Analysis of Scheduling Strategies for Non-Real-Time Class in IEEE 802.16 Networks

Pallavi Grover
M.E Research Scholar/Department of CSE
UIET Panjab University, Chandigarh, India

Akashdeep
Assistant Professor/Department of CSE
UIET Panjab University, Chandigarh, India

***Abstract*— IEEE 802.16 standard is designed to provide services to various types of multimedia applications. It supports five different service classes with diverse quality of service (QoS) requirements .Real-time services such as internet protocol TV, video conferencing, online gaming etc. are always given higher priority by scheduler whereas non-real-time traffic which amounts to majority of the internet traffic is always neglected**

.With such a large volume of traffic, there is a need to develop a dedicated strategy for non-real-time service class. There have been very limited studies in this regard. The aim of this paper is to present survey of scheduling techniques for non-real-time connections in IEEE 802.16 networks.

Index terms - IEEE 802.16 networks, nrtPS, QoS.

I. INTRODUCTION

In recent years, development of wireless technology has taken place at a rapid pace particularly in the field of broadband wireless networks. WiMAX [1] is intended to offer low-cost, high-speed internet access to wide variety of devices. The advantages of WiMAX network are high transmission speed, scalable bandwidth, link layer retransmission, robust security, high peak data rates and an efficient QoS mechanism for data, voice and video. It is very challenging task to assure QoS requirements of every type of traffic in wireless network. These challenges are generated from limitations of wireless networks such as strong attenuation with increasing distance, Rayleigh fading, limited scalability etc. The protocol stack of IEEE 802.16 standard defines characteristics of physical layer as well as media access control layer (MAC) layer. In case of physical layer, standard provides usage of a single carrier physical layer called MAN-SCa, OFDM based physical layer called Wireless MAN-OFDM and an OFDMA based physical layer called Wireless OFDMA. Similarly for MAC layer standard can have different options for duplexing such as (TDD and FDD), multiplexing (burst TDM/TDMA/OFDMA), architecture (Point-to-multipoint, mesh) and frequency band.

WiMAX has defined five service classes: unsolicited grant service (UGS), real-time polling service (rtPS), non-real-time polling service (nrtPS), best-effort service (BE) and extended real-time polling service (ertPS). Each of them is associated with a certain set of QoS parameters such as delay, throughput and jitter. IEEE 802.16 standard does not specify any scheduling algorithm. It is left on the vendors to employ their own scheduling algorithm i.e. how allocation of resources should be done by base station (BS) to subscriber station (SS).

The working architecture of IEEE 802.16 networks is composed of 2 components: BS (base station) and SS (subscriber station). Base station is core element that acts as an interface between infrastructure network such as internet service provider (ISP) with SS. SS then further extend these internet services to its users. SSs connect with BS with an intend to provide differentiated services to its users.

II. RELATED WORK

The IEEE 802.16 standard operates in two modes: Point-to-multipoint (PMP) and mesh mode. PMP is a centralized topology in which BS has responsibility to transmit and coordinate all the traffic between SSs. Any SS who wants to send data to another SS has to communicate with the BS and in the mesh mode all SSs are peers and they can act as a router to relay packets of the neighboring peers. SSs can communicate with each other without the involvement of BS. The scheduling algorithm determines distribution of bandwidth among different users and in what order they will send. There are several ways to classify scheduling mechanisms. Broadly we can classify the scheduling algorithms in three categories: conventional, cross layer and dynamic scheduling algorithms.

A. Conventional algorithms

The literature survey of non-real-time services can be traced back to conventional algorithms such as proportional fairness and modified largest weighted delay first algorithms. Proportional fairness was originally designed for downlink traffic to increase the throughput of the system as well as to provide fairness among multiple queues. It calculates a priority function which is ratio of current rate to average rate and then schedules different queues accordingly. It is very simple and efficient but it does not take into consideration saturated queues of non-real-time traffic.

T. Kim et al. [2] have presented a scheduling policy which is an extension of proportional fair scheduling to satisfy quality of service requirements. This technique adapts scheduling parameters according to number of connections in the network. It first calculates a metric value for each connection which is ratio of rate requested by SS and average rate received by SS.Then average rate is updated according to window size. Parameters to be considered during scheduling are queuing delay, queue length and an operational parameter that signifies intensity of delay requirements. In every frame, selection of connection with higher metric value is done by the scheduler. Moreover admission control procedure is implemented by the means of scheduling ratio estimation.

