WiMAX has been recently labeled as one of the few contending technologies for next generation of wireless networks. Among other special features of this technology, Quality of Service and Radio Resource Management have been in forefront of research activities. Numerous groups have proposed schemes for physical slot, call admission and control, and bandwidth allocation strategies. Scheduling and radio resource management is the heart of QoS for WiMAX networks. This chapter briefs about developments done in this field over the past few years.

2.1 Scheduling in IEEE 802.16

Scheduling is the method by which data flows are given access to system resources (communications bandwidth in this case). This is usually done to load balance a system effectively and/or achieve a target quality of service. WiMAX has two modes of operation: Point to Multi Point (PMP) mode and Mesh mode. PMP mode consists of one BS and multiple SS, communication between different SS takes place only through BS i.e. BS is responsible for both uplink and downlink communication where as in mesh mode different SS may have the ability to communicate among themselves. Every SS can act as Base Station. A central node connected to outside world may be considered as mesh BS which may provide connectivity to external world. In Mesh mode, scheduling is either central in which mesh BS is responsible for scheduling all SSs or it may be distributed ie each SS can participate in decision making process of bandwidth allocation but transmissions in two hop neighborhood are coordinated to avert collisions. This study focuses on PMP mode only and has implemented scheduler based for PMP mode operation.

2.1.1 Issues and Challenges for IEEE 802.16 Scheduling

The goal of this section is to provide understanding of current issues for the design of schedulers. Since WiMAX has to deal with heterogeneous traffic therefore the major design issues concerning development of schedulers may be stated as under

- **Guaranteeing certain degree of fairness** to subscribers running different applications. However ensuring fairness to every node or flow may not be always
easy as it may be conflicting with efficiency. Time requirements of real time applications are more stringent than non-real time applications which also conflicts with fairness.

- **Guaranteed delivery of QoS requirements**: Scheduler shall be able to provide guaranteed delivery of QoS requirements that are negotiated at the time of connection establishment.

- **Effective Channel utilization**: It may be measured in terms of throughput. In order to improve channel utilization several other factors like AMC, MIMO techniques and fragmentation mechanism may be needed to be explored.

- **Complexities**: associated with the implementation of algorithm shall be small. The implemented algorithm shall have very small time complexities as real time applications have their own stringent time requirements.

- **Good bandwidth-request strategy**: Efficient bandwidth request strategy is also very important for scheduler as it could add substantial burden to the resources if not handled carefully. Algorithm shall be able to choose whether to piggyback, multicast and broadcast or send stand alone messages to request more bandwidth.

- **Efficiency**: in terms of delay, throughput, scalability, robustness.

- **Graceful degradation of scheduler performance under heavy network loads**

- **Focus of schedulers is being shifted towards exploitation and use of instantaneous information from other layer to improve over all scheduling process.**

Scheduling in WiMAX is main component of MAC layer that helps to assure QoS to various service classes. Job of scheduler is to schedule packets in frame and allocate resources to various SS. Resources can be allocated in number of slots. Slot is the smallest unit for bandwidth allocation in OFDMA networks. These slots are further mapped to number of sub-channels and time duration (OFDM symbols) where sub-channels are group of multiple physical subcarriers.

Scheduling algorithms consider allocations from two prospective logical and physical. Logical allocation deals with calculation of number of slots based on QoS service classes and physical allocation deals with selection of sub-channels with suitable time intervals. Logical scheduling strategies don’t have knowledge of PHY layer parameters. We had summarized scheduling strategies based on logical scheduling perspective.
There are three distinct scheduling processes in IEEE 802.16 networks: two at the BS - one for downlink and other for uplink and one at the SS for uplink as shown in Figure 2.1. Packets of different applications from higher layer are classified to different queues at BS. These queues are built on per CID basis to prevent blocking of HOL. Queues can be optimized and their number can be controlled. Downlink-BS scheduler serves these queues on basis of QoS parameters and available channel information. The decision in downlink involves calculating the number of SDU to be transferred to different SS. Uplink scheduler present at the base station makes allocation decision based upon bandwidth requests from different SS and quality of service parameters associated with request. Scheduling decisions in the uplink are difficult to make as scheduler at BS does not have an up-to-date knowledge of traffic at various SSs. Third scheduler present at SS utilizes bandwidth grants allocated by UL BS scheduler and further allocates it to various queues present at SS itself. The grants made by BS scheduler are per subscriber basis and therefore mechanism at SS should be efficient enough to maintain desirable performance.
2.2 Uplink Bandwidth Allocation and Request Mechanisms

BS uplink scheduler decides about allocation of slots for traffic to various SSs. Slots are required to enable various SS to send traffic upwards and for requesting more bandwidth. BS downlink scheduler has complete knowledge of traffic such as queue length and packet size which helps it in scheduling decision making. For uplink, SS sends bandwidth request (BWR) packets to BS, BS scheduler decides upon the allocation to be made to different SS on the basis of these requests and available channel conditions. BS scheduler calculates number of slots to be granted to particular SS and informs that SS about allocation using UL-MAP messages. Bandwidth can be allocated to SS in two ways – GPC (Grant per connection) in which BS makes allocations with respect to connections established or GPSS (Grant per subscriber) basis in which grants are made to SS as whole. Further allocation to different data connections is left to SS itself with recent IEEE 802.16 standards permitting only latter approach. Bandwidth request (BWR) made by SS are either incremental or aggregate. The BS adds amount of bandwidth requested by SS to existing request in case of incremental request while it replaces previous request with new one for aggregate. Different BWR/Grant mechanisms supported by IEEE 802.16 are

Un-solicited request:- It consists of dedicated slots reserved for UGS class SSs. Useful for applications that require fixed data rate.

Unicast polling:- Polling is process by which BS allocates bandwidth to SSs for purpose of making bandwidth requests. These allocations may be for individual or group of SS. Polling ensures that SS are able to receive service after some deterministic intervals as per their requirement. This allocation technique is used when un-solicited bandwidth grants are not possible for all SSs. Unicast polling refers to polling of an individual SS. SS indicate requirement of bandwidth by setting its PollMe bit in grant management subheader. No explicit message is transmitted to poll SS but enough bandwidth is allocated to SS so that it can respond with a bandwidth request.

