
ATM Forum Document Number: ATM_Forum/99-0045

Title: Throughput Fairness Index: An Explanation

Abstract: The performance testing document uses a particular function to quantify fairness. This contribution explains the reasons for the selection of the particular function.

Source:
Raj Jain, Arjan Duresi, Gojko Babic
The Ohio State University
Department of CIS Columbus, OH 43210-1277
Phone: 614-292-3989, Fax: 614-292-2911, Email: Jain@ACM.Org

The presentation of this contribution at the ATM Forum is sponsored by NASA.

Date: February 1999

Distribution: ATM Forum Technical Working Group Members (AF-TEST, AF-TM)

Notice:
This contribution 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.

1. INTRODUCTION:

Given n virtual circuits sharing a system, if the measured throughputs are found to be $\{T_1, T_2, \dots, T_n\}$, where the ideal throughputs should be $\{T_1^{\wedge}, T_2^{\wedge}, \dots, T_n^{\wedge}\}$, then according to ATM Forum performance testing specification draft [1] Section 4.4.1, the fairness index of the system is given by:

$$f(x) = \frac{(\sum x_i)^2}{n \cdot \sum (x_i^2)}$$

Where $x_i = T_i/T_i^{\wedge}$ is the relative allocation.

This formula for fairness index was also used in the early discussions on fairness in the traffic management group. During those discussions, it became obvious that problem of fairness can be divided into two parts:

1. Defining a fair criterion and
2. Define a fairness index that quantifies the distance between the goal and the operating point chosen by the switch

There are a number of ways to define the fairness criteria. The ATM Forum Traffic Management Specification [2] lists several different fairness criteria. Equipment vendors may select any one of these. For example, one simple way to divide bandwidth among n users is to distribute it equally among them. Another one is divide it proportional to their minimum cell rates (MCRs). And so on.

Once a fairness criterion has been selected, the second problem of selecting a fairness index is more of a mathematical exercise. In this contribution we address this second part of the fairness issue. We present a formula (fairness index) that measures the closeness of the resource allocation from the desired goal. The formula was first published in the open literature in [3] and is analyzed in detail in [4]. Most of the text of this contribution is based on [4]. The reader should refer to [4] for proofs of the theorems presented here.

Given a system under test (SUT) shared by n users (flows, VCs, Ports), let T_i^{\wedge} be the desired allocation for the i th user while T_i is the actual allocation by the SUT. Let,

$$x_i = T_i/T_i^{\wedge}$$

A totally fair system will allocate resources such that all x_i 's are equal to 1. The fairness index measures the "equality" of the resource allocation and, if not equal, it tells how far the allocation is from equality.

In the rest of this document we refer to "normalized resource allocation" x_i as simply the "allocation."

2. DESIRED PROPERTIES OF A FAIRNESS INDEX

In our search for a good fairness index, we set a number of goals. We found that the indices proposed in the literature did not satisfy some of these. The fairness measures proposed in the literature are:

1. Variance: $1/(n-1) \sum[(x_i - m)^2]$, where $m = 1/n \sum(x_i)$
2. Coefficient of Variation = $\text{Sqrt}(\text{Variance})/\text{Mean}$
3. Min-Max ratio = $\text{Min}(x_i)/\text{Max}(x_i)$

For example, if a SUT gives three VCs throughputs of 1, 3, and 5 Mbps, then the fairness of this SUT is:

1. Variance = 4, Mean $m = 3$
2. Coefficient of Variation = 0.667
3. Min-max ratio = $1/5$

We feel the desired properties of the "fairness index" should be the following:

1. Population Size Independence: The index should be applicable to any number of users, finite or infinite. In particular, it should be applicable to two users sharing a resource. This requirement is satisfied by the above three indices.
2. Scale and Metric Independence: The index should be independent of scale, i.e., the unit of measurement should not matter. For example, the above SUT allocates 1000, 3000, and 5000 kbps to the three VCs, the unfairness measured by variance now is 4000,000, which seems a million times as bad as before even though we have not changed the SUT. Thus, this property rules out the use of variance as a fairness index. Notice that the coefficient of variation and min-max ratio are unaffected by scale.
3. Boundedness: The index should be bounded between 0 and 1, i.e., a totally fair system should have a fairness of 1 and a totally unfair system should have an index of 0. This allows fairness to be expressed as a percentage. For example, a SUT with a fairness of 0.1 could be shown to be fair to 10% of the VCs and unfair to 90%. This helps in intuitive understanding of the index.

The coefficient of variation can be anywhere between 0 and infinity. It is not easy to interpret what level of fairness is implied by a COV of 359, for instance. The min-max ratio is better in this respect in that it does satisfy this requirement.

