Flow Sampling Under Hard Resource Constraints

Nick Duffield Carsten Lund Mikkel Thorup
AT&T Labs–Research
180 Park Avenue, Florham Park, NJ 07932, USA
E-mail: {duffield,lund,mthorup}@research.att.com

Abstract— **Many network management applications use as their data traffic volumes differentiated by attributes such as IP address or port number. IP flow records are commonly collected for this purpose: these enable determination of fine-grained usage of network resources. However, the increasingly large volumes of flow statistics incur concomitant costs in the resources of the measurement infrastructure. This motivates sampling of flow records.**

This paper addresses sampling strategy for flow records. Recent work has shown that non-uniform sampling is necessary in order to control estimation variance arising from the observed heavy-tailed distribution of flow lengths. However, while this approach controls estimator variance, it does not place hard limits on the number of flows sampled. Such limits are often required during arbitrary downstream sampling, resampling and aggregation operations employed in analysis of the data.

This paper proposes a correlated sampling strategy that is able to select an arbitrarily small number of the "best" representatives of a set of flows. We show that usage estimates arising from such selection are unbiased, and show how to estimate their variance, both offline for modeling purposes, and online during the sampling itself. The selection algorithm can be implemented in a queue-like data structure in which memory usage is uniformly bounded during measurement. Finally, we compare the complexity and performance of our scheme with other potential approaches.

I. Introduction

A. Background

The rational management and control of IP networks requires detailed knowledge of the demand and consumption of network bandwidth. Specifically, this entails measuring bandwidth usage—as measured in packets or bytes—broken down by attributes such as source and destination IP address or prefix, and TCP/UDP port numbers. Such a segmentation into traffic classes is used to understand, e.g., the amount of use of different applications, the amount of traffic generated by different customers or organizations, or the origins of traffic at a congested network interface. In operating an IP network, there are typically thousands of such classes of concurrent interest. Some further applications are described in [13].

On the other hand, the ever increasing speeds of transmission links and volume of traffic that traverses them, presents formidable challenges for measurement. To see why this is case, consider that most prevalent traffic measurements entail routers or other network elements constructing reports either at the level of individual packets, or at the level of packet flows. (A flow is a set of packets observed within some time window, and that share a common attribute, e.g, those mentioned in the previous paragraph). These reports are exported to a collector where the measurements are collated and stored for use by applications.

The first challenge is at the point of measurement; the measurement functions of routers may not be able to operate at line speed given the resources available to them. The second challenge is that the volume of complete measurements of all traffic would necessitate large costs, both in the bandwidth required to transmit them to the collector, and the resource needed to store and process the measurements at the collector.

These considerations motivate using some form of data reduction. Candidate techniques include filtering (the selection of subsets of measurements according to packet attribute), aggregation (the amalgamation of data across ranges of attributes) and sampling (the random selection of of a subset of representatives from the data). Of these techniques, sampling allows the retention of measurement detail down to the finest attribute level. This is essential, for although aggregation can be useful for providing reports on well understood network behavior, it inhibits the discovery of new network behavior whose detail is hidden within the aggregates. Thus while the collector typically must support a large number of different aggregators (over key ranges and time periods) for routine reporting, it must also allow arbitrary aggregate to be defined on an ad-hoc basis for exploratory studies on the sampled flow records.

Several forms of sampling for IP measurement have been deployed or proposed. Sampling at the measurement point, by statistical [18] or hash-based [7] methods is under consideration for inclusion in packet sampling standards [6]. In Cisco's Sampled NetFlow [5], a router forms flow reports from a sampled substream of all packets passing through it in order to control the amount of fast and expensive memory necessary to cache flow information while the statistics are formed. Sample and Hold [12] is a generalization of this idea that focuses attention on longer flows that, in practice, account for the bulk of the network traffic [14].

Sampling within the measurement infrastructure is also used to further reduce data volumes. Most closely related to the work of the present paper is size dependent sampling of completed flow reports, first proposed in [9], [10], also known as *smart sampling*. The motivation comes from the empirical fact that the distribution of bytes per flow is heavy tailed [14]. This makes uniform sampling a poor choice, since random omission of a report on a large flow can cause a large fluctuation in the perceived volume of traffic. Instead, a report on a flow of x bytes with a probability $p(x) > 0$ depending on x. When a flow happens to be sampled, the size of the flow is estimated as $x/p(x)$, i.e., the estimated size is renormalized upwards in order to compensate for its chance of not begin sampled. This quantity is an unbiased estimator of the actual byte usage x.

In [9] it was shown that given a *sampling threshold* z, the choice $p(x) = p_z(x) := \min\{1, x/z\}$ is optimal in the sense that it minimizes the cost $V_X + z^2 N$. Here, V_X is the variance of the estimator of the byte total X of a set of flows, and N is the average number of flows that are sampled from the set. Thus z expresses the preference in the trade-off between the opposed goals of having small estimator variance (if z is small) and selecting a small number of samples (if z is large). All flows of size $x \geq z$ are selected with probability 1; this avoids the high estimation variance associate with uniform sampling of large flows. Flows smaller than z are selected with a probability proportional to their size.

The variance spoken of above is entirely due to sampling. The flow sizes are viewed as fixed quantities to be estimated; no statistical properties of the flow sizes is modeled or assumed. Flow sampling is conditionally independent in the sense that once flow sizes are fixed, the sampling decisions are independent. This viewpoint gives considerable formal simplification since the time series of flow sizes is believed to exhibit some dependence. If one tried to model and exploit statistical properties of a set of flow reports, size dependent sampling would be dependent between flows.

Estimator variance is controlled through selection of the sampling threshold z. It can be show that the variance of the unbiased estimate $\widehat{X}$ of the total bytes X in a set of flows is bounded above as $\mathsf{Var}(X) \leq zX$, regardless of the distribution of X amongst the constituent flows. Thus the relative error in estimating X decreases inversely as the square root of of X. On the other hand, the number of samples taken is not further controlled beyond the optimality of the choice of p_z.

B. Motivation

This is paper is motivated by the need to sample flow records under hard resource constraints in the measurement infrastructure. As noted above, a fixed sampling threshold enables one to bound estimation variance: we will speak of this as *variance constrained sampling.* But the number of selected flow records depends on the rate at which flow records are offered, and the distribution of flow lengths that they represent. To see this, note that from a set of flow records indexed by $i \in I$ and with lengths $\{x_i : i \in I\}$, the expected number of sampled records is $\sum_{i\in I} p_z(x_i)$.

As the rate and composition of the traffic varies, even the average number of samples selected can vary considerably. Such variations can have a number of origins, for example:

- Periodic variations in the traffic rate, e.g., over a day or a week;
- Sudden and relatively short lived increases in network traffic due to network attacks, or flash crowds: typically sudden widespread interest by users in a web site;
- Long-term or fluctuating changes in traffic patterns due to routing changes, e.g., around network outages.

It is important that the sampling strategy be able to limit the number of flows that are sampled in order to control resource usage. If resources available to transmit and process flow records are exceeded, then the flow records will be dropped indiscriminately. As pointed out, in [8] this can be regarded as an uncontrolled uniform sampling of the flow records. This would play havoc with the estimation of usage. Because of the observed heavy tailed nature of flow lengths, see [14], usage estimates are very sensitive to the omission of even a single large flow. This observation motivates using smart sampling rather than uniform sampling for flow records.

On the other hand, it is important to ensure sufficient samples are taken in order to satisfy the accuracy requirements of applications when estimating usage in each of the large number of traffic classes. Smart sampling with fixed threshold z may cause over or undersupply of sampled records, depending on the offered traffic.

One approach to control the sampling rate is to dynamically adjust the sampling threshold z in order to match the rate of selection of flow reports to a target. When z is increased, fewer reports are sampled, while when z is decreases, more reports are sampled. A multiplicative control scheme to adjust z has been considered in [10]. However, such control schemes cannot fulfill hard constraints on the selection under arbitrary statistical and systematic variation in the offered traffic. Furthermore, it is necessary to systematically undersample (i.e. arranging for the mean sampling rate to be lower than the target) in order that unexpected variations in the offered reports do not cause the selection rate to frequently exceed the target. In the face of large traffic fluctuations, such undersampling may lead to certain classes of traffic having too few samples.

These considerations motivate the development of a sampling strategy that is able to sample a fixed number m of flow records in each measurement interval, independent of the offered load (assuming that it exceeds m), out of which unbiased estimates of usage in all classes can be formed.

C. Contribution and Outline

We consider the measurement period to be divided into a number of successive windows. In each window we have target number m of flow reports to be selected. During a given window, n flow reports of sizes $\{x_i : i = 1, \ldots n\}$ arrive. We will always assume that $n > m$. Each flow also carries a key, the values of the attributes that distinguish the packets of one flow from those of another. The key of flow report i is denoted by c_i.

Our problem is as follows. For each key c in the set $C = \cup_{i=1,\ldots n}\{c_i\}$ of keys represented in an interval, we wish to retain, through sampling, sufficient data to form unbiased estimates of $X(c) = \sum_{i:\ c_i=c} x_i$, i.e., the total bytes reported with key c. Furthermore, we wish to do accomplish this while storing information relating to at most m flow keys. (This last requirement precludes simply aggregating flow reports over all keys present without sampling, since one could have $\#C = n$.)

In our solution to this problem we will store m of the flow reports, with sizes renormalized in order to enable unbiased estimation. In applications, these selected flow reports would be forwarded at the end of the measurement window to a collector, where estimation of arbitrary estimates is actually performed. The solution to the problem has two components: (i) an algorithm that specifies which m flows are to be selected, and (ii) an implementation of the algorithm that requires at most only $m + 1$ keys to be retained at any time during the window.