Contention based procedures including broadcast or multicast polling:- The available bandwidth may not be sufficient to individually poll all inactive SSs. Contention based grant-request mechanisms allocate bandwidth request contention slot for each uplink frame (in the FDD mode) or subframe (in the TDD mode). This slot is termed as Request IE and its size is indicated by BS in UL-MAP message. This slot enables SS to request for more slots to
transfer data from BS. BS receives the demand and evaluates SS request by considering its service-level agreement, scheduling algorithm and current network state before making allocations.

Piggybacking bandwidth request opportunities: The piggyback bandwidth request uses grant management subheader transmitted in a generic MAC frame. This can avoid SS transmitting a complete (bandwidth request MAC header) MPDU with the overhead of a MAC header only to request bandwidth.

2.3 Related Work

Authors have analysed various theories available in literature and these have been classified and put forward into five different categories in this study. The classification of papers into these categories is no way exhaustive and there are certain studies that can overlap i.e. may belong to two or more different categories. This classification has been done as per authors understanding of working principles of different available algorithms.

The first WiMAX standard was incepted in 2001; Christian Hoymann[22] and K. Wongthavarawat et. al.[23] were first to propose a scheduling architecture for WiMAX networks. While [23] has proposed a hierarchal structure of scheduling, authors in [22] had simulated WiMAX network using MATLAB. Authors have performed experiments on the interaction of fragmentation and padding of OFDM symbols to evaluate its effects on system capacity. Different MAC layer configurations with different levels of robustness were analyzed to know its impact on desired QoS.

As the standard got popular number of researchers working on this new standard grew. In order to estimate the amount of work done in this field we had performed little survey by searching for number of papers published on this topic over the years. The inputted string was “WiMAX scheduling” and publications in journals of two of world known databases were evaluated and are listed below for reference (as accessed on 14 November 2013). Table 2.1 shows number of publications over period of 2006-2013 in IEEE and Science Direct database. The trend shown in table indicates a linear rise in number of publications over these years. Number in most recent past two years 2012 and 2013 is far large as compared to previous years. This is because of evolvement of IEEE 802.16 networks as one of most popular standards in recent years. Total number of WIMAX deployed networks is increasing day by day and so is the number of applications supported by it.
Spectrum allocation for WiMAX in countries like India, China and Taiwan can also be attributed to rise in this number as major contributions of papers is from these countries.

<table>
<thead>
<tr>
<th>Year</th>
<th>Papers available in IEEExplore database</th>
<th>Papers available in Science Direct database</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>2007</td>
<td>19</td>
<td>38</td>
</tr>
<tr>
<td>2008</td>
<td>15</td>
<td>66</td>
</tr>
<tr>
<td>2009</td>
<td>25</td>
<td>52</td>
</tr>
<tr>
<td>2010</td>
<td>28</td>
<td>85</td>
</tr>
<tr>
<td>2011</td>
<td>29</td>
<td>123</td>
</tr>
<tr>
<td>2012</td>
<td>27</td>
<td>102</td>
</tr>
<tr>
<td>2013</td>
<td>32</td>
<td>138</td>
</tr>
</tbody>
</table>

Table 2.1:- Number of journal publications on WiMAX scheduling from 2006-2013

2.3.1 Classification of Schedulers

WiMAX being an ultramodern technology has support for various types of applications. Role of schedulers in such a system is very crucial as these applications have diverse requirements. This reason together with fact that IEEE had allowed vendors to test and implement schedulers of their choice has led to large number of proposals in this field. Authors in this study have compiled few of these studies applicable to PMP mode in this section. These schemes have been divided into 5 different categories as described below

- Traditional Schedulers
- Hierarchal Schedulers
- Cross Layer Schedulers
- Dynamic Schedulers
- Schedulers based on Soft Computing Approaches

2.3.2 Traditional schedulers

Traditional schedulers consist of all those scheduling techniques which had been borrowed from studies of queuing theory principles like Round Robin (RR) scheduler, WRR, WFQ,
etc. These mostly include studies based on breakdown of queues and their characteristics. These has been proposed and implemented by number of researchers especially at the time when WiMAX standard was gaining popularity.

Ball et al. [12] has proposed a round robin (RR) scheduler. This scheduler instead of assigning priorities to real time applications tries to allocate resources to all connection in a round robin fashion. Round Robin distributes equal channel resources to all the SSs without any priority. This is the simplest of all the schedulers with minimal complexity and it was implemented so that decision time required to be taken to schedule every packet can be nullified. The results indicate that although BE service performs good but performance for real time application is not good enough. The algorithm simplifies complexity as no decision time was required but this technique is not suitable for systems with different levels of priority and systems with varying sizes and types of traffic.

Weighted Round Robin (WRR) scheduler is another variant of RR scheduling in which weights are assigned to every flow/queue. These weights are static and once assigned are never changed. The weights determine the number of packets that will be scheduled for that flow. More weight is assigned to flows with more priority[13] but Sayenko et al. [14] insisted that WRR because of its work conservative behaviour is not fit for IEEE 802.16 networks which has a fixed size frame and secondly weights are floating numbers while slots allotted are integers. Moreover WRR may not perform well for WiMAX which has inherent support for multiple applications.

The scheduling strategy of [14] consisted of the following steps:

- Allocation of minimum number of slots to each connection for ensuring minimum reserved traffic rate. rtPS and nrtPS will be allocated free slots at first instance and BE will be assigned remaining slots.
- Decrease in values of delay and jitter by ordering of slots

Overheads for making allocations to UGS, nrtPS and ertPS was estimated while estimation of overheads for rtPS and BE was not possible. Overheads incurred from fragmentation and scheduling decision have also been considered by authors. However authors had applied an interleaved scheme for slots ordering which contradicts with frame structure of standard. Moreover allocation of bandwidth has been done using GPC(grant per connection) mode where allocations are made to individual connections but standard specifies GPSS, grant per
subscriber station mode, in which aggregate bandwidth is allocated to subscriber as whole and further distribution of it to different connections lies with SS itself.

