

Distributed Dynamic Bandwidth Provisioning in Quality of Service Networks

A. Capone, J. Elias, F. Martignon, G. Pujolle

Abstract— Efficient dynamic resource provisioning algorithms are necessary to the development and automation of Quality of Service (QoS) networks. The main goal of these algorithms is to offer services that satisfy the QoS requirements of individual users while guaranteeing at the same time an efficient utilization of network resources.

In this paper we introduce a new service model that provides quantitative per-flow bandwidth guarantees, where users subscribe for a guaranteed rate; moreover, the network periodically individuates unused bandwidth and proposes short-term contracts where extra-bandwidth is allocated and guaranteed exclusively to users who can exploit it to transmit at a rate higher than their subscribed rate. To implement this service model we propose a dynamic provisioning architecture for intra-domain Quality of Service networks. We develop an efficient bandwidth allocation algorithm that takes explicitly into account traffic statistics to increase the users' benefit and the network revenue simultaneously. We demonstrate through simulation in realistic network scenarios that the proposed dynamic provisioning model is superior to static provisioning in providing resource allocation both in terms of total accepted load and network revenue.

Index Terms: - Dynamic Bandwidth Allocation, Service Differentiation, Service Model.

I. INTRODUCTION

Efficient dynamic resource provisioning mechanisms are necessary to the development and automation of Quality of Service networks. In telecommunication networks, resource allocation is performed mainly in a static way, on time scales on the order of hours to months. However, statically provisioned network resources can become insufficient or considerably underutilized if traffic statistics change significantly [1].

Therefore, a key challenge for the deployment of Quality of Service networks is the development of solutions that can dynamically track traffic statistics and allocate network resources efficiently, satisfying the QoS requirements of users while aiming at maximizing, at the same time, resource utilization and network revenue. Recently, dynamic bandwidth allocation has attracted research interest and many algorithms have been proposed in the literature [1], [2], [3], [4].

In this paper we introduce a new service model that, first, provides a quantitative bandwidth guarantee to users and then

A. Capone is with the Department of Electronics and Information of Politecnico di Milano, E-mail: capone@elet.polimi.it

F. Martignon is with the Department of Management and Information Technology, University of Bergamo, Dalmine (BG) 24044, Italy. E-mail: fabio.martignon@unibg.it

J. Elias and G. Pujolle are with the University of Paris 6, LIP6 Laboratory, 8 rue du Capitaine Scott, 75015, Paris, France. E-mail: {jocelyne.elias,guy.pujolle}@lip6.fr

This work is partially supported by the National Council for Scientific Research in Lebanon

exploits the unused bandwidth individuated periodically in the network to propose short-term guaranteed extra bandwidth. To implement this service model we propose a dynamic provisioning architecture that allows, based on traffic statistics measured on-line, to react automatically to changes in bandwidth availability and to allocate resources efficiently within a service provider's network.

In the rest of this abstract we describe in some detail our proposed service model and provisioning architecture and we evaluate the performance of the proposed scheme in realistic network scenarios.

II. SERVICE MODEL AND DYNAMIC PROVISIONING ARCHITECTURE

In this paper we propose a new service model that provides quantitative per-flow bandwidth guarantees, where users subscribe for a guaranteed transmission rate. Moreover, the network periodically individuates unused bandwidth and proposes short-term contracts where extra-bandwidth is allocated and guaranteed exclusively to users who can better exploit it to transmit at a rate higher than their subscribed rate. To implement this service model we propose a distributed provisioning architecture composed by core and edge routers; core routers monitor bandwidth availability and periodically report this information to ingress routers using signalling messages like those defined in [2]. Moreover, if persistent congestion is detected, core routers notify immediately ingress routers. Ingress routers perform a dynamic tracking of the effective number of active connections, as well as their actual sending rate. Based on such information and that communicated by core routers, ingress routers allocate network resources dynamically and efficiently using a modified version of the max-min fair allocation algorithm proposed in [5]. Such allocation is performed taking into account users' profile and willingness to acquire extra bandwidth based on their bandwidth utility function. The allocation is then enforced by traffic conditioners that perform traffic policing and shaping.

III. NUMERIC RESULTS

In this Section we compare the performance, measured by the average accepted load and network revenue versus the total load offered to the network, of the proposed dynamic bandwidth allocation algorithm with a static provisioning strategy. Network revenue is defined as the average extra utility that derives from extra-bandwidth allocation. We refer to different network scenarios to cover a wide range of possible environments.

