STUDY ON STATISTICALLY-FAIR SERVICE IN STATISTICALLY-MULTIPLEXED NETWORKS*

Cheng Dongnian  Liu Zengji
(National Key Laboratory on Integrated Services Networks, Xidian University, Xi'an 710071)

Abstract A new type of fair service, referred to as Statistically-Fair Service (SFS), is proposed in this paper. The SFS discipline is given based on the SFS criterion. Compared to “strict” fair service available, SFS is mainly characterized by its flexible suitability for the nature of statistically-multiplexed networks. By its statistically-fair service to users, therefore, SFS can ensure well end-to-end QoS requirements on a statistical basis with a benefit of enhancement in network utilization. Two useful properties of SFS is presented. One of them, the property of retaining Exponentially Bounded Burstiness(EBB), can facilitate end-to-end delay estimation of EBB-type traffic. Finally, some numerical results obtained from a simulation study on SFS shows that an SFS-equipped node in steady states will in deed retain the EBB attribute of any input flow.

Key words Fair service/queueing; Quality of Service(QoS); Statistical multiplexing; Exponential probability bounds

I. Introduction

Fair service, an important QoS-supporting facility, serves communication networks for the purpose of ensuring dynamically proper use of reserved network resources by users so as to guarantee each user's end-to-end QoS requirements[1-3]. An ideal type of fair service was proposed by A. Demers, et al.[2] on a fluid-flow traffic basis in 1989. Further studies were made by A. K. Paerkh and R. G. Gallager[3] in 1992, and an inequality characterizing the ideal fair service was given by them as

$$S_j(\tau, t) \geq \frac{\phi_i}{\phi_j}, \quad 0 \leq \tau \leq t, \quad j = 1, 2, \ldots, N, \quad i \neq j$$

for any backlogged session i sharing the same outgoing link with other (N - 1) sessions during any time interval [\tau, t], where $S_k(\tau, t)$ is the served amount of session k data during [\tau, t] and $\phi_k$ is a positive real number representing the service share assigned to session k(k = 1, 2, \ldots, N). Such an inequality is thus termed the ideal fair service criterion. Based on the above researches on fair service, called Fluid-Flow Queueing(FFQ) in literature, varieties of fair service schemes were presented later, such as Weighted Fair Queueing(WFQ)[3], Self-Clocked Fair Queueing(SCFQ)[4], Worst-case Fair Weighted Fair Queueing(WF2Q)[5], etc[6-8]. FFQ-is in nature, however, impractical since service could not be provided in arbitrarily small increments in actual networks, say packet-switched ones[8,9]. Alternatively

*Supported by the Electronics Preresearch Foundation of China Academy of Electronics and Information Technology
said, user traffic cannot be divided into arbitrarily small amount, and hence FFQ is merely an ideal fair service criterion. We notice that for any backlogged session \( i \) in a real packet-switched network node with packets as its basic data units, there may be possibilities that there exists another session \( k \) so that

\[
\frac{S_i(t)}{S_k(t)} < \frac{\phi_i}{\phi_k}, \quad k \in \{1, 2, \cdots, N\}, \quad k \neq i \quad (2)
\]

It is also recognized that although violations of the FFQ criterion may happen, service provided to users will still remain “ideal-approached” fair so long as the violations are limited within some reasonable range. Motivated by such a consideration, this paper proposes a novel approach to defining a fair service criterion, called Statistically-Fair Service (SFS). Contrast to “strict” fair service available, SFS is markedly characterized by its “flexible” fair service and its suitability for the nature of statistically-multiplexed networks, such as ATM ones. Since it lays fair service on an exponential probability bound, SFS can facilitate effectively the combination of traffic statistics, end-to-end QoS requirements, fair service, network features, and resource utilization.

The remainder of this paper proceeds as follows: Section II introduces the criterion and two useful properties as well as the discipline of SFS. Next, Section III shows some numerical results obtained from an SFS simulation study. Finally, Section IV draws some conclusions of the whole paper.

II. Criterion, Discipline and Properties of SFS

The kernel of fair service includes two main parts: (1) to ensure the proper use of reserved network resources by respective users so as to prevent them from inference caused by other misbehaving users; and (2) to distribute dynamically unused parts of the already allocated bandwidth among all backlogged users in proportion of their respective service shares. Since it is too “strict” for a real network node to offer the ideal fair service, a practical fair service criterion should adapted well to statistical features in a statistical multiplexing network and should be defined on a statistical basis accordingly. In other words, a fair service should be targeted for as high as possible resource utilization based on the guarantee of each user’s normal use of its reserved resources. More precisely, a node is said statistically fair for any backlogged session, say \( i \), in it only if

\[
\Pr \left\{ \frac{S_i(t)}{S_j(t)} < \frac{\phi_i}{\phi_j} \right\} \leq \epsilon_i, \quad 0\epsilon_i < 1, \quad j = 1, 2, \cdots, N, \quad i \neq j \quad (3)
\]

holds. Let \( X \) and \( Y \) axes represent the served amount of traffic of user flow \( i \) and \( j \) respectively, serving \( i \) and \( j \) is thus equivalent to “walking” on the \( X-Y \) plane. Clearly, point set \( \{(m, n)|m, n > 0\} \) defines a complete “route” corresponding to serving both \( i \) and \( j \). Thus, a “fair” server should sequentially traverse every point along the “fair route”, i.e., \( \{(r\phi_i, r\phi_j)|r = 1, 2, \cdots\} \), till one of \( i \) and \( j \) finishes its transmission. It also noticed that if serving units (i.e., “atomic” data units) of \( i \) and \( j \) are respectively determined, the number of unfair points on the plane the server must go through from one fair point to its next fair one will be more when \( \phi_i/\phi_j \) or \( \phi_j/\phi_i \) is smaller, i.e., more unfair-service chance will result in this case. Therefore, both unfair service probabilities and the corresponding bounds can be