Ranking flows from sampled traffic

Chadi Barakat, Gianluca Iannaccone, Christophe Diot

***Abstract*— Inverting flow properties from sampled traffic is known to be complex and prone to errors. Previous work has mainly focused on inverting general traffic properties such as flow size distribution, average flow size, or total number of flows. In this work, we study the feasibility of the inversion of individual flow properties. We address this problem by analyzing the detection and ranking of the largest flows from sampled traffic. Surprisingly, our analytical analysis indicates that a high sampling rate (10% and even more) is required. To reduce the sampling rate by an order of magnitude, the ranking must be limited to just a few large flows, or the traffic must consist of several millions of flows. The sampling rate can also be reduced if one is not interested in the relative sizes of the largest flows but just aims at detecting them. We verify our analytical result with trace-driven sampling simulations.**

I. INTRODUCTION

Traffic sampling refers to the action of collecting a subset of the traffic on a network link (using a passive tap or a router line card). The subset is then used to infer the original traffic characteristics. This latter operation is called *inversion* of traffic properties. Sampling is used (1) to monitor very high speed links, where capturing 100% of the traffic becomes a challenge, or (2) to limit the load due to monitoring when resources are shared among multiple tasks (e.g., routers) [7]. Sampling is generally performed either in a periodic fashion (e.g., collecting one packet every time period T) or randomly (e.g., every packet is collected with a constant probability p).

The accuracy of the inversion process is both dependent on the sampling rate and on the property to be inferred. For example, it is easy to infer the total number of packets traversing the link from sampled traffic: the inversion consists in multiplying the number of sampled packets by T or $1/p$, in case of periodic or random sampling, respectively.

Inversion becomes very complex for metrics relative to individual flow properties. The reason is simple. The Internet traffic is known to be composed of a large number of very small flows and a small number of very large flows. Therefore, sampling has more chances to hit the large flows and there is a high probability that many short flows will not appear in the sampled set, which makes the inversion complex. Moreover, if flow duration is defined with a timeout [5], then a flow can be split into multiple subflows if the sampling frequency is too low.

Several works in the literature have studied the inversion problem from sampled traffic. The main focus has been the inversion of aggregate flow properties such as flow size distribution [10], [12], average flow size or total number of flows [9] for a given network link.

As a first step toward the inversion of individual flow properties, we model and analyze how to *detect and rank the largest flows from the sampled traffic*. We choose this metric because it is of particular importance for applications such as traffic engineering [19], [18], detection of traffic anomalies [15], and usage-based pricing [11].

We define the problem as follows. Consider a link monitor that, for a given measurement period, samples packets independently of each other with probability p (random sampling) and classifies them into flows ("sampled flows"). At the end of the measurement period, the monitor sorts all sampled flows, based on their sampled size in packets, and returns an ordered list of the t largest flows (or top t flows), t can be any positive integer. Because of the random nature of sampling, this sampled list may not match the list that could have been obtained without sampling. We try to answer the following question: How well does the list of top t sampled flows match the real list of top t flows?

Different definitions for the match between the two flow lists can be imagined. We start by requiring the two *ordered* lists to be exactly the same. We call the problem that results from this definition the *ranking* of the top t flows. We build an analytical model for this problem and we address it from two different perspectives. First, given a sampling rate p, we identify how accurate the match is. Then, given a desired accuracy, we find the required minimum sampling rate. Surprisingly, our analytical analysis indicates that a high sampling rate is required to obtain a good ranking. Considering a link with thousands of flows with a Pareto flow size distribution, the sampling rate must be above 10% to correctly rank the largest

C. Barakat is with INRIA Sophia Antipolis. This work was done while he was visiting Intel Research Cambridge. Email: chadi.barakat@inria.fr.

G. Iannaccone and C. Diot are with Intel Research Cambridge. Email: gianluca.iannaccone@intel.com, christophe.diot@intel.com.

flows. We find that a sampling rate of 1%[1] allows only the successful ranking of the largest few flows, unless the number of flows on the monitored link is in the order of millions (in which case we can successfully rank more flows at the top). We also study different flow definitions, namely 5-tuple and IP destination prefixes. We find that a coarser flow definition improves the ranking accuracy only if the relative sizes of the largest flows increase as a function of the square root of their sizes. Therefore, contrary to common belief, having larger flow sizes does not always help in accurately detecting and ranking the largest flows.

The poor performance of the ranking motivates us to define a second, less constrained problem: we aim at retrieving the correct list of the largest flows, but, within the list, the flows may be in any order. We call this second problem the *detection* of the top t flows. A new model, based on the ranking one, is built and analyzed. This model confirms that the required sampling rate can be reduced by an order of magnitude if one only focuses on detecting the top t flows without any information on their relative importance.

In this paper, we use packet sampling instead of flow sampling[2] for the following reasons. First, it is the way most of the current monitoring devices operate [4], [14]. Second, flow-based sampling requires to look at the content of the packet at the time of sampling. Flow sampling is therefore very resource intensive and has not yet been implemented given its potential impact on the link monitor performance.

The contributions of this paper are threefold. (1) We perform an analytical study of the problem of ranking two sampled flows and compute the probability that they are *misranked*. We propose a Gaussian approximation to make the problem tractable numerically. (2) Based on the model for the ranking of two flows, we propose two general models to study the ranking and detection problems for any set of flows given a flow size distribution. We define a performance metric for the ranking process and evaluate the impact of each metric's parameter on the accuracy of the ranking. These parameters are the total number of flows on the monitored link, the distribution of flow sizes, and the number of top flows to be ranked. (3) We validate our results on realistic traffic using trace-driven sampling simulations.

The rest of the paper is structured as follows. In the next section we discuss the related literature. In Section III we present and analyze a basic model for the case of two flows. Section IV describes and validates the Gaussian approximation for this model. In Section V we generalize the basic model to the case of ranking the largest flows and we analyze the results in Section VI. In Section VII we address the problem of detecting the largest flows. Section VIII presents a validation with a trace-driven simulation of the ranking process. Section IX concludes the paper.

[1]Today, most router vendors recommend a sampling rate between 0.1% and 1% in order to avoid overloading the routers.

[2]According to flow sampling definition, if a flow is sampled, then all packets belonging to that flow are sampled as well [8], [11].

II. Related work

Some aspects of the inversion of sampled traffic have been studied extensively in the literature. Duffield et al. [9] study the problem of flow splitting and propose estimators for the total number of flows and for the average flow size in the original traffic stream. [10], [12] study the inversion of the flow size distribution with two different methods. They both show that the inversion is not very accurate for the entire distribution, though it can be good for the tail. The inaccuracy raises from the number of flows that are not sampled at all and that need to be estimated with an auxiliary method. [10] shows that periodic and random sampling provide roughly the same result on high speed links, and so random sampling can be used for mathematical analysis due to its appealing features. [3] finds the sampling rate that assures a bounded error on the estimation of the size of flows contributing to more than some predefined percentage of the traffic volume. [8] introduces the idea of smart sampling where the purpose is to isolate flows that contribute considerably to the traffic; this is done by selecting flow records with a probability that increases with the flow size.

The ranking of the largest flows has been studied in the literature, mainly from a memory requirement standpoint [11], [13]. Storing all flows can be a challenge on high speed links. Several methods have been proposed to reduce the memory size while minimizing the impact on the flow ranking. The solution often recommended is to maintain a sorted list of flows in a small memory (re-sorted upon every packet arrival), and to clear some records at the bottom of the list when a new flow appears and the memory is saturated. All the works in the literature assume that if the memory size is well chosen, the largest flows can be detected and ranked with a high precision. However, in the presence of packet sampling, even if the methods rank correctly the set of sampled flows, there is no guarantee that the sampled rank corresponds to the original rank. The problem we address in this paper complements these works by focusing on the impact of sampling on the flow ranking.