In [10], Cicconetti et al. surmise that minimum reserved traffic rate is one of the basic QoS parameter for any flow therefore schedulers in IEEE 802.16 may be implemented by class of latency-rate (LR) scheduling algorithms and only two parameters can determine behaviour of such algorithms namely latency and allocated rate. Authors had chosen deficit round robin (DRR) algorithm for its simpler implementation. Complexity of DRR can be of the order of O(1) if certain allocation constraints are met. Deficit Round Robin [DRR] scheduler was also implemented by M. Shreedhar et al.[15]. DRR works by associating a deficit counter with each flow/queue. Deficit counter is incremented by a fixed amount after each round for every flow. A comparison between deficit counter and length of packet decides whether head of queue will be de-queued or not. In case packet is de-queued deficit counter is decremented by length of packet. DRR provides fair queuing but it requires a minimum rate to be reserved for each packet flow which means even BE class shall be provided with a minimum rate. DRR has another flow that it puts a limitation that only one packet at most can be sent for each flow. It provides fairness for variable length packets but major problem in DRR for WiMAX is calculating size of head of queue packet which was not possible for uplink traffic in WiMAX as this information lies with different SS. In fact BS is only able to estimate the overall amount of backlog of each connection, but not the size of each backlogged packet. Therefore DRR had been implemented for downlink and SS schedulers in most of the studies. Cicconetti et al.[10] has provided simulations for performance of 802.16 systems in terms of delay and throughput and concluded that other metrics such as frame duration, bandwidth request mechanism, distribution of traffic among SS and number of connections within SS also affect performance of 802.16 systems. A new variant of DRR scheduler that handles latency critical applications has been proposed by H. K. Rath et al. [18] namely Opportunistic Deficit Round Robin (O-DRR) scheduler in which BS polls subscribers periodically and includes SSs into a set based upon different conditions and selects one SS for allocation from this set.

WFQ (weighted fair queuing) one of the most common variant of fair queuing has been analysed by studies like[17][20]. WFQ has been utilized as its default scheduler in Qualnet simulator for WiMAX Networks [19]. WFQ is a data packet scheduling technique allowing different scheduling priorities to statistically multiplexed data flows. Both in WFQ and FQ,
each data flow has a separate FIFO queue. In FQ, with a link data rate of R at any given
time, M active data flows are serviced simultaneously, each at an average data rate of R/M.
As opposed to FQ, WFQ allows different sessions to have different service shares. WFQ
takes into account arrival time, size and associated weights of flows to make scheduling
decision. A parameter called virtual finishing time is calculated for a packet every time it
enters a router. Packets are sorted by their virtual finishing times in priority queue and packet
whose finishing time is smallest will be the one selected for output. WFQ has a classifier
which classifies packets into appropriate flows and is responsible for transmitting them to
output queue. Weights of queues in WFQ shall satisfy following constraints:-

\[ \sum_{i=1}^{m} w_i = 1, \quad 0.01 \leq w_i \leq 1 \]  

(2.1)

where \( w_i \) is the weight of \( i^{th} \) service class and indicates the portion of bandwidth that will be
allocated to that flow and \( m \) is the total number of flows. Variations of WFQ are also
available like Worst-case Fair Weighted Fair Queueing WF\(^2\)Q that selects the packet for
transmission that has the smallest finish time among all the packets that have already started
service. It achieves more fairness but adds substantial calculation overhead.

Proportional fairness in WFQ can be achieved by setting the weights to \( w_i = 1/c_i \), where \( c_i \) is
the cost per data bit of data flow \( i \). Loutfi Nuaymi et al. [13] argued that a proportional Fair
(PF) scheduler should in theory result in better throughput than the various variants of DRR
schedulers because PF scheduler assigns slots first to those connections that have the best
ratio of current achievable rate to averaged rate and incorporates the aspect of fairness among
different flows. C.F. Ball et al. [21] specified temporary Removal Scheduler (TRS) that
involves identifying the packet call power, depending on radio conditions, and then
temporarily removing them from a scheduling list for a certain adjustable time period \( T_R \).
Reference of few more scheduling algorithms like Drop tail queue, random early detection
and random early detection with IN/OUT may be found at [22][23][24] in which comparisons
between various scheduling techniques have been carried out. The maximum Signal-to-
Interference Ratio (mSIR) scheduler is based on the allocation of radio resources to
subscriber stations which have the highest signal-to-interference Ratio (SIR) but it may lead
to starvation of the flows having lower SIR as no mechanism had been proposed to deal with
such situations. Suitable mechanism is required for preventing starvation of such connections.
2.3.3 Hybrid Schedulers

Hybrid Schedulers are the schedulers that combine several scheduling techniques to accommodate different service types. They had an advantage as a scheduler appropriate to a specific application type may be chosen. Moreover the left over bandwidth after fulfilling needs of every flow may be effectively utilized in the presence of various options. Resources are normally distributed as first level of hierarchy and different types of techniques are employed later on for scheduling specific service flows. Traditional approaches are combined with a certain level of admission control to avoid starvation. Table 2.2 lists major studies in this direction.

Wongthavarawat et al.\[25\][26] were the first to introduce concept of hierarchal schedulers for scheduling problem of WiMAX. They performed scheduling of different service classes using different algorithms. UGS was allotted fixed time slots, rtPS was scheduled using Earliest Deadline First (EDF), nrtPS uses Weighted Fair Queuing (WFQ) while all BE classes were scheduled using Round Robin(RR) algorithm. The inter-class scheduling follows fixed priority with UGS having highest priority followed by rtPS, nrtPS and BE. Every SS is made to follow a traffic contract while demanding bandwidth to avoid starvation and this contract was included in each SS. The algorithm takes into account service received by connection and queue size information apart from arrival time and deadline requirements of rtPS connections. Only uplink scheduling is considered and no mechanism for dynamically determination of durations of UL and DL subframes is specified. An admission control mechanism based on token bucket was also proposed by authors.

