Abstract— IEEE 802.16 standard and its marketing technology Worldwide Interoperability for Microwave Access (WiMAX) is one of the leading technology that gain benefit of adopting orthogonal frequency division multiple access (OFDMA) as its multiple access technique. OFDMA gives flexibility in resources allocation to accommodate a large number of users supporting several services classes with quality of service (QoS). This loads big challenges on its resource allocation. One of the key performance factors of OFDMA resource allocation is its downlink data packing mechanism. In IEEE802.16 standard based OFDMA the downlink data is packed into down link frames in the form of rectangular slots called burst. The standard leaves details of packing mechanism as an open issue for researchers and manufacturer to implement. Therefore recently several algorithms have been proposed in literature. This paper endeavors to identify key factors and tradeoffs issues associated with designing of downlink data packing algorithm for WiMAX based IEEE802.16 through a competitive survey of recent proposed algorithms.

Keywords—IEEE 802.16, WiMAX, OFDMA, Data allocation .

I. INTRODUCTION

WiMAX is a wireless technology that provides broadband data access for fixed, portable, and mobile users based on OFDMA. This multiple access technique based on OFDM modulation provides significant advantages in terms of high spectrum efficiency, robustness against multipath fading channels, resistance to multiuser interference, simplified equalization, and so on [4]. IEEE802.16 OFDMA divides the available resources (frequency & time) into several orthogonal sub carriers in frequency domain and into several adjacent symbols in time domain. This gives more flexibility in assigning those recourses, serving many users supporting several service classes with QoS. This leads to a significant increase in resource allocation challenges. Scheduling in a way so as to accommodate the largest number of users/services and packing data before delivery to the users are the major factors that affect efficiency of resource allocation, as a whole system capacity and system performance.

In centralized system such as WiMAX system the base station (BS) performs scheduling using Quality of services (QoS) scheduling algorithms and data packing using data packing algorithms. The standards specify only uplink (UL) Packing algorithm and leaves QoS scheduling algorithms and downlink (DL) Packing algorithms as open issues which open the door for researcher and manufacturer to propose new algorithms or adopt some existing algorithms to improve WiMAX system performance and capacity. In designing efficient downlink packing algorithm many factors and tradeoffs issues like mapping overhead, wastage slots, power consumption ... etc have to be considered. In this paper we identify most key factors and tradeoffs issues associated with downlink data packing algorithm design. A competitive survey of recent proposed algorithms and how they tackle the key factors and trade off issues is presented.

The rest of the paper is organized as follows: - section II gives brief overview of WiMAX systems, section III gives details of Downlink sub frame allocations, section IV surveys of DL Packing algorithms, while conclude in section V.

II. WiMAX OVERVIEW

WiMAX is a wireless broadband solution that offers a rich set of features with a lot of flexibility in terms of deployment options and potential service offerings [4] IEEE802.16d, IEEE802.16e and IEEE802.16m are standards for Wireless Metropolitan Area Network (WMAN) [1, 2, and 3]. In parallel the WiMAX forum releases several technical specification profiles [5]. Together they make WiMAX one of the most promising technologies for broadband wireless access solution, as well as a 4G candidate. The major distinguish futures of WiMAX are scalable OFDMA, multiple input multiple output (MIMO) antenna, beam forming and adaptive modulation and coding (AMC), support time division duplexing (TDD) and frequency division duplexing (FDD), space time coding, strong security and multiple QoS classes [8].

A. PHY layer overview

One of the operating band profiles for WiMAX is 3.5GHz with subcarrier spacing of 10.94 kHz allowing 128, 512, 1024 and 2,048 FFT size with channel bandwidth of 1.25MHz, 5MHz, 10MHz, and 20MHz, respectively [4]. The transmission and reception between BS and subscriber station (SS) is based on OFDMA. WiMAX dived the channel into orthogonal sub carrier and group those sub carriers into logical sub channel using distributed permutation mode such as partial
use of subcarriers (PUSC), F full use of subcarriers (FUSC) or adjacent permutation mode adaptive modulation and coding (AMC). Further modulate and code those subchannels adaptively based on SNR Quality to improve overall channel efficiency. Time division duplex (TDD) is favorable duplexing where the frame is divided asymmetrically into downlink subframe followed by an uplink subframe separated by a small gap as shown in Fig.1. On the frame the minimum resource allocation unit is a slot which consists of one subchannel over one or more symbol based on permutation modes. Slot capacity can vary with modulation & coding scheme (MCS) used. Packing data into DL and UL frame is done by BS packing algorithms and it broadcasts detail of packing arrangement in DL-MAP & for all users at the beginning of each frame. The next section gives more detail of DL-MAP.

