An Analytical Study of FairOM: A Fair Overlay Multicast Protocol for Internet-Scale Distributed Systems

Yijun Lu[1] and Xueming Li[2,1]
[1]Department of Computer Science and Engineering, University of Nebraska-Lincoln
[2]School of Computer Science and Engineering, Chongqing University, China
{yijlu, xli}@cse.unl.edu

1. Introduction

Fairness emerges as an important research issue in overlay multicast because spreading the multicasting load evenly among participants can eliminate potential traffic hot spots, thus improving the system's Quality of Service (QoS). FairOM [1, 2] has been proposed to enforce participants to contribute the same proportion of their available outgoing bandwidth to each session. With FairOM, more multicast sessions can be enabled simultaneously that would otherwise be impossible.

In this paper, we analyze FairOM and compare it with non-FairOM approaches from two aspects: tree height and number of sessions that can be supported, which measure FairOM from a single-session's and multiple-session's point of view, respectively. Together, they draw an overall picture of FairOM. In this analysis, we make the following assumptions.

[1] There are n nodes in the overlay network and the multicast should cover all of them. The n nodes are denoted by $N = \{N_1, N_2, \ldots, N_n\}$.

[2] Total available bandwidth of nodes, in terms of number of stripes, are $T = \{T_1, T_2, \ldots, T_n\}$.

[3] The number of sessions is denoted as m and the m sessions are $S = \{S_1, S_2, \ldots, S_m\}$.

[4] There are r stripes in each session.

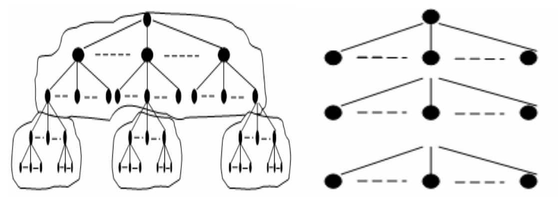
[5] Each node in FairOM contributes α% of its total available bandwidth for each session.

2. Analysis of tree height

In the best case, the nodes organize as a balanced tree, as in Figure 1 (a). In the worst case, the nodes organize in a linear fashion as in Figure 1 (b).

2.1. Analysis of tree height

We classify the nodes with different number of children and use n_i to denote the number of nodes with q_i ($q_i \in T$) children. Suppose that there are k types of capacities and we have



(a) (b)

Figure 1. Approximate best case scenario (a) and worst case scenario (b)

$$\sum_{i=1}^{k} n_i q_i = n \qquad \ldots (1)$$

Based on the theory of balanced tree, if all interior nodes have the same number of r children, the tree height is $\log_r^n$. For simplicity, we approximate the lowest tree height, h_{min}, as the tree height if we restrict the nodes' capacity to be the same as the average capacity, represented by q_min_{avg}, as in Figure 1 (a). Thus in the case of non-FairOM approaches, we have

$$h_{n\_\min} \approx \log_{q\_\min_{avg}}^{n} \qquad \ldots (2)$$

In the worse case, the tree height is determined by the number of interior nodes, so we have:

$$h_{n\_\max} = \sum_{i=1}^{k} n_i \qquad \ldots (3)$$

Suppose the average number of children of all the interior nodes is q_max_{avg}, from formula (1) we have

$$\sum_{i=1}^{k} n_i q\_\max_{avg} = q\_\max_{avg} \sum_{i=1}^{k} n_i = n,$$

we have $\sum_{i=1}^{k} n_i = \frac{n}{q\_\max_{avg}}$

Thus, we finally have

$$h_{n\_\max} = \sum_{i=1}^{k} n_i = \frac{n}{q\_\max_{avg}} \qquad \ldots (4)$$

Now we consider the case of FairOM. In this case, each node is only allowed to allocate at most α% of its total available bandwidth for each session, hence for each stripe. According to the FairOM protocol, the q_min_{avg} and q_max_{avg} would be α% of those values in formula (2) and (4). For this reason, the lowest and highest tree height can be expressed as:

$$h_{f\_\min} \approx \log^{n}_{\alpha\% * q\_\min_{avg}} \quad \ldots (5)$$

$$h_{f\_\max} = \frac{n}{\alpha\% * q\_\max_{avg}} \quad \ldots (6)$$

According to the FairOM protocol, $\alpha\% * q\_max_{avg}$ in (6) must be larger than 1.