III. RANKING TWO FLOWS OF GIVEN SIZES

In this section, we study the probability to misrank two flows of original sizes S_1 and S_2 (in packets). This probability is the basis for the general model for ranking the largest flows. Without loss of generality, we assume $S_1 < S_2$. We consider a random sampling of rate p. Let s_1 and s_2 denote the sizes of both flows after sampling. The two sampled flows are misranked if (i) s_1 is larger than s_2, or (ii) both flows are not sampled, i.e., their sampled sizes equal to zero. By combining (i) and (ii) one can see that the necessary condition for a good ranking is to sample at least one packet from the larger flow (i.e., the smaller of the two flows can disappear after sampling).

The probability to misrank the two flows can then be written as $P_m(S_1, S_2) = \mathbb{P}\{s_1 \geq s_2\}$. We compute and study this probability in the rest of this section. Note that the misranking probability is a symmetric function, i.e., $P_m(S_1, S_2) = P_m(S_2, S_1)$.

Under our assumptions, s_1 and s_2 are distributed according to a binomial distribution of probability p. Hence, we can write for $S_1 < S_2$,

$$P_m(S_1, S_2) = \sum_{i=0}^{S_1} b_p(i, S_1) \sum_{j=0}^{i} b_p(j, S_2). \qquad (1)$$

$b_p(i, S)$ is the probability density function of a binomial distribution of probability p, i.e., the probability of obtaining i successes out of S trials. We have $b_p(i, S) = \binom{S}{i} p^i (1-p)^{S-i}$ for $i = 0, 1, ..., S$, and $b_p(i, S) = 0$ for $i < 0$ and $i > S$.

We do not focus in this section on the case $S_1 = S_2$. We will be accounting for it later when we sum over all possible flow sizes. The probability to misrank two flows of equal size is given by $\mathbb{P}\{s_1 \neq s_2 \text{ or } s_1 = s_2 = 0\} = 1 - \mathbb{P}\{s_1 = s_2 \neq 0\} = 1 - \sum_{i=1}^{S_1} b_p^2(i, S_1)$.

A. Analysis of the misranking probability

The misranking probability depends on the sampling rate p and the sizes of flows S_1 and S_2.

Concerning the dependence on p, it is clear that for any value of S_1 and S_2, the misranking probability tends to 0 when p tends to 1 and to 1 when p tends to 0. Concerning the relative flow sizes, it can be easily proven that the misranking probability becomes smaller when the difference between S_1 and S_2 increases. Indeed, take any two flows of sizes S_1 and S_2, such that $S_1 < S_2$. The flow S_1 can be considered as the aggregation of two flows of sizes $S_1 - k$ and k, $k = 1, 2, ..., S_1 - 1$. If we misrank $S_1 - k$ and S_2, the probability to misrank S_1 and S_2 is equal to 1. Hence, the misranking probability verifies $P_m(S_1, S_2) \geq P_m(S_1 - k, S_2)$, $k = 1, 2, ..., S_1 - 1$. The same reasoning applies to S_1 and $S_2 + k$, $k = S_1 - S_2 + 1, .., 0, 1, ...$, with the misranking probability being a decreasing function of k. Thus, for a flow of size S, the minimum misranking probability is obtained when flow S is compared to a flow of size one packet. This minimum probability is equal to $(1-p)^{S-1}(1-p+p^2 S)$, which tends to zero when S tends to infinity. The maximum misranking probability is reached when S is compared to a flow of similar size.

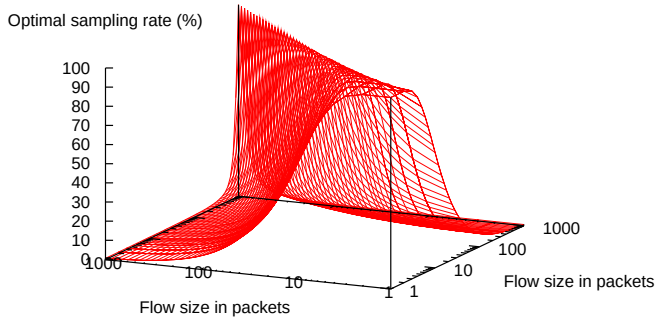

Fig. 1. Optimal sampling rate (log scale)

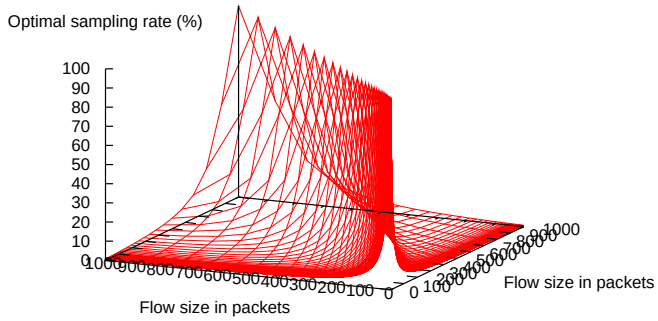

Fig. 2. Optimal sampling rate (normal scale)

B. Optimal sampling rate

For any pair of flow sizes, there is a set of sampling rates that keeps the misranking probability below a certain level. Denote by $P_{m,d}$ the desired misranking probability. When changing p from 1 to 0, the misranking probability increases from 0 to 1. So there exists a sampling rate p_d, function of S_1 and S_2, for which the misranking probability is equal to $P_{m,d}$. Any sampling rate higher (resp. lower) than p_d will lead to a misranking probability smaller (resp. larger) than $P_{m,d}$.

We provide an example on how the sampling rate needs to be adapted to achieve some desired misranking probability. We set $P_{m,d}$ to 0.1% (i.e., we allow the occurrence of one misranking over 1000 trials), then we solve numerically the equation $P_m(S_1, S_2) = P_{m,d}$ for p_d. This gives us the minimum (or the optimal) sampling rate to use, if we want the misranking probability to be below $P_{m,d}$. The results are plotted in Fig. 1 using a log scale for the x and y axis. Clearly, a high sampling

rate is required when flows have similar sizes, and this sampling rate decreases to zero as the difference between the sizes increases. If we take two flows of sizes αS and S with α a real number between 0 and 1, it is clear from Fig. 1 that the optimal sampling rate decreases when S increases (the surface becomes narrower). Generally speaking, the probability of ranking correctly two *large* flows different by α% is higher than in the case of two *small* flows different by the same α%. The conclusion is completely different if we consider two flows of sizes $S-k$ and S, with k a positive integer. Here, the optimal sampling rate increases as S increases, which can be read from Fig. 2, where we plot the optimal sampling rate on a normal scale (the surface becomes wider when S increases). We conclude that it is much more difficult to rank two large flows different by k packets than two small flows different by k packets. This result will be confirmed and analyzed in depth in the next section.

IV. Approximating the misranking probability using Normal distribution

Consider a flow of size S packets that is sampled at rate p. The sampled size follows a binomial distribution. However, it is well known that the binomial distribution can be approximated by a Normal distribution when p is small and when the product pS is of the order of one (flows for which, on average, at least few packets are sampled) [20, pages 108–109]. We assume that this is the case for the largest flows, and we consider the sampled size of a flow as distributed according to a Normal distribution of average pS and variance $p(1-p)S$. Using this approximation, we present a closed form expression for the misranking probability.

