Journal of Network and Computer Applications 39 (2014) 17–25

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Fair uplink bandwidth allocation and latency guarantee for mobile WiMAX using fuzzy adaptive deficit round robin

Ali Mohammed Alsahag [a,*], Borhanuddin Mohd Ali [a], Nor Kamariah Noordin [a], Hafizal Mohamad [b]

[a] Department of Computer and Communications Systems Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[b] Wireless Communications, MIMOS Berhad, Technology Park Malaysia, 57000 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 23 September 2012
Received in revised form
3 March 2013
Accepted 1 April 2013
Available online 18 April 2013

Keywords:
Bandwidth allocation
IEEE 802.16
Fuzzy
QoS
Uplink

ABSTRACT

The explosive demands of rich media applications with their diverse quality of service (QoS) requirements have continuously fuelled the needs for ever more powerful networks. One example of such a network is called WiMAX which is driven by WiMAX Forum based on IEEE 802.16 Wireless MAN standard. One of the issues that still remain open in WiMAX is the scheduling algorithm that goes to meet the QoS requirements. However, QoS provisioning of real-time and non real-time applications are frequently unstable due to insufficient allocation of bandwidth, which leads to degradation in latency guarantee and deterioration of overall system utilization. In this paper, an efficient bandwidth allocation algorithm for the uplink traffic in mobile WiMAX is proposed. Using intelligent systems approach upon the traffic service class information served by the base station (BS), an adaptive deadline-based scheme is designed. The scheme is fully dynamic to guarantee a specific maximum latency for real-time applications, besides improving fairness and throughput, giving due considerations to non real-time applications. The algorithm uses fuzzy logic control which is embedded in the scheduler; its function is to control and dynamically update the bandwidth required by the various service classes according to their respective priorities, maximum latency and throughput. Simulation results show that the proposed algorithm manages to optimize the overall system utilization while at the same time guarantee the maximum latency requirements for real-time traffic.

1. Introduction

The growing popularity of wireless broadband services plus the explosive demands for a diversity of real-time applications such as voice-over-IP (VoIP), video streaming and gaming, have become key driving factors for the deployment of seamless and ubiquitous wireless access networks. The IEEE 802.16 standard for example, defines a broadband wireless access network for metropolitan area, also commercially known as WiMAX (IEEE, 2005, 2006). WiMAX was developed to meet the anticipated growth in the worldwide market for high bandwidth and real-time applications. However, IEEE 802.16 standard does not specify a bandwidth allocation algorithm to guarantee QoS, this is purposely done in order to allow service providers and vendors to innovate in this area and distinguish their products.

In wireless broadband access networks, the channels have to transport a wide variety of multimedia applications and this becomes challenging for network service providers to meet; this is due to the scarce wireless resources available to satisfy all traffic demands with diverse QoS requirements. In WiMAX, the bandwidth resource management is divided into downlink (DL) and uplink (UL) direction controlled by a BS. The service classes' parameters can be differentiated and prioritized in these traffic directions. However, real-time applications such as video conferencing and gaming require a higher bandwidth allocation, which increases delay and reduces efficiency of the overall system. Traffic scheduling and bandwidth management schemes are two key mechanisms in WiMAX that are used to support the required QoS. A number of scheduling algorithms have been proposed to deal with the QoS requirements for the various service class' applications in WiMAX, but it is still necessary to develop appropriate bandwidth allocation schemes to guarantee satisfactory QoS.

In this paper, a fair and efficient bandwidth allocation algorithm for mobile WiMAX in the uplink direction is presented. Towards that end, a fuzzy logic system is developed as an embedded system for a new deadline-aware bandwidth allocation. This is referred here as fuzzy based adaptive deficit round robin scheduling or FADRR. The deadline is controlled and computed based on the input variables, namely maximum latency for real-time traffic and throughput for the non real-time traffic of all service classes. Therefore, the objective of FADRR is to reduce

* Corresponding author. Tel.: +60 176994635.
E-mail addresses: alkazmi_jo@yahoo.com, ali.mansoooor@gmail.com (A.M. Alsahag).

1084-8045/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jnca.2013.04.004

negative impact on the QoS metrics and ensure optimal bandwidth for all service flows, and at the same time to maintain fairness and conserve system resources. To the best of our knowledge, this is the earliest application of embedded intelligent mechanism to manipulate the queue behaviour and grant optimal bandwidth to real-time traffic queues. This algorithm can dynamically distinguish the data of each service type queue; and it adaptively allocates adequate bandwidth for real-time service types, considering their deadline, and improves access for non real-time connections. The performance of the algorithm has been compared by means of extensive simulation scenarios against some representative schemes proposed in this area, namely modified deficit round robin (MDRR) (Cisco), and customized deficit round robin (CDRR) (Laias and Awan, 2010). The results indicate that this scheme manages to provide improvements over both MDRR and CDRR in terms of delay, jitter, throughput and fairness.

