Computer Networks 91 (2015) 654–674

Contents lists available at ScienceDirect

Computer Networks

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

A three level LTE downlink scheduling framework for RT VBR traffic

Satish Kumar*, Arnab Sarkar, Santhosh Sriram, Arijit Sur

Department of Computer Science and Engineering, Indian Institute of Technology, Guwahati, India

ARTICLE INFO

Article history:
Received 30 December 2014
Revised 30 July 2015
Accepted 27 August 2015
Available online 11 September 2015

Keywords:
LTE
QoS awareness
Low overhead scheduling
Overload handling
Radio resource management
RT VBR services

ABSTRACT

In recent literature, it has been observed that obtaining low overhead LTE downlink scheduling mechanisms that ensure a descent QoS is a challenging task especially in the presence of transient overloads. This paper proposes a novel three level LTE downlink scheduling framework (called *TLS*). The three major objectives of the proposed framework includes low overall traffic flow selection and mapping overheads at each Transmission Time Interval (TTI), maintaining acceptable QoS levels for various real-time (RT) variable bit rate (VBR) flows even during transient overloads through graceful degradation, and maximizing spectral efficiency by effectively exploiting multi-user diversity. Both the outermost and the second level of TLS employs a novel adjustable frame-based approach to provide efficient service even in the face of ever changing wireless channel conditions, system workloads, traffic bit rates etc. A low-overhead $O(1)$ look-up based scheme is used in the innermost layer of the framework to physically allocate bandwidth resources to the traffic flows at each TTI. Simulation results clearly reveal that TLS outperforms the existing schedulers with respect to QoS, goodput and spectral efficiency.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Recent advances in wireless broadband technology have spurred an ever increasing demand for diverse data rate traffic flows with varied quality of Service (QoS) requirements ranging from real-time (RT) voice over IP (VoIP), streaming video / audio, online gaming etc. to soft real-time flows like telnet, web-browsing etc. and even non-real-time best-effort data downloads. To meet the challenges imposed by such demands, Long Term Evolution (LTE) systems [1] have been introduced with an important technology shift from circuit switched networks to an all-IP network architecture.

LTE systems provide an elaborate packet scheduling infrastructure which if efficiently harnessed has the ability to support a wide variety of high data rate multimedia and Internet services even in high mobility scenarios. Founded on the Orthogonal Frequency Division Multiplexing (OFDM) mechanism, LTE allows its total radio resource bandwidth (BW) (upto 20 MHz according to LTE release 8) to be simultaneously frequency multiplexed into a fixed number of sub-channels (up to 100 sub-channels of 180 KHz each), each of which may be further time multiplexed at very fine granularities of a Transmission Time Interval (TTI) (1 TTI = 1 ms). A TTI is made up of two time slots of length 0.5 ms. A time/frequency radio resource spanning over one time slot in the time domain and over one sub-channel in the frequency domain is called a Resource Block (RB) and corresponds to the smallest radio resource unit that may be assigned to an user equipment (UE) for data transmission. Along with this, LTE is equipped with features like Channel Quality Indicator (CQI) reporting, link adaptation through Adaptive Modulation and Coding (AMC) and Hybrid Automatic Retransmission Request (HARQ) to support better QoS, low latencies and improved spectral efficiency.

* Corresponding author. Tel.: +918876815314.
E-mail addresses: satish.kr@iitg.ernet.in (S. Kumar), arnabsarkar@iitg.ernet.in (A. Sarkar), s.santhosh@iitg.ernet.in (S. Sriram), arijit@iitg.ernet.in (A. Sur).

http://dx.doi.org/10.1016/j.comnet.2015.08.027
1389-1286/© 2015 Elsevier B.V. All rights reserved.

However, an on-line resource allocation policy which attempts to optimally utilize such an elaborate radio resource allocation infrastructure will pose a prohibitive computational burden at eNodeB (an abbreviation for evolved-NodeB, which refers to the base station that handles the Evolved Universal Terrestrial Radio Access (E-UTRA) mechanism of LTE). Therefore, low-overhead downlink scheduling methodologies that intelligently leverages the attractive features of LTE systems must be designed to effectively serve a variety of heterogeneous UEs (mobile phones, laptops, tablets etc.) such that QoS demands of all flows can be met. Furthermore, these algorithms should simultaneously satisfy other practical constraints / objectives like maximizing spectral efficiency in the face of ever changing wireless channel conditions, handling bursty traffic, dynamic adaptation of QoS for real-time flows in times of transient overloads etc.