Consider two flows of sizes S_1 and S_2 with $S_1 < S_2$. Their sampled versions s_1 and s_2 both follow Normal distributions with averages pS_1 and pS_2, and with variances $p(1-p)S_1$ and $p(1-p)S_2$. We know that the sum of two Normal variables is a Normal variable. So the difference $s_1 - s_2$ follows a Normal distribution of average $p(S_1 - S_2)$ and of variance $p(1-p)(S_1+S_2)$. We have then an approximation for the misranking probability:

$$\begin{aligned} P_m(S_1,S_2) &= \mathbb{P}\{s_1 - s_2 \geq 0\} \\ &= \mathbb{P}\left\{V > \frac{p(S_2-S_1)}{\sqrt{p(1-p)(S_1+S_2)}}\right\} \\ &= \frac{1}{2}\mathrm{erfc}\left(\frac{|S_2-S_1|}{\sqrt{2(1/p-1)(S_1+S_2)}}\right). \quad (2) \end{aligned}$$

V is a standard Normal random variable and erfc(x) $= (\frac{2}{\sqrt{\pi}}) \int_x^\infty e^{-u^2} \mathrm{d}u$ is the complementary error cumulative function. By taking the absolute value for $(S_2 - S_1)$, we make this expression valid for the case $S_1 > S_2$ as well.

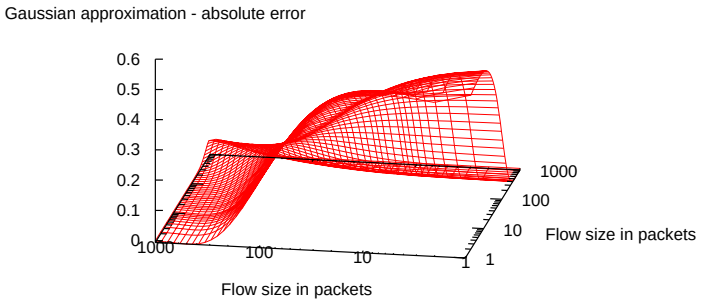

Fig. 3. Gaussian approximation - absolute error, sampling rate 1%

We compare the misranking probability given by the Gaussian approximation to that of the basic model in Eq. (1), and we compute the absolute error. We prefer the absolute error to the relative error because it helps understanding the impact the misranking probability of two particular flows has on the general ranking problem. A large relative error introduced by the Gaussian approximation when the misranking probability is low may have less of an impact on the ranking problem when compared to a small relative error introduced in a region where the misranking probability is high.

We consider different flow sizes and different sampling rates. We find that the absolute error is large when the product pS (i.e., the average number of sampled packets from a flow of size S) is small for both flows (order of 1 or less), which is expected since at this low rate, the Gaussian approximation does not hold. But, when the product pS is large for at least one flow, the absolute error is small and can be neglected. We illustrate this point in Fig. 3, where we plot the absolute error for flow sizes between 1 and 1000 packets and a sampling rate of 1%. We notice how the absolute error is around zero when the size of one flow is larger than 300 packets (pS larger than 3). In that region, the relative error introduced by the Gaussian approximation compared to the basic model can be as small as 1%.

We now use the Gaussian approximation to study how the misranking probability varies with the sizes of both flows, and in particular their difference. Take $S_1 = S_2 - k$, k a positive integer. From (2), the misranking probability increases with S_1 and S_2 (erfc(x) is an increasing function in x). This indicates that it is more difficult to rank correctly two flows different by k packets as their sizes increase in absolute terms. The result is different if we take the size of one flow equal to $\alpha < 1$ times the size of the second, i.e., $S_1 = \alpha S_2$. Here, $(S_1 - S_2)/\sqrt{S_1+S_2}$ is equal to $\sqrt{S_1}(1-\alpha)/\sqrt{1+\alpha}$, which increases with S_1. Hence, the misranking probability given in (2) decreases when S_1 increases. We conclude that, when the two flow sizes maintain the same proportion, it is easier to obtain a

correct ranking when they are large in absolute terms.

We can now generalize the result above. One may think that the larger the flows, the better the ranking of their sampled versions. Our last two examples indicate that this is not always the case. Indeed, the ranking accuracy depends on the relative difference of the flow sizes. In general to have a better ranking, the difference between the two flow sizes must increase with the flow sizes and the increase must be larger than a certain threshold. This threshold is given by Eq. (2): the difference must increase at least as the square root of the flow sizes. This is an interesting finding, indeed. In the context of the general ranking problem, it can be interpreted as follows. Suppose that the flow size distribution has a cumulative distribution function $y = F(x)$. As we move to the tail of the distribution[3], the size of the flows to be ranked increases. The ranking becomes more accurate if the difference between flow sizes increases faster than $\sqrt{x}$. This is equivalent to saying that $\mathrm{d}x/\mathrm{d}y$ should increase with x faster than $\sqrt{x}$. All common distributions satisfy this condition, at least at their tails. For example, with the exponential distribution we have $\mathrm{d}x/\mathrm{d}y \propto e^{\lambda x}$ ($1/\lambda$ is the average), while for the Pareto distribution we have $\mathrm{d}x/\mathrm{d}y \propto x^{\beta+1}$ (β is the shape).

The impact of the flow size distribution on the ranking will be studied in the next sections after we introduce the general ranking problem.

V. Ranking the largest flows

We generalize the previous model to the ranking of the top t flows. Let $N \geq t$ denote the total number of flows available in the measurement interval before sampling. We want the *sorted* sampled list of top t flows to match the list of top t flows in the original traffic. We express in this section the quality of the ranking as a function of the sampling rate p, the flow size distribution, the number of flows to rank t, and the total number of flows N.

A. Performance metric

In order to evaluate how "good" the ranking is, we need to define a performance metric that is easy to compute and that focuses on the top flows. A flow at the top of the list can be misranked (or swapped) with a neighboring large flow or a distant small flow. We want our metric to differentiate between these two cases and to penalize the latter more. We define the metric as follows. We form all flow pairs where the first element is a top t flow and the second element is anywhere in the total list of original flows. The number of these pairs is equal to $N-1+N-2+\cdots+N-t = (2N-t-1)t/2$. We then count the pairs that are swapped after sampling. The number of swapped pairs indicates how good the ranking is at the top of the list. It also allows to differentiate a swap between close flows from that between distant flows. Indeed, if a flow is swapped with its immediate successor in the list, the metric will return a ranking error of 1. Instead, if the same flow is swapped with a distant flow, the number of misranked pairs will become much larger. We have a perfect ranking when the number of swapped flows is equal to zero.

The metric we have described above returns one value for each realization of flow sizes and of the sampled traffic. Given that we want to account for all possible realizations, we define the performance metric as the number of swapped flows *averaged* over all possible values of flow sizes in the original total list and over all sampling runs. We deem the ranking as acceptable when our metric takes a value below one (i.e., on average less than one flow pair is swapped).

B. Computation of the performance metric

Consider a flow of size i belonging to the list of the top t flows in the original traffic (before sampling). We compute the probability that this flow is misranked (or swapped) after sampling with another flow of general size. Denote this probability by $P_{mt}(i)$, where m stands for misranking and t for top. Then, we average over all values of i which gives $\bar{P}_{mt}$.[4] This function gives us the probability that, on average, the top t-th flow is swapped with another general flow. Thus, our performance metric, which is defined as the average number of swapped pairs, is equal to $(2N-t-1)t\bar{P}_{mt}/2$. In the following, we compute the value of $\bar{P}_{mt}$.