The rest of the paper is organized as follows. In Section II we describe the algorithm by which flows are selected, and establish the key property that it yields unbiased estimates $X(c)$ of the $X(c)$. Much of the analysis is couched in term of the theory of the order statistics of nonidentically distributed random variables. Compared with order statistics for identical random variables, there are far fewer general results available in the non-identical case. Some results that we derive in this area are new, as far as we are aware.

In Section III we show how the whole framework of this paper can be extended to selection probabilities stemming from a wider class of cost functions than those considered in [9], [10].

In Section IV we analyze the variance of the unbiased estimators derived in Section II. We perform this from two standpoints: online and offline. In the online version, we construct unbiased estimates of the variances $\mathsf{Var}(X(c))$ themselves, in terms of the measured data alone. Equipped with these, we can directly establish the likely error in estimates as they are made. In the offline version, we calculate the likely variance associated with a given set of flow records for the purpose of dimensioning the measurement infrastructure, and in particular m, during its design. We establish bounds for the estimation variance, and give some limiting forms in the case that the number of slots m and the number of offered flow reports n becomes large. Perhaps surprisingly, it turns out that the offline version is considerably harder than the online version.

In Section VI we turn to describe a queue-based implementation of our algorithm that requires the retention of at most m flow reports during each window. We analyze its performance, and compare with that of other potential approaches to the problem of smart sampling under hard space constraints.

In Section VII we illustrate its performance using measured flow data. Section VIII reviews work related to our proposal. We summarize and conclude in Section IX. The proofs of most of the theorems are postponed until Section X.

II. Sampling & Unbiased Estimation

In this section we show how to sample a fixed number m of flow reports from a population of $n > m$ flow reports. The problem we wish to solve is as follows. The population flow reports have sizes $\{x_i\ :\ i \in I\}$ where $I = \{1, \ldots, n\}$. Each report i contains a key c_i. The problem is to sample m reports while maintaining the ability to form unbiased estimators $X(c)$ of the usage $X(c) = \sum_{i:\ c_i = c} x_i$ for each color c in $C = \cup_{i\in I}\{c_i\}$.

A. Sampling a Fixed Number of Flows

In summary, the solution to the problem is as follows. For each flow i we generate a *random* threshold z_i as follows: generate an independent random variable ω_i uniformly distributed in $(0, 1]$ and set $z_i = x_i/w_i$. We then pick the m items with the largest z_i. Moreover, let z' be the $(m + 1)^{\text{st}}$ largest z_i. For each of the items i picked, we renormalize the usage to be $\max\{x_i, z'\}$. This is the complete scheme. The remainder of this section is devoted to explaining the intuition behind this scheme, and showing that the usage estimates formed from it are unbiased.

Some intuition behind this role of z' can be gained by observing that, knowing the z_i values of $n - m$ unselected flows, z' is the largest known sampling threshold under which the m flows of largest z_i would certainly be selected. Analogous to the case of a fixed sampling threshold, the selected flow sizes must be renormalized through division by an appropriate selection probability in order to create unbiased estimators. Here, if flow i is selected, the usage x_i is renormalized through division by the *random* probability $p_{z'}(x_i)$, yielding $x_i/p_{z'}(x_i) = \max\{x_i, z'\}$. In Theorem 1 below, it will be shown that the resulting usage estimates are unbiased, i.e., the average if the estimates over all possible sampling decisions yields the true usage.

B. Unbiased Usage Estimation

We now define the sampling thresholds more completely, and setup the notation that will be used to state an proved Theorem 1. Let $\omega_i\ :\ i \in I$ be independent random variables identically uniformly distributed in $(0, 1]$. Define $z_i = x_i/\omega_i$, and let $z^{-1}_{(1:I)} \le z^{-1}_{(2:I)} \le \ldots \le z^{-1}_{(I:n)}$ be the order statistics of the collection $\{z_i^{-1}\ :\ i \in I\}$. Since the z_i are continuously distributed, ties occur with probability zero and can be broken by any convenient method. Except for the discussion of variance, the set I is fixed, and we shall often write $z^{-1}_{(i:I)}$ simply as $z^{-1}_{(i)}$. Observe $\mathsf{P}[z_i \ge z] = \mathsf{P}[w_i \le x_i/z] = p_z(x_i)$.

We can consider the ordering of the z_i^{-1} as being realized by a random map π that permutes the elements of I as $z^{-1}_{(i)} = z^{-1}_{\pi(i)}$. Given a quantity such as x_i, with index $i \in I$, we shall use the notation $x_{[i:I]}$ (or simply $x_{[i]}$) to denote quantity $x_{\pi(i)}$ concomitant with the order statistics of z_i^{-1}. Thus $x_{[1]}, x_{[2]}, \ldots, x_{[n]}$ represent the values x_i sorted in decreasing order of z_i. In particular, we write $z_{[i]}$ for $1/(z^{-1}_{(i)}) = z_{\pi(i)}$, i.e,. the i^{th} largest of the z_i.

Let $\chi^I[A]$ denote the indicator function of the event A, i.e., $\chi^I[A] = 1$ when A occurs, and 0 otherwise. Let $A_i^{m:I}$ denote the event that flow report i is selected. Note that $A_i^{m:I} = \{z_i > z_{[m+1,I]}\} \cup B_i$ where B_i is is then event that report i is selected but $z_i = z_{[m+1,I]}$. Events in B_i are due to ties encompassing the $(m+1)^{\text{st}}$ place in the decreasing order of the z_j, with flow i being amongst those selected after tie breaking. Since the distribution of the z_i is continuous, ties occur with probability zero, and hence we can effectively ignore the sets B_i. We abbreviate $\chi^I[A_i^{m:I}]$ by $\chi_i^{m:I}$, and set $I(c) = \{i \in I\ :\ c_i = c\}$ be the set of flows of key c.

Theorem 1: Let $i \in I$, $m < n = \#I$ and $C = \cup_{i\in I}\{c_i\}$,
(i) $\widehat{x}_i = \chi_i^{m:I} x_i/p_{z_{[m+1:I]}}(x_i) = \chi_i^{m:I}\max\{x_i, z_{[m+1:I]}\}$ is an unbiased estimator of x_i, i.e., $\mathsf{E}[\widehat{x}_i = x_i]$.
(ii) For $c \in C$, $\widehat{X}(c)$ is an unbiased estimator of $X(c)$, where

$$\widehat{X}(c) = \sum_{i\in I(c)} \widehat{x}_i \tag{1}$$

$$= \sum_{i=1}^{m} \chi^I\{c_{[i:I]} = c\}\max\{x_{[i:I]}, z_{[m+1:I]}\} \tag{2}$$

Remarks:

- In (i), the estimator is (of course) 0 if flow report i is not sampled, otherwise it is $\max\{x_i, z_{[m+1:I]}\}$.
- The equality in (i) results simply from the fact that $x/p_z(x) = \max\{x, z\}$ for $p_z(x) = \min\{1, x/z\}$.
- Division by $p_{z_{[m+1:I]}}(x_i)$ can be viewed as the appropriate renormalization of flow sizes when sampling threshold is $z_{[m+1:I]}$.
- The second expression in (ii) can be described as the sum of flow sizes x_i with the m largest random thresholds z_i, renormalized by maximization with $z' = z_{[m+1:I]}$, with flows of key

$c_i \neq c$ being omitted from the sum. $\chi^I\{c_{[i:I]} = c\}$ just selects flows with key c.

The proof of Theorem 1 and several other results in this paper depend on a technical result concerning the order statistics of non-identical uniformly distributed random variables, Theorem 2 below. Observe that z_i^{-1} is uniformly distributed on the interval $[0, 1/x_i]$.

For with $0 < r \leq \#I$, let $F_{r:I}$ denote the cumulative distribution function (CDF) of $z_{(r:I)}^{-1}$. Let $q_x(y) = p_{1/y}(x) = \min\{1, xy\}$. Note that $\{z_i^{-1} \leq y\} = \{w_i \leq x_i y\}$ and hence $1/z_i$ has cumulative distribution function (CDF) q_{x_i}. Now $\{z_{(r)}^{-1} \leq y\} = \{\text{at least } r \text{ of the } 1/z_i \leq y\}$, and it follows from the general form given in (2.8.1) of [4] that

$$F_{r:I}(y) = \sum_{i=r}^{\#I} \sum_{S_i} \prod_{\ell=1}^{i} q_{x_{j_\ell}}(y) \prod_{\ell=i+1}^{\#I} \overline{q}_{x_{j_\ell}}(y) \tag{3}$$

where $\overline{q} = 1 - q$ and the summation S_i extends over all permutations $(j_1, \ldots, j_{\#I})$ of I for which $j_1 < \ldots < j_i$ and $j_{i+1} < \ldots < j_{\#I}$. Fortunately, it is not necessary to exploit the particular form of (3) in order to establish our general results.

Theorem 2: Let $J \subset I$ with $\#J < r \leq \#I$, and $0 < w_i \leq y$, $i \in J$. Set $r' = r - \#J$.
(i) The joint CDF of $z_{(r:I)}^{-1}$ and the $\{z_i^{-1} : i \in J\}$ factorizes as

$$\mathsf{P}[z_{(r:I)}^{-1} \leq y, z_i^{-1} \leq w_i : i \in J] = F_{I\setminus J:r'}(y) \prod_{i\in J} q_{x_i}(w_i) \tag{4}$$

(ii) Let h be a real function.

$$\mathsf{E}\left[h(z_{(r:I)}^{-1}) \prod_{i\in J} \chi_i^{r:I}\right] = \mathsf{E}\left[h(z_{(r':I\setminus J)}^{-1}) \prod_{i\in J} q_{x_i}(z_{(r':I\setminus J)}^{-1})\right], \tag{5}$$

when the expectations exist.
(iii) In particular,

$$\mathsf{E}\left[\prod_{i\in J} \frac{\chi_i^{r:I}}{q_{x_i}(z_{(r:I)}^{-1})}\right] = 1. \tag{6}$$

PROOF OF THEOREM 1: (i) follows directly from Theorem 2(iii) on taking $J = \{i\}$ and recalling $q_x(1/z) = p_z(x)$. (ii) From (i) $\sum_{i:c_i=c} \widehat{x}_i$ is an unbiased estimator of $X(c)$. ∎

III. General Single Parameter Probability Functions

This paper is conducted within the framework of the measurement cost function $V_X + z^2 N$, proposed in [10] and discussed in the introduction. Recall that this choice of cost function is motivated by the problem of usage estimation—V_X is the variance in estimating a usage X while N the mean number of samples required for the estimate—and gives rise to the probability p_z that governs much of the work of this paper.