4. Continuity: The index should be continuous. Any slight change in allocation should show up in the fairness index. In the above example, if the three users received 1, 4, and 5 Mbps, respectively, the fairness should obviously be different, yet it is not reflected in the min-max ratio which remains at 1/5.

Thus, we see that none of the known indices satisfy all requirements. This leads us to propose a new index.

3. FAIRNESS INDEX: PROPOSED DEFINITION

If a system allocates x_i to the i th user, the fairness index for the SUT is:

$$f(x) = \frac{(\sum x_i)^2}{n \cdot \sum (x_i^2)}, x_i \geq 0$$

This index measures the "equality" of allocations. If all users get the same allocation, i.e., all x_i 's are equal, then the fairness index is 1, and the system is 100% fair. As the disparity increases, fairness decreases and a scheme which favors only a selected few users has a fairness index near 0.

Notice that the proposed index satisfies all the requirements discussed before. It is dimensionless and independent of scale, it is bounded between 0 and 1, and it is continuous so that any slight change in x_i changes the index.

The fairness index as defined here has a very intuitive interpretation as illustrated by the following example:

Example 1: Suppose we want to allocate 20 Mbps among 100 VCs. Consider two possibilities:

SUT A gives 0.2 Mbps to each of the 100 VCs.

In this case, $x_i = 0.2, i=1,2,\dots,100$

$$\text{Fairness Index} = f_A(x) = f(x) = \frac{(\sum x_i)^2}{n \cdot \sum (x_i^2)} = 1.0$$

SUT A is totally fair.

SUT B selects 10 VCs and gives them 2 Mbps each and nothing to

others.

In this case, $x_i = 2, i=1,2,\dots,10$
and $x_i = 0, i=11,12,\dots,100$

$$\text{Fairness Index} = f_A(x) = f(x) = \frac{(\sum x_i)^2}{n \cdot \sum (x_i)^2} = 0.1$$

SUT B is 10% fair.

If in the case of SUT B, we had selected 20 of the 100 VCs to receive 1 Mbps each, the fairness would have come out to 20%. This intuitive interpretation of fairness index is true in general. That is, given m resources to be allocated to n users, if we select k users and give them m/k resources and give nothing to the rest, the fairness index would be k/n , or the fraction of the users favored. Notice that the index does not depend upon the amount of the resource (m) or the population size (n). In the above example, the population size could have been 2 and the fairness could still be quantified.

4. RELATIONSHIP TO OTHER FAIRNESS DEFINITIONS

If the allocations x_i 's follow a certain random distribution, then the fairness index as defined here can be expressed as a function of the first two moments of x_i 's as shown below:

$$\begin{aligned} f(x) &= \frac{(\sum x_i)^2}{n \cdot \sum (x_i)^2} \\ &= \frac{\{(1/n)(\sum x_i)^2\}}{\{(1/n)\sum (x_i)^2\}} \\ &= \frac{\{E[x]\}^2}{E[x^2]} \\ &= \text{Square of the first moment of } x / \text{second moment of } x \\ &= 1 / \{1 + \text{COV}^2\} \end{aligned}$$

Here, COV is the coefficient of variation defined as the ratio of standard deviation and the mean.

Note that although the relationship between the fairness index proposed here and the coefficient of variation is a simple transformation, it is an important transformation because it gives the fairness index all the desirable characteristics discussed below.

First, the relationship between the COV and fairness is an inverse one. That is, the system is totally fair, if the COV is zero. As the COV increases, fairness goes down. Transformation 2 removes this negativity. Thus, as the fairness increases, the fairness index grows higher. This helps intuitive understanding of the results.

Second, COV is an unbounded quantity. Thus, COV of allocations could be in hundreds or thousands. Transformation 2 makes fairness a bounded quantity. It bounds it between 0 and 1 and allows it to be

expressed as a percentage. As shown earlier, "percent fairness" is not only a matter of convenience, it is also a meaningful interpretation. A SUT which allocates all resources equally to only 10% of the users turns out to have a fairness index of 0.1. The COV in this case is 3, which is difficult to interpret.

5. USER PERCEPTION OF FAIRNESS

One way to rewrite the proposed formula for fairness index is:

$$f(x) = 1/n \sum (x_i/x_f)$$

Where x_f is the "fair allocation mark" computed as follows:

$$x_f = \sum(x_i^2)/\sum(x_i)$$

Thus, each user compares his allocation x_i with the amount x_f and perceives the algorithm as fair or unfair depending upon whether his allocation x_i is more or less than x_f . For the i th user, the algorithm is only x_i/x_f fair. The overall fairness is the average of the perceived fairness of all n users.