Let p_i denote the probability that the size of a general flow is equal to i packets, and P_i denote the flow size complementary cumulative distribution, i.e., $P_i = \sum_{j=i}^{\infty} p_j$. For large number of flows N, we consider safe to assume that flow sizes are independent of each other (see [1]). A flow of size i belongs to the list of top t flows if the number of flows in the original total list, with a size larger than i, is less or equal than $t-1$. Since each flow can be larger than i with probability P_i independently of the other flows, we can write the probability that a flow of size i belongs to the list of the top t flows as $P_t(i,t,N) = \sum_{k=0}^{t-1} b_{P_i}(k, N-1)$, where $b_{P_i}(k, N-1)$ is the probability to obtain k successes out of $N-1$ trials, if P_i is the probability of a success.

[3]Because we are more and more focusing on large flows or because the number of available flows for ranking increases.

[4]Note that the distribution of the size of a flow at the top of the list is different from that of a general flow.

The probability that the t-th largest flow has a size of i packets is equal to $P_t(i) = p_i P_t(i,t,N)/\bar{P}_t(t,N)$. $\bar{P}_t(t,N)$ is the probability that a flow of general size is among the top t in the original total list, which is simply equal to t/N.

Using the above notation we can write the misranking probability between a top t flow of size i packets and any other flow as follows

$$P_{mt}(i) = \frac{1}{P_t(i,t,N)} \left(\sum_{j=1}^{i-1} p_j P_t(i,t,N-1) P_m(j,i) + \sum_{j=i}^{\infty} p_j P_t(i,t-1,N-1) P_m(i,j) \right). \quad (3)$$

In this expression, we sum over all possible sizes of the other flow (the variable j) and we separate the case when this other flow is smaller than i from the case when it is larger than i [5]. $P_m(i,j)$ is the misranking probability of two flows of sizes i and j packets and is given in (1). $\bar{P}_{mt}$ is then equal to $\sum_{i=1}^{\infty} P_t(i) P_{mt}(i)$.

In order to compute $P_m(i,j)$, one can also use the Gaussian approximation given in (2). The Gaussian approximation has the advantage to simplify the numerical analysis since it avoids the two series and the binomial distribution in (1). A second advantage of using the Gaussian approximation is that it transforms the problem from discrete to continuous (flow sizes can be considered as driven by a continuous random variable). This transformation of the problem from discrete to continuous combined with a continuous function for the flow size distribution (e.g., the Pareto distribution) allows us to use classical integral functions. This significantly reduces the computation time to solve the problem (from hours for the original problem to few seconds instead).

However, one has to be careful given that the Gaussian approximation only applies when at least one of the two flows to be compared is large (see Section IV). Since we are interested in the detection and ranking of the largest flows traversing a link, this approximation is appropriate to our case. For this reason, all the results showed in the rest of the paper are obtained using the Gaussian approximation.

VI. Ranking problem: numerical results

We compute the metric $(2N - t - 1)t\bar{P}_{mt}/2$ numerically using (2) and (3). This metric requires as input the flow size distribution p_i. We focus on the Pareto distribution since it is known to be appropriate to model flow sizes in the Internet due to its heavy tailed feature [6]. The Pareto distribution is continuous with a complementary cumulative distribution function given by $\mathbb{P}\{S > x\} = (x/a)^{-\beta}$. $\beta > 0$ is a parameter describing the shape of the distribution and $a > 0$ is a parameter describing its scale. The Pareto random variable takes values larger than a, and has an average value equal to $a\beta/(\beta - 1)$. The tail of the Pareto distribution becomes heavier as β decreases.

We use the measurement results stated in [1] and carried out on the Sprint IP backbone network to set the average flow size. Two flow definitions are considered in [1]: one using the usual 5-tuple made of protocol number, source and destination IP addresses and port numbers, and a second that aggregates packets according to the /24 destination address prefixes. For the first definition, it is stated in Fig. 9 of [1] that the average flow size is equal to 4.8 Kbytes, while for the second definition it is equal to 16.6 Kbytes. To transform the flow size from bytes to packets, we divide by 500 bytes which is a good approximate for the average packet size in the Internet [2]. As for the number of flows in the original list (N), we set it using again the results in Fig. 9 of [1]. It is stated that on the monitored link, the flow arrival rate is equal to 2360 flows/s for the 5-tuple flow definition and to 350 flows/s for the /24 prefix flow definition. We consider a measurement interval larger than one minute, which is a typical value used to report traffic measurements [16]. Then, we use the average number of flows that arrive during this interval to set the total number of flows N. For the case of 5-minutes intervals, this gives a value of N equal to 0.7M flows for the 5-tuple definition and to 0.1M flows for the /24 prefix definition.

In all figures in this section, we plot the ranking metric versus the packet sampling rate p on a log-log scale. We change p from 0.1% to 50%. Every figure shows different lines that correspond to different values of one of the parameters of the model: t, β, N. We are interested in the regions where the value of the metric is below one, indicating that the ranking is accurate on average. To ease the interpretation of results in the figures, the horizontal line of ordinate 1 is plotted as well.

A. Impact of the number of flows of interest

The first parameter we study is the number of ranked flows, t. The purpose is to show how many top flows can be detected and ranked correctly for a given sampling rate. We set β to 1.5 to get a heavy tailed flow size distribution, and we take for the average flow size and total number of flows N the values described previously.

[5] In the case when $j \geq i$, at most $t-2$ flows can be larger than i packets if we want the flow of size i to be in the top t.

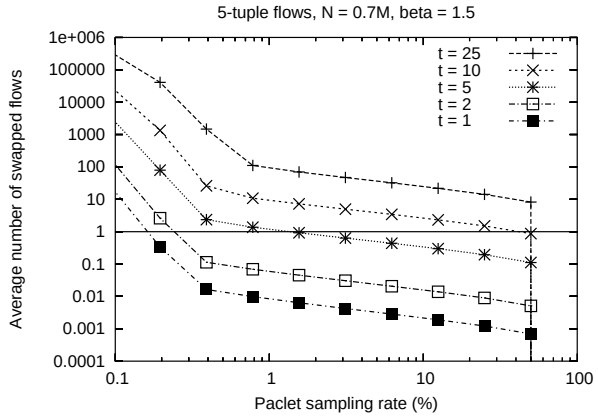

Fig. 4. Performance of sampling with 5-tuple flows varying the number t of top flows of interest

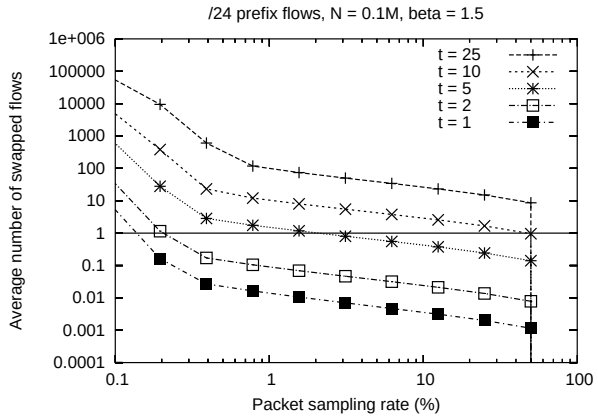

Fig. 5. Performance of sampling with /24 prefix flows varying the number t of top flows of interest