But for other applications, another cost function might be more appropriate. Suppose, for example, that accuracy in estimating the number of flows is as important as estimating the byte usage. For this problem, a cost function of the form $V_X + V_N + z^2 N$ is appropriate, where V_N is the variance in estimating the number of flows. It is natural to ask to what extent the results of this paper can be extended to this or any other single parameter cost function.

Let P denote the space of strictly positive *probability functions* p on $(0, \infty)$ and let $\mathsf{P}_1 = \{p \in \mathsf{P} : \sup_{x>0} p(x) \leq 1\}$. We assume the following:

- A parameter set $\Phi \subset \mathbb{R}$, and a *cost function* $C : \Phi \times \mathsf{P}_1 \to \mathbb{R}$.
- For each $\phi \in \Phi$, the minimizing probability function $p_\phi \in \mathsf{P}_1$ obeying $C(\theta, p_\theta) \leq C(\theta, p)$ $p \in \mathsf{P}_1$ exists, such that $\theta \mapsto p(\theta, x)$ is non-increasing for all $x > 0$.

An example of such a cost function is furnished by the example mentioned above: with cost $V_N + V_X + z^2 N$. We now show that the probability function for this cost function is

$$p(x) = \min\{1, \sqrt{(1+x^2)}\,/z\} \tag{7}$$

To calculate the cost associated with a single usage x and selection probability p, let χ denote the random indicator of selection, i.e. $\chi = 1$ with probability p and $\chi = 0$ otherwise. The usage estimate is $\widehat{x} = \chi x/p$ while the estimate of the number (1) of underlying objects is $\widehat{n} = \chi/p$. Thus the cost is

$$\begin{aligned} C(z,p) &= \mathsf{Var}(\widehat{n}) + \mathsf{Var}(\widehat{x}) + z^2 p & (8)\\ &= (1+x^2)\overline{p}/p + z^2 p & (9) \end{aligned}$$

Since $p \mapsto (1+x^2)\overline{p}/p + z^2 p$ is convex, it takes its minimum value over $p > 0$ at the stationary point $p = \sqrt{(1+x^2)}\,/z$. Thus, restricting to $p \in (0,1]$, the minimum cost is obtained with $p = \min\{1, \sqrt{(1+x^2)}\,/z\}$.

We now return to the general case. With hard resource limitations, we associate a random threshold ϕ_i to each flow size x_i, then select for sampling the m flows of largest threshold. Suppose ϕ_i can be chosen such that

$$\mathsf{P}[\phi_i \geq \phi] = p_\theta(x_i) \tag{10}$$

Then the results of Section II generalize completely upon the notation change $z_i \to \phi_i$ etc. Now we show that a ϕ_i with these properties can always be chosen, at least under the above assumptions. Let $f_x(y) = \sup\{\phi \in \Phi : p_\phi(x) \geq y\}$ denote the left inverse of the nonincreasing function $\phi \mapsto p_\phi(x)$.

Theorem 3: Let ω_i be uniformly distributed on $[0,1]$. Then $\phi_i = f_{x_i}(\omega_i)$ obeys (10).

PROOF OF THEOREM 3: $\mathsf{P}[\phi_i \geq \phi] = \mathsf{P}[f_{x_i}(\omega_i) \geq \phi] = \mathsf{P}[\omega_i \leq p_\phi(x_i)] = p_\phi(x_i)$. ∎

Examples: (i) We first confirm that the framework described in the introduction fits into the scheme described above. Take $\Phi = \mathbb{R}$, $p_\phi(x) = \min\{1, x/\phi\}$. Hence $\phi_i = f_{x_i}(\omega_i) = x/\omega_i$ for $\omega_i \in (0,1]$.

(ii) With the probability function (7), take $\Phi = \mathbb{R}$, $p_\phi(x) = \min\{1, \sqrt{(1+x^2)}\,/\phi\}$. Then $\phi_i = f_{x_i}(\omega_i) = \sqrt{(1+x_i^2)}/\omega_i$ for $\omega_i \in (0,1]$.

IV. Variance and its Estimation

Theorem 1 shows that $\widehat{x}_i$ is an unbiased estimator of the size x_i. For many applications one needs to know what reliability can be attached to estimators, or sums $X(c)$ of such estimators representing the total byte usage in a given key. We characterize the estimators through their variance. This characterization is most appropriate for sums of large number of random variables, since deviations from the mean from sums of independent or weakly dependent random variables, suitably rescaled, follow an asymptotically Gaussian distribution, as follows from the central limit theorem and its variants.

This approach matches well to applications, since it is typically most important to know the accuracy of estimates of usage in keys consuming large amount of network resources. Typically, usage in a given key is large if either or both of the following hold:
(i) there a flow records with size x larger than the sampling threshold z. However, these are always sampled and hence incur no estimation error.
(ii) there are a large number of flow records with size x small than the sampling threshold z. In this case we are in the domain of the central limit theorem, and characterization of accuracy through variance is appropriate.

However, there remains one potential obstacle. Since the threshold z' depends on the collection of z_i, the selection of various flows is, in general, correlated. Intuitively, the selection of a given flow should make the selection of a different flow less likely. An extreme example is when two flows ($n = 2$) compete for one slot ($m = 1$): in this case selection of one flow precludes selection of the other. However, it turns out that, when $m \geq 2$, the random variables $\{\chi_i^{m:I}/p_{z'}(x_i) : \; i \in I\}$ are uncorrelated, i.e,. their mutual covariance is zero. This enables use to write the variance of a the sum $\widehat{X}(c) = \sum_{i:\, c_i=c} \widehat{x}_i$ simply as a sum of the variance of the $\widehat{x}_i$.

The next theorem establishes that the mutual correlations of the $\widehat{x}_i$ vanish, and gives expressions for their variances. Bounds and approximations in applications of these results are discussed in Section V below. Let x^+ abbreviate $\max\{x, 0\}$.

Theorem 4: Assume $m \geq 2$.
(i) $\mathsf{Cov}(\widehat{x}_i, \widehat{x}_j) = 0$ when $i \neq j$.
(ii) $\mathsf{Var}(\widehat{x}_i) = \mathsf{E}[x_i(z_{m:I\setminus\{i\}} - x_i)^+]$
(iii) $\mathsf{Var}(\widehat{x}_i) = \int_0^{1/x_i} dF_{m:I\setminus\{i\}}(y)\, x_i(y^{-1} - x_i) =$
$x_i \int_0^{1/x_i} dy F_{m:I\setminus\{i\}}(y)\, y^{-2}$.

Thanks to Theorem 4, the variance of the estimate $\widehat{X}(c)$ total usage attributable to any key c can be decomposed as a sum:

$$\mathsf{Var}(\widehat{X}(c)) = \mathsf{Var}(\sum_{i\in I(c)} \widehat{x}_i) \quad (11)$$

$$= \sum_{i,j\in I(c)} \mathsf{Cov}(\widehat{x}_i, \widehat{x}_j) \quad (12)$$

$$= \sum_{i\in I(c)} \mathsf{Var}(\widehat{x}_i) \quad (13)$$

In practice, the actual usage X is unknown after sampling; one has only the estimates $\widehat{X}$ to work with. How, then, should the variance of usage estimates be estimated from the samples themselves? The way to do this for variance constrained smart sampling was described in [8]. In the present case, it follows from:

Theorem 5: $\chi_i^{m+1:I} z_{[m:I\setminus\{i\}]}(z_{[m:I\setminus\{i\}]} - x_i)^+$ is an unbiased estimator of $\mathsf{Var}(\widehat{x}_i)$.

Note that, conditioned on flow i being selected (i.e. $\chi_i^{m:I} = 1$), $z_{m:I\setminus\{i\}} = z_{m+1,I}$. Hence, the practical content of Theorem 5 is that when a flow of size x is selected and the random sampling threshold is z', the estimated variance associated with the usage estimate $\max\{x, z'\}$ should be taken as $z'(z' - x)^+$. In forming the estimate of the variance of $\widehat{X}(c)$, we add up all such contributions due to sampled flows of key c.

V. Properties of the Variance

A. Strategies for Variance Determination

In view of Theorem 5, online estimation of the variance of usage estimators is straightforward: one need only add a contribution $z'(z' - x)^+$ for each sampled usage component x in an aggregate. On the other had, for the purposes of dimensioning measurement subsystems, it is useful to know how the (actual) estimator variance depends on the flow data and on the cache size.

A prime challenge to determining this variance stems from the relative complexity of calculating it, at least exactly. This is principally due to the the nature of the problem. The variance in Theorem 4(ii) is a (negative) moment of an order statistic; there are very few general expressions for such quantities available, and these only in special cases; see, e.g., [1]. Given this constraint, there appear to be two possible approaches that might useful in practice:
(i) Derive bounds and approximations for the variance, preferably as a functional of relatively simple characteristics of the offered flows. These would have the advantage of being robust with respect to variations in the details of the offered flows. Such a functional could be applied directly to traces of flow statistics.
(ii) Use a Monte-Carlo approach in which sampling is simulated, and approximate the variance by the unbiased estimator of Theorem 5. This approach is computationally straightforward, but is likely of limited predictive power beyond the range of flows and parameters used in the simulations. For this reason we favor the first approach.
In view of Theorem 4, finding bounds and approximations for variance entails doing the same for the mean sampling threshold. We start by finding and exploiting the relationship between the stochastic sampling threshold z' under a hard volume constraint, and the fixed sampling threshold z in the related variance constrained sampling problem.

