Asian Journal of Information Technology 15 (10): 1596-1606, 2016
ISSN: 1682-3915

HMM and Fuzzy Based Scheduling and Resource Allocation for Downlink Traffic in WiMAX OFDMA Networks

K. Valarmathi
Panimalar Engineering College, Chennai, India

Abstract: In WiMAX OFDMA networks, the existing resource allocation and scheduling techniques results in more overhead and delay. In order to overcome this problem, in this study, we propose HMM and fuzzy based scheduling and resource allocation for Downlink Traffic in WiMAX OFDMA networks. In this technique, new users are admitted after checking the predicted free slots using Hidden Markov Model (HMM) against the estimated slots of each user. Then the exact number of slots for each user is estimated using fuzzy logic inference considering QoS requirement such as bandwidth, delay and transmission rate of users over each subchannel. The scheduling of slots on each sub channel level is performed based on user priority which is estimated again using fuzzy logic inference considering the channel quality, queue length and urgency factor. By simulation results, we show that the proposed technique enhances the network throughout and minimizes the overhead.

Key words: WiMAX, OFDMA, networks, HMM, QoS

INTRODUCTION

IEEE 802.16 WiMAX is signified as a standard for accessing wireless broadband networks. It is thought up by Institute of Electrical and Electronics Engineers (IEEE) to satisfy various end users requirements. It exploits Orthogonal Frequency Division Multiple Access (OFDMA) with the spectrum ranges between 1.25 and 28 MHz. It benefits both Time Division Duplexing (TDD) and Frequency Division Duplexing (FDD).

Beam forming and Multiple Input Multiple Output (MIMO) are some of the sophisticated techniques used by WiMAX. Apart from this, it makes use of advanced coding techniques such as space-time coding and turbo coding.

WiMAX is extensively used in data, telecommunications (VoIP) and IPTV services (triple play). The bandwidth range of WiMAX makes available broadband connectivity through variety of devices to overall the world. WiMAX adds significant enhancements:

- It improves NLOS coverage by utilizing advanced antenna diversity schemes and Hybrid Automatic Repeat Request (HARQ)
- It adopts dense sub-channelization, thus increasing system gain and improving indoor penetration
- It uses Adaptive Antenna System (AAS) and Multiple Input Multiple Output (MIMO) technologies to improve coverage
- It introduces a downlink sub-channelization scheme, enabling better coverage and capacity trade-off (Arhaif, 2011; So *et al*., 2010; Rakesh and Dalal, 2010)

Need for resource allocation: In the present time we have been witnessing a dramatic growth in network-based video applications which includes on-demand video streaming and also distance learning. These network based applications require transmitting video over communication channels such as the internet or wireless channels that can reveal a wide changeability in throughput, delay and packet loss.

An IEEE 802.16 wireless network is one of the standards for Broadband Wireless Access (BWA) in Metropolitan Area Networks (MANs). This 802.16 wireless networks can support various IP based high speed broadband wireless services such as streaming video, FTP, email, chatting and so on. WiMAX is one of the recent broadband wireless communication technologies which supports resources allocation services. To increase the coverage and throughput, relay stations were introduced. To deal with the eventually increasing of the more traffic to compete with the limited QoS ad been a huge challenge.

To handle these increasing traffics, WiMAX supports adaptive modulation and coding such that the distance between the Subscriber Station (SS) and the Base Station (BS) determine the type of QoS to be used for the video transmission. Allocating the resources for video transmission dynamically in wireless network is highly complex and non-linear. It becomes even more complex when the wireless network is designed for heterogeneous traffics with different Quality of Service (QoS) requirement like WiMAX network (Shuaibu *et al*., 2011; So-In *et al*., 2009; Hua *et al*., 2010; Katsaggelos *et al*., 2005).

Downlink resource allocation: In this downlink resource allocation 802.16 WiMAX networks employ multiple closely spaced sub-carriers which are grouped in sub-channels. These sub-carriers will form a sub-channel which need not be adjacent. In the downlink allocation, a sub-channel may be intended for different users depending on their channel conditions and data requirements. Then the destination node can allocate resources to user devices with lower Signal-to-Interference-and-Noise Ratio (SINR) per sub-channel with less resources to user devices with higher SINR (Wang and Dittmanno, 2007).

Advantages:

- By resource allocation it is possible to supply exactly required resources for the transmission of video frames by which the network's energy can be conserved
- Through resource allocation management we can allocate the resources for the video frames in an efficient manner and get the desired throughput

Issues: During the video transmission resources allocation is the important factor in the wireless networks, since the dynamic resource allocation, the allocation of the resources is dynamically changed as the channels change. During the transmission of these video frames, if any frames get lost or dropped, to retransmit these frames the network consume lot of power. The main aim of the resource allocator is the allocating the resources required for the video transmission i.e., number of slots, for each user in each WiMAX frame. It is difficult to achieve the maximum throughput in the WiMax networks since each transmission demands different quality of services and this is possible to achieve only through the resource allocator in the networks. We require this resource allocator in order to use the resources in an efficient manner (Shuaibu *et al*., 2011; So *et al*., 2010a, b).

Problem identification: A Fuzzy Logic User Identifier (FLUI) was designed to intelligently differentiate users from urgent and regular which considers the QoS fulfillment, priority value and QoS measure tendency of users (Chung *et al*., 2012). In a fuzzy based adaptive scheduling schema was proposed where the scheduler is dynamically selected based on the current traffic context such as the number of flows of each QoS class (Seo *et al*., 2011). In a fuzzy logic partition-based call Admission Control (FZ CAC) was proposed (Shuaibu *et al*., 2011). The scheme primarily partitions the total link bandwidth into three which corresponds to Constant Bit Rate (CBR), Variable Bit Rate (VBR) and Hand Over (HO) services. The fuzzy logic admission control scheme was implemented in the HO portion to intelligently keep dropping probability as low as possible based on the available bandwidth.