Thus we can roughly compare the tree height between the FairOM and non-FairOM approaches. The main result is as follows.

$$ratio_{\min} = \frac{h_{f\_\min}}{h_{n\_\min}} = \frac{1}{1 + \log^{\alpha\%}_{q\_\max_{avg}}} \quad \ldots (7)$$

$$ratio_{\max} = \frac{h_{f\_\max}}{h_{n\_\max}} = \frac{1}{\alpha\%} \quad \ldots (8)$$

To get a numerical sense about the two ratios, we set α% as 25% and q_max_{avg} as 16. The results are that $ratio_{min}$ equals to 2, and $ratio_{max}$ equals to 4.

2.2. Push tree height toward the lower bound

We propose two mechanisms to push the tree height of FairOM toward the lower bound. First, we prevent the worse case, the linear structure, from happening as early as possible by monitoring the tree construction process. Whenever a linear structure is discovered, it randomly picks other nodes as children rather than the current ones. Second, realizing that the first optimization can slow down the forest building process, we use threshold to strike a balance. The optimization process is active when the current expected final tree height is longer than the threshold (thus improvement is needed) and is inactive otherwise.

3. Analysis of the protocols' capacity

Recall that each node in FairOM uses *a%* of its bandwidth for each session. Given the node with the smallest capacity, denoted as T_{min}, in terms of the stripes it can forward, the following constraint must apply since a stripe is the smallest unit of transmission: $T_{min} * a\% \geq 1$ *and an integer.* Therefore, the number of sessions that FairOM can support is:

$$NF = \left\lfloor \frac{100}{a} \right\rfloor \leq \lfloor T_{\min} \rfloor \quad \ldots (9)$$

The metric of comparison is the probability that a non-FairOM approach can support *NF* sessions. The rationale is that, if a non-FairOM approach is very unlikely to match the number of sessions FairOM can support, it will have an even smaller probability to support more sessions than FairOM.

For a non-FairOM approach, we examine a given node *i* that has a bandwidth of T_i. For each session, the contribution of node *i* in terms of the number of stripes it forwards, is an integer between 1 and T_i, since in a non-FairOM approach a node contributes arbitrary amount of its capacity. Suppose it can support *NF* sessions, we denote its contribution to the *NF* sessions as C_1, C_2, … C_{NF}. Because each contribution has T_i options, the number of all possible combination is:

$$Total_i = T_i^{NF} \quad \ldots (10)$$

In all these combinations, not all the possible combinations satisfy the requirement that the sum of all contribution is less than or equal to T_i—to make the forwarding load within node *i*'s capacity. We assume that the number of feasible combinations (i.e., combinations that can make the forest feasible) is F_i. To calculate F_i, we treat T_i as T_i 1s and *NF* contributions as *NF* bins. Because contribution has to be at least 1, we first pick *NF* 1s and put them to the *NF* bins to make them non-empty, then randomly put the remaining 1s to the *NF* bins. Thus, C_i is determined by the placement of the (T_i – *NF*) remaining 1s. Because C_i has at most (T_i – *NF*) options, we have

$$F_i \leq (T_i - NF)^{NF} \quad \ldots (11)$$

In this formula, (T_i – *NF*) is positive because of formula (1). Clearly, we have $F_i < Total_i$. Then the probability that node *i* can support *NF* sessions is:

$$P_i = \frac{F_i}{Total_i} \leq \left(\frac{T_i - NF}{T_i}\right)^{NF} \quad \ldots (12)$$

Suppose there are *n* nodes in the multicast group and P_{max} is the maximum value of P_i ($i = 1, 2, \ldots, n$), the probability that a non-FairOM approach can support the same number of sessions as FairOM is:

$$P_{All} \leq (P_{\max})^n \quad \ldots (13)$$

Please notice that P_{All} depends on the value of *n* that is usually very large. Even when P_{max} is very close to 1, P_{All} can still be very small with even a small number of *n*. For example, when P_i is 0.99 and *n* is 500, P_{All} is 0.0066. Thus, we believe that FairOM has a much larger capacity than non-FairOM approaches because it can support more simultaneous multicast sessions.

References

[1] Y. Lu, and H. Jiang, Design and evaluation of a new and effective fairness scheme for multicasting in Internet-scale distributed systems, In Proc. of HPDC-14, Research Triangle Park, NC, July 24-27. pp. 285-286.

[2] Y. Lu, H. Jiang, and D. Feng, FairOM: Enforcing proportional contributions among peers in Internet-scale distributed systems, In Proc. of ISPA 05, Nanjing, China, Nov. 2-5, 2005.