Call Capacity Model and QoS Routing Algorithms on Multi-Channel Multi-Radio Wireless Meshes

Sunghun Kim
The Attached Institute of ETRI
Daejeon, Korea
Email: walmuns@ensec.re.kr

MyungIn Ji
ETRI
Daejeon, Korea

Joong Soo Ma
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, Korea

Abstract—The capacity of the conventional wireless mesh network (WMN) with single channel single radio is limited due to co-channel interference. To resolve this capacity limitation problem, multi-channel multi-radio (MCMR) protocols have been proposed, and hybrid multi-channel protocol (HMCP) is a representative MCMR protocol. However, supporting delay sensitive realtime VoIP applications over high capacity MCMR-WMNs is still challenging problem. To support VoIP service over an MCMR WMN, we need the accurate call capacity model to estimate the feasible capacity region, and route computation algorithms for allocating the routes of voice calls within the feasible region to meet quality-of-service (QoS) constraints and improve the call capacity. In this paper, we introduce a new voice call capacity model of HMCP on MCMR WMNs. Both experimental and simulation results demonstrate that the proposed call capacity model accurately estimates the voice call capacity for G.711 and G.729 codecs within 5% of the actual call capacity. Also, we propose two QoS routing algorithms for finding feasible routes to meet QoS constraints as well as to improve the call capacity of network by utilizing the proposed call capacity model (i.e., feasibility considerations). By extensive simulations, our proposed QoS routing algorithms effectively protect voice calls and increase the call capacity.

I. INTRODUCTION

Voice over Internet Protocol (VoIP) services have been proliferated, and the widespread wireless LAN (WLAN) supports VoIP services in homes and offices [1]. For example, Skype, one of the most popular VoIP applications, is used by over 521 million users worldwide [12]. However, the WLAN has some limitations, such as small service coverage and difficulty of deployment. To resolve these limitations, the emerging wireless mesh network (WMN) is currently attracting.

It has been revealed that the capacity of a single channel single radio multi-hop wireless mesh network can not scale up with the network size, due to the co-channel interference [2]. A multi-channel multi-radio (MCMR) technique has been widely considered as an efficient approach to increase the wireless network capacity [2]–[5]. In [5], Kysasenu and Vaidya studied how the capacity of a multi-channel network scales as the number of nodes, channels, and radios increases. Their results imply that it may be possible to build capacity-optimal multi-channel networks with a small number of radios per node. Also, they proposed hybrid Multi-channel protocol (HMCP) which boosts the network capacity linearly as increasing the number of channels with only two radios per node.

HMCP boosts the network capacity, however, supporting delay sensitive realtime VoIP applications over WMNs is still challenging. This because, the call or network capacity threshold is reached, delay increases dramatically and R-score goes down (the detailed results in our previous work [12]). To maintain the network performance of an MCMR WMN stable and to provide good VoIP services, we must do call admission control with the accurate call capacity model. The accuracy of network quality management highly depends on how well the call capacity of the wireless network is inferred. However, estimating voice call capacity of a given network configuration is challenging problem in this context because we should consider wireless medium characteristics and HMCP operations in a multi-hop multi-channel multi-radio environment. We develop a call capacity model for HMCP in a multi-hop MCMR WMN. We also discuss about routing algorithms to allocate the route of a new arriving call within feasible regions. The major contributions of our work can be summarized as follows:

- We design the voice call capacity model of hybrid multi-channel protocol with constant bit rate (CBR) and variable bit rate (VBR) traffics on a multi-hop multi-channel multi-radio WMN to do call admission control. Also, we validate this model with multi-hop scenarios via both
testbed experiments and NS-2 simulations [9].

- We address the routing computation problem to find a feasible route for a VoIP call with quality of service (QoS) constraints. We develop simple but effective routing algorithms with feasibility considerations. Extensive simulation results demonstrate that our algorithm effectively protect voice calls and increase served call capacity.