From the above literature review, we can say that none of the works used fuzzy logic for resource scheduling and resource allocation.

Hence in this proposal, we propose a HMM and Fuzzy Based Scheduling and Resource Allocation technique for Downlink Traffic in WiMAX OFDMA networks. It consists of 3 phases: Admission Control, Resource Allocation and Scheduling.

Literature review: Meddour *et al*. (2012) have proposed we propose a cross layer optimizer named XLO between scalable video streaming application and IEEE 802.16 MAC layer. The main objective is to allow video streaming applications to adapt its parameters according to 802.16 MAC layer conditions and resource availability. XLO uses the existing service flow management messages exchanged between a Base Station (BS) and a Subscriber Station (SS) and make them available to the video streaming application via a specific XLO interface. The authors have also introduced an enhanced admission control function at the BS that takes into account video adaptability property. The advantage of this approach is that it has effectiveness for better quality of services with the resources available.

She *et al*. (2011) have proposed a MAC-layer Active Dropping (AD) scheme for achieving effective resource utilization and maintaining application-level quality of services in real-time video streaming over the emerging wireless broadband access networks based on Time-Division Multiple Accesses (TDMA). By proactively dropping the MAC-layer Protocol Data Units (MPDUs) of a video frame that are unlikely to meet its application-layer delay bound, the proposed scheme releases the precious transmission resources to the subsequent frames or other competing service streams. The advantage of this approach is that it leads to more efficient resource utilization than that by the conventional prioritization-based cross-layer approaches which simply manipulate transmission and/or retransmission priority of each MPDU.

Liang *et al*. (2010) have defined an Energy-conserved Uplink Resource Allocation (EURA) problem in 802.16j networks under the transparent mode which asks how to arrange the uplink resource to satisfy mobile stations (MSs) requests and minimize their energy consumption. Objective 1 is necessary while objective 2 should be achieved when objective 1 is met. The above bi-objective problem is especially important when the network is non-saturated. The EURA problem is NP-hard and they proposed a heuristic with two key designs. First, they have exploit relay stations to allow more concurrent uplink transmissions to fully use the frame space. Second, they have reduced MS transmission powers by adjusting their rates and paths. The advantage of this approach is that it can save up to 80% of MS energy as compared with existing work.

Nascimento and Rodriguez (2009) have proposed a Dynamic Resource Allocation (DRA) strategy that can provide operators the flexibility to deliver broadband traffic with high spectral efficiency. The DRA unit constitutes a scheduler, Link Adaptation (LA), Resources Allocation (RA) and Hybrid Automated Repeat Request (HARQ) components inter-working seamlessly. The potential for the DRA to deliver QoS is achieved through service classification lists, where higher priority is given towards retransmitted packets and subsequently to first-time transmitters with packet delay (Nazhad,). The advantage of this approach is that DRA scheme has the capacity to provide enhanced coverage for NRTV (Near Real Time Video) services in Wide Area Networks (WANs).

Hou *et al*. (2009) have proposed an efficient yet simple design framework for achieving flexible resource allocation and packet scheduling for non-real-time Polling Service (nrtPS) traffic in IEEE 802.16 networks. By jointly considering the selective automatic repeat request mechanism at the media access control layer as well as the adaptive modulation and coding technique at the physical layer, this proposed framework enables a graceful tradeoff between resource utilization and packet delivery delay while maintaining the minimum throughput requirements of nrtPS applications. The advantage of this approach is that inter-service time, delivery delay, good put and resource utilization are investigated for performance evaluation.

MATERIALS AND METHODS

Overview: In this study, we propose HMM and Fuzzy Based Scheduling and Resource Allocation for Downlink Traffic in WiMAX OFDMA networks. In this technique, new users are admitted after checking the predicted free slots using Hidden Markov Model (HMM) against the estimated slots of each user. Then the exact number of slots for each user is estimated using fuzzy logic inference considering QoS requirement such as bandwidth, delay and transmission rate of users over each subchannel. The scheduling of slots on each sub channel level is performed based on user priority which is estimated again using fuzzy logic inference considering the channel quality, queue length and urgency factor.

Estimation of metrics

Bandwidth: The available bandwidth for downlink frame is computed using Eq. 1:

$$AB = (1-S_d/S_t)b\ S_t/t_f \tag{1}$$

Where:

S_d = Number of allocation downlink slots in one frame
S_t = Total slots in downlink sub-frame
t_f = Duration of the frame
b = Number of bits transmitted in downlink slots

Delay: The packet delay is estimated based on the following Eq. 2:

$$D = D_s + D_q + D_m + D_t \tag{2}$$

Where:

D_s = Scheduling delay
D_q = Queuing delay
D_m = Mapping delay
D_t = Transmission delay (Seraphin and Ramesh, 2013)

Transmission Rate (TR): The expected transmission time reveals the multiple transmission rate capacity onto the routing metric (Verma *et al*., 2008):

$$ETT\ (\text{or TR}) = ETX \times (M/R_{tx}) \tag{3}$$

Where:

ETX = Expected transmission count
M = Size of the data packet
R_{tx} = Raw data transmission rate of the link

Here, ETX is estimated based on the forward and reverse packet delivery ratios.

