQ}) \quad (17)$$

F. Throughput

The throughput of T class of traffic can be obtained as follows:
$$\gamma_T^{sys} = \left(1 - P_{dT}^{sys}\right)\alpha_T \quad (18)$$

G. Average packet delay

Thus, the average packet delay of T class is given by,
$$E(D_T^{sys}) = \frac{E\left(n_T^{sys}\right)}{\left(1 - P_{dT}^{sys}\right)\alpha_T} \quad (19)$$

V. RESULTS

We have written discrete event simulation programs in MATLAB environment following M/M/3/N [9] procedures. Equal buffer length (of 50 packets) was kept for each queue. RT and NRT packets are assumed to be 512 bytes [3] whereas the BU packets are assumed to be 64 bytes. The service rates of the B, N and R-queues are kept 27, 75 and 132 packets/sec which is proportional to service rates of multi-band routers [3]. We ran each simulation for 20 trials having different traffic class arrival rates as follows:
$$\lambda_B(i) = \{\ i\ \},\ \lambda_N(i) = \{\ 3i\ \},\ \lambda_R(i) = \{\ 18i\ \}$$

where i = 1, 2, 3, ..., 20. We have run simulations with increased arrival rates of all types of traffic to observe the impact of heavy traffic on the multi-band system. The arrival rate of B-queue and N-queue are increased slowly in each trial whereas the RT traffic arrival rate are increased at a much higher rate. This eventually saturates the R-queue and we explain the impact of this overflow on different performance parameters of our proposed system and typical existing system.

A. Validation of analytical model

The validation part has been presented in the technical report [11] where we find that the analytical and simulation results are very close to each other.

B. Comparison between proposed and typical architectures

In this subsection, we present results comparing the performance results of our proposed (band-sharing) architecture and typical (non-sharing) architecture.

1) Utilization: Utilization is a performance measurement that indicates how efficiently bands are used and whether there is any unused capacity of the system. As shown in Fig. 6, when packet arrival rates are low (in trials #1 through 7) compared to the capacity of each queue, all the queues have somewhat similar and low utilization for both the architectures. However, for trials #8 through 20, the utilizations of B-queue and N-queue are much higher for proposed architecture than for typical one. This is because increased number of RT packets are dropped in typical architecture (see Fig. 8) whereas in proposed one, they are accommodated in B-queue and N-queue, thereby improving their utilizations and maximizing system performance.

2) Average packet delay: As shown in Fig. 7, the delay for RT traffic (for trails # 8 through 20) in proposed architecture is higher than the typical one. This is because excessive RT packets are immediately dropped from the system in typical architecture and these lost packets do not come into account in delay calculations. On the contrary, in proposed architecture overflowed RT packets get chances to be enqueued in B and N-queue before being dropped.

3) Packet drop probability: For both the approaches, drop rate of BU and NRT are low and similar (see Fig. 8) since arrival rates are lower than the service rates. When the arrival rates for RT increases (in trials # 7 through 20), RT packet drop rate gradually goes up for typical (non-shared) architecture.
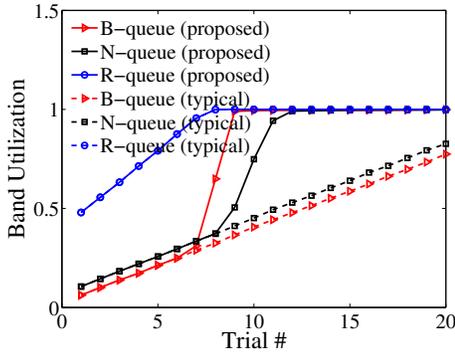
Fig. 6. Band utilizations for proposed and typical architectures for different simulation trials.
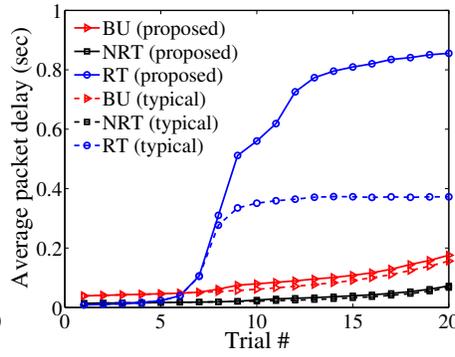
Fig. 7. Average packet delay for proposed and typical architectures for different simulation trials.
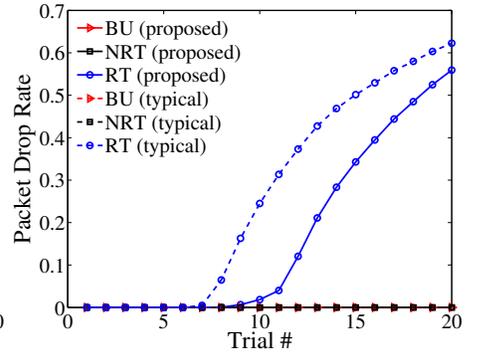
Fig. 8. Average packet delay for proposed and typical architectures for different simulation trials.