B. Variance Constrained Threshold

We now recall from [9], [10] the relationship between the sampling threshold z and the number of flow selected in variance constrained sampling. Let N_z denote the random number of samples that would be taken with a fixed sampling threshold z. Now flow i is sampled with probability $p_z(x_i)$ and hence the mean number of sampled flows is

$$\mathsf{E}[N_z] = m(z) := \sum_{i\in I} p_z(x_i) \quad (14)$$

Now $z \mapsto m(z)$ is clearly a strictly decreasing continuous function on from $(\min_i x_i, \infty)$ onto $(0, \#I]$. Hence for each target number of sampled flows m there is a $z_* = z_*(m) > 0$ such that $m(z^*) = m$, i.e, the mean number of flows is equal to the target number.

What then, is the relationship between z_* and z', the random sampling threshold use to sample *exactly* m flows? If such a relationship can established, we can exploit an algorithm presented in [10] that is able to quickly compute z_* for a specified target m, given a set of flow sizes $\{x_i : i \in I\}$.

C. Asymptotic Equality of Thresholds

Since flow i is sampled with probability $p_z(x_i) = \mathsf{P}[z_i > z]$, N_z has the the same distribution as $\widehat{N}_z = \#\{i \in I : z_i > z\}$. Since

$$\{\widehat{N}_z \geq m\} = \{z < z_{[m+1]}\} \tag{15}$$

we can derive properties of $z_{[m+1]}$ in terms of those of N_z. In this section we bound the deviations of $z_{[m+1]}$, and show that it converges to z_* in a certain asymptotic regime.

Let c is the cumulant generating function of N_z, i.e.,

$$c(\theta, z) = \log[E^{\theta N_z}] = \sum_{i \in I} \log(e^\theta p_z(x_i) + \overline{p}_z(x_i) \tag{16}$$

and let J denote the convex rate function defined as the Legendre transform of c w.r.t. its first argument, i.e.,

$$J(m, z) = \sup_\theta (m\theta - c(\theta, z)) \tag{17}$$

Theorem 6: (i) For $m_- < m(z) < m_+$,

$$\mathsf{P}[N_z \leq m_-] \quad \leq \quad e^{-J(m_-, z)} \tag{18}$$

$$\mathsf{P}[N_z \geq m_+] \quad \leq \quad e^{-J(m_+, z)} \tag{19}$$

(ii) $J(m, z) = 0$ if and only if $m = m(z)$ or, equivalently, if and only if $z = z_*(m)$.
(iii) For $z_- < z_*(m) < z_+$,

$$\mathsf{P}[z_{[m+1]} \leq z_-] \quad \leq \quad e^{-J(m, z_-)} \tag{20}$$

$$\mathsf{P}[z_{[m+1]} \geq z_+] \quad \leq \quad e^{-J(m, z_+)} \tag{21}$$

Suppose we scale up our system by an integer $\lambda > 1$. Thus there λ flows of each size x_i, and we increase the number of slots for samples to λm. The cumulant generating function becomes $c_\lambda = \lambda c$, and the corresponding rate function is $J_\lambda(m, \beta) = \lambda J(m/\lambda, \beta)$. Thus $z_{[\lambda m+1]}$ obeys bounds of the form of Theorem 6, but with rate function $mJ(m, z)$. It follows that as $\lambda \to \infty$, $z_{[\lambda m+1]}$ converges in probability to z_*, i.e., for every $\varepsilon > 0$, $\mathsf{P}[|z_{[\lambda m+1]} - z_*| < \varepsilon] \to 1$ as $\lambda \to \infty$. More generally, we can say that $z_{[\lambda m+1]}$ obeys the large deviation upper bound: for every closed set B,

$$\lim_{\lambda \to \infty} \lambda^{-1} \log \mathsf{P}[z_{[\lambda m+1]} \in B] \leq -\inf_{z \in B} J(m, z). \tag{22}$$

A matching lower bound can also be established.

D. Bounds on Estimation Variance

Theorem 6 says that the stochastic threshold z' is close to the deterministic threshold z_* when the number of flows and number of slots both become large. Informally at least, we expect estimator variance to be roughly the same in both cases. However, this approach has two weaknesses. Firstly, it applies in an asymptotic sense, and second, the closeness is not qualified. In principal, closeness of z' and z_* be determined from the rate function J. In practice, it is difficult to obtain quantitative features of J. This motivates obtaining explicit bounds on estimator variance that hold in all regimes.

Define $X^-_z = \sum_{i>0: x_i<z} x_i$; $X = X^-_\infty$ is the total size of all flows in I. Given $z > 0$, The number of flows of size less than z is $n^-_z = \#\{i > 0 : x_i < z\}$; the number of size not less than z is $n - n^-_z = n^+_z = \#\{i > 0 : x_i \geq z\}$. The average size of a flow of length less than z is $x^-_z = X^-_z / n^-_z < z$. Lemma 2, found in Section X, collects together some properties of these quantities that we shall exploit. As usual, e denotes the base of natural logarithms.

Theorem 7:

$$\frac{\mathsf{Var}\, \widehat{x}_i}{x_i} \quad \leq \quad (z_* - x_i)^+ + u(\min\{z_*, x_i\}) \tag{23}$$

where $u(x) = \inf_{\alpha > 0} u_\alpha(x)$, with

$$u_\alpha(x) \quad = \quad \frac{X^{1+\alpha}}{\alpha m^{1+\alpha} x^\alpha} \, q(n^-_x, m - n^+_x, 1 + \alpha) \tag{24}$$

and

$$q(n, m, \beta) \quad = \quad \frac{(1 - \beta/m)^{m-\beta}}{(1 - \beta/n)^{n-\beta}} \tag{25}$$

Theorem 8: Suppose $n, m, X \to \infty$ with $X/m \to x_\infty$. Then

$$\limsup_{n,m,X \to \infty} u(x) \leq \begin{cases} e x_\infty / x \, \log(x_\infty / x), & x < x_\infty \\ 0 & x \geq x_\infty \end{cases} \tag{26}$$

E. Identical Flow Sizes

Some useful intuition can be obtained from the special case that all flow sizes are equal. This is amenable to explicit calculation through the simplification of the CDF from (3). As before we write $\#I = n$.

$$F_{I \setminus \{1\}:m}(y) = \sum_{i=m}^{n-1} \binom{n-1}{i} q^i_x(y) \, (1 - q^{n-1-i}_x(y)) \tag{27}$$

Using Theorem 4(iii) one finds after a change of variable $y \to y/x_1$ then

$$\mathsf{Var}(\widehat{x}_1) \quad = \quad x_1^2 \int_0^1 dy \sum_{i=m}^{n-1} \binom{n-1}{i} y^{i-2} (1-y)^{n-i-1} \tag{28}$$

$$= \quad x_1^2 \, (n-1) \sum_{i=m}^{n-1} 1/(i(i-1)) \tag{29}$$

$$= \quad x_1^2 \frac{n-m}{m-1} \tag{30}$$

This is the same expression as would be obtained with uniform sampling of each of n objects into the m slots, except with $m-1$ in place of m in the denominator. This reflects the loss of one degree of freedom involved in determining the sampling threshold z' from the data.

Note that the variance diverges for $m = 1$: estimating the total usage through one sample results in infinite estimator variance. By contrast, in variance constrained sampling, to sample one flow *on average*—so that $m(z) = 1$ in (14)—one takes $z = nx$ resulting in variance $x_1^2(n-1)$.

A similar phenomena can be constructed when picking any number of k items. Suppose we add $k-1$ enormous items, say of size 2^n, to the above n items of size x_1. With very high probability, the $k-1$ enormous items are going to be picked, and then we left with the previous task of picking one out of the n items of size x_1. Hence the associated estimation error in selecting the k items will be enormous.

F. A Conjecture

From Section V-C we know that the thresholds $z_*(m)$ and $z' = z_{[m+1]}$ for variance and resource constrained smart sampling are asymptotically equal in a regime of a large number of offered flows n and slots m available for them. Note from Theorem 4(ii) we have the simple bound $\mathsf{Var}(\widehat{x}_i) \le x_i \mathsf{E}[z']$. Indeed, in variance constrained sampling with threshold z,a corresponding bound $\mathsf{Var}\,\widehat{x}_i \le x_i z$ holds. Thus bounding z' in terms of z, or at least in expectation, would provide us with a simple bound for estimator variance in the present case.

We now state a conjecture on the relationship between them. We have been able to establish this conjecture in the special case of uniform flow sizes, although we do not provide the proof here. The idea behind the conjecture is that compared with variance constrained sampling for a mean number m of flow record, that of we allow slightly more space for our samples, the variance in resource constrained sampling will not exceed that of the variance constrained case. For example, expanding the room available for samples ameliorates the high variance problem discussed at the end of Section V-E.

In the conjecture we write $\widehat{x}_i(m)$ to emphasize the dependence of the threshold and estimators on the target number of samples m.

Conjecture 1: Assume $m \ge 10$. Then $z_{[m+\lceil 2\sqrt{m}\rceil]} \le z_*(m)$ and hence $\mathsf{Var}(\widehat{x}_i(m+\lceil 2\sqrt{m}\rceil)) \le x_i z_*(m)$.
The intuition behind adding a multiple $2\sqrt{m}$ to the number of samples taken, is that it roughly encompasses the likely error in the actual number of samples taken in variance constrained smart sampling.

G. Large and Small Flows

Given the importance of estimating the large contributions to usage, it is interesting to determine how estimation error for large flows depends on their prevalence in the mix, and the space m available for sampling. (Note this is a different question to that examined at the end of the previous paragraph, where we dealt with estimating a small flow against the background of larger flows).