Channel quality: The signal to noise interference ratio is computed using the following Eq. 4, (Kakali *et al*., 2011):

$$\sigma_x = \frac{\rho_s}{\rho_o - \mu_0}.\frac{\nu_{tot}}{\nu_{ds} + \omega.\nu_{ps}}.\lambda_x \tag{4}$$

Where:

ρ_s = Noise power density of the subcarrier
ρ_0 = Noise power density of the other cells
μ_0 = Thermal noise power of the receiver
V_{tot} = Total number of subcarriers

V_{ds} = Total data subcarriers
V_{ps} = Total pilot subcarriers
ω = Pilot to data subcarrier power ratio
λ_x = Subcarrier gain
λ_x = Derived using the following Eq. 5:

$$\lambda_x = \left| \sum_{a=1}^{N} M_y \sigma_y e^{j\theta_y} e^{-2y\lambda_x T_d} \right|^2 \quad (5)$$

Where:
a = Index of multiple path
σ = Average power amplitude of yth path
λ_x = Frequency of the subcarrier x
T_d = Time delay of path a

Queue length: Let q_{ij} (t) be the queue length of aggregated traffic flow of service category j (Karunkuzhali and Tomar, 2012). The queue length is evaluated using in Eq. 6:

$$q\,(t) = B_{ij}\,(t)\,ITF\,(t)\text{-}(1\text{-}PER_{ij})\,BW_{ij}\,(t)\;\eta_{ij}\,\forall i \in \{1,2,...,z\}, j \in \{1,2\} \quad (6)$$

Where:
$B_{ij}(t)$ = Number of base stations
ITF (t) = Input traffic flow to the subscribed base station
PER_{ij} = Average packet error rate for extracting the channel quality
η_{ij} = Average spectral efficiency (in bits/s/Hz)
$Bw_{ij}(t)$ = Bandwidth allocated for queue depletion
$Bw_{ij}(t)\,\eta_{ij}$ = Queue depletion rate

Urgency factor: The Urgency Factor (UF_i) of a service flow is estimated using the following Eq. 7, (Lin *et al*., 2009):

$$UF_i = (S_i \times (Pr + 1))/f_i \quad (7)$$

Where:
S_i = Number of slots
Pr = Flow priority
f_i = Deadline of bandwidth request

Thus, Urgency factor considers latency, priority and fairness.

Timeslot: The timeslot for transmitting the data packet of frame i available in the queue at time t in the scheduling cycle j is given using Eq. 8 (She *et al*., 2010):

$$TS\,(i,t,j) = D\text{-}T_{w\,(i,t,j)} \quad (8)$$

Where:
t = Universal time detected in the BS (Counting starts from zero)
j = Current scheduling cycle index (starts from one)
D = Delay (estimated in section 3.2.2)

The waiting time of the data packet in the queue ($T_{w(i,\,t,\,j)}$) = Defined using the following Eq. 9:

$$T_{w(i,\,t,\,j)} = (j\text{-}i)\times\omega + (n+1\text{-}(j\text{-}i))\times T_{ist} + T_{tx}(t\,(i,\,t,\,j) + T_c\,(i,\,t,\,j) \quad (9)$$

Where:
$(j\text{-}i)\times\omega+(n+1\text{-}(j\text{-}i))$ = Number of scheduling cycles experienced by frame i in the queue at time t
ω = Scheduling cycle
T_{ist} = Inter-service waiting time for the frame
$T_{tx}(t\,(i,\,t,\,j)$ = Time utilized for transmission of data packet of previously cached frames in scheduling cycle j
$T_c\,(i,\,t,\,j)$ = Total timeslots consumed for transmission (either success or failure) of data packets of frame i in the scheduling cycle j

Hidden Markov Model (HMM): A hidden markov model includes persistent finite-state Markov Chain, an variables indicating the output and a distribution for every transition over that variable in the Markov Chain. It is defined as a stochastic process of moving among states and process of emitting an output sequence. But, the series of state transitions are hidden and observed only through the sequence of emitted symbols.

The main elements of HMM are as follows; fixed state sequence of length T:

$$N = n_1, n_2, ..., n_T \quad (10)$$

Where:
N = The fixed state sequence of length T
W = The set of hidden states
O = The set of observation symbols per state
X = The state transition probability distribution
Y = The observation symbol probability distribution in state j
π = Denotes the initial state distribution

A number of hidden states:

$$W = \{W_0, W_1, ..., W_N\} \quad (11)$$

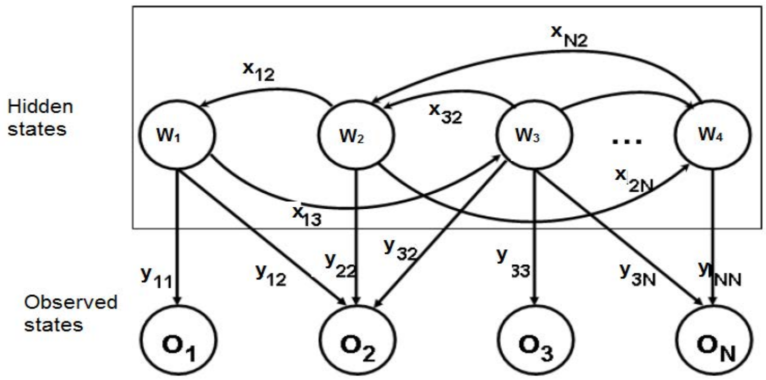

Fig. 1: Hidden Markov Model

Where N = hidden states in the system. A set of observation probability distributions reflecting random variables or stochastic processes:

$$O = \{O_1,...,O_2,O_N\} \quad (12)$$

The initial state distribution is illustrated below:

$$\pi = \{\pi_i\} \quad (13)$$

π_i = probability of initiating in state i. A transition probability distribution (X) based on the earlier state and indicates a new state after each time t:

$$X = \{x_{i,j}\} \quad (14)$$

$x_{i,j}$ = probability of moving from state i to j:

$$x_{i,j} = P\left[v_k = W_j \mid v_{k-1} = W_i\right] \quad (15)$$

A observation state probabilities is given by Y:

$$Y = \{y_{ik})\} \quad (16)$$

$$y_{ik} = P\left[o_k \text{ at } t|v_k = W_j\right] \quad (17)$$

The entire model is precisely shown using Eq. 18:

$$\tau = (X, Y, O) \quad (18)$$

Proposed solution: Our proposed technique includes three phases:

