# Comparison of Sampling Methods for Non-Intrusive SLA Validation

Tanja Zseby
Fraunhofer Institut FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin
zseby@fokus.fraunhofer.de

*Abstract*-- ***This paper introduces an approach for non-intrusive SLA validation based on sampling techniques. Suitable sampling methods are compared with regard to sampling effort and achievable accuracy. Mathematical modeling is supplemented by tests with traffic traces from distributed online gaming with multiple players in Berlin, Madrid and Kawasaki.***

***Index terms*-- network measurements, statistics**

## I. NON-INTRUSIVE QUALITY VALIDATION FOR STATISTICAL SLAs

Non-intrusive measurements provide an elegant way for investigating the quality of existing flows without burdening the network with test traffic. Based on existing customer traffic, they generate exactly the quality statements required for the validation of guarantees in service level agreements (SLAs). Nevertheless, increasing data rates make it costly and sometimes impossible to provide exact quality statements for all observed packets. Furthermore, costs and required resources for the SLA validation should be limited to a small fraction of the service provisioning costs itself.

Therefore we propose to formulate ***statistical SLAs*** that are based on quality estimations, instead of exact measurements. Many applications do not require exact values for quality parameters and can tolerate a few packets that violate the SLA. Furthermore, reduced measurement costs allow providers to offer services at lower prices, giving an incentive for customers to use statistical SLAs. Additionally, the expected accuracy, which should be included in the SLA, could be adapted to customers' demands.

A comparison of passive and active measurements and approaches for passive multipoint measurements can be found in [GrDM98] and [ZsZC01]. Sampling methods for investigations of different metrics have been presented in [ClPB93], [DuGr00], [DuLT02] and [ChPZ02]. Standardization of schemes is done within the IETF [PSAMP].

In this paper we introduce an approach for efficient non-intrusive SLA validation based on sampling. We model the SLA validation problem as estimation of the percentage of packets that violate the SLA and use one-way delay as example metric. We investigate the tradeoff between sampling effort and achievable accuracy and discuss solutions for sampling synchronization for multipoint measurements. Empirical tests with traces from distributed online gaming supplement the theoretical results.

## II. SAMPLING TECHNIQUES

### *A. Classification of Sampling Schemes*

In accordance to PSAMP terminology introduced in [ZsMR04] and [Duff04] we distinguish ***random sampling***, where packet selection is based on random functions and ***systematic sampling***, based on deterministic functions. The selection decision can depend on packet arrival time (***time-based***), packet count (***count-based***) or parts of the packet content (***content-based***). Table 1 gives an overview of the categorization and a short description of the basic schemes.

Figure 1 shows a classification of schemes with regard to the ***sampling effort*** and ***required traffic information***. Systematic time-based sampling requires no traffic information and is very simple to realize by periodically enabling/disabling the packet capturing function. The effort for systematic count-based sampling is equally small, but the scheme requires packet counters. Filtering needs access to packet content. Probabilistic and n-out-of-N sampling require the generation of random numbers and therefore require a higher effort. n-out-of-N requires slightly more effort due to the maintenance of random number lists. Non-uniform probabilistic schemes require functions for probability calculation based on arrival time, packet count or content. Since packet counters are widely available in routers and other network devices, we concentrate on count- and content-based schemes.

| | Scheme | Description |
|---|---|---|
| Random | n-out-of-N | Random selection of n elements out of a population of N |
| | Uniform probabilistic | Each packet selected with probability p |
| | Non-uniform probabilistic (probabilistic time-, count- or content-based) | Like simple probabilistic, but selection probability depends on temporal or spatial packet position or packet content |
| System. | Systematic count-based | Deterministic selection based on the spatial packet position |
| | Systematic time-based | Deterministic selection based on the packet arrival time |
| | Filtering (systematic content-based) | Deterministic selection based on packet content (includes hash-based methods) |

**Table 1: Description of Sampling Schemes**

*B. Sampling for Multipoint Measurements*

Sampling for multipoint measurements must ensure that the same packets are captured at all involved measurement points. A synchronization of sampling processes can be realized by ***hash-based methods*** as proposed in [DuGr00]. This method allows deployment of sampling at all observation points, but provides only a pseudo random selection. Furthermore, it requires packet processing, which is costly compared to other methods.

A further approach is the use of a ***heterogeneous measurement infrastructure***. A few high performance measurement points are positioned close to servers, provide full measurements and serve as collectors. Less sophisticated measurement devices based on sampling are used at client sides and transfer results to the collector [Zseb02]. This allows the applicability of arbitrary sampling schemes, i.e. also very cost efficient schemes. Since full measurements are required at a few points, this approach is only cost efficient, in scenarios where many users communicate with a few servers.