In the first scenario we gauge the effectiveness of the proposed traffic-based bandwidth allocation algorithm. We consider, in line with [1], [2], a scenario that consists of a single-bottleneck with 2 core nodes, 24 end nodes (12 source-destination pairs) and traffic conditioners at the edge. All links are full-duplex and have a propagation delay of 1 ms. The capacity of the links connecting the two core nodes is equal to 3 Mb/s, and that of the links connecting the end nodes to core nodes is 2 Mb/s. We use 12 Exponential On-Off traffic sources; the average On time is set to 200 s, and the average Off time is varied in the 0 to 150 s range to simulate different traffic load conditions while at the same time varying the percentage of bandwidth left unused by every connection. Six sources have a peak rate of 40 kb/s and a subscribed rate of 100 kb/s while the remaining sources have a peak rate of 1 Mb/s and a subscribed rate of 300 kb/s. The algorithm updating interval is set to 20 s. We assume, for simplicity, that all users have the same utility function proposed in [3], [6], $U(x) = 1 - e^{-\frac{x^2}{x+h}}$, that models the perceived utility of real-time elastic traffic for an allocation of x bandwidth units. The parameter h is set as in [6].

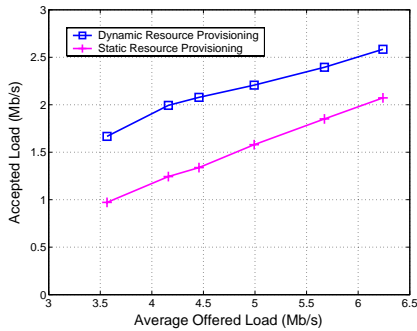


Fig. 1. Average total accepted load versus the average total load offered to the network in the single-bottleneck topology

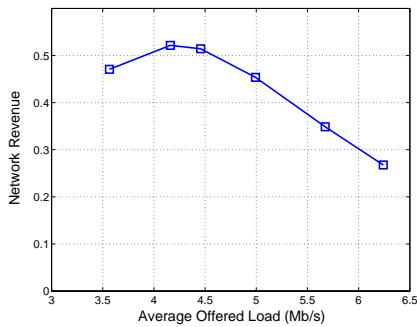


Fig. 2. Average total network revenue obtained versus the average total load offered to the network in the single-bottleneck topology

Figures 1 and 2 show, respectively, the average total load accepted in the network and the corresponding total revenue as a function of the average total load offered to the network. It can be observed that our dynamic provisioning algorithm is very efficient in resource allocation compared to a static provisioning algorithm for all values of the offered load, providing improvements up to 60% in the total accepted traffic.

We then considered a more realistic scenario that consists of

6 nodes and 8 bidirectional links, all having a capacity equal to 2 Mb/s and propagation delay of 1 ms. In this topology, 6 Exponential On-Off traffic sources are considered: 3 sources have a peak rate of 100 kb/s and a subscribed rate of 250 kb/s; 2 sources have a peak rate of 1 Mb/s and a subscribed rate of 500 kb/s and the last one has a peak and subscribed rate of 1 Mb/s. All other parameters are set as in the previous scenario. In this scenario, various connections compete for network capacity with different connections on different links.

Also in this scenario the dynamic allocation algorithm outperforms static allocation, as shown in Figures 3 and 4, thus proving the benefit of the proposed scheme. These results verify that our allocation algorithm allows service providers to increase network capacity utilization and consequently network revenue with respect to static provisioning techniques.

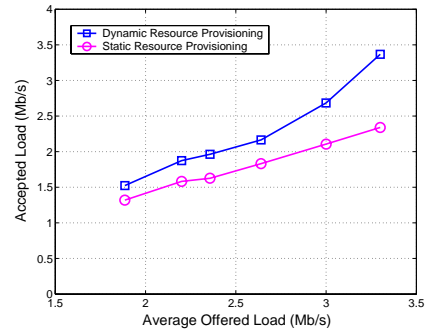


Fig. 3. Average total accepted load versus the average total load offered to the network in the second topology

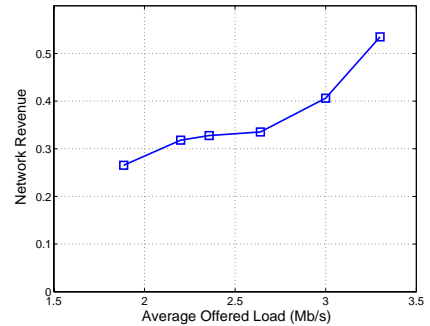


Fig. 4. Average total network revenue using dynamic bandwidth allocation versus the average total load offered to the network in the second topology

REFERENCES

- [1] A. T. Campbell and R. R.-F. Liao. Dynamic core provisioning for quantitative differentiated services. *IEEE/ACM Transactions on Networking*, pages 429–442, vol. 12, no. 3, June 2004.
- [2] H. Schulzrinne and X. Wang. Incentive-compatible adaptation of internet real-time multimedia. *IEEE Journal on Selected Areas in Communications*, pages 417–436, vol. 23, no. 2, February 2005.
- [3] Z. Cao and E. Zegura. Utility max-min: An application-oriented bandwidth allocation scheme. In *Proc. IEEE Infocom'99*, New York, USA, March 1999.
- [4] F. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, pages 33–37, vol. 8, 1997.
- [5] D. Bertsekas and R. Gallager. *Data Networks, 2nd Edition*. Prentice-Hall, 1992.
- [6] L. Breslau and S. Shenker. Best-effort versus reservations: a simple comparative analysis. In *Proc. ACM SIGCOMM*, pages 3–16, September 1998.