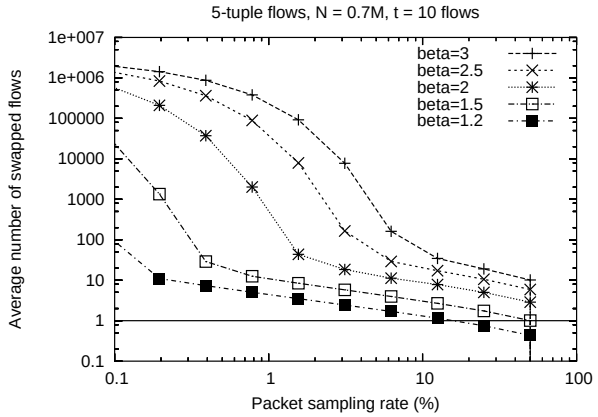

Fig. 6. Performance of sampling with 5-tuple flows varying the flow size distribution

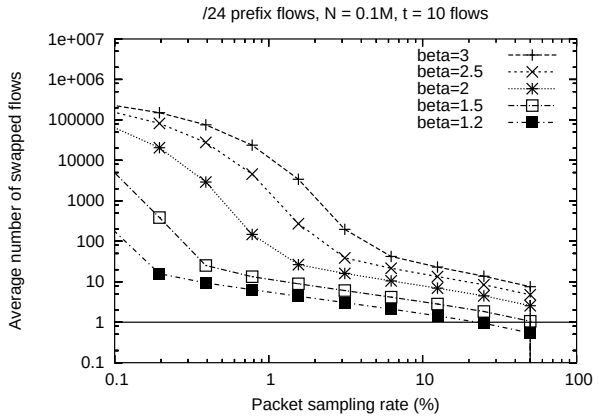

Fig. 7. Performance of sampling with /24 prefix flows varying the flow size distribution

We set the measurement interval to 5 minutes. The performance of ranking the top t flows is shown in Fig. 4 for the 5-tuple flow definition and in Fig. 5 for the /24 destination prefix definition.

We observe that the larger the number of top flows of interest, the more difficult it is to detect and rank them correctly. In particular, with a sampling rate as small as 1%, it is possible to rank at most the top 5 flows. However, as we look at larger values of t, the required sampling rate to get a correct ranking increases well above 10%. Note that with a sampling rate of 0.1%, it is impossible to get a correct flow ranking.

Furthermore, a coarser definition of flows (the /24 destination prefixes) does not provide a significant gain in the ranking accuracy. Even if flows are on average much larger, a sampling rate around 1% is still needed to rank correctly the top few flows.

B. Impact of the flow size distribution

We consider the ranking of the top 10 flows over a 5-minutes interval varying the shape parameter for the Pareto distribution among five distinct values: 3, 2.5, 2, 1.5 and 1.2. Note that for $\beta \leq 2$, the Pareto distribution is known to be heavy tailed (infinite variance). The values taken by our metric are shown in Fig. 6 for the 5-tuple flow definition and in Fig. 7 for the /24 prefix one. We can make the following observations from the figures:

- Given a sampling rate, the ranking accuracy improves as β becomes smaller, i.e., the flow size distribution becomes more heavy tailed. Indeed, when the distribution tail becomes heavier, the probability to obtain larger flows at the top of the list increases, and since it is simpler to rank larger flows (for distributions satisfying the square root condition described in Section IV), the ranking becomes more accurate.
- The ranking is very inaccurate unless the sampling rate is very high. One needs to sample at more than 50% to obtain an average number of misranked flow pairs less than one for a value of β equal to 1.5 (i.e, heavy tailed distribution), and at more than 10% for a value of β equal to 1.2 (i.e., pronounced heavy tailed distribution). For larger values of β (i.e., lighter tail), the sampling rate needs to be close to 100%. Note that for non-heavy tailed distributions,

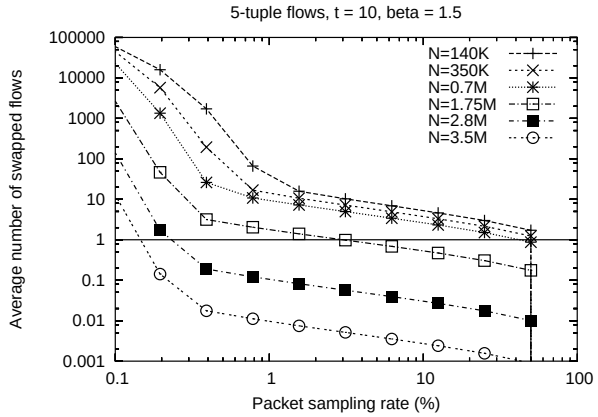

Fig. 8. Performance of sampling with 5-tuple flows varying the total number of flows

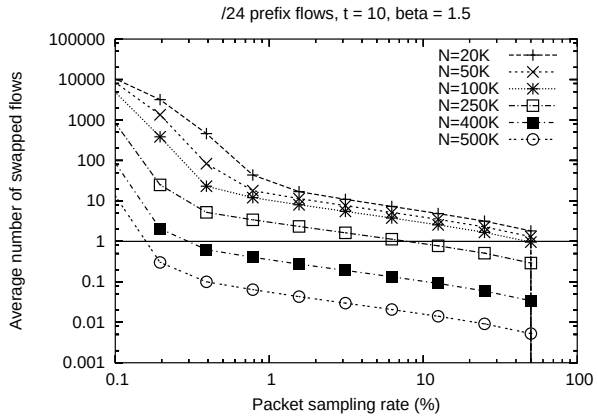

Fig. 9. Performance of sampling with /24 prefix flows varying the total number of flows

the accuracy of the ranking deteriorates very rapidly as the sampling rate goes below 10%. For heavy tailed distributions, the same phenomenon happens when the sampling rate is below 1%.

C. Impact of the total number of flows

Another important parameter in the ranking problem is the total number of flows in the analyzed link/traffic (N). When the total number of flows increases, the flows at the top of the list should become larger, and therefore as we saw in Section IV, the ranking accuracy should improve for flow size distributions satisfying the square root condition (in particular the Pareto distribution we are considering). N varies with the utilization of the monitored link (the higher the utilization, the larger the number of flows). N can also vary with the duration of the measurement interval (the longer we wait before ranking and reporting results, the larger the number of flows).

We study the impact of the value of N on the ranking metric. We take the same value of N used in the previous sections and computed over 5-minutes measurement interval (0.7M flows for the 5-tuple and 0.1M flows for the /24 destination prefixes), then we multiply it by some constant factor ranging from 0.2 (5 times fewer flows) to 5 (5 times more flows). Results are shown in Fig. 8 and Fig. 9. The lines in the figures correspond to a factor value equal to: 0.2, 0.5, 1, 2.5, 4, and 5. In these figures, we consider the ranking of the top 10 flows with β set to 1.5.

The ranking accuracy improves as N increases. For small values of N (140K for 5-tuple and 20K for /24 destination prefixes), a sampling rate of 50% or higher is required for a correct ranking. On the other hand, when N is very large (3.5M for 5-tuple or 500K for /24 destination prefixes), the ranking is very accurate even with a sampling rate of the order of 0.1% (one packet over 1000!). Indeed, with a very large number of flows, the top flows are very large and, thus, easier to rank.

D. Summary of results

We can summarize our main observations as follows: (1) Ranking flows from sampled traffic is not very accurate, unless a very high sampling rate is used (10% and even more). A sampling rate of the order of 1% allows to detect and rank the top few flows. (2) The heavier the tail of the flow size distribution, the better the ranking. (3) The ranking performance improves when there are more flows in the analyzed traffic. For millions of flows, a 1% sampling rate gives good results. (4) There is no significant difference in the ranking performance between the 5-tuple and the /24 destination prefix flow definitions.