II. SYSTEM MODEL

As shown in Fig. 1, mesh nodes are a basic element of a WMN, and they have a routing functionality to connect to each other. A mesh access point (AP) has an additional function to the mesh node for serving clients. A mesh portal is also a gateway of the Internet. A call admission controller can be located in the mesh portal. The call admission controller uses a call capacity model to decide acceptance of the arriving call with a route between source and destination based on interference map and traffic load information. This centralized configuration enables to manage stable VoIP call management easily, and many commercial WLAN and WMN platforms have followed a centralized approach [7], [8]. Each node in an MCMR WMN operates as HMCMP. As shown in Fig. 2, an HMCMP node uses two radios to communicate with its neighboring nodes. One is a fixed radio to receive data from its neighboring nodes via its own fixed channel. The other is a switchable radio to transmit data to its neighboring nodes via the neighboring nodes’ fixed channel. Each HMCMP node selects the two-hop least used channel as its fixed channel [5].

An HMCMP node advertises its fixed channel information via broadcasting HELLO packets over all available channels and utilizes this information to lookup the fixed channel of the destined node of the packet. If an HMCMP node has traffic for two or more neighbors with different fixed channels, then channel switching occurs to transmit data. Current IEEE 802.11 chipset supports channel switching within 25–35 μs [10]. In this paper, terms interfaces and radios are used interchangeably.

III. CALL CAPACITY UTILIZATION MODEL

To model voice call capacity with HMCMP, we first observe behaviors of HMCMP. As shown in Fig. 4, a HMCMP node consumes its time to transmit data with deferring other transmissions due to CSMA/CA mechanism. Also, the HMCMP node should switch its switchable radio channel to the fixed channel of the intended neighbor node of the packet.

We define the normalized capacity utilization at node i, \((c_i)\), which is the total bits/second traffic transmitted, received, heard, or channel switching at node i, normalized to the nominal link capacity. For brevity, we write “capacity utilization” for “normalized capacity utilization.” Throughout this paper, all voice traffic flows are bi-directional constant bit rate (CBR) and variable bit rate (VBR) traffic flows.

(a) Interference Map: Modeling call capacity utilization starts from characterizing the interference between nodes and building an interference map. We use the carrier sense factor (csf) to characterize the interference between a given pair of nodes. For the given nodes x and y, \(csf_{xy}^x\) (carrier sense factor of x with y) is defined by \(csf_{xy}^x = \frac{x's\ actual\ transmission\ rate\ without\ y}{x's\ actual\ transmission\ rate\ without\ y}\), where both x and y attempt to transmit data with the maximum possible rate. This \(csf_{xy}^x\) has values between 0.5 and 1. If x and y sense each other’s carrier perfectly, \(csf_{xy}^x\) is 0.5. This is because they have equal transmission opportunities by MAC protocol operations. On the contrary, \(csf_{xy}^x\) is 1 if x and y cannot sense each other’s carrier at all.

(b) Original Offered Load: An original offered load (I) at each node is normalized by the call link capacity between nodes. In an MCMR environment, the call link capacity between nodes x and y can be expressed as

\[ I_{x,y} = \frac{1}{R_{avg}(I_{voice,pkt}) + (T_{slot} \cdot CW_{min}/2)} \]

where \(T_{voice,pkt}\) is \(T_{payload} +\) short inter-frame space (SIFS) + \(T_{ack} +\) distributed coordination function inter-frame space (DIFS). \(T_{payload}\) is the time to transmit voice data consisting of a physical layer convergence protocol (PLCP) header, a preamble, a MAC header, a frame check sequence, real-time transport protocol/user datagram protocol/IP headers, and a payload. \(T_{ack}\) is the required time for sending an acknowledgment (ACK) frame, and \(T_{slot}\) is a slot time. \(R_{avg}\) is the average number of packets generated at each node per second, and \(R_{avg}\) can be represented as \(R \times R_{voice,active, ratio}\). R is the number of packets generated by per second with CBR mode, \(R_{voice,active, ratio}\) is the voice active ratio (its typical value is 0.385 in ITU-T conversation model [11]). Also, \(CW_{min}\) is the minimum contention window size. The value of \(I_{x,y} \in [1, 2]\) is determined by the configuration of the channel assignment.
in the MCMR-WMN. \( f_{x,y} \) is 1 if node \( x \)'s fixed channel is different from node \( y \)'s fixed channel. Otherwise, \( f_{x,y} \) is 2 because a VoIP call consists of two reverse-direction flows, and these flows interfere each other.