III. MATHEMATICAL MODELS

*A. Proportion vs. Percentile Estimation*

The goal of SLA validation is to check whether packets in a data stream are conformant to the QoS guarantees given in an SLA. An estimation of the whole distribution of the metric of interest (here delay) is difficult and contains much more information than needed in this context. The estimation of ***mean*** and ***standard deviation*** of the delay values can give first insights about the quality situation for the application, but are inadequate to validate the conformance to an SLA.

***Percentiles*** of the delay distribution provide valuable parameters to assess the general network situation [ChMC03]. E.g. the $95^{th}$ percentile provides the information that 95% of all delays are below the percentile value. But if the $95^{th}$ percentile lies above the delay threshold in the SLA, we get no idea what percentage of packets really violated the contract.

Therefore we propose a different approach. Instead of estimating percentiles, we estimate the ***percentage*** of packets that violate the contract. With this we model the validation task as an estimation of the proportion of packets that exceed a given delay limit.

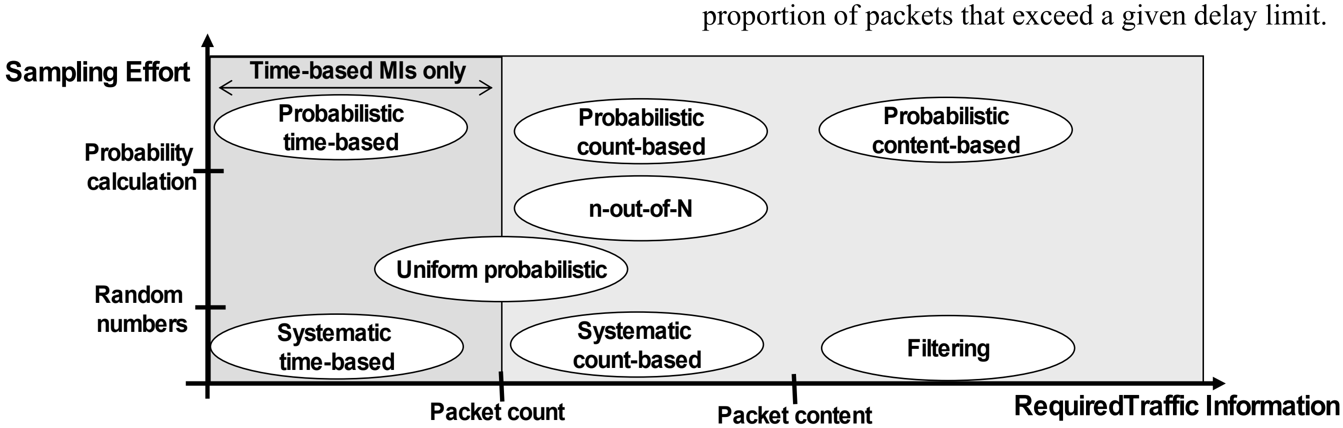


**Figure 1: Overview of Sampling Schemes**

A packet with delay $d > d_{max}$ is considered as a violator (hit,1), packets with delay $d \leq d_{max}$ are considered as conformant (no hit, 0). Then we can model the number of non-conformant packets in a measurement interval as binomial distributed random variable.

*B. Estimation Accuracy*

An assessment of the estimation quality is done by looking at the distribution of the estimate (i.e. how the estimate would evolve if we perform infinite sampling runs). We need to consider two important quality criteria: The ***bias*** quantifies how far the mean of all estimates (for infinite sampling runs) lies from the exact value and is measured by the expectation. The ***precision*** quantifies how the estimates from multiple experiments scatter around the mean and is measured by the variance. For an easier comparison of schemes we use the relative standard error.

$$StdErr_{rel}\left[\hat{P}\right] = \frac{\sqrt{V\left[\hat{P}\right]}}{P}. \quad (1)$$

| Notation | Description |
|---|---|
| N | Number of packets in population (measurement interval) |
| $n_R$ | Real sample size: real number of packets in sample |
| $n_T$ | Target sample size: number of packets in sample that is aimed at (fixed value) |
| M | Number of violators in measurement interval |
| m | Number of violators in sample |
| P=M/N | Violator proportion in measurement interval |
| $\hat{P} = m/n_R$ | Estimated violator proportion from sample |
| $f_R = n_R/N$ | Real sample fraction |
| $f_T = n_T/N$ | Target sample fraction (equals the selection probability used in probabilistic sampling) |
| $K = N/n_T$ | Sample period for systematic sampling (rounded off in experiments) |

**Table 2: Notation**

*C. n-out-of-N Sampling*

n-out-of-N sampling selects exactly n packets out of the population of the N packets observed in the measurement interval.