- Phase 1: admission control
- Phase 2: resource allocation
- Phase 3: scheduling

Admission control: When the new users arrive at the Base Station (BS), the free slots per frame are predicted using the Hidden Markov Model (HMM) against the estimated slots of each user. This estimation is based on current traffic load of the BS (Fig. 1).

When the user arrives at BS, initially it estimates the number of free slots and passes the values it to the HMM model. Where TS is taken as observation state O Observation probability (O) for a given sequence N is given as:

$$U(O \mid N,\tau) = \prod_{t=1}^{T} P\left(TS_t|n_t,\tau\right) = y_{n1}\left(TS_1\right)\times y_{n2}\left(TS_2\right),...,y_{nT}\left(TS_T\right) \quad (19)$$

The probability of the state sequence is given by:

$$P(N|\tau) = \pi_{n1}\, x_{n1}\, x_{n2}\, x_{n2n3} \ldots x_{nT-1nT} \quad (20)$$

Based on the above two Eq. 19 and 20, the probability of observations is estimated:

$$P(O|\tau) = \sum_{N} P(O \mid N_1\tau)P(N \mid \tau) = \sum_{n_1 \ldots n_T} \pi_{n1}\, y_{n1}(TS_1)x_{n1n2}\, y_{n2}(TS_2) \ldots x_n T - 1nT^y\, nT(TST) \quad (21)$$

In order to determine the single optimum state sequence for an observation sequence n_1, a Viterbi algorithm is used. The probability of most probable timeslot is estimated using the Eq. 22:

$$\beta(i) = \max_{n_1,n_2,n_3,\ldots,n_{T-1}} P(n_1,n_2...n_T = W_i,TS_1,TS_2...TS_t|\tau) \quad (22)$$

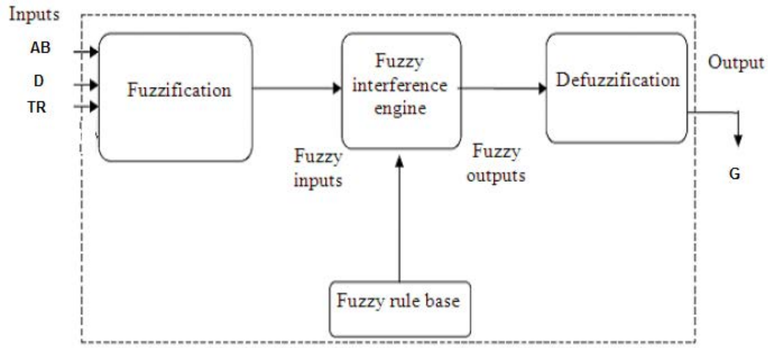

Fig. 2: Fuzzy logic inference for resource allocation; AB = Required Bandwidth; D = Delay; TR = Transmission Rate; G = optimum number of slots

The extreme probable state is estimated using the following Eq. 23:

$$\hat{n}_T \arg \max_{1\le i\le N} [\beta_T(i)] \tag{23}$$

The sequence of states is initiated again as the pointer in every state. State sequence backtracking is given using the following Eq. 24:

$$\hat{n}_T = [\beta_{T+1}(\hat{n}_{t+1}), t] = T-1, T-2, \ldots, 1 \tag{24}$$

Where:

- β = Additional matrix of size N×T
- T = Sequence length time
- β = Introduced in Viterbi algorithm to estimate the user slots. This backtracking gives the required set of states

Resource allocation: Following the estimation of free slots per frame using HMM (described in previous study), it is required to estimate the optimal number of slots (G) to be allocated for each user. This is performed using the following two steps:

- BS examines the QoS requirement of the users that includes user priority, minimum required bandwidth, minimum delay, transmission rate over each sub-channel. (Described in section 3.2.1-32.3)
- BS invokes Fuzzy Logic Inference (FLI) engine that generates the optimal number of slots (G) to be allocated for each user. This is explained here

Fuzzy logic inference based resource allocation: Fuzzy logic is referred to as a logical system that indicates the classical logic for reasoning under ambiguity. It helps while dealing with inexact and uncertain information as the network behaves dynamically.

A fuzzy engine is typified by the inference system that includes the system rule base, input membership functions that fuzzify the input variables and the output variable de-fuzzification process (Fig. 2).

The steps involved in the fuzzy logic technique are detailed below:

- Fuzzification
- Inference with rule base
- Defuzzification

Fuzzification: This involves the conversion of crisp inputs into linguistic values. Each of these inputs values is denoted using fuzzy set. This fuzzy set is related to a membership function which is used to describe the method by which the crisp input belongs to the set.

Our proposed resource allocation process takes three input parameters such as bandwidth, minimum delay, transmission rate into consideration. Using the inputs fuzzy parameters and inference engine, optimal number of slots (G) is obtained as the output.

As each of the input fuzzy parameters possesses the minimum and maximum boundary condition, it is represented using triangular membership function. The membership to each of the fuzzy variables is assigned using intuition technique as it reduces computation complexity.