The original offered load for VoIP calls at node \( i \) to node \( k \) using fixed channel \( c \) is expressed as

\[
i_{i,k}(c) = \frac{n_{i,k}}{f_{i,k} \cdot N_{i,k}},
\]

where \( n_{i,k} \) is the number of calls through the link from node \( i \) to node \( k \) using fixed channel \( c \). Thus, the total original offered load at node \( i \) can be represented by \( l_i = \sum_c \sum_k l_{i,k}(c) = \sum_c l_c(c) \), where \( c \in \{ \text{available channels} \} \) and \( k \in \{ \text{neighbors of node } i \} \).

(c) Hypothesized Traffic Load: This traffic term models the amount of time for “D” and “C” periods in Fig. 4. Each node attempts to transmit the amount of an original load. However, the node can transmit extra traffic due to retrieessions by packet collisions. For VoIP traffic, the RTS/CTS mechanism is typically disabled because the VoIP payload size is small, and RTS/CTS is an overhead. Collisions occur at a receiver if the receiver has an interferer in its carrier sensing range, which is outside the carrier sensing range of the transmitter (hidden terminal problem). To model this hidden terminal problem, we can justify conditions for the hidden terminal problem. Node \( i \) with an intended receiver \( k \), and node \( j \) are hidden from each other if \( csf_{i,k} = 1 \) and \( csf_{j,k} < 1 \). The proportion of node \( j \)'s traffic that reached node \( k \) is \( 2 \cdot (1 - csf_{j,k}) \). Thus, we approximate the probability of packet collisions at the receiver \( k \) with the transmitter \( i \) and the interferer \( j \) to be

\[
i_{i,k}(c) \cdot l_j(c) \cdot 2 \cdot (1 - csf_{j,k}).
\]

The probability that a collision occurs twice or more for the same packet is relatively small, so we approximate the number of retrieessions of the MAC frame as 1. Finally, a hypothetized traffic load \( \bar{l}_i(c) \) on the given channel \( c \), which consists of the original offered load and extra traffic due to retrieessions, can be expressed as:

\[
i_{i}(c) = \sum_k l_{i,k}(c) \cdot (1 + \sum_j l_j(c) \cdot 2 \cdot (1 - csf_{j,k})),
\]

where \( k \in \{ \text{neighbors of node } i \} \), and \( j \) satisfies \( csf_{j,k} = 1 \) and \( csf_{j,k} < 1 \). Total hypothetized traffic load at node \( i \) can be expressed as \( \bar{l}_i = \sum_c l_c(c) \).

(c) Traffic Deferred and Switching Delay: The remaining factors of the call capacity utilization at a node are traffic deferred and switching delay (B and S periods in Fig. 4). In a multi-channel environment, each node transmits data over the neighbor’s fixed channel with channel switching. The fraction of time for staying on each channel at the node (\( t_d(c) \)) is proportional to the hypothetized traffic load on each channel; thus, we can model the time for traffic deferred at node \( i \) on the given channel \( c \) as

\[
o_i(c) = \sum_{c \neq j} t_d(c) \cdot 2 \cdot (1 - csf_c) \cdot t_i(c),
\]

where \( i \) and \( j \) satisfy \( csf_c < 1 \).

