Available online at www.sciencedirect.com

ScienceDirect

Computer Communications 31 (2008) 2257–2269

computer communications

www.elsevier.com/locate/comcom

Network utility maximization for triple-play services

Lei Shi *, Changbin Liu, Bin Liu

Department of Computer Science and Technology, Tsinghua University, Room 9-416, East Main Building, Beijing 100084, PR China

Received 22 April 2007; received in revised form 14 February 2008; accepted 15 February 2008
Available online 29 February 2008

Abstract

It is well known that Next-Generation Network (NGN) will inevitably carry triple-play services (i.e. voice, video and data) simultaneously. However, the traditional strict-priority based scheduling algorithm intensively used in current Internet cannot maximize the overall network utility for NGN, instead brings significant global welfare loss. In this paper, we study how to achieve Network Utility Maximization (NUM) in NGN running triple-play services. By investigating the characteristics of most of its traffic classes, we explicitly present their utilities as the function of allocated bandwidth. We further formulate the NUM objective as a nonlinear programming problem with both inequality and equality constraints. A solution using Lagrange Multiplier is given on the simplified problem with only equality constraints, which indicates the major distinction from strict-priority based scheduling, the existence of a turning point for IPTV users. Simulations are also carried out using LINGO on the original complicated problem. Several useful results are presented on the new features of the NUM-based scheduling. We also discuss the methods to alleviate the impact of turning point and the consequent unstable bandwidth allocation.

Keywords: Network utility maximization; Triple-play services; Next generation network; Bandwidth allocation

1. Introduction

The Internet has been evolving in recent years to adapt with the emerging abundant applications. Specifically, in the near future, the voice, video and data traffic (herein called triple-play services), which are previously forwarded by separate networks such as Public Switched Telephone Network (PSTN), the cable television network and the original Internet, will be carried on a single converged network, i.e. the Next Generation Network (NGN). NGN must be able to natively conduct triple-play services, which means that all traffic classes of voice, video and data should be managed to meet their particular Quality of Service (QoS) requirements, such as strict packet delay, jitter and loss guarantees. It is believed that the deployment of NGN and the provisioning of triple-play services will eventually not only benefit the Internet users with richer contents, but also increase ISP revenues by acquiring much higher per-subscriber profit.

While several Internet Service Providers (ISPs) have proposed their architecture and detailed specifications to support triple-plays in NGN, they all have to deal with a critical issue that how to schedule traffic and allocate bandwidth for triple-play services at both backbone networks and access links. Due to the efficiency consideration that NGN cannot be designed with over-provisioning technique to avoid congestion, more advanced congestion-phased traffic scheduling algorithms are essentially required to compromise the benefit of all the traffic classes. Designing such a scheduling (bandwidth allocation) algorithm is exactly the premier issue this paper tries to settle.

Prior to the study presented here, numerous related works have been published on this issue. In industry designing NGN [1,3], the strict-priority based scheduling algorithm is the most widely adopted one in carrying out bandwidth allocation. However, this solution rigidly favors the voice and

* Corresponding author. Tel.: +86 10 62773441.
E-mail address: shijim@gmail.com (L. Shi).

0140-3664/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2008.02.016

video traffic without flexibility, thus can only be deemed as a conservative method when no better one is available.