The membership function for these input parameters and output is represented as f (AB), f (D), f (TR) and f (G) (Fig. 3-5).

Inference mechanism: This method is based on the fuzzy rules that link the input and output parameters (fuzzy rule base) and the membership functions for inputs and output parameters.

In order to generate the inference engine, initially the membership functions for input and output parameters are developed.

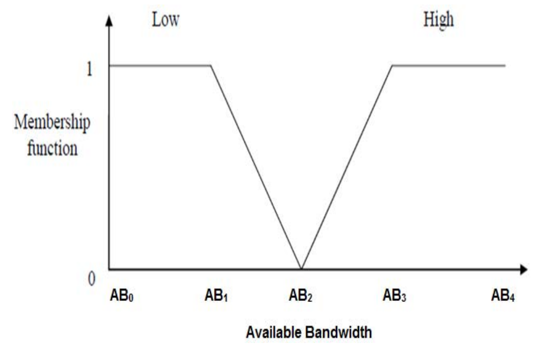

Fig. 3: Membership function for available bandwidth

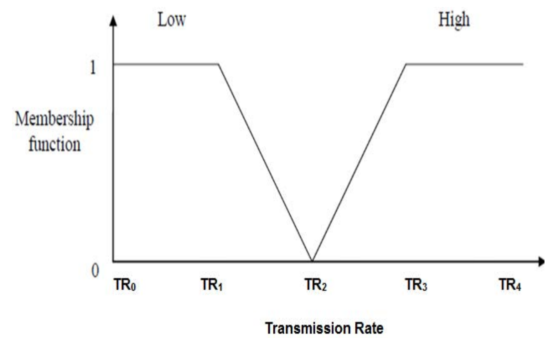

Fig. 4: Membership function for transmission rate

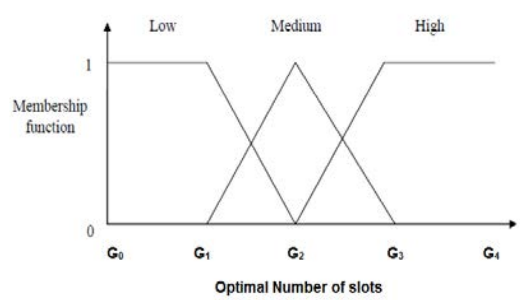

Fig. 5: Membership function for optimal number of slots

The fuzzy inference system is designed based on 9 rules described in Table 1. To demonstrate the fuzzy inference system, the following rule is considered to show the function of inference engine and outputs of each rule are each rule are combined for generating fuzzy decision.

Rule 1: If (AB = low) and (D = high) and (RSS= high) Then, G = high, end if.

Defuzzification: It is the method by which a crisp value is extracted from a fuzzy set as illustration value. During fuzzy decision making, the centroid of area technique is taken into account for defuzzification. The Defuzzifier is based on Eq. 25:

Table 1: Fuzzy rules

Bandwidth (AB)	Delay (D)	Transmission Rate (TR)	Optimal number of slots (G)
Low	Low	Low	Low
Low	Low	High	Low
Low	High	Low	Medium
Low	High	High	High
High	Low	Low	Low
High	Low	High	Medium
High	High	Low	Low
High	High	High	Medium

$$F\_priority = \sum_{All\_rules} u_i \times \delta(u_i)]/[\sum_{all\_rules} \delta(u_i) \quad (25)$$

Where:

F_priority = Degree of decision making
u_i = Fuzzy variable
$\delta(u_i)$ = Membership function

The output of the fuzzy priority function is altered to the crisp value based on the above defuzzification method.

Scheduling: The incoming downlink traffic is first classified and grouped according to its service class. Each user may have different types of service classes. The IEEE 802.16 defines the following five different Quality of Service (QoS) classes:

- Unsolicited Grant Service (UGS): It is designed in order to maintain the constant bit rate real time traffic (CBR)
- Extended Real Time Polling Service (ertPS): It is designed for supporting Voice Over IP (VoIP) traffic.
- Real Time Polling Service (rtPS): It is designed for supporting Variable Bit Rate (VBR) real-time
- Non Real Time Polling Service (nrtPS): it supports non-real-time VBR traffic and it only guarantees minimum throughput for an application
- Best Effort Service (BE): the BE class allocates resources to subscribers if and only if there are left-over resources after allocating the resources to other QoS classes of higher priority. This QoS class guarantees neither delay nor throughput

The priority among QoS classes is defined as follows: UGS >ertPS >rtPS >nrtPS >BE. The scheduling of the traffic is performed after obtaining the above mentioned traffic class priority. The process involves fuzzy logic inference which is briefed below. For each user for each sub channel the BS measures the following

- Channel quality (or) channel condition using the Signal-to-Interference-and-Noise-Ratio (SINR)
- Queue length of each connection (Q)
- Urgency factor of each service class (UF)

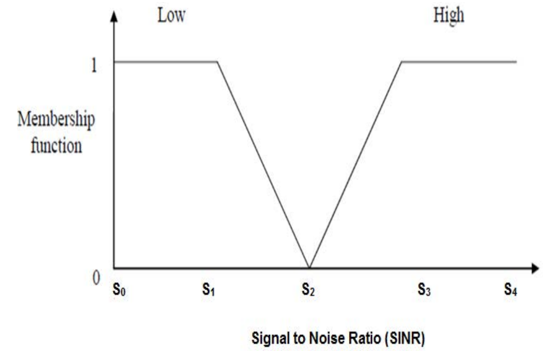

Fig. 6: Membership function for channel quality (SINR)