J. Sun et al. [27] proposed two different schedulers for BS and SS. Allocation to UGS and opportunities for bandwidth request in case of rtPS and nrtPS were made at connection establishment. Data grants to rtPS, nrtPS classes were made considering bandwidth request information and their minimum requirements. The residual bandwidth was distributed in accordance to pre-assigned weights. Fixed priority scheme was implemented at SS assigning priorities of 1,2,3 and 4 for for BE, nrtPS, rtPS and UGS service classes respectively. UGS is allocated guaranteed bandwidth at the first instant, deadlines for rtPS packets were then calculated based on arrival time and tolerated delay and were scheduled on the basis of approaching deadlines. For nrtPS packets, a virtual number was calculated to guarantee minimum reserved bandwidth and was assigned to each packet. All BE classes were scheduled in FIFO manner.
<table>
<thead>
<tr>
<th>Study</th>
<th>Scheduling</th>
<th>Phase</th>
<th>UGS</th>
<th>rtPS</th>
<th>nrtPS</th>
<th>BE</th>
</tr>
</thead>
<tbody>
<tr>
<td>25,26</td>
<td></td>
<td>1</td>
<td>Fixed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>EDF</td>
<td>WFQ</td>
<td>RR</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>BS</td>
<td>1</td>
<td>Guaranteed Bandwidth</td>
<td>Grant Bandwidth Request Opportunity at connection setup</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Guarantee min reserved rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>WFQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SS</td>
<td></td>
<td>Fixed bandwidth</td>
<td>Fixed Priority</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>BS</td>
<td>1</td>
<td>Assigned to high priority queue</td>
<td>Intermediate queue</td>
<td>Low priority queue</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>migrate to high priority queue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>SS</td>
<td>1</td>
<td>slots given to each time sensitive connection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Remaining slots distributed in Round robin fashion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>BS Tier 1</td>
<td>Fixed Bandwidth</td>
<td>Priority based on queue length</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>Tier 2</td>
<td>Fixed Bandwidth</td>
<td>Fairness Queuing</td>
<td>WRR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>per class flow</td>
<td>Tier 3</td>
<td>-----------</td>
<td>EDF</td>
<td>Shortest length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Scheduler 1</td>
<td>EDF used for UGS, rtPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scheduler2 (nrtPS)</td>
<td>WFQ(bandwidth request)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scheduler 3(BE only)</td>
<td></td>
<td>WFQ (traffic priority)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Phase I</td>
<td>Modified SFQ(start time fair queuing)</td>
<td>MCFQ with start/finish time stamping</td>
<td>BE served only when no traffic of other classes. Starvation avoided using admission control and traffic shaping. WRR used among various BE flows.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phase II</td>
<td>Priority scheduler with priority in order UGS/ertPS &gt; rtPs/nrtPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2:- Comparison of hybrid scheduling strategies
N. Liu et al. [28] presented another hierarchical scheme that uses combination of three different schedulers to satisfy different traffic class requirements. UGS and rtPS having time sensitive requirement were served by scheduler one employing EDF algorithm. WFQ was used to schedule minimum bandwidth reserving flows like nrtPS where weights were in proportion to their bandwidth requirements. BE was scheduled by third scheduler employing WFQ scheduling. Weights in scheduler three correspond to traffic priorities of each BE connection and these schedulers were served in fixed priority order with scheduler one being assigned highest priority.

Perumalraja et al. [29] had proposed a multimedia supported uplink scheduling technique which includes PF (Proportional fair) BS scheduler. A different scheduler at SS was proposed known as earliest due date scheduler. Resources were first allocated to UGS service by BS scheduler and then were allocated to those SS which have at least one non-UGS connection. This was accomplished by polling different SS and reserving one slot in each frame for each SS having nrtPS and rtPS connections. SS having only BE service connections were allocated one slot after every three frames. The residual bandwidth if any was allocated to different SS in proportion to their bandwidth requests received. EDD scheduler present at SS calculates deadlines of each packet in their respective queues and serves these queues on basis of that deadline without considering type of service into account.

Chen et al. [30] proposed a scheduling algorithm for TDD mode taking both uplink and downlink allocation in account. Deficit Fair Priority Queuing (DFPQ) will be used at first level of hierarchy to avoid starvation of low priority flows. First level of scheduling is based on these two techniques:

1) transmission based policy where DL was given higher priority

2) Priorities was assigned in the order  rtPS> nrtPS>BE.

These two policies were combined by strict priority scheme which had assigned priorities in order: DL-rtPS, UL-rtPS, DL-nrtPS, UL-nrtPS, DL-BE, and UL-BE. UGS connections were allocated fixed bandwidth. Different algorithms were used for scheduling at second level of hierarchy namely EDF for rtPS, RR for BE and WFQ for nrtPS. Authors have also provided a new connection control algorithm to check whether system will be able to handle new connection by considering minimum reserved rate of connection against available capacity.
Only BE connections were accepted without checking as they don’t have any specific requirements.

Maode Ma et al. proposed in [31] a three-tier scheduling framework in which loads for downlink and uplink could be unbalanced. The size of DL sub-frame was computed in beginning which was not the case in Chen et al. [30]. Size of each sub-frame was computed based on queue lengths and on the basis of bandwidth requests. They divided scheduling scheme into 3 tiers, Tier 1 scheduling scheme exists at BS only. It performs bandwidth allocation coarsely across service class and across SS. Tier 2 scheduler works at SS and determines number of time slots granted by BS for different connections within each service class at each SS. Tier 3 scheduling is to determine the transmission priorities of packets in each connection at each SS.

Juliana Freitag et al. [32] used the concept of high, intermediate and low priority queues to handle varying types of traffic. High priority queue is used to handle flows that must be scheduled in next frame which includes UGS packets and uni-cast request opportunities for rtPS and nrtPS flows. Intermediate and low priority queues were used to handle rtPS, nrtPS and BE flows respectively. Queues were served using strict priority however starvation was avoided as request whose deadline is going to expire is migrated to high priority queue. Although approach seems to be very simple but it adds significant overhead as queues need to be checked for deadline expiry.

Mohammad Masri et al. [33] further extended concepts of Juliana et al. [32] to incorporate issue of scalability and used GPSS(Grant per subscriber Station) mode for data grants instead of GPC(grant per connection) which the authors thought to be of major hindrance to scalability issue. The scheduler used was at SS only whose job was to distribute the granted slots to different connections. Slots were first given to each time sensitive connection and remaining slots were distributed in round robin fashion, starting from UGS, rtPS flows needing additional airtime, then to nrtPS connections and finally to BE connections.

Settembre et al. [34] proposed a scheduler that considered fragmentation, packing capabilities of the packets before making scheduling decision. The scheme combined fixed priority among different service categories and used following queuing principles for different traffic classes: fixed bandwidth for UGS, WRR for rtPS and nrtPS and RR for BE. Weights in WRR were determined as per guaranteed bandwidth. Initial allocations inWRR/RR assumes presence of most robust profile while actual available profile was used for making bandwidth
allocations. This guaranteed enough bandwidth for existing flows even in worst cases however unjustified high blocking rate and low link utilization may be caused if the channel is good. The authors did not provide any admission control mechanism to guarantee minimum bandwidth.