Next, we consider the channel switching delay ("S" periods in Fig. 4) of HMCP. Accoring to HMCP behaviors in Fig. 4, if a node has traffic to neighboring nodes with different fixed channels, then the node should tune its switchable radio’s channel to its neighbor’s fixed channel to transmit data. At this time, channel switching occurs, and so does delay. This channel switching delay with a high data rate network system is not negligible. For instance, the channel switching delay is 35 \( \mu \)s, and \( T_{voice.pkt} \) is 37.96 \( \mu \)s in IEEE 802.11a with a data rate of 54Mbps, a G.729 codec, and a sample period of 20 ms.

Each HMCP node manages packets classified according to transmitting channels by using the channel based queue (CBQ). Packets in the CBQ are served as round robin fashion. If there is no packet in certain CBQ entry, then the HMCP node skips that CBQ entry. Thus, the HMCP node only stays channels having queued packets. To guarantee fairness of packet transmissions, HMCP defines a variable \( T_{max} \). \( T_{max} \) is the maximum time to stay on the current operating channel for transmitting data packets. After \( T_{max} \), a node immediately switches its switchable radio’s channel to transmit data packets in other queue entry. To model the time for staying on the specific channel to transmit data, let \( T_{k}(c) \) be the time for staying on the channel \( c \) at node \( i \) in k-th round. We can easily predict that \( T_{k}(c) \) equals to the time for transmitting and deferring packets arrived during \( (k-1) \)-th round. For example, node A in Fig. 4 stays on channel 2 in k-th round when \( T_5(2) = T_{A}(2) + \Delta S_5(2) \). Also, \( T_{A}(2) \) should be smaller than or equals to \( T_{max} \). Based on this example, thus, we can express \( T_{k}(c) \) as following:

\[
T_{k}(c) = \min\{\sum_s T_{k}(c) + \Delta S_s(c) + o(c), T_{max}\},
\]

where \( \Delta S_s(c) \) is the channel switching delay and \( \Delta \) is the channel switching delay in \( T_{k}(c) \).

Let \( T_{k}(c) \) be the asymptotic value of \( T_{k}(c) \) when \( k \) increases. Then, the time period of one round can be expressed as \( \sum_{c \in F_i} T_{k}(c) \), where \( F_i \) is a set of data channels used by node \( i \) to transmit data to its neighbor nodes. The number of channel switchings during one round is \( |F_i| \). Finally, we can get the time for switching channels at node \( i \) represented by

\[
s_i = \left\{ \begin{array}{ll}
\sum_{c \in F_i} T_{k}(c), & |F_i| \geq 1 \\
0, & |F_i| = 1
\end{array} \right.
\]

where \( F_i \) is a set of data channels used by node \( i \) to transmit data to its neighbor nodes. If \( |F_i| = 1 \), then there is no channel switching.

By summing up all factors of the call capacity utilization at node \( i \), we can represent \( c_i \) as:

\[
c_i = \sum_c n_i(c) + \sum_c o_i(c) + s_i.
\]

Based on (1), we can obtain the call capacity utilization at each node in an MCMR WMN. To check feasibility for supporting VoIP calls with the given route, the capacity utilization of each node, \( c_i \), should be smaller than 1.

IV. QoS ROUTING ALGORITHMS

In this section, we mention about QoS routing algorithms for supporting VoIP service based on our call capacity model.
Even if we have the accurate call capacity and easily check the feasibility of given call and network configurations such as channel allocation and nodes’ position, allocating QoS routes for calls to meet the feasibility as well as the network-wide call capacity.

First of all, we focus on finding feasible routes. In wireless network environment, one link can interfere to any links round it, because all links share the same medium. Thus, when we allocate the route for a new voice call, we must check the feasibility of whole network (\(c_i < 1\), for all node \(i\) in the network). To find route paths satisfying this feasibility condition, we develop a heuristic algorithm with Graph theory. We define the symbols related with our routing computation algorithm. The HMCP network can be represented by graph \(G = (V, E)\), where \(V\) is a set of nodes in the network, and \(E\) is a set of wireless links in the network. Also, active VoIP calls in the network can be represented by \(P = \{p_1, p_2, ..., p_k\}\), where \(k\) is the number of current active calls. \(P\) is a set of paths for active calls, and each \(p_i\) is a set of nodes along the path for the call \(i\).