We can gain some insight through a model in which all flows are either large, having size x_+, or small, having size x_-. Write the ratio of the sizes as $\alpha = x_-/x_+ < 1$. Let there be h large flows and hence $n-h$ small flows. Suppose $x_1 = x_+$. Then it follows from Theorem 4(iii) and (3) that:

$$\frac{\mathsf{Var}(\widehat{x}_1)}{x_+^2} = \sum_{i=m}^{n-1} \sum_{j=0}^{\min\{h,i\}} \binom{h}{j}\binom{n-1-h}{i-j}\alpha^{i-j} \qquad (31)$$

$$\times \int_0^1 dy\, y^{i-2}(1-y)^{h-j}(1-\alpha y)^{n-1-h-(i-j)} \qquad (32)$$

We have not been able to obtain an explicit expression for the above. However, one can identify the dominant terms for small α, i.e., those with the lowest power of α. These are obtained by making i as small as possible and j as large as possible. Note the α in the function u is not important in this determination since for small α, $1-\alpha y \approx 1$ for $y \in [0,1]$.

Theorem 9: Assume the above two size model, with $x_1 = x_+$ and $\alpha = x_-/x_+ < 1$. The estimator variance for the larger flow length has the asymptotic behavior

$$\mathsf{Var}\,\widehat{x}_1 \sim x_+^2 K_{n,m,h}\,\alpha^{(m-h)^+}, \qquad \alpha \to 0 \qquad (33)$$

where

$$K_{I,m,h} = \begin{cases} \binom{n-1-h}{m-h}/(m-1) & \text{if } h \le m \\ (h+1-m)/(m-1) & \text{if } h > m \end{cases} \qquad (34)$$

We can interpret Theorem 9 as follows. During sampling, the large flows are more likely to be sampled than the small flows because they have (stochastically) larger thresholds x_i/ω_i. Once the space for sampling can accommodate all the large flows, i.e., $m \ge h$, the variance falls of as α^{m-h} for small α. On the other hand, if there is not room to sample all the large flows, i.e., $m < h$, the estimation error persists even for small α.

We illustrate the approximation for the case of $n = 11$ flows, out of which $m = 5$ are to be sampled. We show the exact and approximate values of the standard error in $\widehat{x}_1$ as a function of h, the number of large size flows in Figure 1, for $\alpha = 0.01, 0.1$ and 0.5. Note for the left hand plot ($\alpha = 0.01$) the knee in the curve as h passes through m. The variance falls off sharply for small m. This behavior is also exhibited to a lesser degree for $\alpha = 0.1$. In these two cases, the approximation is very good. It is less good for $\alpha = 0.1$, being only within a factor of about 2. This illustrates a caveat to using the asymptotic behavior of Theorem 9 as an approximation: the proof was confined to the regime $(1-\alpha)^{n-1-m} \approx 1$. In the three examples $(1-\alpha)^{n-1-m} = 0.95, 0.60$ and 0.03 respectively, for $\alpha = 0.01, 0.1$ and 0.5.

VI. Implementation and Complexity

A. Streaming Algorithm for Flow Selection

So far we have posed our problem in the abstract of choosing m of n offered flow records. In applications, the flow records will typically be presented sequentially in order of creation. However, it is not desirable to maintain all n flow records in storage while the selection of m representatives is made. Nor is it necessary. We now present an algorithm which processes the offered flow records sequentially, storing no more than $m+1$ of them at any time.

Each flow record i is regarded as a triple (c_i, x_i, z_i) where c_i is the flow key, x_i the flow size, and z_i the random thresholds. The algorithm maintains a min-priority queue of at most $m+1$ flow records, where random threshold is the priority. Thus the queue contains the $m+1$ flow records of largest threshold seen so far.

Specifically, when a new flow record of size x is presented, a random number ω is generated from a uniform distribution in $(0,1]$, and the threshold $z = x/\omega$ is calculated. If the queue is not full, the new record is inserted. If the queue is full, the flow

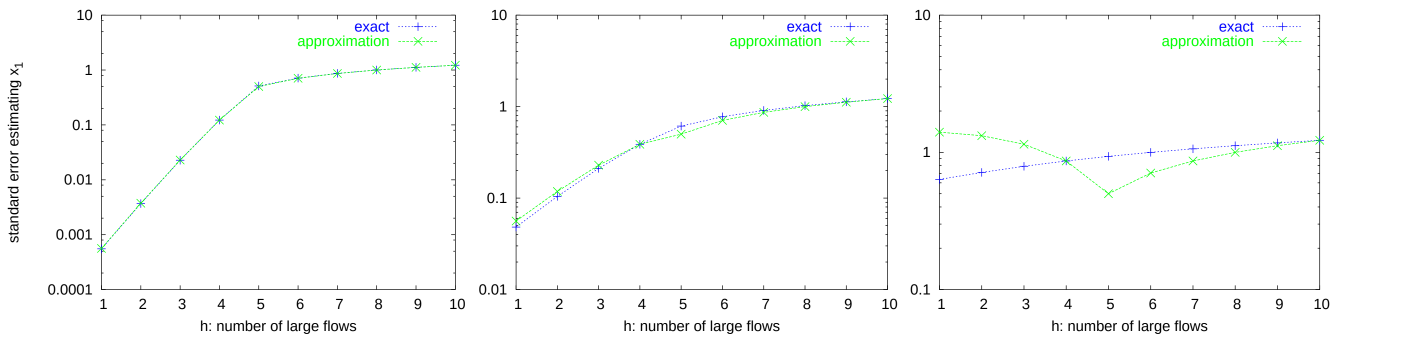

Fig. 1. Relative estimation error for large flow, as function of number of large flows, according to ratio $\alpha = x_-/x_+$. Left: $\alpha = 0.01$. Middle: $\alpha = 0.1$. Right: $\alpha = 0.5$. Exact and approximation. Note better accuracy of approximation for smaller α.

of lowest priority (i.e. minimum threshold is extracted from the queue). Its priority is compared with that of the new flow record. Of these two, the record with highest priority is inserted back into the queue, while the record of lowest priority is discarded. This completes the processing of a single flow. Note that for each flow, only one random number is needed.

At the end of the measurement period, or when all offered flows have been processed, the flow of lowest priority is selected. This has threshold $z_{[m+1]}$. The other flows have thresholds $\{z_{[i]} : i = 1, \ldots, m\}$. These flows are reported with renormalized sizes $\{\max\{x_{[i]}, z_{[m+1]}\} : i = 1, \ldots, m\}$.

B. Complexity, Implementation & Variants

During the processing of each new flow, both the insertion operation and deletion of the flow with lowest priority take $O(\log m)$ time using a standard heap [3]. In fact, since z_i will be an integer or floating point number, this cost can be brought down to $O(\log\log m)$; see [21].

If we are willing to set aside space for $2m + 2$ flow records, the average cost per item can be bought down to a constant using a variant of the above implementation. Insertion of flow records is performed until the buffer contains $2m + 2$ records. At this point a cleanup is performed to reduce the occupancy down to $m + 1$. Using a standard selection algorithm [3], the records of $(m + 1)^{\text{st}}$ largest threshold can be identified, and items of smaller threshold deleted, all in time linear in m. The cleaning is executed once for every $m + 1$ arrivals to the queue; hence the cost is $O(1)$ per item processed. Deletion is likewise linear because no item is deleted twice. After cleanup, we resume filling the buffer with fresh arrivals. Note that in practice, the time between successive arrivals should be bounded below by the above amortized $O(1)$ processing time per item.

A further modification processes every item in constant time without having to wait for the cleanup to execute. Two queues of capacity $2m+2$ are used, one queue being used for collection while the other is cleaned down to $m+1$ items. Then each record is processed in constant time, plus $O(m)$ at the end of the measurement period in order to find the $m+1$ items of largest threshold from the union of the contents of the two buffers. Thus, provided the between successive arrivals should be bounded below by the $O(1)$ processing time per item, the processing associated with each flow can be completed before the next flow arrives.

C. Alternate Sampling Strategies

Previous work on variance constrained smart sampling [9], [10] proposed a dynamic algorithm to control mean sample volume. As discussed in the introduction, such approaches cannot satisfy a hard constraint on the number of samples, even under stationary conditions. Moreover, the number of samples selected responds to systematic as well as statistical variations in the offered traffic. The performance studies of Section VII will substantiate these statements. This motivates a more detailed description of the previous work.

In the previous approaches, time is divided into windows. The set of flows arriving at window is used to determine some value of the sampling threshold z, which is then used predictively as the sampling threshold for the next window. One sees immediately that, whatever the method by which z is determined, this approach is not robust w.r.t variations in the offered load that take place at shorter timescales than the window width.

Within this framework, several variants are possible. One approach is to find the threshold z_* that would makes the mean number of samples taken from the flows arriving in the window equal to the target. [10] presents a probabilistic algorithm to find the root in a time that is linear in the number of flows offered in the window. However, this necessitates storing the sizes of all the flows until the end of the window, which may run up against memory constraints.

A simpler approach is to iteratively adjust the thresholds $z^{(k)}$ in the window k according to the discrepancy between the target number m and the actual number $N^{(k)}$ of flows selected in the windows:

$$z^{(k+1)} = z^{(k)} \frac{N^{(k)}}{m} \tag{35}$$

Further variations on this approach include:

- *Emergency Control:* the window is terminated early if the number of selected flows exceeds m, and one takes $z^{(k+1)} = z^{(k)}\tau/T$ where T is the width of the full window, and τ is the time from the start of the window at which m is exceeded. This gives some robustness to systematic changes in the offered load.
- *Variance Compensation:* the target value m is adjusted downwards in order to accommodate upward statistical fluctuations in the number of sampled. For example, we make use $m - s\sqrt{m}$ in place of m, where s can be through of as the number of standard deviations that are to be accommodated.