On the other hand, researches in academia mainly concentrated on utility-based solutions. Shenker [16] for the first time discussed traffic classifications in IP network from the viewpoint of user utility. He further investigated the characteristics of several well-known traffic classes, including TCP elastic traffic, UDP hard real-time traffic, delay-adaptive traffic, as well as rate-adaptive traffic. However, no explicit expression of utility function was given. Later, Kelly et al. presented some first works [11,12] in applying utility-based method from economics to the area of scheduling and bandwidth allocation in the objective of Network Utility Maximization (NUM). Significantly, he showed that both centralized and decentralized pricing algorithms were capable to achieve NUM. In [8], Dharwadkar et al. studied the utility functions from the point of their shapes. They categorized the utility functions into three general types: step, linear and concave, and based on these features proposed a heuristic scheduling algorithm that executed dynamic bandwidth allocation and achieved Network Utility Maximization (NUM). Zimmermann and Killat argued in [17] that the utility function stands for the user's preference of bandwidth, which can be modeled as an increasing, strictly concave, and continuously differentiable function, perfectly fitted by the logarithm function. The utility function of HTTP-like traffic class was studied in [7] by Chang et al. They derived a close-form expression for the utility function of HTTP traffic from the behavior of HTTP applications and underlying TCP connections. Harks and Poschwatta [10] proposed scheduling algorithms under the "utility fair" assumption, where bandwidth is allocated such that each user is offered with equalized utility to guarantee fairness. Massoulie and Roberts [15] generalized three objectives for bandwidth allocation in network links: max–min fairness, proportional fairness and minimum potential delay. They developed corresponding scheduling algorithms for each of them respectively.

Although the previous works contributes a lot in building up the basic theoretical framework of the utility-based scheduling and bandwidth allocation, as well as the concerning pricing strategy, at this time, no single work has emphasized on the practical issue of scheduling triple-play services under the background of NGN. Motivated by the desire to bridge such a gap between theory and reality, we work through this issue with the well-known NUM objective. By classifying NGN traffic into five categories according to their diversified utility functions, we explicitly formulate this issue into a nonlinear programming problem with both inequality and equality constraints. After some safe approximations, we further translate it into a nonlinear programming problem with only equality constraints, which can be solved accurately by adopting the well-known Lagrange Multiplier method.

We discuss the solution using our theoretical method in a simplified scenario where there only exist IPTV users and TCP elastic users. A new feature for the NUM-based scheduling is discovered that the IPTV users will face a turning point in bandwidth allocation, before which they actually gain no bandwidth at all.

Simulations on the original scheduling problem under two network scenarios are carried out using nonlinear programming software LINGO. Some bandwidth allocation results under NUM-based scheduling are observed: (1) In both network scenarios, the utilities of VoIP users and other low-throughput real-time UDP users are well guaranteed regardless of the network provisioning conditions, since they are the most cost-effective traffic for bandwidth allocation; (2) IPTV users give up all the bandwidth at first when network provisioning is below a turning point and after that step directly to nearly half of its maximal bandwidth requirement; (3) TCP elastic and interactive users are provisioned nearly proportionally except around IPTV user's turning point where the bandwidth allocation is rather unstable; (4) NUM-based scheduling in general achieves at least 25% utility gain over the strict-priority based scheduling.

We also discuss two measures to alleviate the impact of IPTV's turning point: one is to increase the penetration rate of IPTV service and the other is to elevate IPTV user's maximal utility. Through simulations, we find that the latter method is more practical under current Internet environment.

Compared with the strict-priority scheduling deployed extensively in industry, our results demonstrate that while offering highest strict priority for VoIP users is indeed the best choice, assigning IPTV users the second-highest strict priority actually does not accord well with the objective of NUM. In highly congested networks, the utility gain in allocating bandwidth to IPTV users is rather limited since IPTV user has a considerably high bandwidth threshold to be well provisioned.

The rest of this paper is organized as below. In Section 2, we present our NGN traffic classifications and formulate the utility function for each class. In Section 3, we solve the equivalent nonlinear programming problems. In Section 4, we compute numerical results by LINGO. Section 5 discusses some limitations of this paper and points out future directions. Finally, Section 6 concludes the paper.

2. NGN user classifications and their utility functions

Due to the remarkable distinction in QoS requirements among NGN users, it is important for designers to understand their classifications and treat different traffic classes disparately to achieve global welfare maximization. In this paper, we partition NGN users into five categories according to their explicit QoS requirements: (1) voice over Internet Protocol (VoIP) users previously using traditional PSTN; (2) emerging Internet Protocol Television (IPTV) users; (3) traditional TCP elastic users, including those relying FTP or P2P to download files; (4) web users and other TCP interactive users; (5) other streaming and gaming users generating UDP traffic.

To measure user satisfaction degree, we introduce the well-known concept of user utility, which is first invented