Specifically, our route computation algorithm is to find feasible routes with given source, destination, and active calls’ configuration (i.e., \(P\)). To the end, we first find all possible paths from source to all nodes in the network. This process is performed only one time for each source node, if the network configuration does not change. Secondly, among possible paths from source, the feasible paths can be determined by checking the feasibility with given source, destination and \(P\).

Algorithm 1 describes about whole process for finding all possible paths from the given source node to all destination nodes. This algorithm is inspired by breath first search (BFS) algorithm, which is a one of well-known algorithms to traverse all vertices once in the graph. Since the original BFS algorithm only visits vertices at one time, we can not find all possible paths from the given source node in the network by the original BFS algorithm. To allow revisit each vertex at most its node degree, we can find all possible paths from the given source node (see line number 4, 10, and 24 in Algorithm 1). Algorithm 1 requires the network graph \(G\) with source node \(src\) and returns possiblePathFromSrc, which is a set of possible path from source node \(src\). After finding all possible paths from source node \(src\), the feasibility of each possible path from \(src\) is checked. Algorithm 2 implements the procedure of finding all feasible paths from source node \(src\) to destination node \(dst\). This algorithm requires \(G\), \(src\), \(dst\), \(P\), and possiblePathFromSrc, and returns
Algorithm 1 FindPossibleRoute(G, src)
1: for all $u \in V[G]$ - [src] do
2: VISIT[$u$] $\leftarrow$ 0 // Initialize visit counter
3: end for
4: VISIT[src] $\leftarrow$ VISIT[src] + 1
5: Q $\leftarrow$ $\phi$ // Initialize Queue structure: EMPTY marker
6: ENQUEUE(Q, src)
7: while $Q \neq \phi$ do
8: $u$ $\leftarrow$ DEQUEUE(Q)
9: for all $v \in \text{Adj}[u]$ do
10: if VISIT[v] $\geq$ Count(Adj[v]) then
11: continue
12: end if
13: if $u = \text{src}$ then
14: AddPossiblePath([u $\rightarrow$ v])
15: // Insert [u $\rightarrow$ v] to possiblePathFromSrc
16: end if
17: else
18: for all $r \in$ possiblePathFromSrc do
19: if $r$’s last vertex = u AND CheckNoCycle([r $\rightarrow$ v]) then
20: AddPossiblePath([r $\rightarrow$ v])
21: end if
22: end if
23: ENQUEUE(Q, v)
24: end for
25: end while
26: RETURN possiblePathFromSrc

Algorithm 2 FindFeasibleRoute(G, src, dst, P)
1: for all $r \in$ possiblePathFromSrc do
2: if $r$’s first vertex = src AND $r$’s last vertex = dst then
3: if CheckFeasibility(G, r, P) then
4: AddFeasibleRoute(r)
5: // Insert $r$ to feasibleRouteFromSrcToDst
6: end if
7: end if
8: end for
9: RETURN feasibleRouteFromSrcToDst

feasibleRouteFromSrcToDst, which is a set of feasible paths from source node src to destination node dst. Based on feasibleRouteFromSrcToDst, the feasible route for the given nodes src and dst can be easily determined. If there are multiple candidates for determining the feasible route from source node src to destination node dst, then we choose one feasible path among possible feasible paths according to routing policies. One simple routing policy is to choose the shortest feasible path among possible feasible paths. We call this scheme as shortest feasible path (SFP) routing algorithm. With consideration of the traffic load on each node, we can choose the feasible path with the minimum summation of utilizations of nodes involving the path (called as MUF).
delay performance of G.729 codec with 20 sample period on three hop counts. Based on this figure, the actual call capacity of this configuration is 36 which means the target delay for a good call, and the call capacity obtained by our call capacity model is 38. However, at the point of this analytical call capacity, delay is over 300 ns. This is because our analytical call capacity model is based on throughput constraints, but actual call capacity in this configuration is constrained by delay performance. We can deal with this overestimation at QoS routing computation phase. We also analyzes the impact of radio switching delay to the capacity, and the result is in Fig. 8. The call capacity highly depends on channel switching delay. The call capacity decreases by about 20% as increasing the channel switching delay by 1 ms.