However, proposed architecture does not allow RT traffic to drop as long as the excessive RT packets can be queued in the empty buffer spaces at B-queue and N-queue.
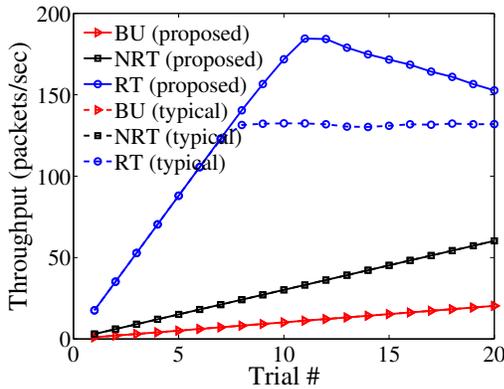
Fig. 9. Throughput for proposed and typical architectures for different simulation trials.

4) Throughput: As shown in Fig. 9, the throughput of NRT and BU class are increased with the increase of their arrival rates for both architectures. However, in case of RT class and for the typical architecture, the throughput is saturated at μ_R (= 132 pkts/sec) when the RT arrival rate reaches this value. However, the RT class throughput for the proposed architecture go much higher (due to sharing of other under-utilized bands) and reaches its peak value in trial #11. After that it starts to decrease slowly due to the impact of increased arrival rates of other queues (B and N-queue).

C. Summary of results

Results show that the proposed architecture attempts to maximize system utilization (Fig. 6) through band sharing. This affects on the average queue occupancy and delay (Fig. 7) of RT traffic. However, the packet drop and throughput (Figs. 8 and 9) are significantly improved in the proposed band-sharing architecture.

VI. CONCLUSION

In this paper, we have proposed a novel scheduling algorithm for multi-band mobile routers that exploits band

sharing. We have developed an analytical model to perform queuing analysis of the proposed multi-band system and derived various performance metrics that have been validated by extensive simulations. Our results show that the proposed architecture can ensure maximum possible utilization through the sharing of capacities among the bands. Our proposed scheduling algorithm and related analytical model can help network engineers build next generation mobile routers with higher throughput and utilization ensuring least packet loss for different classes of traffic.

REFERENCES

[1] I. F. Akyildiz, D. M. Gutierrez-Estevez, and E. C. Reyes, "The evolution to 4G cellular systems: LTE-advanced," *Physical Communication*, vol. 3, pp. 217–244, March 2010.

[2] H. Singh, J. Hsu, L. Verma, S. S. Lee, and C. Ngo, "Green operation of multi-band wireless LAN in 60 GHz and 2.4/5 GHz," in *Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, Jan 2011, pp. 787–792.

[3] E. Perahia, C. Cordeiro, M. Park, and L. L. Yang, "IEEE 802.11ad: defining the next generation multi-gbps Wi-Fi," in *7th IEEE Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, Jan 2010.

[4] S. Singh, R. Mudumbai, and U. Madhow, "Distributed coordination with deaf neighbors: Efficient medium access for 60 ghz mesh networks," in *IEEE INFOCOM*, San Diego, CA, March 2010.

[5] Y. bing Lin, W. ru Lai, and R. jaye Chen, "Performance analysis for dual band PCS networks," *IEEE Transactions on Computers*, vol. 49, pp. 148–159, Feb 2000.

[6] K. Doppler, C. Wijting, T. Henttonen, and K. Valkealahti, "Multiband scheduler for future communication systems," *I. J. Communications, Network and System Sciences*, vol. 1, no. 1, pp. 1–9, Feb 2008.

[7] L. Verma and S. S. Lee, "Multi-band Wi-Fi systems: A new direction in personal and community connectivity," in *IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, Jan 2011, pp. 665–666.

[8] M. S. Hossain, M. Atiquzzaman, and W. Ivancic, "Scheduling and queue management for multi-class traffic in access router of mobility protocol," in *IEEE International Conference on High Performance Computing and Communications (HPCC)*, Melbourne, Australia, Sep 1-3, 2010.

[9] D. Gross, J. Shortle, J. Thompson, and C. M. Harris, *Fundamentals of Queueing Theory*. Wiley-Interscience, Aug 2008.

[10] K. E. Avrachenkov, N. O. Vilchevsky, and G. L. Shevlyakov, "Priority queueing with finite buffer size and randomized push-out mechanism," *Performance Evaluation*, vol. 61, pp. 1–16, Jun 2005.

[11] M. S. Hossain, H. S. Narman, and M. Atiquzzaman, "Scheduling and queue management of multi-band mobile routers," University of Oklahoma, Technical report, TR-OU-TNRL-13-101, Feb. 2013, http://cs.ou.edu/~netlab/.