L. F Chan et al. [35] proposed scheduling algorithm (2TSA) which works as two tier system and uses satisfaction levels of different connections. Starvation avoidance and fair allocation of residual bandwidth were the main aim of the algorithm. Allocations to UGS were fixed in every frame. All connections were classified into three different categories as satisfied, unsatisfied or over satisfied depending upon the allocated bandwidth and their minimum requirements. A weight was assigned depending upon the category to which the connection belongs. After making fixed allocations to UGS, first tier scheduler assigns highest priority to unsatisfied connections followed by other categories. Second tier scheduler is applied to share residual bandwidth based on weights of connections.

Some authors have proposed application specific scheduling strategies which are independent of all scheduling structures. It has an advantage that the complexities involved in these strategies is very small and suffers from less delay. Lee et al.[36] has argued that VoIP services has some problems on UGS and rtPS classes and had proposed an algorithm to solve these problems. They argued that making fixed allocations to UGS leads to waste of resources during silence period and bandwidth request mechanism of rtPS connections cause delay and overhead which is not tolerable for VoIP services. The strategy proposed by them assumes that SS has a silence detector in higher layers that will inform BS about their voice state transitions. One of the reserved bits of MAC generic header was used to avoid MAC overhead. Simulations also justify the usefulness of the study.

Yaseer P. Fallah et al. [37] proposed different scheduling architectures for BS and SS. They proposed combination of scheduling schemes for Real–time multimedia support in IEEE 802.16 networks. The scheduling process is divided into two phases. Phase 1 uses three different scheduling algorithms where flows with similar characteristics use same type of scheduler. UGS and ertPS uses modified start-time fair queuing to increase temporal fairness; rtPS and nrtPS classes require BW guarantee and uses multi-class fair queuing algorithm whereas BE traffic can use any scheduler since no QoS guarantee is required for BE. BE is scheduled only when there is no QoS for other classes and may at times face starvation. Starvation for BE class was avoided using admission control and traffic shaping.
also employed among similar BE flows. Phase II was used to handle packets selected in phase-I using a priority based scheduler where UGS and ertPS enjoy higher priority than rtPS and nrtPS. The drawback of this method in the inherent complexity associated with this algorithm and starvation of BE flows which at times may account for large amount of traffic.

2.3.4 Cross Layer schedulers

The objective of cross layer scheduling techniques is to optimize resource utilization with help of communication among various layers of network architecture. These studies can be further grouped as those belonging to communication between MAC-PHY, MAC-IP and MAC-APP layer. There had been some papers discussed in the previous section that have taken into account AMC characteristics of PHY layer which can be referred as MAC-PHY communication.

Since the standard provides a link adaptation framework based on which the MCS can be adapted to the channel conditions therefore there is need for MAC-PHY communication and scheduler can be defined to utilize this capability. Noordin et.al.[38] proposed and implemented a cross layer architecture for WiMAX system and justified this cross layer communication with help of simulations. The optimizer was implemented as an interface between MAC and PHY layers and it was used to extract and tune required parameters. The authors were of the opinion that application layer is not needed in the cross-layer architecture as application requirements will be considered by Quality of Service provisioning at MAC level.

Liu et al. [39] [40] introduced a priority-based scheduler at the medium access control (MAC) layer for multiple connections with diverse QoS requirements and where each connection employs adaptive modulation and coding (AMC) scheme at the physical (PHY) layer. A priority function (PRF) was defined for each connection admitted in the system and a factor called normalized channel quality factor was calculated based on SNR and PRF. It was updated dynamically depending on the wireless channel quality, QoS satisfaction, and service priority across layers. PRF was dependent on coefficient for each class together with:

- normalized channel quality for BE connections
- normalized channel quality and rate performance for nrtPS connections,
- normalized channel quality, and delay requirements for rtPS connections.
Time-slots to UGS were fixed and all non-UGS connections were assigned slots on basis of priority in order of rtPS followed by nrtPS and BE. Priority Function (PRF) was used to allocate left over time-slots. The technique proposed is simple and easy to implement since it does not depend on any specific traffic model. It has drawback that only one non-UGS connection per frame can be scheduled that can cause delay and may also lead to starvation for low priority flows. This can also cause delays to even high priority flows also at the time of network overload. Therefore algorithm shall be combined with an efficient CAC mechanism.

Unlike [39][40] who had restricted their cross layer architecture to layer 1 and layer 2 authors of [41] [42] emphasized the need of involvement of layer 3 and layer 2 for better QoS service since some very important information is available at these layers that can help in scheduling process. They included mapping between Layer 3 and Layer 2 QoS where Integrated and differentiated services are mapped to 802.16 MAC service classes as shown in table 2.3. Apart from defining mapping authors had also defined admission control scheme which allows a connection only if requested bandwidth is low as compared to remaining link capacity and a fragmentation control mechanism to group fragments of same IP packet to be treated as single unit by MAC layer.

<table>
<thead>
<tr>
<th>IP QoS</th>
<th>MAC 802.16 QoS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guaranteed service(GC)</td>
<td>Unsolicited Grant Service(UGS)</td>
</tr>
<tr>
<td>Controlled Load(CL)</td>
<td>Real time polling service(rtps)</td>
</tr>
<tr>
<td>Expedited forwarding(EF)</td>
<td>Non-Real time polling service(nrtps)</td>
</tr>
<tr>
<td>Assured Forwarding(AF)</td>
<td></td>
</tr>
<tr>
<td>Best effort (BE)</td>
<td>Best Effort(BE)</td>
</tr>
</tbody>
</table>

Table 2.3:- Mapping rule from IP QoS to 802.16 QoS

One more cross layer mechanism that communicates with application layer has been proposed by Triantafyllopoulou et. al.[43][44]. The proposed scheduler works at both BS and SS. Algorithm tries to take advantage of adaptation capabilities existing at both PHY and application layers. Multi rate features of multimedia applications and AMC capability of physical layer are combined in scheduling process. The optimization process collects abstraction of layer-specific information and informs corresponding layers of the required
changes. These changes were instructed based on a decision algorithm that decides about the MCS and traffic rate for each SS. Although technique benefits in terms of QoS and system capacity but it adds a lot of complexity at BS.