In the example of 20 Mbps distributed among 100 VCs using SUT B:

$$x_i = 2, i=1,2,\dots,10 \text{ and } x_i = 0, i=11,12,\dots,100$$

$$\text{Fair allocation } x_f = \sum(x_i^2)/\sum(x_i) = 2$$

Thus the 10 users that received 2 Mbps each consider the SUT to be 100% fair while the 90 users who didn't receive anything perceive the SUT to be 0% fair. The overall fairness is 10%. All users with $x_i > x_f$ are said to be "favored" users and those with $x_i < x_f$ are "discriminated" users.

6. PROPERTIES OF THE PROPOSED FAIRNESS INDEX

In this section, we present a few simple results regarding behavior of the fairness index as allocations are changed. The proofs of the theorems are in [4].

1. Resource Exchange: If we take a small amount of resource from a user and give it to another user, the new allocation should be more fair if the receiver has less resource than the giver and vice versa. The proposed index does satisfy this property as the following theorem states:

Theorem 1: If dx resource is taken away from the k th user and given to j th user, then the fairness:

1. increases iff $dx < x_k - x_j$
2. remains same iff $dx = x_k - x_j$
3. decreases iff $dx > x_k - x_j$

2. Additional Allocation: If each user is given an additional amount of resource, we would expect individual perception of fairness to rise and the overall fairness to go up. On the other hand, if only a single user is given additional resources, others may become dissatisfied if the user is already a "favored" user. These properties of the fairness index are stated by the following theorem:

Theorem 2a: If each user is given an additional amount c of the resource, their individual perception of fairness increases and so does the overall fairness index:

$$f(x_1+c, x_2+c, \dots, x_n+c) \geq f(x_1, x_2, \dots, x_n)$$

Theorem 2b: If a single user j is given a small additional allocation dx_j without changing other allocations, the new allocation is more (less) fair than before iff j is a discriminated (favored) user, i.e.,

$$f(x_1, x_2, \dots, x_j+dx_j, \dots, x_n) > f(x_1, x_2, \dots, x_j, \dots, x_n) \text{ if } x_j < x_f$$

$$f(x_1, x_2, \dots, x_j+dx_j, \dots, x_n) < f(x_1, x_2, \dots, x_j, \dots, x_n) \text{ if } x_j > x_f$$

Here, x_f is the fair allocation mark defined earlier.

3. Maximization: The fairness index has a bell-shaped behavior curve with respect to each individual allocation. As a particular user's allocation is increased from zero, the fairness first increases. It then reaches a maximum. Any further allocation to that user results in other users perceiving unfairness and the overall fairness decreases.

Theorem 3: If we vary single user j 's allocation, while not affecting other users' allocations, the fairness reaches maximum when

$$x_j = x_g$$

Where x_g is the fair allocation mark for the remaining $n-1$ users. i.e.,

$$xg = \sum(x_i^{**2}, i \text{ not} = j) / \sum(x_i, i \text{ not} = j)$$

7. OTHER FAIRNESS FUNCTIONS

The proposed fairness index is not the only function that satisfies the requirements listed in Section 2. In fact, any function of the form:

$$f(x) = \{[(1/n)\sum(x_i)]^{**r}\} / \{(1/n)\sum[(x_i)^{**r}]\}$$

seems to satisfy all the requirements, including those of continuity, boundedness and dimensionlessness. However, the proposed index (with $r=2$) is still a preferred measure of fairness because of its "linear behavior" with respect to the fraction of favored users in Example 1. If m resources are to be distributed among n users and we favor k users giving them m/k each and discriminate against $n-k$ users, then the above function will give

$$f(x) = (k/n)^{**}(r-1)$$

That is, favoring 10% would result in a fairness index of $0.1^{**}(r-1)$. Obviously, $r=2$ seems to be the correct choice.

8. SUMMARY

The problem of fairness can be broken down in two parts: selection of a fairness criterion and quantification of equality. For quantification of equality, a fairness index function was proposed. The proposed index has many desirable properties which other formulae do not satisfy. The proposed index is independent of scale of the allocation metric. It is bounded between 0 and 1 so that it can be meaningfully expressed as a percentage. Finally, it is continuous so that any change in allocation changes the fairness also.

9. REFERENCES

- [1] F. Kaudel, "ATM Forum Performance Testing Specification Draft," ATM Forum/BTD-TEST-TM-PERF.00.10, December 16, 1998.
- [2] ATM Forum, "Traffic Management Specification 4.0," af-tm-0056.0, 1996
- [3] R. Jain, "The Art of Computer Systems Performance Analysis," Wiley, 1991, pp. 36.

[4] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer System," Digital Equipment Corporation, Technical Report, DEC-TR-301, September 26, 1984, 40 pp., available at <http://www.cis.ohio-state.edu/~jain/papers/fairness.htm>