Further Dong-Hoi Kim et al. [3] proposed modifications in existing proportional fair scheduling algorithm for non-real-time service in which standard serves the requirement of minimum bit rate. It's main objective is to determine a priority metric that will increase the number of non-real-time users to whom service is provided. It has integrated per sub channel bit rate into priority metrics. It has made a significant increase in outage performance but overall throughput is equivalent to existing algorithms.

Jeong et al. [4] presented a packet scheduling algorithm for non-real-time services which have soft QoS requirements i.e QoS requirements can be degraded to a little extent presenting a trade-off between packet delay and throughput requirements. In this additional delay is introduced for the users that are under bad channel conditions and scheduler keeps on serving SSs which are in good channel conditions untill QoS degradation is accepted by non-real-time service users. The proposed technique aims at maximizing system throughput but it shows degradation in outage performance of individual user.

B. Cross layer scheduling algorithms

Cross layer algorithms are those which use information of top or lower layers while taking scheduling decision. For example: Scheduling at MAC layer is done with the help of information received from physical layer. A few of them are listed as follows:

Qingwen Liu et al. [5] proposed a cross layer scheduling algorithm providing QoS support in IEEE 802.16 networks. A priority is assigned to each connection which is updated dynamically on basis of wireless channel quality and QoS satisfaction. For each nrtPS connection minimum reserved rate (ŋ) is defined. The proposed algorithm ensures average transmission rate should be greater than ŋ. Scheduling of nrtPS connections is done with the help of priority function which is dependent on nrtPS-class coefficient and rate satisfaction indicator. Rate satisfaction indicator is ratio of average transmission rate over minimum reserved rate. If value of indicator is greater than 1 then requirement is satisfied otherwise packets should be sent as soon as possible. This scheduler offers flexibility, scalability and low implementation complexity. The major drawbacks include fairness issue among same service class and imperfect channel condition arising due to error in estimation and feedback latency.

Authors Fen Hou et al. [6] presented a simple scheduling structure for non-real-time services in IEEE 802.16 networks. It is a cross layer algorithm which considers selective automatic repeat request mechanism at MAC layer and adaptive modulation and coding scheme at the physical layer. It tries to ensure minimum throughput requirements of nrtPS class and at the same time maintains flexibility between resource allocation and packet scheduling. To achieve this flexibility two parameters m and n are defined where m represents number of SSs selected in each MAC frame and n represents bandwidth which is granted to SS when it is being serviced. In the beginning of each frame m SSs which have superior channel conditions are selected and n amount of resources are given to them. When the value of m=1 then it can be opportunistic scheduling leading to maximum resource utilization and when m= total number of SSs then it leads to minimum resource utilization but lesser delivery delays. This paper focused on scheduling unicast nrtPS applications but there is dire need to concentrate on multicast multimedia applications.

Authors Subramanyam Y et al. [7] studied different uplink scheduling algorithms giving more emphasis on Latest time limit first with reserved bandwidth (LTL-RB). In this algorithm the uplink scheduler consists of three modules: information module, database module and service assignment module. Information module first categorizes packets on mobile station (MS) basis and then extracts information related to queue size and then pass on this information to scheduling database module. This module provides information to all MSs in the network and then service assignment module in which authors determine uplink subframe allocation in terms of number of bits per SS which are then transformed to number of time slots i.e unit of information element used in UL-MAP. LTL-RB algorithm reserves minimum amount of bandwidth for each service class during each frame time and distributes surplus bandwidth among different classed in high to low priority: UGS, rtPS, ertPS, nrtPS and BE. One more cross-layer scheduling architecture is proposed in which we have channel state information(CSI) module and queue state information module (QSI).CSI represents channel fading states between BS and k^{th} user and QSI represents number of untransmitted packets in the buffer of the user and hence the packets are scheduled using this information. This algorithm results in fairness, high frame utilization and lesser average delay.

Hua Wang et al. [8] proposed a downlink QoS scheduling mechanism in OFDMA based WiMAX networks. It consists of two modules: dynamic resource allocator module (DRA) and connection admission control (CAC) module. DRA module helps in efficient utilization of radio resources maintaining guaranteed QoS parameters of all the connections at the same time .The algorithm takes into consideration three main aspects: backlogged traffic, average modulation efficiency and satisfying QoS. It first evaluates bandwidth required on basis of backlogged traffic and average modulation efficiency rate. Then this estimation is further increased or decreased according to levels of QoS satisfaction. For scheduling of nrtPS connections a token bucket is associated with each queue. Tokens keep on arriving at a constant rate which serves minimum throughput requirement. After scheduling of PDU, we deduct these tokens from token queue. It also defines a binary random variable that indicates subchannel allocation and another variable which denotes time slots acquired on this subchannel.