2.3.5 Dynamic schedulers

Mukul et al. [45] argued that process of bandwidth request and scheduling can be adaptive and proposed a stochastic adaptive scheduler for rtPS traffic based on the prediction of the rtPS packets arrival. BS allocates bandwidth for rtPS traffic after receiving a bandwidth request. During this period of request and grant it is possible that SS may receive from upper layers new rtPS packets which may aggregate bandwidth requirement. However these packets need to wait till next request is sent. The basic idea is to predict arrival of more rtPS packets so that subscriber can request time slots for currently present rtPS data and also for data which can arrive during this request-response time. A staircase function is proposed and network calculus method has been used to analyze proposed method. The authors had tried to reduce delay and length of queue however for simulations data flow has been assumed to be generic and no statistic on data entry was taken.

Z. Peng et al. [46] tried to improve upon the work proposed by Mukul and others [45] by considering queue length factor and Lagrange’s Interpolation function to estimate value of time width and data arrival rate. Authors had proposed a variable to estimate proportion of values of this function. The value of this variable changes dynamically however the simulations are based on the same model and only rtPS class has been considered.

Jin-Yup Hwang et al. [47] divided traffic models into two types namely NRTV for real time and FTP for non real time traffic. Authors had stated adaptive traffic allocation scheduling scheme that provides priorities to traffic class group and to SSs belonging under that class group. This priority is based on traffic type and maximum allowed delay time. The Real time traffic is scheduled using RR and non real time using Proportional Fair algorithms, similarly real time packets are preferentially allocated.

N. Ruangchaijatupon et al. [48] tried to impart fairness to adaptive scheduling scheme by using concept of priority queuing and deficit queuing and used an adaptive deficit quantum to handle priority queue. This quantum was based on current queue size and channel capacity and algorithm assigns this quantum to a particular flow depending upon whether the traffic is in burst or non-burst state.
In [49] an adaptive queue-aware algorithm is proposed for uplink bandwidth allocation and rate control mechanisms in SS for polling services in a GPSS system. This scheme helps to adjust amount of bandwidth allocated for polling service dynamically as per variations in traffic load, channel quality, and queue length at SS at the same time maintaining QoS performances such as protocol data unit delay and protocol dropping probability at desired level. However approach draws no boundaries between real-time and non real-time services and fails to exploit QoS factors like latency in scheduling.

The authors of [50] Raghu et al. proposed a queue based algorithm in which adaptability is implemented by defining a parameter X defined as ratio of maximum time a rtPS or nrtPS MPDU can wait in queue (i.e. max_mpdu_delay) to the maximum latency specification of the real-time flows. This parameter was used to control QoS given to real-time and non real-time services and was varied to obtain the desired delays for real-time and non real-time traffic flows.

S. Kim et al.[51] argued that a bandwidth-request grant mechanism used in IEEE 802.16 may not be effective for TCP flows because of the dynamic nature of sending traffic and had proposed a scheme for TCP traffic that does not need any bandwidth request process for allocation. Instead, it estimates amount of bandwidth required for a flow based on its current sending rate.

One of the most recent work in the field of dynamic scheduling has been done by M. Fathi et al. [52] where a joint scheduling and CAC method is proposed. The whole process is divided into two stages, in stage one weighted fair queuing is used to assign initial weights to different traffic classes in order of rtPS to nrtPS to BE. Bandwidth allocated is calculated as function of packet dropping probability, average arrival and departure rates of a class. Law of moving averages was employed to calculate new rates and new portion of bandwidth was allocated to any flow at stage II. An appropriate scheduler based on type of traffic was then used to schedule these packets.

2.3.6 Soft Computing based

These categories of scheduling strategies tend to formulate scheduling problem as an optimization problem that aiming to optimize resource allocation to different SS/service flows subject to certain constraints. Soft computing techniques like Genetic Algorithm, Neural Networks, game theory etc are potential candidates for solving optimization problems.
and have been successively applied to resource allocation problem in WiMAX. This section explores studies on WiMAX schedulers based upon these principles.

To get a solution to optimization problem, V. Sharma et al. [53] had used the concept of dynamic programming. A linear programming based approach with a complexity of $O(n^3 m^3 N)$ where $N$, $n$, and $m$ denote the number of slots, number of SSs and number of sub-channels is proposed. Based on developed approach, two different algorithms were defined for BS and SS. Authors suggested that designing a joint allocation scheme for both uplink and downlink may be more advantageous but doing so is not possible in TDD mode. Therefore different classes were scheduled separately for uplink and downlink on basis of their priorities. Priorities are defined in order of UGS, rtPS, nrtPS to guarantee minimum requirements, and finally allocations were made to satisfy their demands. Choosing a particular algorithm depends upon available resources and channel conditions. Algorithm at SS was implemented considering overall system performance and fairness to users. For algorithm at BS same steps were followed and two models were proposed: packet model which prohibits fragmentation for UGS and rtPS and byte model which allows use of fragmentation. However, the complexities involved with this approach were still demanding therefore heuristic algorithms with complexity of $O(nmN)$ were proposed. The authors proved that proposed algorithms optimize the overall system performance but may result in unfairness and modified it using proportional-fair concept.

The authors of [54], A. Mohammadi et al., had defined problem of scheduling as achievement of two goals namely maximizing number of UGS and total packets sent. They tried to formulate scheduling problem as 0-1 Knapsack problem which is NP hard problem and therefore argued that concept of dynamic programming can be applied to optimize such problem. To achieve second goal authors assigned more priority to packets belonging to UGS class. However only mathematical induction based theorems has been stated as a proof for applicability of dynamic programming metaphor.

In [55], Niyato and Hossain has defined a queuing theoretic and optimization based model for resource management for system working in TDMA/TDD mode. A utility function was defined which depends upon amount of allocated bandwidth, average delay, throughput, and admission control decision for UGS, rtPS, nrtPS, and BE, respectively. Minimum ($b_{min}$) and maximum ($b_{max}$) limits were defined for bandwidth allocation to each connection. Sharing of bandwidth was made using threshold based partitioning approach and a threshold was also
defined for each service class. To optimize value of threshold authors have suggested two solutions iterative and optimal approach. Optimal approach has a complexity of $O(M \times \Delta b)$ where $M$ denotes the number of ongoing and incoming connections and $\Delta b = b_{\max} - b_{\min} + 1$. An exorbitant complexity associated with optimal algorithm makes it less suitable from implementation point. Therefore iterative approach based on water filling mechanism was used which is more implementation friendly with far less complexity without sacrificing system performance. A queuing and analytical model was developed to analyse connections and packet level performance. Continuous Time Markov Chain model was employed at connection level model to calculate number of outgoing connections and blocking probability. An optimization problem was formulated from these parameters whose aim is to maximize system revenue maintaining blocking probability at target level.