VII. Detection of the largest flows

In this section, we relax the requirement for the ranking in order to reduce the required sampling rate. Previously, we considered the problem of identifying the t largest flows and, at the same time, to list them in the correct order. Here, we focus on the identification of the top t flows, alone. We use the same mathematical tools as in the previous sections, but we need to define a different performance metric. We are now interested in the ranking between any flow in the top t list with only those flows *outside* this list. We expect the required sampling rate to be lower.

A. Performance metric

We define the metric for the detection problem as being the average number of swapped flow pairs, where the first element of a pair is in the list of top t flows and the second element outside this list. We compute this metric as follows. Consider the probability that a flow among the top t is swapped with a flow that does belong

to the top t. Let $\bar{P}^{\star}_{mt}$ denote this probability. Following the same approach described in Section V, we can write

$$\bar{P}^{\star}_{mt} = \frac{1}{\bar{P}^{\star}_{t}} \sum_{i=1}^{\infty} \sum_{j=1}^{i-1} p_i p_j P^{\star}_{t}(j,i,t,N) P_m(j,i).$$

In order to obtain $\bar{P}^{\star}_{mt}$, we sum over all possible size values for the flow in the top t (index i) and all possible values for the other flow not among the top t (index j). In this expression, p_i and p_j represent the probability that the size of a flow is equal to i or j packets, respectively. $P_m(j,i)$ is the probability that two flows of sizes i and j are misclassified – we use again the Gaussian approximation described in (2). $P^{\star}_{t}(j,i,t,N)$ is the joint probability that a flow of size i belongs to the list of top t flows while another flow of size j does not belong to it (i.e., it is in the bottom $N-t$ flows). $\bar{P}^{\star}_{t}$ is the joint probability that a flow of any size belongs to the list of top t flows while another flow of any size does not belong to this list. It is equal to $t(N-t)/(N(N-1))$.

We now compute $P^{\star}_{t}(j,i,t,N)$ for $j < i$, i.e., the probability that flow i belongs to the top list while flow j does not. The number of flows larger than i should be smaller than t, while the number of flows larger than j should be larger than t. The probability that a flow size is larger than i is $P_i = \sum_{k=i}^{\infty} p_k$. The probability that it is larger than j is $P_j = \sum_{k=j}^{\infty} p_k$. The probability that a flow size is between j and i given that it is smaller than i is $(P_j - P_i)/(1-P_i)$. We call it $P_{j,i}$. It follows that:

$$P^{\star}_{t}(j,i,t,N) = \sum_{k=0}^{t-1} b_{P_i}(k, N-2) \sum_{l=t-k-1}^{N-k-2} b_{P_{j,i}}(l, N-k-2).$$

The first sum accounts for the probability to see less than t flows above i packets. The second sum accounts for the probability to see more than t flows above j given that k flows ($k < t$) were already seen above i. For $t = 1$, $P^{\star}_{t}(j,i,t,N)$ is no other than $P_t(i,t,N-1)$, and both $\bar{P}^{\star}_{mt}$ and $\bar{P}_{mt}$ are equal (i.e., the ranking and the detection problems are the same).

Once $\bar{P}^{\star}_{mt}$ is computed we multiply it by the total number of flow pairs whose one element is in the top t and the other one is not, which is equal to $t(N-t)$. This is our metric for the detection problem. As for the ranking problem, we want this metric to be less than one for the detection of top t flows to be accurate.

B. Numerical analysis

To illustrate the difference between ranking and detection, we consider the same scenario as in Section VI-A. We plot the detection metric as a function of the sampling rate, for different values of t (the number of large flows of interest) and for the two flow definitions (Fig. 10 and Fig. 11). A comparison between these

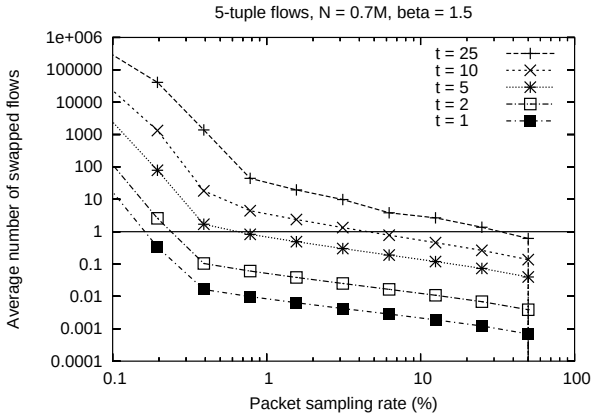

Fig. 10. Detection of the largest flows (5-tuple)

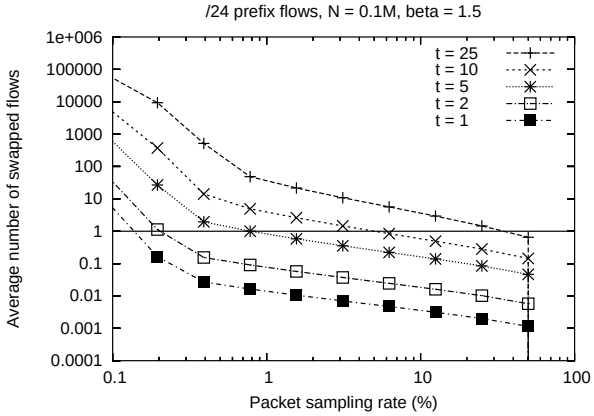

Fig. 11. Detection of the largest flows (/24 prefix)

results and their counterparts in Fig. 4 and Fig. 5 shows a significant gain in the detection case. All plots are shifted down. For example, in the case of ranking, the required sampling rate was higher than 50% to rank correctly the top 10 flows for a value of the shape parameter for the flow size distribution equal to 1.5. Now this can be done with a sampling rate in the order of 10%. The same gain can be observed if we reconsider the other scenarios in previous sections (not presented for lack of space). We can also observe in the figures how the aggregation of flows using the /24 destination prefix does not improve the accuracy of top flow detection.

VIII. Trace-driven simulation

We have been studying the ranking problem by computing the average value of our performance metric under the assumption that flow size distributions are perfect Pareto and that all the packets of the flows are available for sampling and ranking. In reality, this may not be the case. The distribution of flow sizes may deviate from Pareto. In addition, flows can be truncated if they last for more than the measurement interval. Indeed, some network operators may choose to use a "binning" method, where packets are sampled for a time interval,