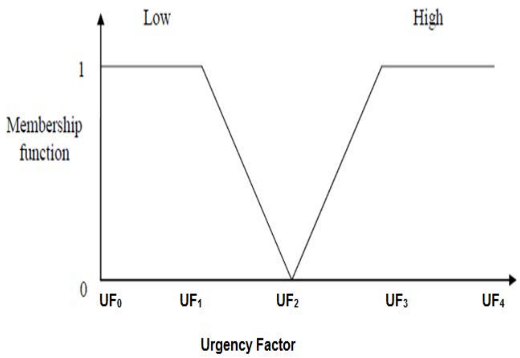

Fig. 7: Membership function for queue length

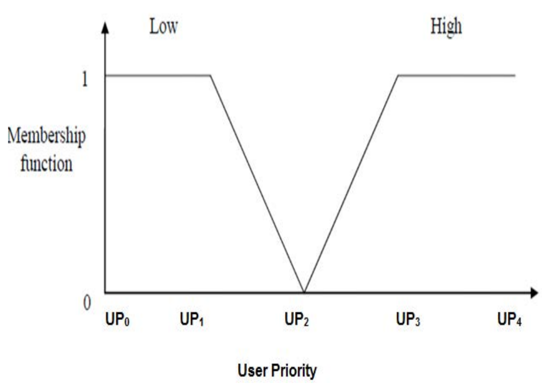

Fig. 8: Membership function for urgency factor

The above parameters are estimated. Utilizing the above 3 parameters, the Fuzzy Logic Inference (FLI) engine forms fuzzy member ship functions and grades with low, medium and high values.

Fuzzification: Our scheduling scheme involves three inputs parameters such as SINR(S), Q and UF. Using the inputs fuzzy parameters and inference engine, optimal number of slots (G) is obtained as the output.

The membership function for these input parameters and output is represented as f (SINR), f (Q), f (UF) and f (UP) shown in Fig. 6-8.

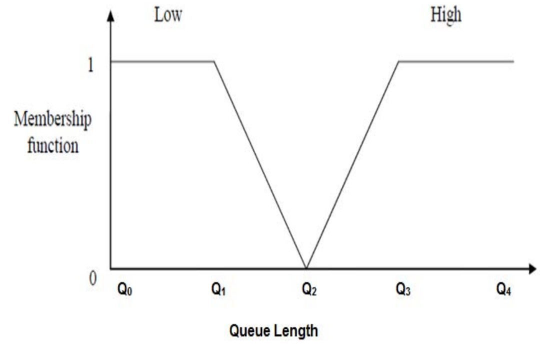

Fig. 9: Membership function for user priority

Table 2: Fuzzy rules

SINR	Queue length	Urgency factor (UR)	User Priority (UP)
Low	Low	Low	Low
Low	Low	High	Medium
Low	High	Low	Low
Low	High	High	Medium
High	Low	Low	Low
High	Low	High	Medium
High	High	Low	Medium

Inference mechanism: The fuzzy inference system is designed based on 9 rules described in Table 2. To demonstrate the fuzzy inference system, the following rule is considered to show the function of inference engine and outputs of each rule are each rule are combined for generating fuzzy decision (Fig. 9).

Rule 1: If (SINR = high) and (Q = high) and (UF = high) Then, UP = high, end if.

The obtained output value is defuzzified similar to study. As an outcome of fuzzy decision, we can obtain the priority of each user. Then scheduling of slots on each sub channel level is performed based on the obtained priority of users.

RESULTS AND DISCUSSION

Simulation model and parameters: Network Simulator (NS2) (25) is used evaluate performance of the proposed HMM and Fuzzy based Scheduling and Resource Allocation Technique (HFSRA). The proposed scheme is implemented over IEEE 802.16 MAC protocol. In the Simulation, clients (SS) and the Base Station (BS) are deployed in a 1000×1000 m region for 50 sec simulation time. All nodes have the same transmission range of 500 m. In the simulation, both the CBR and video traffic are used. CBR is used for non-real time traffic and Video is used for real-time traffic.

There are 8 downlink traffic flows from BS-SS. Among the 8 flows, 4 CBR and 4 Video flows are used.