Fig.1. OFDMA TDD frame structure

B. MAC layer overview

WiMAX MAC layer consist of three sublayers namely the convergence sublayer, the common part sublayer and the security sublayer. Convergence sublayer handle interface with upper layer protocols like ATM, Ethernet and IP protocol. Security sublayer secures the air transmission and protects the data from theft of the services using encapsulation protocol and a privacy and key management protocols. MAC common part performs several essential functions like access controls, MAC PDU construction, mobility support, Automatic Repeat Request (ARQ), power saving, call admission control, QoS provisioning, connection managements and packet scheduling [6].

- **MAC PDU construction**

The incoming MAC services data unit (MSDUs) form the higher layer are assembled to form MAC protocol data unit (MPDUs) where single MSDU fragments into two or more MPDUs. Or multiple MSDUs aggregate into single MPDU so it can be efficiently transmitted over the air. Farther multiple MPDUs that share same PHY parameter are concatenated into single burst to reduce MAP overhead. Each MPDU start with a 6 byte generic MAC header (GMH) followed by variable length payload information and optional 4byte CRC [34]. Fragments, packing, and concatenation are shown in Fig.2.

![MPDUs Fragments, packing, and concatenation.](image)

Fig.2. MPDUs Fragments, packing, and concatenation.

- **QoS scheduler**

WiMAX is connection oriented, where each uplink or downlink session application is associated with a 16 bit logical ID called Connection Identifier (CID). A WiMAX service flow is a unidirectional flow of packets with a particular set of QoS parameters, and is identified by a Service Flow Identifier (SFID) [7]. Services classes supported by WiMAX consist of Unsolicited Grant Service (UGS), Real Time Polling Service (rtPS), Non Real Time Polling Service (nrtPS), Extended Real Time Polling Service (ertPS) and Best Effort (BE). WiMAX implements QoS by BS QoS scheduler which classifies the traffic into flows and maps each connection CID into associated SFID that define its QoS parameters. The QoS scheduler maintains this parameter by its scheduling police.

- **Call Admission Control (CAC) & Bandwidth allocation**

Call admission control is a mechanism of controlling new connections that want to connect to the network where the BS decides whether to accept or reject new connections based on the available resource and QoS requirements. The BS will grant the bandwidth to SS in two modes. First mode is Grant per Connection (GPC) where each connection is treated separately and bandwidth is allocated after BS received SS request. Second is Grant per Subscriber Station (GPSS) where all connections from a single SS are treated as single unit and bandwidth is granted accordingly by the BS on a per SS basis [33].

III. WIMAX FRAME ALLOCATION

In TDD WiMAX the frame is divided into DL subframe and UL subframe. DL-MAP & UL-MAP messages are used by BS.
to control access to the air frame. DL-MAP & UL-MAP relevant time is shown in Fig.3. In the uplink direction, data is allocated horizontally to minimize number of subchannel. The allocation starts with lowest numbered subchannel in the lowest numbered OFDMA symbol and continues till the last subchannel is reached. It is then continued from the lowest numbered subchannel in the next OFDMA symbol [2]. In the downlink direction data are allocated in rectangular shape to minimize both number of subchannels and number of OFDMA symbols which required an efficient two dimensional packing algorithm [15]. Fig.4. shows as example of DL &UL data mapping.

\[
\text{DL\_MAP\ header} = 104 \\
\sum_{i=1}^{n} (44+16 \text{No CID})
\]