VII. Performance and Examples

In this section we investigate the performance of the resource constrained selection algorithm proposed in this paper. The primary element of this study is to demonstrate that estimator accuracy is reasonable, and is in any case not appreciable worse than the earlier proposal for dynamically controlled variance constrained smart sampling proposed in [10] and summarized briefly in Section VI-C.

Such studies could be regarded as sufficient for our purpose, since resource usage can be determined at the user's will, simply by setting the space m available for samples. This can be done *regardless* of the offered traffic. However, it is also interesting to compare m with the actual number of samples taken by the dynamic control algorithm targeting an *expected* number of samples m. This is particularly interesting in the face of statistical and systematic variation in rates and composition of the stream of offered flow records. In general, the number of records actually taken by the dynamically controlled algorithm will vary around the target value m during the collection period.

In some sense, the comparison of algorithm performance on a given set of flows demonstrates only a lower bound on the effectiveness of the current proposal relative to dynamic control. This is because another traffic set could conceivably have worse properties that would lead to greater variation—i.e. less control—of the number of samples taken by the dynamic control algorithm. The beauty of our resource constrained algorithm is that the user need not worry that some unexpected traffic fluctuation will overwhelm their measurement system.

A. Flow Data

We used two sets of NetFlow records for our experiments. The first set, AGGREGATE, comprised flow records collected in an aggregation network of a major service provider, representing traffic generated by several thousand users during one day. The average flow arrival rate was about 45 per second.

The second set, GATEWAY, comprised flow records collected at an Internet gateway of an enterprise. The average flow production rate was about 186 per second.

B. A Measure of Estimation Accuracy

In the experiments described below, we will compare actual usages $\{X(c) : c \in C\}$ over a key set C, with their corresponding estimates $\{\widehat{X}(c) : c \in C\}$. We wish to characterize the accuracy of estimation through a single summary statistic. For a given color c, one measure is the relative error

$$\delta(c) = \frac{|X(c) - \widehat{X}(c)|}{X(c)} \tag{36}$$

In the summary statistic, we wish to reflect the fact that estimation in keys with large usage is more important than for keys of small usage, at least in many network management applications. This motivates averaging the relative error $\delta(c)$ with rates proportional to the actual usage $X(c)$ to be estimated. Thus, our summary statistic is the Weighted Mean Relative Error:

$$\text{WMRE} = \frac{\sum_{c\in C} |X(c) - \widehat{X}(c)|}{\sum_{c\in C} X(c)} \tag{37}$$

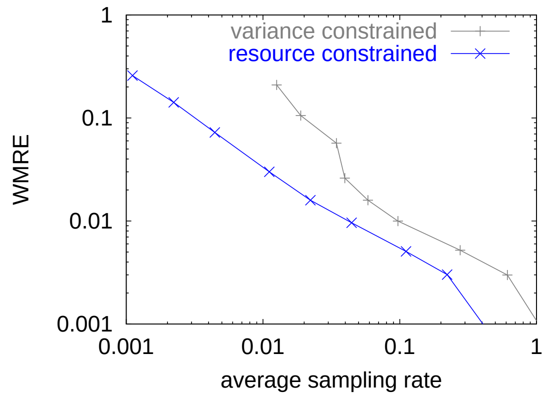

Fig. 2. Summary estimation error (WMRE) for resource vs. variance constrained sampling, as a function of the sampling rate. Note uniformly better accuracy of resource constrained sampling.

C. Experiments: Accuracy vs. Resource Use

In our first experiment, outbound traffic in trace AGGREGATE was aggregated by key according to user side IP address. For each sampling method, accuracy was summarized through the WMRE of the total usage during the entire trace for each key.

One question is how to meaningfully set the duration time window for estimation. On the one hand, applications working with highly aggregated data in the collector might tolerate windows of a few minutes or even an hour. Applications that need a fast response time, such as for security monitoring, might require responsiveness down to a few seconds. We used a variety of time windows in our experiments; here we report using an intermediate value of one minute.

We conducted resource constrained sampling with a target number of slots m ranging from $m = 3$ to $m = 3,000$. This represented an average flow sampling rate ranging from about 0.1% to nearly 1. In Figure 2 we display the WMRE and a function of the sampling rate for two sampling methods: the resource constrained method proposed in this paper, and dynamically controlled variance constrained sampling from [10]. The WMRE is displayed as a function of the average sampling rate. The average sampling period is obtained by dividing the number of sampling slots used by the average number of flows arriving per window. For resource constrained sampling, the number of slots is exactly m. For variance constrained sampling, it is the maximum number samples taken in any window, when the target mean number of samples is m. Numbers from some initial windows were discarded due to transience in the sampling threshold. We did not perform variance compensation: while this lowers the maximum rate, it would also just correspondingly decrease estimator accuracy.

Note from Figure 2 that the present proposal performs uniformly better than variance constrained sampling, with a reduction in resource usage of about half an order of magnitude over a range of sampling rates. The absolute estimation accuracy is also reasonable, being e.g. about 1% for sampling rate of 1%. We believe such accuracy is sufficient for many network management applications.

The previous experiment investigated spatial accuracy of usage estimation over a set of keys. We now look a accuracy of estimation of the total traffic over the set of time windows. Using trace GATEWAY, again using a window of 1 minute duration, we compared the two sampling methods. The time series of actual and predicted 1 minute rates are shown in Figure 3(left). The target number of sampled flows per window was $m = 10$, yielding an average sampling rate of about 1 in every 1,000 flow records. Note that the curves are almost coincident the whole trace.

Resource usage by the dynamically controlled variance constrained algorithm is shown in Figure 3(right). Note for $m = 10$ there are excursions of up to about four times the target value; even for a higher sampling rate $m = 100$, there are excursions up to about twice the target.

VIII. Related Work

As is clear from the introduction, the present work builds on the proposal for smart sampling with a static threshold by Duffield, Lund and Thorup [9], [10], and provides a harder resource bound than the dynamic control scheme presented in [10].

More closely related to the latter approach are several pieces of work concerning dynamic control of network measurements. Choi, Park and Zhang [2], focus on maintaining accuracy of estimates of the short-term traffic load at a router under varying traffic rates. The motivation of this study was the identification of changepoints in the traffic load; the accuracy of rate estimates determines the resolution at which changes can be detected. Hernandez, Chidester and George [15] use a predictive approach to anticipate variations in the offered load and adjust the sampling interval accordingly in order to meet sampling volume constraints. Their method combines linear prediction with a fuzzy logic approach that classifies the broad dynamics of the load.

Non-uniform sampling, and size dependent sampling in particular, has been proposed by Horwitz and Thompson [17]. Related variance estimators have been discussed by Thompson and Seber [20]. Although smart sampling can be viewed as somewhat similar in approach, the details are substantively different. In the standard approaches it is customary to determine the selection probabilities of items of size x_i as a deterministic function $x_i/\sum_i x_i$ of the data. So in distinction with variance constrained smart sampling, all data have a chance to be not selected, and all the data must be gathered before any can be selected. In distinction with the present approach, there is no constraint on the number of samples taken.

The problem of deriving flow length distributions from those of uniformly sampled flows has been examined by Hohn and Veitch [16]. Note that the problem of recovering the distribution is better behaved that that of recovering the usage, since the usage involves summing the (absolute) frequencies f_i of flows of length i against the length, i.e. forming $\sum_i i f_i$. Thus, in a corresponding estimation problem one could have estimated frequencies be close to the actual frequencies, while at the same time the estimated usage could be very different from the actual usage. The related problem of deriving flow length distribution from packet sampled flow statistics (such as those produced by sampled NetFlow) was considered by Duffield, Lund and Thorup; see [11].

IX. Summary and Conclusions

This work was motivated by the need to sample within the measurement infrastructure, the flow statistics that a currently collected by network routers and switches. The heavy tailed nature of packets and bytes per flow forms a challenge for sampling, since omission of a record of single large flow can seriously degrade the accuracy of estimates of network traffic usage. The recent proposal for smart sampling, i.e., size dependent sampling of flow records, forms the starting point for our work.

Our work addresses the need to bound resource usage (bandwidth, storage, aggregation caches) by the sampled flows. Any uncontrolled overflow of these resources seriously impairs estimation accuracy for the reasons just described.

We proposed a simple algorithm that allows smart sampling from a population of flow records into a buffer of fixed size m, while maintain the ability to form unbiased usage estimates. The algorithm is simple, requiring the generation of one random number per flow, and can be implemented in constant time per flow, independent of m. In a trace-based experiment, the performance of the algorithm—in terms of accuracy vs. resource usage—was found to be uniformly better than a recent proposal based on dynamically adjusting the sampling threshold of the original smart sampling proposal. Moreover, with the method proposed here, the selected samples always fit within a hard constraint, *whatever* the offered traffic, unlike the previous proposal.

Although the method is primarily envisaged for use today within the measurement infrastructure, it could in the future be deployed in routers to sample flows prior to export, resources permitting.

The main future work is to attempt to substantiate the conjecture of Section V-F concerning the relationship between estimator variance for the present algorithm, and for the original variance constrained smart sampling. Such a relationship will simplify the task of dimensioning of the measurement infrastructure to meet goals in estimator accuracy.

X. Proofs of Theorems

Proof of Theorem 2: (i) When all the $z_i^{-1} \leq y$, the event $z_{(r)}^{-1} \leq y$ occurs if and only if at least r' of the other z_i^{-1} (i.e. those with index i in $I \setminus J$) are not greater than y; in other words, if and only if $z_{(r':I\setminus J)}^{-1} \leq y$.