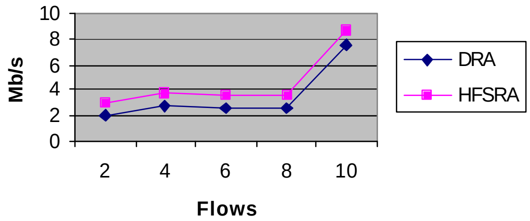

Fig. 10: Flows vs bandwidth

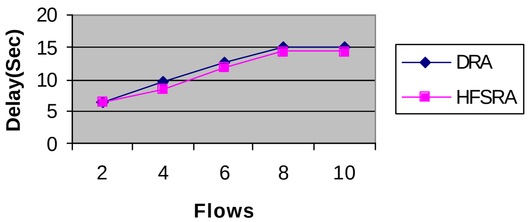

Fig. 11: Flows vs delay

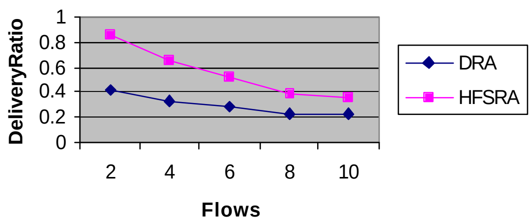

Fig. 12: Flows vs delivery ratio

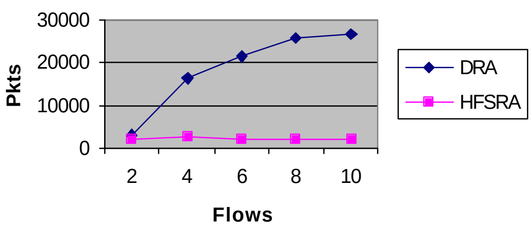

Fig. 13: Flows vs drop

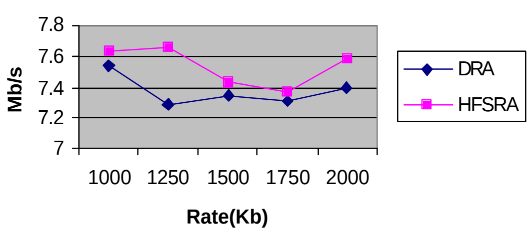

Fig. 14: Rate vs bandwidth

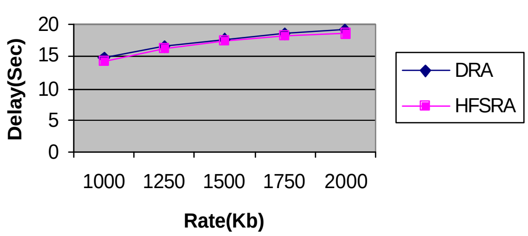

Fig. 15: Rate vs delay

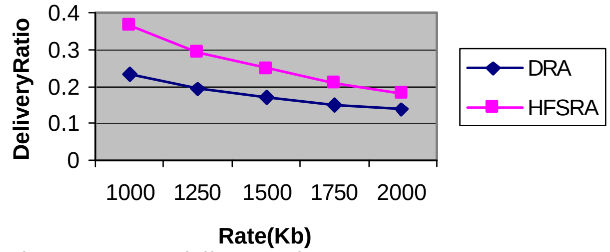

Fig. 16: Rate vs delivery ratio

The simulation settings and parameters are summarized in Table 3.

Performance metrics: We compare our proposed HFSRA scheme with the DRA [ ] scheme. We mainly evaluate the performance according to the following metrics:

Aggregated bandwidth: We measure the received bandwidth (in Mb sec^{-1}) for CBR traffic of all flows.

Average end-to-end delay: The end-to-end-delay is averaged over all surviving data packets from the sources to the destinations.
Delivery ratio: It is the number of packets successfully received by the receiver.

Packet drop: It is the number of packets dropped during the transmission. The parameters are extracted from the trace file. The performance results are presented in the next section.

Effect of varying the traffic flows: In the first set of graphs, both type of traffic flows are varied from 2-8 and kept as the X-axis.

From Fig. 10, we can see that the aggregated bandwidth of our proposed HFSRA is higher than the existing DRA technique. From Fig. 11, we can see that the delay of our proposed HFSRA is less than the existing DRA method. From Fig. 12, we can see that the delivery ratio of our proposed HFSRA is higher than the existing DRA technique. From Fig. 13, we can see that the packet drop of our proposed HFSRA is less than the existing DRA technique.

Effect of varying transmission rates: In the first experiment, we vary the rate as 1000, 1250, 1500, 1750 and 2000 Kb for the 8 flows. The results are given as follows.

From Fig. 14, we can see that the aggregated bandwidth of our proposed HFSRA is higher than the existing DRA technique. From Fig. 15, we can see that the delay of our proposed HFSRA is less than the existing DRA method. From Fig. 16, we can see that the delivery ratio of our proposed HFSRA is higher than the existing DRA technique.

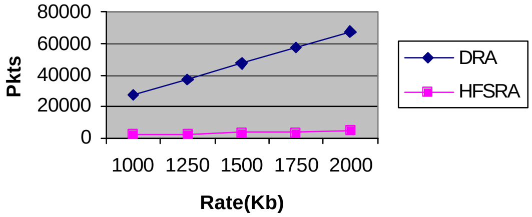

Fig. 17: Rate vs drop

Table 3: Simulation settings

Parameters	Values
Area size	1000×1000
Mac	802.16
Clients	10
Radio range	500m
Simulation time	50 sec
Routing protocol	DSDV
Traffic source	CBR and video
Physical layer	OFDM
Video trace file	JurassikH263-256k_trace.dat
Channel error rate	0.01
Packet size	1500 bytes
Frame duration	0.005
Transmission rate	1000, 1250, 1500, 1750, 2000Kb
No. of flows	2, 4, 6, 8

From Fig. 17, we can see that the packet drop of our proposed HFSRA is less than the existing DRA technique.

CONCLUSION

In this study, we have proposed HMM and Fuzzy Based Scheduling and Resource Allocation for Downlink Traffic in WiMAX OFDMA networks. In this technique, new users are admitted after checking the predicted free slots Using Hidden Markov Model (HMM) against the estimated slots of each user.

Then the exact number of slots for each user is estimated using fuzzy logic inference considering QoS requirement such as bandwidth, delay and transmission rate of users over each subchannel.

The scheduling of slots on each sub channel level is performed based on user priority which is estimated again using fuzzy logic inference considering the channel quality, queue length and urgency factor. By simulation results, we have shown that the proposed technique enhances the network throughout and minimizes the overhead.

REFERENCES

Arhaif, M.S., 2011. Comparative study of scheduling algorithms in WiMAX. Intl. J. Sci. Eng. Res., 2: 1-7.

Chung, Y.H., C.J. Chang and Z. Zhang, 2012. A user-differentiation-based resource allocation scheme for OFDMA downlink system. Proceedings of the 2012 IEEE International Conference on Communications (ICC), June 10-15, 2012, IEEE, Ottawa, Canada, ISBN: 978-1-4577-2052-9, pp: 4658-4662.