We can estimate the number of violators in the measurement interval from the number of violators in the sample as follows:

$$\hat{M} = \frac{1}{f_T} \cdot m = \frac{1}{f_T} \cdot \sum_{i=0}^{n_T} x_i \quad \text{with} \quad x_i \sim Be(P) \quad (2)$$

The random variable $x_i$ denotes the conformance of the sampled packets to the SLA ($x_i = 0$ if packet delay $d \leq d_{max}$ and $x_i = 1$ if $d > d_{max}$). $x_i$ can be modelled as Bernoulli distributed random variable (RV) with probability of success P=M/N. The number of violators m in the sample can be modelled as number of hits in an experiment with $n_T$ trials. Since we cannot select a packet that we once selected again, we have to consider a selection without replacement, i.e. m can be considered as RV with a hyper geometric distribution. The proportion P of violators in the measurement interval is estimated by the proportion of violators in sample.

$$\hat{P} = \frac{\hat{M}}{N} = \frac{1}{N \cdot f_T} \cdot m = \frac{m}{n_T} \quad (3)$$

We get an unbiased estimate $\hat{P}$ (see appendix) with the following variance:

$$V\left[\hat{P}\right] = V\left[\frac{m}{n_T}\right] = \frac{1}{n_T^2} \cdot V\left[m\right] = \frac{1}{n_T} \cdot \frac{N - n_T}{N - 1} \cdot P \cdot \left(1 - P\right) \quad (4)$$

With N-1≈N we get the following relative standard error (see derivation in appendix (section **Error! Reference source not found.**)):

$$StdErr_{rel} = \sqrt{\frac{(1-P) \cdot (1-f_T)}{P \cdot n_T}} \quad (5)$$

The estimation accuracy depends on the sample fraction and on the real violator proportion. The real violator proportion is unknown and has to be approximated from the sample or replaced by worst case parameters in order to make an accuracy prediction in advance. For small sample fraction ($f_T <$ 5%), we can neglect the finite population correction and get the following simplified formula:

$$StdErr_{rel} = \frac{1}{P} \cdot \sqrt{\frac{P \cdot (1-P)}{n_T}} = \sqrt{\frac{(1-P)}{P \cdot n_T}} \quad (6)$$

*D. Probabilistic Sampling*

Probabilistic sampling makes a selection decision per packet. That means each packet is selected with a given probability p regardless of the fact how many packets already have been selected before. The real sample size $n_R$ varies for each run, and usually differs from the target sample size $n_T$ ($n_R \neq n_T$). In [DuLT02] the number of packets that belong to a specific flow is estimated by modelling the probabilistic selection process with a Bernoulli distributed random variable $\omega_i$ with success probability $f_T=n_T/N$. The variability of $n_R$ is neglected. $\omega_i$ is defined to be 1 if the packet is selected and 0 if the packet is not selected. If we consider the packet property "violate SLA" instead of "belong to flow f", we can apply the same model and get the following estimate for the number of violators[1].

$$\hat{M} = \frac{1}{f_T} \cdot m = \frac{1}{f_T} \cdot \sum_{i=0}^{M} \omega_i \quad \text{with} \quad \omega_i \sim Be(f_T) \qquad (7)$$

We get an unbiased estimate and calculate the variance and standard error for the estimate $\hat{P}$ in accordance to [DuLT02] as follows (see appendix):

$$V\left[\hat{P}\right] = V\left[\frac{\hat{M}}{N}\right] = \frac{V\left[\hat{M}\right]}{N^2} = \frac{M}{N^2} \cdot \left(\frac{1}{f_T} - 1\right) = \frac{P}{N} \cdot \left(\frac{1}{f_T} - 1\right) \qquad (8)$$

$$StdErr_{rel}\left[\hat{P}\right] = \frac{1}{P} \cdot \sqrt{\frac{P}{N} \cdot \left(\frac{1}{f_T} - 1\right)} = \sqrt{\frac{1 - f_T}{P \cdot n_T}} \qquad (9)$$

*E. Systematic Sampling*

For independent packet delays we can apply the n-out-of-N model. But if correlations occur, the systematic selection process can interfere with periodicities in the packet sequence. We may get a non-representative accumulation of packets with specific properties (e.g. packets with high delays) in the sample and with this a biased estimation. The nature of this bias heavily depends on the specific traffic mix. Therefore we cannot derive a generic model (valid for arbitrary traces) as for random methods.

[1] Please note that for a consistent notation throughout this document we use a different notation than [DuLT02].

