
ATM Forum Document Number: ATM Forum/94-0881

Title: Fairness: How to Measure Quantitatively?

Abstract: Fairness is an important criterion in the design of congestion control schemes for ATM Networks. However, the forum has not yet selected a standard definition for a quantitative measure. We propose an Index of Fairness that we have been using successfully for over ten years. This index has several desirable properties. In particular, it always lies between 0 and 100% and so quantitative statements about fairness of various proposals can be made.

Source:

Raj Jain
The Ohio State University

Raj Jain is now at Washington University in Saint Louis, jain@cse.wustl.edu <http://www.cse.wustl.edu/~jain/>

Date: September 26-29, 1994

Distribution: ATM Forum Technical Working Group Members (Traffic Management)

Notice: This contribution is sponsored by National Institute of Standards and Technology (NIST). It has been prepared to assist the ATM Forum. It is offered to the Forum as a basis for discussion and is not a binding proposal on the part of any of the contributing organizations. The statements are subject to change in form and content after further study. Specifically, the contributors reserve the right to add to, amend or modify the statements contained herein.

INTRODUCTION

ATM Forum traffic management group has been working aggressively on comparing various congestion control alternatives. In comparing these alternatives, the statements about the fairness are mostly qualitative, for example, "the scheme X is unfair" or "it becomes fair with round-robin scheduling." Some people have used variance or standard deviation of throughput as a measure of fairness but these are not good measures as explained below.

The ideas presented here are from Jain, Hawe, and Chiu (1983) [1] and are also presented in [2]. We propose to divide the problem of fair allocation of resources in two parts. First part consists of determining the "ideal" or "optimal allocation" and second part consists of quantifying the deviation from this ideal. An

example of the first part is the max-min optimality, which was accepted as the default for optimal allocation in the July meeting. We will use the max-min optimal allocation as the desired goal, although the discussion here applies equally well to any other definition of optimality.

Suppose given a network configuration, with n VCs, the throughput of the i th VC is Z_i under max-min optimality. Now if a scheme results in a throughput Y_i , instead, its fairness can be measured by first taking a ratio of $X_i = Y_i/Z_i$. An ideal scheme will give all VCs their optimal allocation and all X_i 's will be equal to 1. However, any real scheme will result in ratios slightly different from 1 and its fairness is given by the following fairness index:

$$\text{Fairness Index} = ((\sum(X_i))^{**2}) / (n * \sum(X_i^{**2}))$$

Example: Suppose it is determined that the max-min optimal allocation of a 155 Mbps bottleneck link among three sources is 100 Mbps, 40 Mbps, and 15 Mbps. A congestion control scheme results in 50 Mbps, 30 Mbps, and 75 Mbps, respectively for the three sources. The scaled throughputs for the sources are 50/100, 30/40, and 75/15 or 0.5, 0.75, and 5. Substituting these in the fairness index, we get, $((0.5+0.75+5)^{**2}) / (3 * (0.5^{**2} + 0.75^{**2} + 5^{**2}))$ or 0.504. The scheme is 50.4% fair.

This index has several desirable properties:

1. Population Size Independence: The index is applicable to any number of VCs finite or infinite. Note that variance or standard deviation are statistical measures and require a large number of VCs to be valid.
2. Scale Independent: As defined above X_i 's were ratios of throughput. Other alternatives such as raw throughput, response times, or their ratios (power) can be used instead. The formula applies equally well to all metric (and so do the variance and deviation for that matter). However, variance and standard deviation are scale dependent in such cases. For example, if we use variance of throughput to measure fairness and the throughput is measured in Cells per millisecond, the variance is measured in Cells-squared per millisecond-squared. If the unit is changed, the value changes. For example, a scheme that results in a throughput variance of 10 cells-squared per millisecond-squared can be said to have variance of 10^{**7} cells-squared per second-squared. Such scaling make it difficult to interpret the fairness and also give opportunities for exaggregation (or hiding) of unfairness.

The fairness index has no units and regardless of the throughput unit used, the value always remains same.

3. Boundedness: The variance and standard deviation have no bound. Highly unfair schemes can have infinite variance. The fairness index, on the other hand, is bounded between 0 and 1. In practice, we found it useful to express it as a percentage. The index lies between 0 and 100%. Most people find it easier to understand the statement "The scheme is 90% fair" than "The scheme has a variance of 35364 cells-squared per second squared."

4. Direct Relationship: The relationship between fairness and variance is an inverse relationship. As the fairness goes down, the variance goes up and vice versa. An unfair scheme has high variance. The fairness index as defined above has a direct

relationship to the fairness. As the fairness goes up so does the index.

5. Continuity: The index is continuous. Even a slight change in a VC's allocation is reflected as a change in the index. Some indexes such as the ratio of maximum throughput to minimum throughput do not have this property.

6. Intuitive: We have been successfully using the fairness index as the fairness measure for the last 10 years and have found it easy to explain to non-statisticians. It results in intuitive values for several simple cases. For example, if the bandwidth is divided among contending VC's such that it is divided equally among 80% of the VCs while the remaining 20% of the VCs are starved, the fairness index comes out 80%. This applies for any other percentage as well. In particular, if $X_i=1$ for $i=1,2,\dots,a$ and $X_i=0$ for $i=a+1, a+2,\dots,n$ where a is any fraction between 0 and 1, the fairness index is a .

Based on the above arguments, we propose that ATM Forum adopt the fairness index as the quantitative measure of fairness in comparing congestion control alternatives in future.

In the discussion so far we ignored the fact that throughput, response time and other metrics are dynamic quantities in the sense that they vary with time. The fairness is, therefore, also a dynamic quantity. Fairness at time t can be computed based on throughput (or any other metric) at time t . Overall fairness can be computed based on overall throughput.

REFERENCES

- [1] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer System" DEC Technical Report DEC-TR-301, September 26, 1984. Available from Jain@ACM.Org.
- [2] R. Jain, "The Art of Computer Systems Performance Analysis," John Wiley and Sons, New York, 1991.