### Table 1 DL-MAP fields

<table>
<thead>
<tr>
<th>Field Type</th>
<th>Field Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL-MAP header</td>
<td>104 bytes</td>
</tr>
<tr>
<td>Fields of DL-MAP</td>
<td>requiring 104 bits / Subframe</td>
</tr>
<tr>
<td>Message type (8bits)</td>
<td>to specify type of the message</td>
</tr>
<tr>
<td>PHY synchronization</td>
<td>(32bits) for frame duration and number.</td>
</tr>
<tr>
<td>DCD Count (8bits)</td>
<td>to count change in DCD, which describes the downlink burst profiles.</td>
</tr>
<tr>
<td>BS ID (48bits)</td>
<td>to identifying the BS.</td>
</tr>
<tr>
<td>No Symbols (8bits)</td>
<td>to specify number of OFDMA symbols used</td>
</tr>
<tr>
<td>No CID (8bits)</td>
<td>to specify number of connections on this burst.</td>
</tr>
<tr>
<td>CID (8bits)</td>
<td>to represents the assignment of the IE to a broadcast, multicast, or uncast addresses.</td>
</tr>
<tr>
<td>Symbol offset (8bits)</td>
<td>to specify starts of burst measured in OFDMA symbols.</td>
</tr>
<tr>
<td>Sub channel Offset</td>
<td>(6bits) to specify lowest index OFDMA subchannel used for carrying the burst, starting from sub channel 0.</td>
</tr>
<tr>
<td>No Subchannel (6bits)</td>
<td>to specify number of OFDMA Sub channel used to carry the burst.</td>
</tr>
<tr>
<td>No Symbols (7bits)</td>
<td>to specify number of OFDMA symbols used to carry the burst.</td>
</tr>
<tr>
<td>RCI (2bits)</td>
<td>Repetition Coding Indication to indicate the repetition code used inside the allocated burst.</td>
</tr>
<tr>
<td>DIUC (4bits)</td>
<td>to specify type of burst profile.</td>
</tr>
</tbody>
</table>

Fig.3. Minimum time relevance of DL-MAP (TDD)

Fig.4. DL & UL data mapping

### A. DL-Map and it’s overhead

WiMAX assigns slots to users in downlink in a rectangular form called a burst where a burst contains data for a single or multiple CID that share the same physical parameters. Each user is informed about its burst allocation by broadcasting the DL_MAP massage with the most reliable MCS at the beginning of the DL subframe. Table (1) describes the fields of DL-MAP. The fields consist of two main groups. The first group which required 104 bits once per DL Subframe consists of Message Type, PHY Synchronization, DCD Count, BS ID and No Symbols. The second group consists of No CID, CID, Symbol Offset, Subchannel Offset, Boosting, No Subchannel, No Symbols, Repetition Coding Indication and DIUC [2]. This group requires 44+16 No CID bits once per burst to define a two-dimensional allocation pattern of the burst and is called Downlink MAP Information Element (DL_ MAP_ IE). The total DL_MAP header is equal to summation of required bits as in equation (1) [10].

Now let see how it work in an example scenario where we have 10 users running VoIP application. Each user received (42 bytes in each frame [9]). The total data to be sent in each frame is \((10 \times 42) = 420\) byte. If each user has its own burst then the size of DL-MAP is \(((104+ 60x10)/8) = 88\) byte. To broadcast DL-MAP massage we have to use most reliable MCS so it may required \(88/3 = 29.333 \approx 30\) slot. Higher MCS can be used to unicast VoIP data such as 12 byte on average per slot so it required \(420/12 = 35\) slot. Therefore to sending 35 slots of data we spend 30 slots wastage as overhead. Though this may not always be the case, but it does illustrate
how crucial DL_MAP overheads can be. In WiMAX release 1.5 proposes periodically allocation to overcome this problem.

B. Wastages slots in downlink frame

The wastage in the down link frame not only because of DL-MAP overhead but also because of packing process. To illustrate this wastage a simple packing example is given hereafter. We consider (10 subchannels x 10 symbols, 50 slots) DL subframe with PUSC mode and four requests (303, 235, 90, and 55) bytes it has to be send in this DL subframe. First step is covert those request into slot as per MCS (403/24 = 16.79 ≈ 17 slot, 235/16 = 14.6 ≈ 15 slot, 90/9 = 10 slot, 47/6 = 7.33 ≈ 8). Second step shape those request into best rectangle that can fit into DL subframe as following (17+1 = 18 into 3 x 6, 15 into 3 x 5, 10 into 2 x 5, 8 into 4 x 2). Finally fit those rectangles in DL subframe as shown in Fig. 5.

![Fig. 5. arbitrary data packing example](image)

The are several wastages in this packing example. Firstly few bit approximation wastage when we convert request from bytes into slot, secondly some request cannot form a rectangle so in need to allocate extra slot like 17+1=18. Thirdly bad fitting may cause unused slot and final last request will not get enough space due to previous wastage so it return to QoS scheduler this may affect QoS requirements. DL-MAP packing algorithm and its overhead are crucial to system performance as it affect spectrum efficiency, capacity and QoS of the system.