Jinchang Lu et al. [9] combined cross layer design with opportunistic scheduling to serve QoS requirements in WiMAX point-to-multipoint networks. Authors have named proposed algorithm as Holistic Opportunistic Scheduling (HOS). It takes queue, channel and traffic QoS characteristics into consideration. In this scheme scheduling is a two-step process. First stage is performed at each SS in which scheduling priority is set for each individual packet.It is determined through dynamic priority index(DPI), channel specification index (CSI), normalized time delay satisfaction index(NTDSI) and normalized predictive starvation index(NPSI).After the calculation of scheduling priorities they are arranged in descending order. SSs will pick maximum value of scheduling priority (Max_SP) for competing with other SSs in order of uplink transmission. The second stage of scheduling is done at BS which sorts Max_SP(i) received from each SS in descending order. It then selects the one with highest value of Max_SP and allocates time slots according to its BW_request. After this scheduler at SSs extracts the information of allocated time slots and then distributes it among packets from highest to lowest priority.

C. Dynamic schedulers

Jin-Yup Hwang et al. [10] proposed an adaptive scheduling technique in which traffic is classified into 2 classes: real-time (VoIP and NRTV) and non-real-time (FTP).The algorithm assigns priority to a traffic class group and to all SSs under that group. Priority is calculated using type of traffic and maximum allowable delay time. The scheduling of real-time traffic is done through round robin (RR) whereas that of nrtPS through proportional fair scheduler (PF). Real-time traffic is preferentially allocated over non-real-time traffic. The major drawback of proposed algorithm is that it decreases the data rate of non-real-time traffic to satisfy the delay requirements of real-time traffic.

Ruangchaijatupon et al. [11] introduced fairness in the concept of adaptive scheduling as discussed above. It is a centralized scheme which consists of two modules: admission control and scheduling module. Admission control module checks whether to accept a connection or reject it on basis of minimum reserved traffic rate Scheduling module combines priority queuing and deficit based scheduling. It classifies traffic into three priority queues: first queue is for UGS, second for ertPS, rtPS and nrtPS and third for BE. The second priority queue which includes nrtPS is our main consideration. It is scheduled on the basis of round robin with an adaptive quantum. Value of quantum is calculated on basis of current queue size and capacity of channel. This algorithm ensures fairness and equally low delay to all traffic classes

K.R. Raghu et al. [12] proposed a queue length based scheduling for real-time as well as non-real-time traffic. The scheduling is done on the basis of number of MPDUs present at the start of uplink subframe. Adaptability of this algorithm is implemented by defining a parameter α which is ratio of maximum time nrtPS MPDU can wait in queue to maximum latency of real-time flow. α is a design parameter which controls QoS given to real-time and non-real-time traffic. This algorithm helps in providing excess resources to non-real-time packets and also takes queue length into consideration while guaranteeing QoS.

Mohammad Fathi et al. [13] proposed a combination of dynamic scheduling and CAC scheme. It distinguishes among different service classes by assigning a priority function. It is calculated with the help of service weights and arrival rate. We use priority function as basis for allocating bandwidth. The proposed algorithm is implemented in two stages. In first stage each service queue is associated with priority weights as defined by standard in order of rtPS>nrtPS >BE and bandwidth to be allocated is defined in terms of priority function which includes packet dropping probability ,average arrival and departure rates. In second stage it obtains priority function from stage 1 and then uses local schedulers to transmit packets to individual classes. Authors have used weighted fair queuing (WFQ) for scheduling nrtPS class. Usage of appropriate schedulers for different service classes leads to better network performance. Proposed algorithm takes connection arrival time into consideration while allocating bandwidth. This helps to avoid network instability.

D. Optimization techniques

Ali Mohammed Alsahag et al. [14] proposed fuzzy based adaptive deficit round robin uplink scheduler that adjusts weights of service queue for real-time as well as non-real-time applications. The allocation of bandwidth is done on the basis of deadline based scheme. To compute the deadline authors have used maximum latency for real-time and throughput for non-real-time as input variables. The main aim of the paper is to ensure fairness and use the system resources efficiently. The overall mechanism can be divided in three fundamental phases: fuzzification, fuzzy inference and defuzzification. In fuzzification process we use two input variables as real-time maximum latency (RT_{ML}) and non-real-time throughput (NRT_{THR}). These input variables are processed with the help of a rule base in fuzzy inference phase and then finally in defuzzification phase crisp numerical values are obtained which determines weights to be used as an indication for priority. In bandwidth assignment process, several queues are maintained which are associated with a DC value. In each round DC is incremented by α value which is determined by fuzzy system keeping in consideration overall capacity of the system. Transmission of queue takes place when DC value is equal to amount of requested bandwidth. The proposed algorithm optimizes the overall system utilization but gives more consideration on satisfying maximum latency requirements of real-time traffic.