(ii) From part (i), the joint probability density of g of $z_{(r)}^{-1}$ and $\{z_i^{-1} : i \in J\}$ factorizes as

$$g(y, \{w_i\}) = F'_{r':I\setminus J}(y) \prod_{i\in J} q'_{x_i}(w_i) \tag{38}$$

when $0 < w_i \leq y$, $i \in I$. Hence we can write

$$\mathsf{E}\left[h(z_{(r)}^{-1}) \prod_{i\in J} \chi_i^{r:I}\right] \tag{39}$$

$$= \int_{y=0}^{\infty} dy\, h(y) \left(\prod_{i\in J} \int_{w_i=0}^{y} dw_i\right) g(y, \{w_i\}) \tag{40}$$

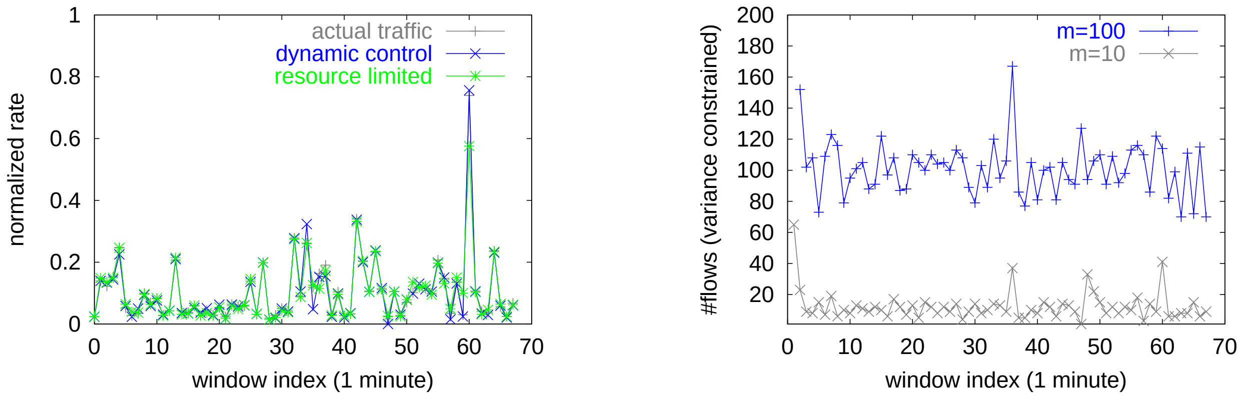

Fig. 3. Performance Relative to Dynamically Controlled Variance Constrained Sampling. Left: time series actual and estimated byte rates, aggregated over 1 minute window, sampling approximately 1 in 1,000 flows. Right: time series of number of samples taken in dynamic controlled algorithm, for a given expected number of samples $m = 10$ and 100. Observe excursions of up to four times (for $m = 10$) and two times (for $m = 100$) the target mean value m.

$$= \int_{y=0}^{\infty} dF_{r':I\setminus J:}(y) h(y) \prod_{i\in I} \int_{w_i=0}^{y} dq_{x_i}(w_i) \quad (41)$$

$$= \mathsf{E}\left[h(z^{-1}_{(r':I\setminus J)}) \prod_{i\in J} q_{x_i}(z^{-1}_{(r':I\setminus J)}) \right] \quad (42)$$

(iii) This follows by taking $h(y) = \prod_{i\in J} 1/q_{x_i}(y)$ in (ii). ■

Proof of Theorem 4: (i) By Theorem 2(iii), for $i \neq j$,

$$\frac{\mathsf{E}[\widehat{x}_i \widehat{x}_j]}{\mathsf{E}[\widehat{x}_i]\mathsf{E}[\widehat{x}_i]} = \frac{\mathsf{E}[\widehat{x}_i \widehat{x}_j]}{x_i x_j} = \mathsf{E}\left[\frac{\chi_i^{m+1:I} \chi_j^{m+1:I}}{q_{x_i}(z^{-1}_{(m+1)}) q_{x_j}(z^{-1}_{(m+1)})} \right] = 1 \quad (43)$$

(ii) Using Theorem 2(ii):

$$\begin{aligned} \mathsf{Var}(\widehat{x}_i) &= x_i^2 \,\mathsf{Var}\left(\frac{\chi_i^{m+1:I}}{q_{x_i}(z^{-1}_{(m+1:I)})} \right) && (44)\\ &= x_i^2 \left(\mathsf{E}\Big[\frac{\chi_i^{m+1:I}}{q^2_{x_i}(z^{-1}_{(m+1:I)})}\Big] - 1 \right) && (45)\\ &= x_i^2 \mathsf{E}[1/q_{x_i}(z^{-1}_{(m:I\setminus\{i\})}) - 1] && (46)\\ &= \mathsf{E}[x_i (z_{m:I\setminus\{i\}} - x_i)^+] && (47) \end{aligned}$$

The integral form (iii) follows after integration by parts. ■

Proof of Theorem 5: by Theorem 2(ii)

$$\begin{aligned} &\mathsf{E}[\chi_i^{m+1:I} z_{[m+1:I]} (z_{[m+1:I]} - x_i)^+] && (48)\\ &= \mathsf{E}\left[\frac{q_{x_i}(z^{-1}_{(m:I\setminus\{i\})})(z_{[m:I\setminus\{i\}]} - x_i)^+}{z^{-1}_{(m:I\setminus\{i\})}} \right] && (49)\\ &= \mathsf{E}[x_i (z_{m:I\setminus\{i\}} - x_i)^+] && (50)\\ &= \mathsf{Var}(\widehat{x}_i) \quad ■ && (51) \end{aligned}$$

Proof of Theorem 6: (i) Applying Chernoff bounds:

$$\mathsf{P}[N_z \geq m] \leq e^{-\sup_{\theta>0}(\theta m - c(\theta, z))} \quad (52)$$

Now $c(\cdot, z)$ is convex and has derivative $\partial c(0, z)/\partial\theta = m(z)$. Hence for $m \geq m(z)$ the supremum in (52) can be extended to all $\theta \in \mathbb{R}$. A similar argument bounds $\mathsf{P}[N_z \leq m]$ for $m \leq m(z)$.

(ii) By a standard standard convex analysis (see [19]), $\partial J(m, z)/\partial m$ is the inverse function of $\partial c(\theta, \beta)/\partial\theta$. Hence, since $\partial c(0, z)/\partial\theta = m(z)$, then $\partial J(m(z), z)/\partial m = 0$. As a supremum of convex functions, $J(\cdot, z)$ is convex and hence takes its minimum value uniquely at $m(z)$. By convexity, $c(\theta, z) \geq \theta m(z)$, and hence $J(m(z), z) = 0$.

(iii) follows from (i) using (15) ■

Lemma 1: Let g be a positive function, and h a differentiable and strictly positive function on R^+ that decreases to 0 at infinity.

$$\int_{v_0}^{\infty} dv\, g(v) \leq h(v_0) \sup_{v\geq v_0} \frac{g(v)}{-h'(v)} \quad (53)$$

Proof of Lemma 1: By a change of variables

$$\begin{aligned} \int_{v_0}^{\infty} dv\, g(v) &= \int_{h(v_0)}^{\infty} dy\, \frac{g(h^{-1}(y))}{h'(h^{-1}(y))} && (54)\\ &= h(v_0) \sup_{y\in(0,h(v_0)]} \frac{g(h^{-1}(y))}{-h'(h^{-1}(y))} && (55)\\ &= h(v_0) \sup_{v\geq v_0} \frac{g(v)}{-h'(v)} \quad ■ && (56) \end{aligned}$$

Lemma 2: (i) $z \mapsto n_z^+ + X_z^-/z$ is a decreasing function of $(0, \infty)$ onto $(0, n)$. Hence when $m < n$, the equation

$$n_z^+ + X_z^-/z = m \quad (57)$$

has a unique positive solution for z, which we denote by z_*.
(ii) $z \mapsto X_z^-/(m - n_z^+)$ is nondecreasing on $[z_*, \infty)$.
(iii) $m > n_z^+$ for $z > z_*$.

Proof of Lemma 2: (i) $z \mapsto n_z^+ + X_z^-/z$ is continuous at each x_i, and decreasing otherwise.

(ii) $X_z^-/(m - n_z^+)$ is continuous from the left and piecewise constant between successive ordered values in Ω. Let $y < z$ be two such successive values. We show that $X_y^-/(m - n_y^+) > X_z^-/(m - n_z^+)$ leads to a contradiction. Since $n_z^+ = n_y^+ + 1$ and

$X_z^- = X_y^- + y$, this would yield $n_y^+ + X_y^-/y > m$, contradicting (i) since $y > z_*$.

(iii) by (i), $X_z^-/z + n_z^+ < m$ for $z > z_*$, whence $m > n_z^+$. ■

PROOF OF THEOREM 7: Set $k = m-1$ and $n = \#I - 1$. According to Theorem 4,

$$\frac{\mathsf{Var}(\widehat{x}_i)}{x_i} = \mathsf{E}[(z'-x)^+] \tag{58}$$

$$= \int_{x_i}^{\infty} dz\, \mathsf{P}[z' > z] = \int_{x_i}^{\infty} dz\, \mathsf{P}[N_z > k] \tag{59}$$

$$\leq (z_* - x_i)^+ + \int_{\min\{x_i, z_*\}}^{\infty} dz \mathsf{P}[N_z > k] \tag{60}$$

where we have bounded the integrand above by 1 for $z \leq z^*$. In the remainder of the proof, we establish an upper bound on the last term in (60) through a number of steps.

(i) *A Chernoff Bound.* $\mathsf{P}[N_z > k] \leq e^{-\theta k}\mathsf{E}[e^{\theta N_z}]$ for any $\theta \geq 0$. Now

$$\mathsf{E}[e^{\theta N_z}] = \prod_{i\in I}(1 + (e^\theta - 1)p_z(x_i)) \tag{61}$$

$$= e^{\theta n_z^+} \prod_{i:\, x_i < z} (1 + (e^\theta - 1)x_i/z)) \tag{62}$$

$$\leq g_z(\theta) := e^{\theta n_z^+}(1 + (e^\theta - 1)x_z^-/z)^{n_z^-} \tag{63}$$

where the inequality follows from the concavity of $x \mapsto \log(1 + (e^\theta - 1)x/z)$.