IV. SURVEY OF DOWNLINK PACKING ALGORITHMS

After The BS QoS scheduler select the data to be transmitted in the DL subframe, the downlink packing algorithm shapes and places the data into the two dimension DL subframe for transmission. This may seem similar to the will known two-dimensional pin packing problem [11] but there are many constrains in downlink Packing which make it more difficult. There are several factors and tradeoffs which have to be considered to efficiently pack the data into down link sub frame. These include unused slots, extra allocated slots, fragment overhead, power consumption, efficient use of subchannel, complexity and QoS requirements. Through a survey of recent work we are going to discuss those factors and see how they are related and also discuss contradictions.

A. Survey of packing algorithms based on the most important factor addressed.

- Algorithms concentrating on packing efficiency with fragments.

Ben-Shimol et. al. address the problem of two-dimensional mapping for wireless OFDMA in [12], where a Raster Algorithm is proposed. The algorithm sorts the selected requests data in descending manner and packs them row by row from left to right and from top to bottom. It starts with the largest request that can be fitted without fragments. It then breaks the other requests into more bursts which lead to an increased DL_MAP overhead. Also grouping requests with same MCSs is not considered. Allocation by rows increases SS active time leading to increases the power consumption which is not suitable in mobile application and allocation from left fixed the size of DL-MAP.

Xin et. al. Propose Mapping with Appropriate Truncation and Sort (MATS) Algorithm [13]. A MAT tries to improve Raster Algorithm by limit fragments of large request into one using truncation and then mapping. But it still produces many fragmentations which result in large overhead.

- Algorithms concentrating on packing efficiency without fragments.

The Weighted less flexibility first (WLFF) Algorithm was proposed by Wang et. al. [14]. The algorithm defines the flexibility on calculated terms and allocates the requests based on the flexibility which may require more computation. Also one of flexibility criteria is to sort the requests from large to small which may contradict with the requests QoS order. WLFF reserves the first column for DL_MAP and this may limit the number of allocated bursts.

One Column Stripling with non-increasing Area first mapping (OCSA) proposed by So-In et. al. [15] which was further enhanced in [16] as (eOCSA) by reduce it complexity from O (n3) to O (n2) at the cost of assigning some extra slot also problem of unused slot still exist and the algorithm does not consider grouping allocations with same MCSs which increase the map over head. The efficiency of the algorithm highly degraded under some traffic sizes.

Zhu et. al. Allocate the burst in columns of identical width then shuffle allocated bursts to combine the scattered unused space in the frame to form a large space that can accommodate more bursts [17]. Orientation-Based Burst Packing (OBBP) algorithm present by Eshanta et. al. [18] was similar to [15]. The algorithm is based on burst factorization and pre-
arranging them using matrices. This simplify finding optimal
column or rows to minimize the unused slot and avoids
padding the algorithm shows good efficiency at heavy load but
the efficiency goes down at low loads. Also the algorithm
efficiency is dependent on burst size ratio.

- Algorithms concentrating on reduce DL_MAP
  overhead.

Sample Data Region Allocation Algorithm (SDRA) was
proposed by Bacioccola et. al. [19]. The algorithm first
groups the SS data with same MCS into so called date regions
and then assigns data regions by columns from bottom to top
and from right to left. This minimizes DL_MAP by reducing
the number of IEs. But the SS has to decode the entire data
region to get its own data from listed SSs in the region which
increases the SS power consumption.

Oh seksi et. al. [20] propose an algorithm that first prepare a
bucket of one time slot with more than one sub channel to
construct data from different SSs with the same MCS in
combined columns. It starts as one column, and if the buckets
grow it expands by filling another column which may not be
fully utilized. The algorithms maintain QoS requirements as it
group the requests sequentially as order by the QoS scheduler.

Shiann et. al. [21] Aim to reduce DL_MAP overhead by
minimizing the number of bits required for a DL_MAP IE.
Instead of sending complete IE indexing they construct the
relation among bursts and send portion of IE indexing and
leave it for the SSs to reconstruct complete IE indexing.
Piggybacking IE on data so that it is transmitted with higher
MCS was proposed by Ju-Yeop et. al [22]. Both [21, 22] need
modifications in the standards.

- Algorithms concentrating on reducing the power consumptions.