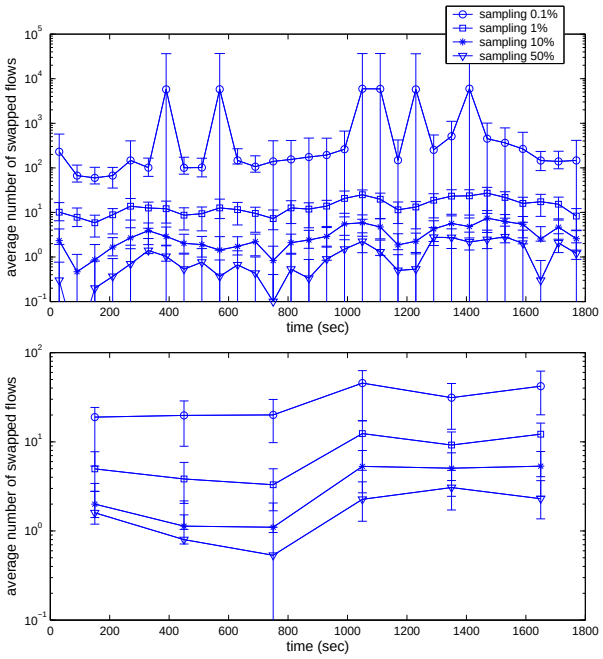

Fig. 12. Performance of ranking vs. time, 5-tuple, top 10 flows

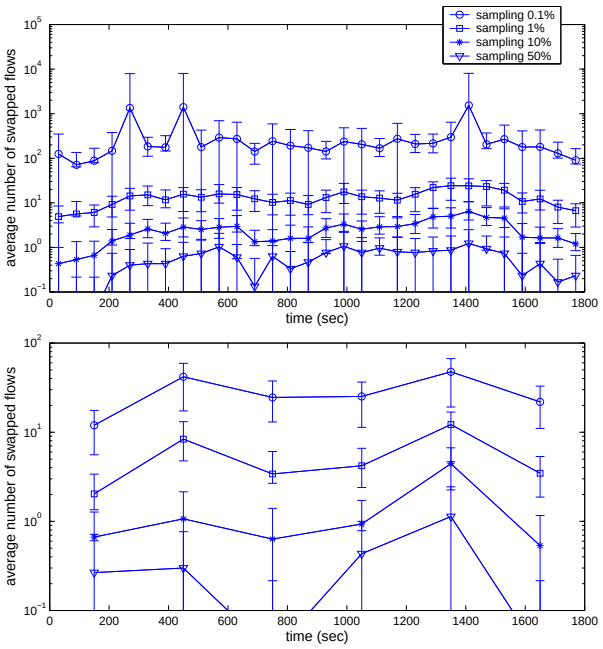

Fig. 13. Performance of ranking vs. time, /24 prefix, top 10 flows

classified into flows, ranked, and then reported. At the end of the interval, the memory is cleared and the operation is repeated for the next measurement interval. With this binning method, all flows active at the end of the measurement interval are truncated, so that not all sampled packets of the truncated flow are considered at the same time for the ranking. The truncation may, therefore, penalize large flows and alter the tail of the flow size distribution (where flows are of large size and probably last longer than the measurement interval).

We run real-time trace-driven simulations using the binning method to illustrate that the assumptions we made in the models do not change our results. Furthermore, we want to show that these results hold when we sample and rank real flows, collected on an operational network. Finally, we want to study how the performance metric deviates from its average over several measurement intervals as well as over multiple sampling runs.

A. Simulation setup

We consider the trace used to plot the Fig. 9 in [1]. This trace is collected on a OC-12 (622 Mpbs) link in Sprint IP backbone late 2001. It is a 30-minutes trace where we have for both definitions of flows (5-tuple and /24 destination address prefixes) the sizes of all flows, the durations of all flows and their starting times. The monitored link carries an average traffic of 90Mbps. The information we have on this trace does not include however the instants of arrivals of individual packets within each flow neither their sizes. We then need to transform the flow level trace into a packet level trace in order to study the ranking problem. To this end, we generate artificially the packets of a flow during its lifetime. We take all packets equal to 500 bytes, a typical value for the average packet size in the Internet [2]. For a flow of size S, duration D and starting time T (all given by the trace), we compute first the number of packets for this flow, then we distribute these packets uniformly in the interval $[T, T+D]$ (for long flows this is equivalent to saying that packets are the realization of a homogenous Poisson process). The multiplexing of the instants of arrivals of packets for all flows gives the packet level trace we are looking for. Note that the packet arrival process is not a crucial aspect of the simulation given that we are looking at large measurement intervals (1 to 5 minutes) when compared to the average flow duration of 13 seconds [1]. However, we have also validated our results on NLANR packet-level traces [17] and briefly present some results in section VIII-C.

Once we have generated the trace, we sample it at various sampling rates, we cut it into bins (a bin is a measurement interval), we collect flows within each bin, and finally we rank them. We do the same process (classification and ranking) for the trace before sampling. By comparing the ranking before and after sampling, we are able to compute our metric for each bin.

For each sampling rate we conduct 30 runs over which we compute the average of the metric as well as its standard deviation. The average and standard deviation are then plotted versus the bin number (or time) for different bin values and for different sampling rates. The standard deviation is plotted as an error bar around the

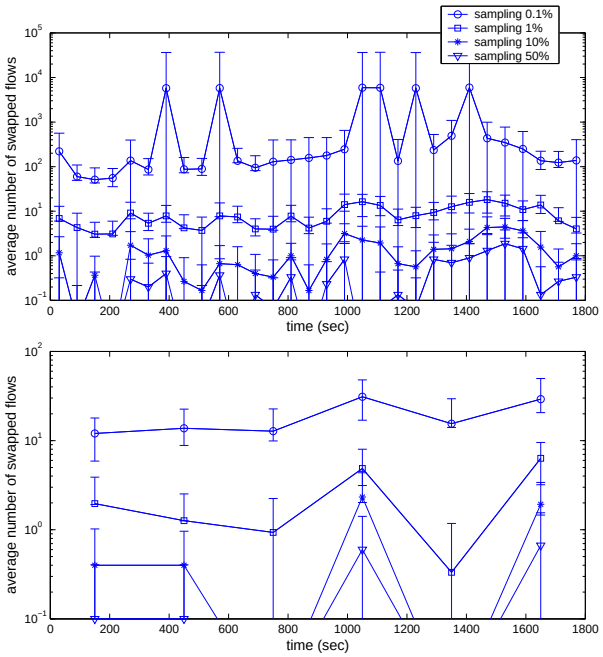

Fig. 14. Performance of detection vs. time, 5-tuple, top 10 flows

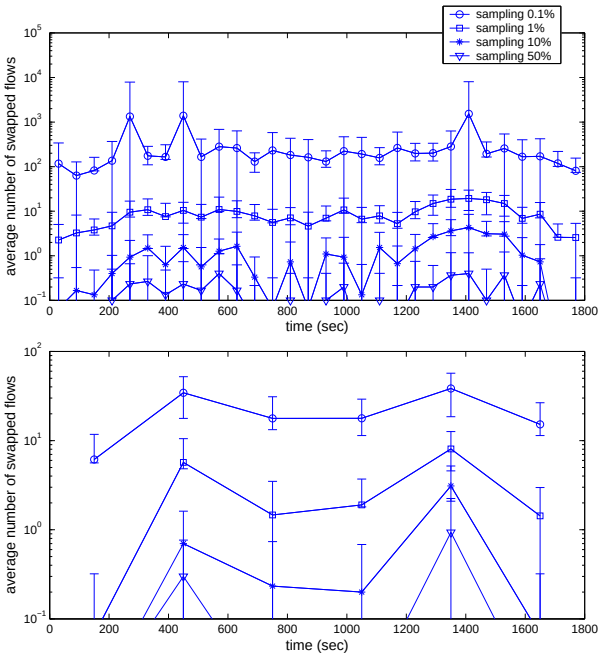

Fig. 15. Performance of detection vs. time, /24 prefix, top 10 flows

average; it tells us how stable the ranking of the largest flows is when we run the simulation multiple times at the same sampling rate.

B. Simulation results

We show the results for the detection and ranking of the top 10 flows and two bin values: 1 and 5 minutes. The results for the ranking are plotted in Fig. 12 for the 5-tuple and in Fig. 13 for the /24 destination prefixes. We plot the results for the detection of the top 10 flows in Fig. 14 for the 5-tuple and in Fig. 15 for the /24 destination prefixes. Each line in the plots corresponds to a different sampling rate.

