Estimating the Volume of Elephant Flows under Packet Sampling

Tatsuya Mori, Masato Uchida, Ryoichi Kawahara

I. Introduction

Recently, a number of papers have revealed that distributions of flow size are heavy-tailed, and that these characteristics hold on several time scales. This suggests that relatively few flows, each carrying a large number of packets (i.e., "elephant flows") account for a large part of total aggregated traffic on a certain time scale. Since the contributions and impacts of such flows on network operation are quite large, it is essential to keep statistics about them. This paper focuses on estimating the volume of elephant flows under packet sampling. We evaluate the accuracy and cost of the estimation theoretically and discuss its feasibility using measured packet trace.

II. Theoretical Analysis

We define a set of N packets as the population. The packet sizes are denoted by $x_1, x_2, ..., x_N$. From the population, we randomly sample n packets, whose sizes are denoted by $y_1, y_2, ..., y_n$. Then, we classify packets into the flows to which each belongs. Let the jth flow contain N_j packets in the population, and let n_j be the number of packets in a random sample of size n that belong to this flow. Here y_{jk} $(k = 1, 2, ..., n_j)$ expresses the measured packet sizes sampled on the jth flow. Then, for every packet in the population, we define the variate x'_{ji} $(i = 1, 2, ..., N)$ as

$$x'_{ji} = \begin{cases} x_i & \text{if the packet is in the } j\text{th flow,} \\ 0 & \text{otherwise} \end{cases}$$

The sum of the packet sizes for the jth flow is calculated as $X_j = \sum_{i=1}^{N} x'_{ji}$. The estimates of X_j and its standard error are calculated as follows [1].

$$\hat{X}_j = \frac{N}{n} \sum_{k=1}^{n_j} y_{jk} \quad (1)$$

$$\sigma(\hat{X}_j) = \frac{N S'_j}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} \quad (2)$$

, where $S'_j = \sqrt{\sum_{i=1}^{N} {x'_{ji}}^2 - {X_j}^2/N} \Big/ \sqrt{N-1}$, which is the population standard deviation of x'_{ji}. In repeated samples of size n, $\hat{X}_j$ is an unbiased estimate of X_j. The coefficient of variation of $\hat{X}_j$ is

$$cv(\hat{X}_j) = \frac{\sigma(\hat{X}_j)}{E\left[\hat{X}_j\right]} = \frac{\sigma(\hat{X}_j)}{X_j} \quad (3)$$

, which shows how the estimate of X_j varies relative to the true value. Assuming that packet sizes are set to $x_i = M$ [1] , we can approximate $cv(\hat{X}_j)$ as

$$cv(\hat{X}_j) = \frac{\sigma(\hat{X}_j)}{N_j M} = \sqrt{\frac{N-n}{Nn(N-1)}} \sqrt{\frac{1}{N_j} - \frac{1}{N}} \quad (4)$$

, which shows that the coefficient of variation is inversely proportional to ${N_j}^{1/2}$ when N is sufficiently large to N_j, where $N_j \gg 1$.

III. Evaluation

The packet trace for this study was collected on the one of the international backbone lines for the WIDE project [2]. The line is a 100-Mbps Ethernet and carried sufficient volume of traffic during the measurement period. The trace contained $1,000,000$ packets from one-way traffic (US-to-Japan). This trace was used as the population for sampling (i.e, $N = 1,000,000$). The trace was about 377 seconds long. We define an IP flow as a group of IP packets having the same combination of source IP address, destination IP address, source port, destination port, and protocol field. With this definition of flow, the trace contained packets for 172,074 flows. We investigated N_j and X_j for each flow. Figure 1 shows the log-log complementary cumulative distribution (LLCD) plots of N_j and X_j; we see that both distributions are highly heavy-tailed. That is, most of the flows were quite small, while a few flows were quite large. For instance, the top 100 flows carrying large numbers of packets ($N_j \geq 1222$) accounted for 53.9% of the total aggregated traffic. In this study, we call these flows "elephant flows".

We consider the probability $P_j[k]$ that k packets belonging to the jth flow are included among n sampled packets. This is

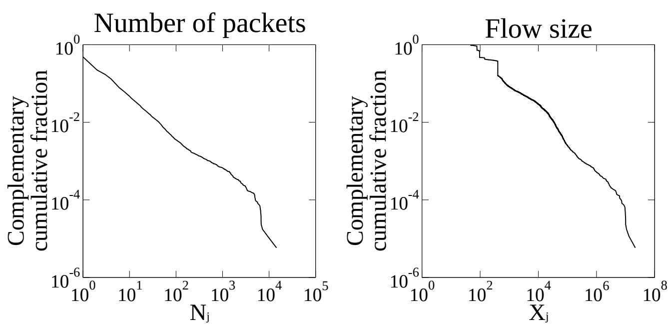

Fig. 1. LLCD plots for N_j and X_j.

The authors are with the NTT Service Integration Laboratories, NTT Corporation, Tokyo, Japan. E-mail: tatsuya@nttlabs.com, {uchida.masato, kawahara.ryoichi}@lab.ntt.co.jp.

[1] This approximation makes the coefficient of variation smaller, especially for flows with fewer packets, as we will show later.

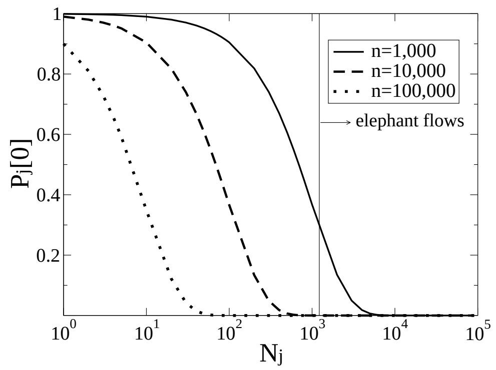

Fig. 2. Probability $P_j\left[0\right]$ given $N = 1,000,000$.