D.David Neels Pon Kumar et al. [15] proposed a neural network based fuzzy priority scheduling algorithm. Fuzzy is used to calculate priority which itself is composed of primary fuzzy scheduler and dynamic fuzzy scheduler. Primary fuzzy scheduler takes as input Expiry time (E), Waiting time (W), Queue length (Q), Packet size (P) and gives priority index as output. Then this priority index is feed as an input to dynamic fuzzy scheduler which calculates final priority by taking type of service into consideration. Now scheduling is done using artificial neural network based on prioritized input received from DPFS (dynamic priority fuzzy scheduler).Artificial neural network (ANN) consists of three layers: input layer, modified Kohonen layer and Grossberg layer. The input layer processes prioritized output received from DPFS and arrange them in order of their priority .Now the output is fed to modified kohonen layer which checks whether the value is in the range of threshold. If it is so it is given as an input to grossberg layer otherwise rejected. This algorithm improved fairness and prevents the starvation of low priority traffic but it has not considered bursty traffic conditions. The major drawback is that it has increased processing time.

Dusit Niyato et al. [16] presented survey of game theory techniques for management of radio resources in different wireless networks and also proposed bandwidth allocation and connection admission control scheme for IEEE 802.16 networks. Bandwidth allocation is described with the help of a non-cooperative game in which three parameters are considered: players, strategies and pay off's. Players refer to rtPS and nrtPS connection, strategies implies amount of bandwidth to be allocated to a new connection and payoff's refer to the total utility of currently running rtPS and nrtPS connections plus utility of new connections. The solution of game is provided with the help of Nash equilibrium which is calculated through best response function. It maximizes the payoff of BS which defines QoS requirements of connections. The simulation results of adaptive scheme shows that it is not able to satisfy delay and throughput necessities when load is high.

Dusit Niyato et al. [17] also proposed a non cooperative game for allocation of bandwidth and entry into the network. The players include BS and SS which needs services from BS. Whenever a new connection requests services from BS it calls the proposed algorithm. The new incoming connection tells BS about traffic class (rtPS, nrtPS, BE) and QoS requirements. Then, BS makes strategies and calculates expected pay off for each strategy. It calculates nash equilibrium and decides about accepting or rejecting the connection. If it accepts it also makes decision of how much bandwidth should be allocated to new connection. The main advantage of this technique is that it helps in achieving desired levels of QoS satisfaction.

III. Conclusion and Future Scope

This paper has presented scheduling strategies for nrtPS connections in IEEE 802.16 networks. Different scheduling strategies can be classified into three categories: conventional, cross layer and dynamic schedulers. In addition to these optimization techniques for scheduling such as neural networks, neuro fuzzy based scheduling and game theory based techniques are also explored. Survey of these theories shows that scheduling approaches for nrtPS connections are still in growing stages of development and there exists a lot of scope for further improvement. A lot of techniques can be applied for improvement of nrtPS traffic class. One such technique can base its scheduling decision on the basis of queue length of nrtPS class. The algorithm works by determining all nrtPS bandwidth requests and storing these requests in a virtual queue at the BS. Bandwidth requests are sorted by the lowest queue length and a counter is assigned to each virtual queue. Once the allocation is done, the algorithm verifies if bandwidth request made by connection was satisfied or not. If not, the algorithm checks availability of more symbols to be allocated. If there are more symbols, these symbols will be allocated to the connection and its assigned counter will be decreased. Then algorithm will again schedule the connections based on the lowest queue length and the lowest counter. Proposed algorithm tries its best to avoid starvation of nrtPS connections.

References

[1]. IEEE Std. 802.16n 2013. IEEE standard for local and metropolitan area networks Part 16: Air interface for fixed and mobile broadband wireless access systems—amendment 2: Higher Reliability Networks. IEEE Computer Society and IEEE Microwave Theory and Techniques Society. p. 1–132. ISBN 978-0-7381-8425-8, 2013.

[2]. Kim T, Lim J.T. Quality of service supporting downlink scheduling scheme in worldwide interoperability for microwave access wireless access systems, IEEE Journal of Institute of engineering and technology, communications 2010; 4(1):32-38.