The trace-drive sampling simulations confirm our analytical results. A sampling rate of the order of 50% is required to accurately detect and rank the top 10 flows (the average plus standard deviation of the metric should be below the horizontal line of ordinate 1). A sampling rate of the order of 10% works sometimes, whereas a sampling rate of 1% never works. Only detecting the top 10 flows is somewhat "easier": a sampling rate of 10% gives very good results.

We observe that the accuracy of the detection and ranking varies over time. It also varies over multiple runs of the simulation at the same sampling rate. Indeed, moving from one bin to another, the sizes of flows change as well as the total number of flows. This leads to a different ranking error value. Also, when we rerun the simulation, the sampled sizes of flows change (different realization of a stochastic process) and hence the result of the ranking process.

Concerning the increase in the bin length (from 1 to 5 minutes), we notice a slight improvement in the performance, which is also in agreement with our analytical results (first and third lines from top in Fig. 8 and 9). The simulations also show that the aggregation of flows using the /24 destination prefixes does not improve significantly the performance of detection and ranking, even though aggregated flows become larger.

C. Additional validation on the Abilene network

To further validate our analytical results, we sample and rank the 10 largest flows on a 30-minutes trace collected by NLANR on an OC-48 link in the Abilene backbone network (Abilene-I [17]). In contrast to the Sprint trace, the Abilene trace provides full information on the instants of arrivals of packets and the flows to which they belong, so we do not need to generate the packet arrivals as before.

This trace provides additional insights on the performance of ranking from sampled traffic. The link from which the trace has been collected has a higher utilization as well as a larger number of flows, and the flow size distribution exhibits a short tail feature.

In Fig. 16, we plot the ranking performance metric for the top 10 flows as a function of time for one minute bins and for different sampling rates. Again, we average the results over 30 runs and show the standard deviation.

The figure confirms the difficulty of ranking the largest flows with a low sampling rate. The performance is worse than with the Sprint trace: a sampling rate above

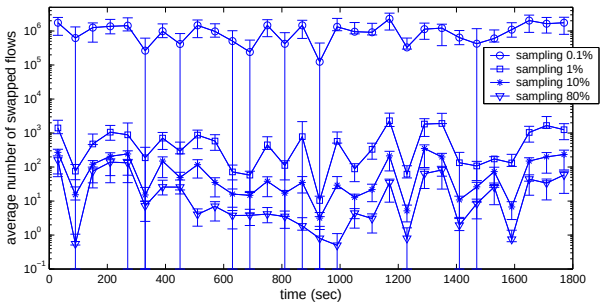

Fig. 16. Performance of ranking vs. time, 5 tuple, top 10 flows, Abilene network

50% is required. This poor performance is the consequence of the short tail of the flow size distribution. This is again in agreement with our analytical results (Section VI-B). We observe that the error increases very fast when the sampling rate drops below 1%.

IX. Conclusions

We study the problem of detection and ranking the largest flows from a traffic sampled at the packet level. The study is done with stochastic tools and trace-driven simulations. We find that the ranking accuracy is strongly dependent on the sampling rate, the flow size distribution, the total number of flows and the number of largest flows to be detected and ranked. By changing all these parameters, we conclude that ranking the largest flows is not that accurate in presence of traffic sampled at the packet level. For example, to detect and rank the top 10 flows, a sampling rate higher than 10% is required. A sampling rate of the order of 1% works well when the total number of flows is very large (of the order of millions), or when we focus on the very few top flows. Our study also shows that the required sampling rate decreases by an order of magnitude if one only wants to detect the largest flows and is not interested in their relative importance.

We are currently exploring three possible future directions for this work. First, we want to study the accuracy of the ranking when the sampled traffic is fed into one of the mechanisms proposed in [11], [13] for sorting flows with reduced memory requirements. A second direction is to study the feasibility of refining the ranking using some protocol-specific information carried in the headers of the sampled packets. For example, one can imagine the use of the TCP sequence numbers to better estimate the size of the sampled flows. Such method will improve the ranking of the largest flows, although its main drawback is the lack of generality (e.g., it does not apply to flows defined based on the address prefixes, or when the protocol headers are encrypted). Finally, the third direction we are exploring is the use of adaptive schemes that set the sampling rate based on the characteristics of the observed traffic.

References

[1] C. Barakat, P. Thiran, G. Iannaccone, C. Diot, and P. Owezarski. Modeling Internet backbone traffic at the flow level. *IEEE Transactions on Signal Processing (Special Issue on Signal Processing in Networking)*, 51(8):2111–2124, Aug. 2003.

[2] CAIDA: Cooperative Association for Internet Data Analysis. http://www.caida.org.

[3] B. Y. Choi, J. Park, and Z. Zhang. Adaptive packet sampling for flow volume measurement. Technical Report TR-02-040, University of Minnesota, 2002.

[4] Cisco Systems. NetFlow services and applications. White Paper, 2000.

[5] K. C. Claffy, H.-W. Braun, and G. C. Polyzos. A parameterizable methodology for internet traffic flow profiling. *IEEE JSAC Special Issue on the Global Internet*, Mar. 1995.

[6] M. Crovella and A. Bestravos. Self-similarity in the World Wide Web traffic: Evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835–846, Dec. 1997.

[7] N. Duffield. A framework for passive packet measurement. IETF PSAMP WG Internet Draft, Dec. 2003. draft-psamp-framework-05.

[8] N. G. Duffield and C. Lund. Predicting resource usage and estimation accuracy in an IP flow measurement collection infrastructure. In *Proceedings of ACM Sigcomm Internet Measurement Conference*, Oct. 2003.

[9] N. G. Duffield, C. Lund, and M. Thorup. Properties and prediction of flow statistics from sampled packet streams. In *Proceedings of ACM Sigcomm Internet Measurement Workshop*, Nov. 2002.

[10] N. G. Duffield, C. Lund, and M. Thorup. Estimating flow distributions from sampled flow statistics. In *Proceedings of ACM Sigcomm*, Aug. 2003.

[11] C. Estan and G. Varghese. New directions in traffic measurement and accounting. In *Proceedings of ACM Sigcomm*, Aug. 2002.

[12] N. Hohn and D. Veitch. Inverting sampled traffic. In *Proceedings of ACM Sigcomm Internet Measurement Conference*, Oct. 2003.

[13] J. Jedwab, P. Phaal, and B. Pinna. Traffic estimation for the largest sources on a network, using packet sampling with limited storage. Technical Report HPL-92-35, HP Laboratories, Mar. 1992.

[14] Juniper JunOS. http://www.juniper.net.

[15] A. Lakhina, M. Crovella, and C. Diot. Diagnosing network-wide traffic anomalies. In *Proceedings of ACM Sigcomm*, Aug. 2004.

[16] M. Lambert. A model for common operational statistics. RFC 1857, Oct. 1995.

[17] NLANR: National Laboratory for Applied Network Research. http://www.nlanr.net.

[18] K. Papagiannaki, N. Taft, and C. Diot. Impact of flow dynamics on traffic engineering design principles. In *Proceedings of IEEE Infocom*, Hong Kong, China, Mar. 2004.

[19] A. Shaikh, J. Rexford, and K. G. Shin. Load-sensitive routing of long-lived IP flows. In *Proceedings of ACM Sigcomm*, Sept. 1999.

[20] M. Spiegel. *Theory and Problems of Probability and Statistics*. McGraw-Hill, 1992.