Niyato et al. [56] [57] utilized concepts of non-cooperative game theory for admission control and scheduling in IEEE 802.16 networks. Players in game are the rtPS and nrtPS connections that want to maximize their QoS performance while total utility of both ongoing connections is regarded as playoff. The problem is to find equilibrium point between two types of connections so that a new connection may be offered bandwidth while meeting the QoS requirements of both ongoing and new connection.

The concepts of Genetic Algorithm as a solution to scheduling problem have been proposed by [58]. Authors proposed a cross layer APP-MAC-PHY scheduling algorithm based on genetic algorithm that uses information at the application layer together with AMC properties of WiMAX aiming to provide optimal scheduling. The algorithm works by having current rate allocations as initial population of two chromosomes at time $t$, defined as

$$\bar{X}_1(t) = \{x_1^1(t) \ldots \ldots x_1^n(t)\} \text{ and } \bar{X}_2(t) = \{x_2^1(t) \ldots \ldots x_2^n(t)\}$$ (2.2)

and weight of each user was taken as a function of modulation scheme index, packet error rate (PER), SNR and QoS parameters which depend upon different types of services. Fitness function taken is defined as

$$f(\bar{X}(t)) = \min_{x_s \leq M_s} \sum_s -w_s \log (x_s)$$ (2.3)

where $M_s$ is the rate limit and aim is to minimize this value. Priority based selection operator and a suitable crossover and mutation operator have been used to allocate bandwidth to different users. However simulations are provided only for a small number of nodes and only BE traffic class has been considered for performing simulations.
R. Gunasekaran et al.[59] had also utilized genetic metaphor to solve broadcast scheduling problem in WiMAX networks. They had represented computer network as graph with nodes representing stations and edges as connection between nodes. A set of nodes is found such that all nodes in that set could transmit at same time without any conflict. Authors have tried to find an optimal TDMA frame represented in the form of M*N matrix where M is number of time slots in frame and N is the number of nodes based on satisfaction of following constraints:-

- Each node must be activated at least once.

- A primary constraint that no node can receive and transmit data at same time and secondary constraint that data cannot be received from two nodes at same time.

In case of multiple solutions utilization index defined as ratio of total number of nodes activated to total number of slots available is considered. The genetic algorithm was applied to maximize value of utilization index where chromosomes are represented as M*N matrix having values [0, 1] where each row corresponds to time slot and column represents a node. Value of 1 indicates that corresponding node is active at specified time. Chromosome population is generated by converting different permutations into required 2-D matrix by assigning suitable position. Two chromosomes having good fitness scores are selected for crossover with a predefined crossover probability. Mutation is applied with a probability .005 to incorporate randomness into the solution.

Chenn Jung et al. [60] had proposed an autonomic resource management scheme for forced termination of multimedia handoffs and optimize utilization of bandwidth in heterogeneous wireless communication systems. Genetic algorithm attempts to fit need of newly arrived connection by using bandwidth allocation strategy based on fitness function. Bandwidth is forecasted by using probability of disconnection for each station during next period. A call admission control mechanism accordingly determines whether a newly arriving call or a handoff should be accepted. The results have been tested on WiMAX and wifi heterogeneous environment.

Authors of [61] had used customized particle swarm optimization aided algorithm to maximize the capacity of OFDMA by adaptively assigning sub channels to users. Capacity is increased by assigning sub-channel to user with best gain and by distributing power using water filling algorithm. The PSO algorithm evolves to approach an optimal solution from
population set. The customized algorithm works for discrete particle positions unlike classical PSO algorithm which is valid for only continuous particle positions. The authors argue that PSO aided sub-channel allocation is able to provide significant gain in capacity even with very small population size and number of iterations. The method is advantageous as sub-channel allocation and assignment was done simultaneously thereby improving complexity of the method to be of order O(N). Authors believe that this is a crucial factor in systems like WiMAX where wireless channel changes rapidly.

Neural networks have also been proposed to solve bandwidth allocation and scheduling problem by [62]. A feed forward neural network with a single scalar output had been chosen to make decision of allocating bandwidth needs of different users. The authors had defined an algorithm which is based on real measures processed by neural networks in presence of heterogeneous flows. The algorithm had been implemented using request-grant feature of 802.16. The approach seems to be novel but seems to lack in providing validity as only nrtPS connections had been studied.

Yao-tien et al.[63] had utilized ANN for performing load balancing in mobile agent. The approach aims to maximise number of tasks served so that high performance clusters can be made that can be later used for scheduling. As per authors such a system is needed with growth of service types and number of users in mobile networks. Although the authors had not specifically applied this to any network but such a system is also required in standards like WiMAX. The authors had applied back propagation algorithm to calculate load of a host dynamically by taking CPU, memory, I/O, queue-length as input. The ANN classifies a host into under loaded, moderately loaded and over-loaded categories and thereafter calls appropriate routine to load balance it to.

Raliean et al. [64] had used theory of neural networks to predict traffic characteristics in WiMAX. ANN has been associated with Stationary Wavelet Transform to predict traffic time series. The main focus of their study is to compare quality of forecasting obtained using different configurations of ANN and testing these configurations using real traffic data from a WiMAX Network developed by Alcatel. This is rarest of the work in which data taken is real world data. Comparisons with previous techniques are drawn at the end to show usefulness of technique.

The authors of [65], D.D.N.P Kumar et al. have used neuro-fuzzy based methods to provide QoS and solve scheduling problem. They divided the scheduling problem in two stages. In
first stage fuzzy logic is used to provide priorities to different services based on queue size and second stage uses a multi layer neural network for scheduling. The input to the first layer of neural network is the output of fuzzy network while layer two and three comprises of Kohenen and Grossberg neural layers respectively. The role of the fuzzy logic was to provide a normalized input to ANN and was used as input layer.