[3]. Kim D.H, Ryu B.H, Kang C.G. Packet Scheduling Algorithm Considering a Minimum Bit Rate for Non-real-time Traffic in an OFDMA/FDD-Based Mobile Internet Access System, Journal of Electronics and Telecommunications Research Institute 2004; 26(1): 48-52.

[4]. Lee JH, Kim S, Kang CG, Lee SC, Tcha Y. Opportunistic packet scheduling algorithm for non-real-time service with a soft QoS requirement in a mobile broadband wireless access system, Journal of Telecommunication Systems 2008; 39(1): 15–24.

[5]. Liu Q, Wang X, Giannakis G.B. A Cross-Layer Scheduling Algorithm With QoS Support in Wireless Networks, IEEE

Transactions on vehicular technology 2006; 55(3), 839 – 847.

[6]. Hou F, Ho P.H. A Flexible Resource Allocation and Scheduling Framework for Non-real-time Polling Service in IEEE 802.16 Networks, IEEE transactions on Wireless Communications 2009; 8(2): 766 – 775.

[7]. Y S, Y V. WiMAX Base Station Scheduling Algorithms, white paper, Tata Consultancy Services,2009.

[8]. Wang H, Dittmann L. Downlink resource management for QoS scheduling in IEEE 802.16 WiMAX networks, Journal of Computer Communications 2010; 33(8):940-953.

[9]. Lu J, Ma M. Cross-layer QoS support framework and holistic opportunistic scheduling for QoS in single carrier WiMAX system, Journal of network and computer applications 2011; 34(2):765-773.

[10]. Hwang JY, Han Y. An adaptive traffic allocation scheduling for mobile WiMAX. In: Proceedings of IEEE 18th international symposium on personal, indoor and mobile radio communications, 2007. PIMRC 2007. Athens. p. 1–5.

[11]. Ruangchaijatupon N, Ji Y. Adaptive scheduling with fairness in 802.16e networks. In: Proceedings of international conference on wireless communications, networking and mobile computing, Shangai, 21–25 September 2007. p. 1976–1979.

[12]. Raghu KR, Bose S. Queue based scheduling for IEEE 802.16 wireless broadband. In: Proceedings of 6th international conference on information, communications & signal processing, 10–13 December, 2008 Singapore. p. 1– 5.

[13]. Fathi M, Rashidi B, Samaneh, Alireza K, Taheri Hassan H. A dynamic joint scheduling and call admission control scheme for 802.16 networks, Journal of Telecommunication Systems 2013; 52(1): 195-202.

[14]. Alsahag A.M, Ali B.M, Noordin N.K, Mohamad H. Fair uplink bandwidth allocation and latency guarantee for mobile WiMAX using fuzzy adaptive deficit round robin, Journal of network and computer applications 2014;39:17-25.

[15]. Kumar D.D.N.P, Murugesan K, Raghavan S, Suganthi M. Neural Network based Scheduling Algorithm for WiMAX with improved QoS Constraints In: Proceedings of international conference on Emerging trends in electrical and computer technology, 23-24 March, 2011,Tamil Nadu. p. 1076 – 1081.

[16]. Niyato D, Hossain E. Radio resource management games in wireless networks: an approach to bandwidth allocation and admission control for polling service in IEEE 802.16, IEEE Journal of Wireless Communications 2007;14(1):27–35, Feb.

[17]. Niyato D, Hossain E. QoS aware bandwidth allocation and admission control in IEEE 802.16 broadband wireless access networks: a non-cooperative Game theoretic approach, Journal of Computer Networks 2007;51(11):3305–21.

Authors Profile

Pallavi Grover received the **B.Tech.** degree in Computer Science and Engineering from the D.A.V Institute of Engineering and Technology, Jalandhar, Punjab Technical University, Punjab, India, in 2012.Currently doing **M.E.** in Computer Science and Engineering in UIET, Panjab University, Chandigarh, India. Her research interest includes wireless communication (**WiMax**), QoS in WiMAX, Communication networks.

Akashdeep received the **B.Tech.** degree in Computer Science and Engineering from Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib,Punjab Technical University, Punjab, India and has done **M.Tech.** in Computer Science and Engineering from Punjabi University, Patiala.Presently, he is perusing **Ph.D**. from Guru Nanak Dev University, Amritsar and serving as an Assistant professor in UIET, Panjab University,Chandigarh,India. His research interest includes WiMAX, Soft computing, Neural Networks and fuzzy logic in WiMAX.