2. IEEE 802.16 concept and overview

IEEE 802.16 is a wireless broadband standard that offers higher data rate over a metropolitan area of up to 70 km (IEEE, 2010). The standard identifies the frequency band, medium access control (MAC) and physical (PHY) layers for wireless broadband access. It uses orthogonal frequency division multiplexing access (OFDMA) as the multiplexing technology that distributes the bandwidth into numerous frequency sub-carriers. It manipulates the frequency range of the channel by the modulation code and modulate the information to the sub-carriers before transmission (Bacioccola et al., 2010). The available resources in OFDMA are known in the time domain (TD) as symbols and in the frequency domain (FD) as sub-carriers, which integrate into a sub-channel system. OFDM is used to provide the multiplexing of user's data streams on both uplink and downlink transmissions whose scope is beyond the scope of this paper.

To support QoS in wireless broadband networks, the MAC layer has to size up the downlink and uplink traffic. The received flows at the MAC layer are classified and related with the corresponding service class. Five service classes have been defined in IEEE 802.16. They are unsolicited grant service (UGS), extended real-time polling service (ertPS) and real-time polling service (rtPS), which known as real-time applications, while non real-time polling service (nrtPS) and best effort (BE) services support non real-time applications. The bandwidth allocation algorithm must ensure the QoS for different traffic types real-time and non real-time, at the same time utilizing efficiently the available bandwidth. While the DL scheduler in a BS easily allocate DL data to subscriber stations (SSs), the uplink traffic transmission relies on request and grant mechanism where the SSs can use the polling mechanism BW-REQ messages periodically to connect to the BS. Throughout the scheduling algorithm, the BS reserves the required bandwidth based on the number of slots required for each request in the uplink sub-frame. Along with all service classes, excluding the UGS and ertPS which are granted with a fixed bandwidth, rtPS must be taken into account due to the nature of variable packet sizes and also must guarantee the maximum-latency to satisfy the QoS requirement. The available bandwidth in the system is allocated to SSs based on the scheduling mechanism; this has the impact of satisfying the diverse QoS requirements while at the same time increase bandwidth utilization to improve the system capacity. The bandwidth allocation algorithms performed by the BS provides more significant end user satisfaction, therefore these algorithms must be designed and used cautiously to optimize the system utilization.

3. Related works

Scheduling and bandwidth allocation algorithms are two of the critical mechanisms to provide the required QoS in a packet network. A number of bandwidth allocation algorithms have been studied to overcome the drawbacks that affect service class applications in WiMAX. A comprehensive survey and taxonomy of scheduling in IEEE 802.16e WiMAX networks can be found in Msadaa et al. (2010). This section provides overview of some key scheduling algorithms that have been proposed for WiMAX networks.

In Shreedhar and Varghese (1996), an amendment to both round robin (RR) and weighted round robin (WRR) which is known as deficit round robin (DRR) was proposed, where the scheduling process requires only $O(1)$ complexity to process a packet in the queue in addition to the implementation simplicity, at hardware with a reasonably low cost. However, real-time packets will have to wait until the scheduler completes serving other queues in this round including BE packets, the effect of which is that real-time packets may miss their deadline.

The method in Wongthavarawat and Ganz (2003) proposed a two-tier hierarchical algorithm. In the first-tier deficit fair priority queuing (DFPQ) allocates the total available bandwidth for DL and UL traffic, while at the second tier, it handles different queues by different conventional algorithms. The drawback is that DFPQ does not guarantee QoS for real-time services.

A scheduling algorithm for the rtPS was proposed in Ball et al. (2005). This algorithm manipulates a scheduling list that contains all the SSs that can be served at the next frame. However, the algorithm specifies that the SSs that has low transmission quality is suspended temporarily from the transmission list for a period of time. This mechanism is repeated periodically for all SSs. If the transmission quality is still low, the scheduler grants another suspended period of time.

In Ball et al. (2006), RR scheduler was investigated, whereby it allocates the available resources to all the SSs in a round-robin fashion without giving due priority to real-time applications. This algorithm is uncomplicated and simply manipulates the available resources among the SSs. Therefore it is considered inappropriate for traffic with diverse characteristics and QoS requirements. In order to distinguish the required bandwidth for each queue the authors in Cicconetti et al. (2006) handled the traffic in two ways; first, for traffic that does not require QoS, and next for traffic that do require QoS. However, fairness in this case was not considered. In Rath et al. (2006), the opportunistic deficit round robin (O-DRR) scheduler proposed an analytical method for getting an optimal polling interval for uplink data flow via the polling interval mechanism, the BS polls service flow periodically to make sure that the traffic delays are achieved. The system considering several situations, for instance, the SSs must ensure that the queue should not be empty as well as the receive SNR must exceed the threshold value. However, the allocation mechanism of the O-DRR algorithm leads to an additional overhead at the BS because it requires the manipulation of quantum size and a deficit count for each SSs, repeatedly. A modification to DRR known as modified deficit round robin (MDRR) was proposed in Cisco. It implements a quantum ϕ and a deficit counter (DC) for each service type queue. In each round the scheduler assigns the service type queue by DC value that is added by the value in every round; the scheduler transmits the packets till the DC empties or when the packet queue length is greater than the DC, and then move to the next queue. However, real-time packets will experience severe delay when the traffic in the system is heavy.

A latency and modulation aware bandwidth allocation algorithm called highest urgency first (HUF) was proposed in Lin et al. (2009). While HUF translates the data bytes into slots to investigate the effect from various adaptive modulation and coding scheme (MCS), it does not take full benefit from the MCS variations. Additionally, with HUF the request is discarded whenever its deadline is less than a frame duration, which is treated as a violation of maximum latency requirement.