(ii) *Least upper bound.* Assertion: when $z \geq z_*$, $g_z(\theta)$ takes its minimum at $\theta_z > 0$ where

$$e^{\theta_z} = \frac{(k - n_z^+)(z - x_z^-)}{x_z^-(n-k)} \tag{64}$$

yielding the following expression for $g_z(\theta_z)$:

$$\left(\frac{1 - x_z^-/z}{1 - (k - n_z^+)/n_z^-}\right)^{n_z^- - (k - n_z^+)} \left(\frac{x_z^-/z}{(k - n_z^+)/n_z^-}\right)^{k - n_z^+} \tag{65}$$

Proof: $g_z''(\theta) = e^\theta(z - x_z^-)X_z^-/((e^\theta - 1)x_z^- + z)^2 > 0$ and hence g_z is convex, while $g_z'(\theta) = 0$ requires $e^\theta = (k - n_z^+)(z - x_z^-)/(x_z^-(n-k)) \geq 1$. By Lemma 2 (iii), since $k > n_z^+$, this requires $z > X_z^-/(k - n_z^+)$, or, in other words, $z > z_*$. (65) then follows by substitution of (64) into g_z.

(iii) Using Lemma 1 with $h(x) = x^{-\alpha}$, $\alpha > 0$ and $v_0 = x$:

$$\int_x^\infty dz\, \mathsf{P}[N_z \geq k] \leq \frac{1}{\alpha x^\alpha} \sup_{z \geq x} z^{1+\alpha} g_z(\theta_z) \tag{66}$$

$$\leq \frac{1}{\alpha x^\alpha} \sup_{z\geq x}\left(\frac{X_z^-}{k - n_z^+}\right)^{1+\alpha} q(n_z^-, k - n_z^+, 1+\alpha) \tag{67}$$

$$\leq \frac{1}{\alpha x^\alpha}\left(\frac{X}{k}\right)^{1+\alpha} q(n_x^-, k - n_x^+, 1+\alpha) \tag{68}$$

The last two inequalities use Lemma 3 with the substitutions $n \to n_z^-$ and $k \to k - n_z^+$, together with $n_z^- = n - n_z^+ > k - n_z^+$ since $n > k$. ■

PROOF OF THEOREM 8:

$$\limsup_{n,m,X\to\infty} \inf_{\alpha>0} u_\alpha(x) \leq \inf_{\alpha>0} \limsup_{n,m,X\to\infty} u_\alpha(x) \tag{69}$$

$$= \inf_{\alpha>0} \frac{x_\infty^{1+\alpha}}{\alpha x^\alpha} \tag{70}$$

When $x \geq x_\infty$, $x_\infty^{1+\alpha}/(\alpha x^\alpha)$ decreases to 0 as $\alpha \to \infty$. Otherwise, it takes it minimum when $\alpha = 1/\log(x_\infty/x)$. ■

Lemma 3: Let $m > \beta > 0$.

(i)

$$\left(\frac{1 - x/z}{1 - m/n}\right)^{n-m}\left(\frac{x/z}{m/n}\right)^m z^\beta \leq \left(\frac{xn}{m}\right)^\beta q(n, m, \beta) \tag{71}$$

(ii) When $n \geq m$, $k \mapsto q(n-k, m-k, \beta)$ is increasing for $0 \leq k < n - b, m$.

(iii) When $n \geq m$, $\beta \mapsto q(n, m, \beta)$ is increasing for $0 \leq k < m$.

PROOF OF LEMMA 3: (i)Some algebra shows that the LHS of (3) is equal to the RHS times the quantity

$$\exp((n-\beta)f(x/z, (m-\beta)/(n-\beta))) \tag{72}$$

where $f(a,b) = (1-b)\log((1-a)/(1-b)) + b\log(a/b)$. But $\partial^2 f(a,b)/\partial b^2 = -1/b(1-b) < 0$ and hence $b \mapsto f(a,b)$ is concave, and takes its maximum value, 0, when $\partial f(a,b)/\partial b = 0$, which occurs when $a = b$.

(ii) Using $\log x \geq 1 - 1/x$,

$$\frac{\partial \log q(n-k, m-k, \beta)}{\partial k} \tag{73}$$

$$= \frac{b(m-n)}{(n-k)(m-k)} + \log\left(\frac{1 - b/(n-k)}{1 - b/(m-k)}\right) \tag{74}$$

$$\geq \frac{b^2(n-m)}{n-b-k} \geq 0 \tag{75}$$

(iii)

$$\frac{\partial \log q(n-k, m-k, \beta)}{\partial \beta} = \log\left(\frac{1 - \beta/n}{1 - \beta/m}\right) > 0 \quad ■ \tag{76}$$

PROOF OF THEOREM 9: Suppose first that $h \leq m$. Then the smallest power of α is obtained in the single term with $i = m$ and $j = h$. This term is

$$\alpha^{m-h}\binom{n-1-h}{m-h} u_{I,m}(\alpha) \tag{77}$$

where

$$u_{I,m}(\alpha) = \int_0^1 dy\, y^{m-2}(1 - \alpha y)^{n-1-m} \tag{78}$$

From preceding remarks, it is clear we can take $\alpha = 0$ in the integrand (only) without changing the dominant behavior as $\alpha \to 0$. In this limit we use $u_{I,m}(0) = 1/(m-1)$. More generally, this remains a good approximation provided $(1-\alpha)^{n-1-m} \approx 1$, e.g., if $\alpha(n-1-m) \ll 1$.

Now suppose $h > m$. Then each of the terms with $i = j$ for $i = m, \dots, h$ gives rise to a term in α^0. Thus the dominant terms combine to form

$$\int_0^1 dy \sum_{i=m}^{h} \binom{h}{i} y^{i-2}(1-y)^{h-i}(1-\alpha y)^{n-1-h} \tag{79}$$

As before we can setting $\alpha = 0$ in the integrand, in order to obtain the behavior $h/(m-1) - 1$ as $\alpha \to 0$. ∎

References

[1] B.C. Arnold and N. Balakrishnan, "Relations, Bounds and Approximations for Order Statistics", Lecture Notes in Statistics, vol. 53, Springer, New York, 1988.

[2] B.-Y. Choi, J.Park, Zh.-L. Zhang, "Adaptive Random Sampling for Load Change Detection", ACM SIGMETRICS 2002 (Extended Abstract).

[3] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, "Introduction to algorithms", 2nd Edition, MIT Press, McGraw-Hill, 2001.

[4] H.A. David, "Order Statistics", Second Edition, Wiley Series in Probability and Mathematical Statistics, Wiley, New York, 1981

[5] Cisco NetFlow; for further information see http://www.cisco.com/warp/public/732/netflow/index.html.

[6] N.G. Duffield (Editor), "A Framework for Passive Packet Measurement", Internet Draft, October 2003, draft-ietf-psamp-framework-04, work in progress.

[7] N. G. Duffield and M. Grossglauser, "Trajectory Sampling for Direct Traffic Observation", *IEEE/ACM Transactions on Networking*, vol. 9, pp. 280-292, 2001. Abridged version in Proc. ACM Sigcomm 2000, Computer Communications Review, Vol 30, No 4, October 2000, pp. 271–282.

[8] N.G. Duffield and C. Lund, "Predicting Resource Usage and Estimation Accuracy in an IP Flow Measurement Collection Infrastructure", ACM SIGCOMM Internet Measurement Conference 2003, Miami Beach, Fl, October 27-29, 2003

[9] N.G. Duffield, C. Lund, M. Thorup, "Charging from sampled network usage," ACM SIGCOMM Internet Measurement Workshop 2001, San Francisco, CA, November 1-2, 2001.

[10] N.G. Duffield, C. Lund, M. Thorup, "Learn More, Sample Less: Control of Volume and Variance in Network Measurement", Technical Report.

[11] N.G. Duffield, C. Lund, M. Thorup, "Estimating flow distributions from sampled flow statistics", ACM Sigcomm 2003, Karlsruhe, Germany, August 25-29, 2003.

[12] C. Estan and G. Varghese, "New Directions in Traffic Measurement and Accounting", Proc SIGCOMM 2002, Pittsburgh, PA, August 19–23, 2002.

[13] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, J. Rexford, F. True, "Deriving traffic demands for operational IP networks: methodology and experience", In Proc. ACM Sigcomm 2000, Computer Communications Review, Vol 30, No 4, October 2000, . 257–270.

[14] A. Feldmann, J. Rexford, and R. Cáceres, "Efficient Policies for Carrying Web Traffic over Flow-Switched Networks," *IEEE/ACM Transactions on Networking*, vol. 6, no.6, pp. 673–685, December 1998.

[15] E.A. Hernandez, M.C. Chidester, A.D. George, "Adaptive Sampling for Network Management", Technical Report, Department of Electrical and Computer Engineering, University of Florida.

[16] N. Hohn, D. Veitch, "Inverting Sampled Traffic", ACM SIGCOMM Internet Measurement Conference 2003, Miami Beach, Fl, October 27-29, 2003

[17] D.G. Horvitz and D.J. Thompson, "A Generalization of Sampling without replacement from a Finite Universe", J. Amer. Statist. Assoc. Vol. 47, pp. 663-685, 1952.

[18] Inmon Corporation, "sFlow accuracy and billing", see: http://www.inmon.com/PDF/sFlowBilling.pdf

[19] R.T. Rockafellar, "Convex analysis", Princeton University Press, Princeton, NJ, 1970.

[20] S.K. Thompson and G.A.F. Seber, "Adaptive Sampling", Wiley Series in Probability and Statistics, Wiley, New York, 1996.

[21] M. Thorup, "Equivalence between Priority Queues and Sorting", Proc. 43rd IEEE Symposium on Foundations of Computer Science (FOCS), pp. 125–134, 2002.