Hou, F., J. She, P.H. Ho and X. Shen, 2009. A flexible resource allocation and scheduling framework for non-real-time polling service in IEEE 802.16 networks. Wirel. Commun. IEEE. Trans., 8: 766-775.

Hua, Y., Q. Zhang and Z. Niu, 2010. Resource allocation in multi-cell OFDMA-based relay networks. Proceedings of IEEE INFOCOM, March 14-19, 2010, San Diego, CA, USA., pp: 1-9.

Kakali, V.L., P.G. Sarigiannidis, G.I. Papadimitriou and A.S. Pomportsis, 2011. A novel HMM-based learning framework for improving dynamic wireless push system performance. Comput. Math. Appl., 62: 474-485.

Karunkuzhali, D. and D.C. Tomar, 2012. Fuzzy based dynamic buffer management in WiMAX 16 m networks. J. Comput. Sci., 8: 1946-1956.

Katsaggelos, A.K., Y. Eisenberg, F. Zhai, R. Berry and T.N. Pappas, 2005. Advances in efficient resource allocation for packet-based real-time video transmission. Proc. IEEE., 93: 135-147.

Liang, J.M., Y.C. Wang, J.J. Chen, J.H. Liu and Y.C. Tseng, 2010. Efficient resource allocation for energy conservation in uplink transmissions of IEEE 802.16 J transparent relay networks. Proceedings of the 13th ACM International Conference on Modeling Analysis and Simulation of Wireless and Mobile Systems, October 17-21, 2010, ACM, Bodrum, Turkey, ISBN: 978-1-4503-0274-6, pp: 187-194.

Lin, Y.N., Y.D. Lin, Y.C. Lai and C.W. Wu, 2009. Highest Urgency First (HUF): A latency and modulation aware bandwidth allocation algorithm for WiMAX base stations. Comput. Commun., 32: 332-342.

Meddour, D.E., A. Abdallah, T. Ahmed and R. Boutaba, 2012. A cross layer architecture for multicast and unicast video transmission in mobile broadband networks. J. Netw. Comput. Appl., 35: 1377-1391.

Nascimento, A. and J. Rodriguez, 2009. Dynamic Resource Allocation for IEEE 802.16e. In: Mobile Lightweight Wireless Systems. Granelli, F., C. Skianis, P. Chatzimisios, Y. Xiao and S. Redana (Eds.). Springer, Berlin, Germany, ISBN: 978-3-642-03818-1, pp: 147.

Rakesh, J. and U. Dalal, 2010. A survey of mobile WiMax IEEE 802.16 m standard. Prepr., Vol. 8,.

Seo, S.S., J.M. Kang, N. Agoulmine, J. Strassner and J.W.K. Hong, 2011. FAST: A fuzzy-based adaptive scheduling technique for IEEE 802.16 networks. Proceedings of the 12th IFIP/IEEE International Symposium on Integrated Network Management (IM), May 23-27, 2011, IEEE, Dublin, Ireland, ISBN: 978-1-4244-9219-0, pp: 201-208.

Seraphin, G.B. and S. Ramesh, 2013. QoS provisioning in vertical handover of WiMAX/WLAN overlay network for fixed and mobile users. Wirel. Commun., 5: 158-162.

She, J., F. Hou, B. Shihada and P.H. Ho, 2010. MAC-layer active dropping for real-time video streaming in 4G access networks. Syst. J. IEEE., 4: 561-572.

She, J., F. Hou, B. Shihada and P.H. Ho, 2011. MAC-layer active dropping for real-time video streaming in IEEE 802.16 networks. GESJ. Comput. Sci. Telecommun., 2: 1-12.

Shuaibu, D.S., S.S. Yusof and N. Fisal, 2011. Dynamic resource allocation in mobile WiMAX using particle swarm optimization techniques. Intl. J. Phys. Sci., 6: 1009-1014.

So I.C., R. Jain and A.K.A. Tamimi, 2010a. Resource Allocation in IEEE 802.16 Mobile WiMAX. In: Orthogonal Frequency Division Multiple Access Fundamentals and Applications. Lingyang, T.J. and Y. Zhang (Eds.). CRC Press, Boca Raton, Florida, USA., ISBN: 978-1-4200-8824-3, pp: 189-234.

So, I.C., R. Jain and A.K.A. Tamimi, 2010b. Generalized weighted fairness and its application for resource allocation in ieee 802.16 e mobile WiMAX. Proceedings of the 2nd IEEE International Conference on Computer and Automation Engineering (ICCAE), February 26-28, 2010, IEEE, Singapore, Asia, ISBN: 978-1-4244-5585-0, pp: 784-788.

So-In, C., R. Jain and A.K. Tamimi, 2009. Scheduling in IEEE 802.16e mobile WiMAX networks: Key issues and a survey. Inst. Electr. Electron. Eng. J. Select. Areas Commun., 27: 156-171.

Verma, L., S. Kim, S. Choi and S.J. Lee, 2008. Reliable, low overhead link quality estimation for 802.11 wireless mesh networks. Proceedings of the 5th IEEE Annual Communications Sensor Workshops on Mesh and Ad Hoc and Networks, June 16-20, 2008, IEEE, San Francisco, California, ISBN: 978-1-4244-2562-4, pp: 1-6.

Wang, H. and L. Dittmann, 2007. Adaptive radio resource allocation in hierarchical QoS scheduling for IEEE 802.16 systems. Proceedings of the IEEE Global Conference on Telecommunications, November 26-30, 2007, IEEE, Washington, DC., USA., ISBN: 978-1-4244-1042-2, pp: 4769-4774.