Fuzzy logic has been employed by Tarek Bchini et. al. [66] and Jaraiz Simon et. al[67] in handover algorithms. Use of fuzzy logic for implementing inter-class scheduler for 802.16 networks had been done by Yaseer Sadri et al.[68]. Authors had defined fuzzy term sets according to two variables $d_{rt}$ which means latency for real time applications and $t_{nrt}$ meaning throughput for non real time applications. The set was defined as

$$ T(d_{rt}) = \{ \text{Low, Low Medium, Medium, Medium High, High} \} = \{ L, LM, M, MH, H \}, $$

$$ T(t_{nrt}) = \{ \text{Low, Medium, High} \} = \{ L, M, H \}. $$

Fuzzy rule base formed had dimensions $|T(d_{rt})| \times |T(t_{nrt})|$. Fuzzy system works by selecting an appropriate queue to transmit data. Authors argue that the fuzzy system responds well to delay and latency requirements.

Shuaibu et al. [69] has developed intelligent call admission control (CAC) in admitting traffics into WiMAX. They had proposed a fuzzy logic partition-based call admission control (FZ-CAC) which partition total link bandwidth into three classes corresponding to constant bit rate (CBR), variable bit rate (VBR) and handover (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. Authors believe that proposed approach outperforms both partition-based CAC (PB-CAC) and conventional bandwidth allocation CAC (CB-CAC).

Mohammed Alsahag et al. [70] had utilized uncertainty principles of fuzzy logic to modify deficit round robin algorithm to work dynamically on the basis of approaching deadlines. The Fuzzy based scheduler dynamically updates bandwidth requirement by different service classes according to their priorities, latency and throughput by adjusting weights of respective flows. The system takes two inputs: maximum latency (ML) of HOL packet for the corresponding queues such as UGS, ertPS and rtPS class services, denoted as $R_{ML}$ and throughput for non-real time classes such as nrtPs and BE denoted by $N_{Thru}$. Fuzzy system
employs single variable output structure and works whenever a request arrives from SS. A quantum variable defined by equation 2.5 is calculated

$$\phi_i = \gamma \ast \left( \frac{s_i \ast \left( \frac{w_i}{W} \right)}{D} \right) + M_i$$  \hspace{1cm} (2.5)

$M_i$ is a fixed increment by one maximum packet length to guarantee the scheduler to transmit at least one packet in every round.

$D$ is deadline of packet.

$w_i$ the flow weight.

$W$ is total weight of all flows.

$\gamma$ is optimal bandwidth ratio obtained from embedded fuzzy logic procedure in order to provide adequate bandwidth required to transmit requests without missing their deadline as well as guarantee fairness in the system.

$\Phi_i$ is quantum value indicating number of packets to be transferred for real time traffic.

Aim of method is to calculate optimal bandwidth for both real and non real type traffic which was found by calculating weight of the flow using fuzzy logic. This weight was used to calculate amount of bandwidth to be allocated to real time connections. Authors had used 5, 3, 5 membership function for throughput, latency and output weight respectively. The fuzzy database consists of 15 different rules for input to output mapping. Although amount of bandwidth to be allocated to real time traffic is provided but no mechanism has been suggested for amount of allocations to non-real time traffic. DRR has been used which has its own problems for implementing scheduler for uplink traffic as information lies with SS. There is considerable overhead related with calculating latency for HOL packets in DRR but this has been neglected. The number of rules in databases is less as only 15 rules have been considered.

We have tried to put together algorithms available in various dimensions in one study. There have been similar attempts in this direction like [71-75] but they fail to contribute at certain issues as stated below. While Miray et al. [71] has covered schedulers in mesh mode, the main focus of other studies [72-74] is on traditional schedulers like RR, WRR, FQ, WFQ only where as there are large number of studies available on other working principles also.
<table>
<thead>
<tr>
<th>Study</th>
<th>Origin &amp; amendments to WiMAX standard</th>
<th>Factors affecting scheduling performance</th>
<th>Comparison of WiMAX with similar technologies</th>
<th>Traditional Approaches like (RR, WRR, FQ, WFQ etc)</th>
<th>Hierarchal Approaches</th>
<th>Dynamic/Channel Aware schedulers</th>
<th>Cross Layer Schedulers</th>
<th>Soft Computing Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>73</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>74</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Only proportional fair scheduler</td>
<td>No</td>
<td>No</td>
<td>General approach presented; no study explored</td>
<td>No</td>
</tr>
<tr>
<td>75</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 2.4:– Comparison of studies on WiMAX scheduling surveys
Chapter 2
Literature Survey

Lamia Chhari et al. [75] has covered very few hierarchal and dynamic schedulers while none of them had considered scheduling as a cross layer approach or as an optimizing problem. Present chapter focuses on factors effecting scheduling in WiMAX and includes all the categories in which scheduling algorithms can be divided. Cross layer and soft computing (neural network, fuzzy logic etc.) based schedulers along with traditional, hierarchal and dynamic schedulers have been covered and compared to previous studies. Some of the theoretically and practically proven scheduling techniques from other domains like [76] [77] which can suitably be applied/migrated to WiMAX are also included. Table 2.4 justifies usefulness of the survey done by authors in present work. Table summarizes gaps in previous studies on IEEE 802.16 scheduling surveys.

2.4 Chapter Summary

This chapter has talked about process of bandwidth allocation and scheduling in IEEE 802.16 networks. Standard offers different types of grant-request mechanisms and equipment manufacturers are free to design their own scheme. There are three different schedulers required in WiMAX networks and their specific roles in IEEE 802.16 scheduling problem have been discussed. Design of uplink scheduler is a difficult task as information available with BS is not up to date. Development of different scheduling techniques over past years had been covered. The scheduling schemes are divided into five classes, so called traditional, hierarchal, cross layer scheduling, soft computing based and dynamic schedulers. Differences of these techniques and advantages of one over others are also discussed. Authors have tried to demonstrate that focus of today’s work in scheduling of IEEE 802.16 networks is getting shifted from traditional or hierarchal algorithm to more efficient dynamic algorithms. Schedulers based on soft computing techniques like neural networks, genetic algorithm or fuzzy logic have started to evolve. Although number of studies in this direction are still not aggressive but future belongs to these studies. New studies that work on fusion of soft computing techniques have started new trends of development. There has been number of survey papers on WiMAX schedulers and a comparative analysis of these has also been tabled at the end.