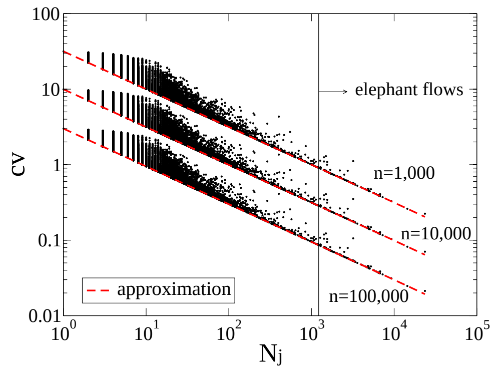

Fig. 3. Coefficient of variation of $\hat{X}_j$.

given by

$$P_j\left[k\right] = \binom{N_j}{k}\binom{N-N_j}{n-k}\bigg/\binom{N}{n} \qquad (5)$$

, which is a hypergeometric distribution. The probability that the jth flow is *not* sampled is given by $P_j\left[0\right]$. Figure 2 illustrates $P_j\left[0\right]$ given $N = 1,000,000$. For instance, in case of $n = 10,000$, $P_j\left[0\right] > 0.95$ for flows with $N_j < 6$, and $P_j\left[0\right]$ is almost zero for elephant flows. Thus, most of the flows with fewer than 6 packets will probably *not* be sampled, and almost all few elephant flows will be sampled. Since most of the flows comprise few packets as shown in Fig. 1 (about 90% of the flows comprise fewer than 6 packets), this leads to great savings in the size of flow tables to be kept; while flows with a larger number of packets will most probably be sampled.

Next, we calculate $cv(\hat{X}_j)$ from the trace. Results given in Fig. 3, the dots are calculated from (3), and the lines correspond to the approximations calculated from (4).

For larger values of N_j, we find that $cv(\hat{X}_j)$ is small enough for good accuracy of estimation. For example, for $n = 10,000$, $cv(\hat{X}_j) < 0.41$ for elephant flows, while $4.45 < cv(\hat{X}_j) < 9.95$ for flows with fewer than 6 packets. We also find that for larger N_j, values for the actual $cv(\hat{X}_j)$ and its approximation are close to each other. For flows with smaller N_j, the variation in packet size makes the coefficient of variations larger than the approximation.

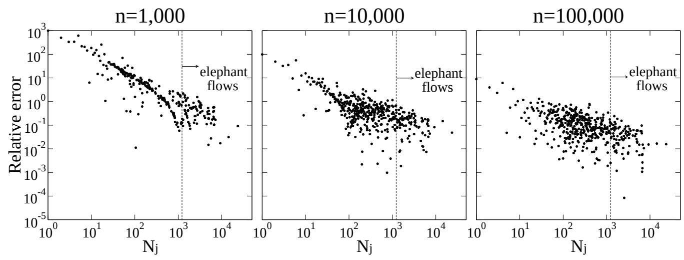

Fig. 4. N_j vs. relative error of $\hat{X}_j$.

TABLE I
PROPORTION OF FLOWS SAMPLED.

	$n = 1,000$	$n = 10,000$	$n = 100,000$
total	0.43%	3.12%	21.39%
elephant	90.00%	100.00%	100.00%

Finally, we give an actual example of estimation and examine its accuracy and cost. Due to restrictions of space, we only give an example using simple count-based packet sampling. Only one sampling trial was done for $n = 1,000$, $10,000$, and $100,000$. Fig. 4 shows the relative error of $\hat{X}_j$ for the flows that are represented in the sample. The relative error is calculated as $\left|\hat{X}_j - X_j\right|/X_j$. We find that the relative error decreases with increasing n_j. This result agrees with that shown in Fig. 3. The average relative errors for elephant flows are 0.47 ($n = 1,000$), 0.15 ($n = 10,000$), and 0.04 ($n = 100,000$), while those for non-elephant flows are 257.01 ($n = 1,000$), 30.85 ($n = 10,000$), and 3.87 ($n = 100,000$). Increasing n thus also improve the accuracy of estimation. We also see that the estimates are much better for elephant flows than for other (non-elephant) flows. Table I shows the proportions of all flows and of elephant flows sampled at different n values. In all cases, most of the elephant flows are sampled, while most of the other smaller flows are *not* sampled. For instance, in case of $n = 10,000$, total number of flows sampled (5,377 flows) is only 3.12% of all flows (172,074 flows), while all elephant flows are sampled. Thus, packet sampling can greatly reduce memory resource for keeping flow statistics. The result is in good agreement with the one shown in Fig. 2. The characteristics we describe may also be useful for identifying elephant flows efficiently.

IV. SUMMARY

We discussed the accuracy and cost of a method for the estimation of traffic volumes in elephant flows, Given N, n, and the distribution of N_j we can use (4) to evaluate the accuracy of estimation and then use (5) to evaluate the cost. The heavy-tailed distribution of flow size means both that we can esitmate the volumes of elephant flows accurately and that we can greatly reduce memory resource for keeping flow statistics.

REFERENCES

[1] W. G. Cochran, Sampling Techniques (3rd ed.), New York: John Wiley & Sons, 1977.
[2] Widely Integrated Distributed Environment Project, http://www.wide.ad.jp.