Secondly, SFP and MUFP are implemented based on NS-2 to evaluate performance of QoS routing schemes. We use only G.711 codec with 20 ms packetization interval to evaluate routing algorithms, because we focus on routing algorithms. We deploy nodes as 6x6 grid network in a 500x500 square meter area. Each node can generate a call intended to any nodes. We add calls with a given hop count to the grid topology network until the call is rejected or dropped (i.e., the maximum number of calls). If there is no feasible path for a newly generated call, then the call is rejected, and if R-score of an active call is below 70, then the active call is dropped. Fig. 9 shows that our simple routing algorithm with call capacity model (SFP) to meet QoS constraint supports more number of current voice calls. As shown in this figure, though the number of calls supported is decreasing as increasing the path length, there are more opportunities for finding better paths, and hence the network with routing algorithm considering feasibility can support 20-35% extra calls comparing to shortest path routing (SP). However, SFP based on our capacity model overestimates the call capacity about 5% of actual capacity obtained by simulations. The overestimation of SFP is inherent from our call capacity model’s overestimation as we mentioned before.

Next, we evaluate two proposed routing algorithms (i.e., SFP and MUFP) in terms of call acceptance and dropping rates. Throughout this performance evaluation, calls are generated as Poisson process with the fixed mean rate and the average duration of a call is exponentially distributed. We set a mean arrival rate of Poisson process to 1/5 calls/sec. and adjust the average call duration to increase the traffic load in the network. Fig. 10 (a) and 10 (b) show the percent of calls rejected or dropped for each routing scheme in grid topology with uniform traffic pattern. The result of routing schemes without consideration of overestimation by setting the feasibility condition as \( c_i < 1 \) for all node \( i \) in the network is in Fig. 10 (a). Fig. 10 (a) demonstrates that MUFP has better performance comparing to SFP about 10\% under heavy traffic load. Also, Fig. 10 (a) shows that SFP and MUFP yield dropped calls due to overestimation of our call capacity model. Dropped calls are worse than rejected calls in terms of user experience, and therefore we should handle this overestimation problem of our call capacity. To deal with overestimation of our call capacity model, we use the feasibility condition expressed as \( c_i < 0.95 \) for all node \( i \) in the network. Fig. 10 (b) demonstrates that SFP and MUFP effectively protect active calls from miss-admittedly calls by setting feasibility condition, but this feasibility condition makes a slight increase on call rejection rate.

VI. CONCLUSION

In this paper, we addressed the challenging issues for providing VoIP service on MCMR WMNs. To maintain QoS of VoIP calls, we must do call admission control. The accuracy of the call admission control highly depends on how well the call capacity over the multi-hop MCMR WMN is inferred. We first develop the voice call capacity model of HMC, which is a representative multi-channel multi-radio protocol on MCMR WMNs. We also validate the proposed capacity model via both experiments and simulations. The performance results show that our call capacity model estimate a tight upper bound of the actual call capacity obtained from experiments on our testbed.

Next, we design QoS routing algorithms to select feasible paths based on the proposed call capacity model to increase number of supported VoIP calls and minimize future call rejections. We proposed two simple and effective route computation schemes: shortest feasible path (SFP) considering feasibility and minimum utilization feasible path (MUFP) considering feasibility and call capacity utilization at nodes. Using more information to compute the routing path for voice calls, we can support voice calls with better voice quality in terms of call reject/drop rate.

Our modeling work is general enough to be extended for newer architectures such as IEEE 802.11n system or IEEE 802.11ad system. We also will address routing algorithms based on our call capacity model with considerations of heterogeneous traffics (i.e., data, video, and other elastic traffic) as our future work.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the NRF funded by the MEST(2012-0002872)

REFERENCES