The two most common traditional downlink scheduling approaches are Maximum throughput (MT) (which selects flows experiencing the best instantaneous channel conditions) and Proportional Fair (PF) [2] (selects flows with the least running average throughput at a given instant). Although, both are QoS unaware and hence, may not directly be applicable for today's multimedia and Internet applications, most of the recent QoS aware schemes [3] employ them to better cell spectral efficiencies and fairness between flows. Authors in [4] proposed a two level resource allocation strategy where the outer level consists of a fixed time window called β over an inner TTI level. Resource allocation is done at successive window junctions by allocating the best available flows (obtained by ordering their decision index values) on a round-robin basis over the sub-channels progressing TTI by TTI until all RBs of all TTIs in β are allocated. There has been several recent contributions [5–7] which prioritize flows primarily based on head-of-line packet delays. In [5], a two step per TTI flow scheduling and mapping scheme has been proposed. At the first step, RBs are allocated to those flows whose packet waiting times are nearing their transmission deadlines. Throughput enhancement for the residual RBs is done at the second step. Yang et al. in [6] formulate the resource allocation problem as a mixed integer nonlinear problem to minimize the average waiting time across all active flows. This problem is known to be NP-hard and therefore, it poses prohibitive computational budget at eNodeB during on-line RB allocation. Further, they have designed a four step heuristic approach as a cost effective sub-optimal solution to the problem. In [8], Wang et al. presents a cross time interference aware evolutionary scheduling algorithm that fairly allocates resources based on an adaptive fairness threshold.

Many schemes that attempt to handle delay sensitive applications have been proposed in [9–11]. Among them, two notable algorithms namely, the LOG rule [9] and EXP rule [10] consider all active flows (including both RT and non-RT flows) together at every TTI and selects flows based on their individual metric values. For non-RT flows, the metric values for both schedulers are obtained based on a proportional fair policy [2]. For RT flows, the metric value for the j^{th} flow on the r^{th} sub-channel is obtained as a trade-off between head-of-line packet delays and spectral efficiency as shown in Eq. 1a and 1b for LOG rule and EXP rule, respectively:

$$m_{j,r}^{LOG-Rule} = b_j log\Big(c + \alpha_j \times HOLD_j\Big) SE_r^j \tag{1a}$$

$$m_{j,r}^{EXP-Rule} = b_j exp\left(\frac{\alpha_j \times HOLD_j}{c + \sqrt{(1/N_{rt})\sum_p HOLD_p}}\right) SE_r^j \tag{1b}$$

where, b_j, c , α_j are tunable parameters, SE_r^j is the spectral efficiency for the j^{th} flow on the r^{th} sub-channel, $HOLD_j$ is the delay of head-of-line packet for the j^{th} flow and N_{rt} denote the total number of active RT flows. Analysis in [3] and also the experimental analysis done in this work, reveal that with its exponential nature, the RT metric values for EXP rule is in general more sensitive to packet delay urgencies as compared to LOG rule and therefore is more robust especially in underloaded situations. EXP rule also attempts to maintain fairness among RT flows by normalizing the delay of a flow over the square root of the mean delay of all RT flows [12].

However, both EXP rule and LOG rule suffer from three principal drawbacks. Firstly, by following a metric based selection policy that considers all active flows (combining both RT and non-RT flows) together at every TTI, these algorithms incur high scheduling complexities. Secondly, the same reason as above (consideration of both RT and non-RT flows together) creates possibilities where non-RT flows may be prioritized over urgent RT flows resulting in packet loss due to missed deadlines. Thirdly, in the absence of overload estimation and dynamic QoS adaptation mechanisms, these algorithms often perform poorly in transient overload situations.

There has been a few works which focus on two-layer resource allocation architectures [13–15] as heuristics towards lowering the complexity of the combined optimization problem of maximizing both QoS and spectral efficiency by splitting the problem and tackling it at distinct layers. In a recent work proposed in [11] (known as EXP-VT-SH), authors conceived an interesting two phase procedure based on cooperative game-theory that performs resource sharing combining the EXP-Rule with a virtual token mechanism.

Piro et al. in [13] propose a frame-based two level packet scheduling algorithm (called FLS) where the first level computes the amount of data to be transmitted by each RT flow (such that their delay constraints may be satisfied) in the next LTE frame (a fixed time interval of duration 10 ms) using discrete time linear control law. The inner level then physically allocates RBs to flows at each TTI using a PF scheduler. To prioritize RT VBR flows over non-RT flows, in the TTIs corresponding to an LTE frame, RBs are first allocated to all the active RT VBR flows until their total amount of data calculated for the entire frame has been transmitted. The remaining RBs are then allocated to the non-RT flows. It may be observed that by computing the total data transmission requirement of each RT flow for an entire frame, employing a PF scheduler for RT flows at each TTI and by distinctly prioritizing RT over non-RT flows, FLS attempts to guarantee strict packet delivery delay bounds for the RT flows. However, in this endeavor, it tends to neglect spectral efficiency and therefore often suffers low resource utilization. A more severe consequence of this drawback is that FLS' performance tends to

Download English Version:

https://daneshyari.com/en/article/450728

Download Persian Version:

https://daneshyari.com/article/450728

Daneshyari.com