Desset et. al. Present algorithm to reduce the SS power
consumption by using binary search tree[23]. The algorithm
minimizes the average duration of a burst within a given
frame, in such a way that a SS needs to be awake for receiving
data during a shorter time, but the complexity limits number of
users to 8. This is no practical. Papers by Abbas et. al. [24,25]
model DL packing problem as a programming model where
the authors formulate objectives, decision variables and
constraints mathematically, and apply optimization solver or
full search to get optimal solution. The algorithm aims to
reduce BS energy consumption by maximizing the use of
resources at the base station while reducing the energy wasted
caused by sending padding bits. But the algorithm requires
more computation time.

- Algorithms concentrating on cross layer designing.

Considering traffic QoS priority Yen-Wen et. al. Propose
Target Side Allocation (TSA) Algorithm [10]. The algorithm
searches for rectangular side that can accommodated the burst
without fragments called target side, from left to right so that
the number of traffic bursts can be minimized. Otherwise it is
placed normally which may increase number of bursts. In
addition they analyze tradeoffs between degrading of
modulation level cost and DL-MAP cost to increase the
resource utilization.

Jia-Ming et. al. combine problem of scheduling and burst
allocation in cross-layer manner and propose that the
scheduler arrange the size of real time and non real time data
traffic as multiples of fixed size called buckets so that they can
be easily packed into DL subframe. But this gives less
freedom to the scheduler [26]. Greedy Scheduling Algorithm
(GSA) introduced by Anatolij et. al.[27] is one of few
algorithms that address efficiency, flexibility, QoS,
interference management and computational load in both
online and offline architecture for DL packing problem. The
algorithm consist of stages each stage concern with one or
more performance factors to optimize the overall performance.
But the complexity of packing stage is $O(n^4)$ which is quite
high.

- Other related work.

Caretti et. al. [28]. Focus on resources consumed by edge SSs.
Edge SSs experience poor channel quality, due to severe path-
loss and high interference from concurrent transmissions in
nearby cells. The authors propose partitioning the sub-carriers
of an OFDMA MAC frame into logical bands with different
power levels to provide edge SSs with extra power to use
more efficient modulations, thus limiting the wastage of
channel capacity while improving the quality perceived by
dges users. But allocating extra power may not solve the
problem. In addition to previous designing factors Joo-Young
et. al. Apply power boosting to improve downlink capacity and
propose a heuristic burst construction algorithm [29] that
constructs bursts in a manner that reduces both resource
wastage and resources usage using power boosting. Yuan-
Cheng et. al. propose search for best corner algorithms
(BCO)[30]. The algorithm avoids request fragments, assign
some extra slot and reduce the unused slot by placing the
request in the best corner in the two-dimension subframe.

B. Some considerations and design factors for downlink
packing algorithm.

- Allocation vertically minimizes number of time slot
  which leads to minimizes SS power consumption but
  it might leads to an efficient utilizing of subchannels.

- Sorting the request gives more efficient backing but it
  might affect the QoS priority.

- Considering the request QoS priority in allocation
  convert the problem from two-dimensional packing
to knapsack problem where each request has size and
value which more complicated.

- Grouping the request minimizes MAP overhead but it
  increase SS power consumption.
• Allowing fragment simplifies the packing but it increases the MAP overhead.
• Full search packing algorithms gives efficient packing but its complexity major problem.
• Considering power boosting or apply optimal block in allocation leads to minimizes the number of subchannel that increase number of symbols and SS power consumption.

V. CONCLUSION

In this survey we study the downlink data packing algorithms for OFDMA WiMAX based IEEE802.16. Process of efficient designing downlink data packing algorithm involve steps that effect, Channel efficiency, Scalability, QoS, SS Power consummation, and System Capacity. Also cause resources wastage in several forms. Unfortunately most of these factors contradict with each other and most of proposed algorithm try to improve some factors at the cost of scaring other factors like, grouping requests with same MCS to minimize DL-MAP overhead at the cost of SS power consumptions or minimize the wastage slots by extensive search for best order and best factor to fit the data at the cost of complexity. That is why it considers as trade off issues which require optimization between those factors to improve the system performance. Many algorithms have been proposed but there is lack of comprehensive comparative analysis of these algorithms under different traffic classes.

REFERENCES


With the submission of this manuscript I would like to undertake that the above mentioned manuscript has not been published elsewhere, accepted for publication elsewhere and that my Institute’s (Dept. of Electronics & Communication, Jamia Millia Islamia) representative is fully aware of this submission.