*F. Filtering and Hash-based Sampling*

Filtering (systematic content-based sampling) is a deterministic function on the packet content. It is quite likely to introduce a bias due to interferences of the traffic mix structure with the deterministic selection process. A special form of filtering is the hash-based selection [DuGr00], which emulates probabilistic sampling. A deterministic function on the packet content is used to calculate a hash value. If the hash value falls in a specific range the packet is selected. If sufficient randomness is achieved (see [DuGr00]), we can apply the probabilistic sampling model.

*G. Comparison of Schemes*

From the mathematical models we observe that the relative standard error for n-out-of-N sampling is equal to the standard error for probabilistic sampling multiplied by a factor $\sqrt{(1-P)}$.

$$StdErr_{nofN} = \sqrt{(1-P)} \cdot StdErr_{prob} \qquad (10)$$

Since $0 \le \sqrt{(1-P)} \le 1$, we can deduce that n-out-of-N sampling provides a smaller standard error and with this a better accuracy than probabilistic sampling. Nevertheless, the difference depends on the violator proportions in the measurement interval and can get very small if there are only few violators.

Figure 32 and Figure 3 show the theoretical standard error for n-out-of-N (approximated and exact formula) and probabilistic sampling for a violator proportion of P=0.2 and P=0.8. It can be seen that for higher violator proportions the difference between the schemes increases. It also can be observed how the approximated n-out-of-N model diverges from the exact model for large sample fractions.

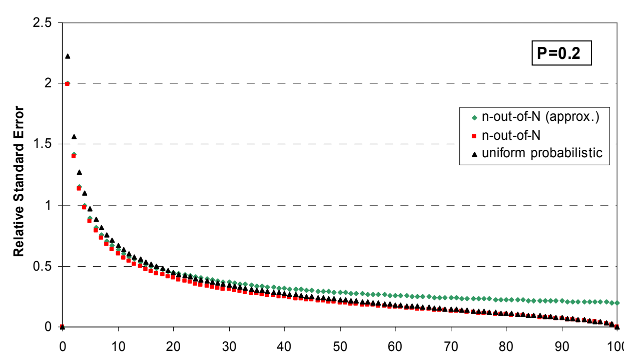


**Figure 2: Theoretical Standard Error over Sample Fraction (for P=0.2)**

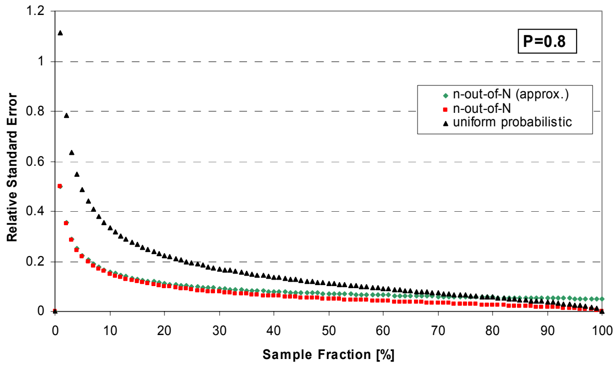


**Figure 3: Theoretical Standard Error over Sample Fraction (for P=0.8)**

## IV. EXPERIMENTS

### *A. Trace Collection*

The traces were collected during a demo event for IPv6 measurement software developed in the 6QM project [6QM]. Players in Berlin, Madrid and Kawasaki (Japan) participated in a distributed gaming event with Quake2 over IPv6. The involved networks were WIDE (Japan), Euro6IX(Spain) and 6WIN (Germany). GPS synchronized measurement boxes were installed at all participating locations to perform passive one-way delay measurements. We collected 6 traces with up to 139,756 packets per trace between the server in Berlin and clients in Berlin, Madrid and Kawasaki.

### *B. Empirical Results*

We split the traces into measurement intervals of N=10,000 packets each and simulated 10,000 sample runs for each scheme. We repeated the experiments for different sample fractions from 1 to 100 %. The diagrams show the relative standard error over the target sample fraction, which may differ from the real sample fraction for probabilistic and systematic sampling.

Figure 4 and Figure 5 compare the theoretical with the empirical results from the Berlin-Kawasaki trace for n-out-of-N and probabilistic sampling With a threshold of 170.00 ms we get a violator proportion P=0.144. Empirical results for both random schemes are very close to the model. Further experiments with the other traces and parameter settings showed the same compliance to the models.

Figure 7 and shows the empirical results for n-out-of-N, probabilistic and systematic sampling for violator proportions P=0.144 and P= 0.913 (realized by a lower threshold). For systematic sampling the start point was chosen randomly to get a variation of

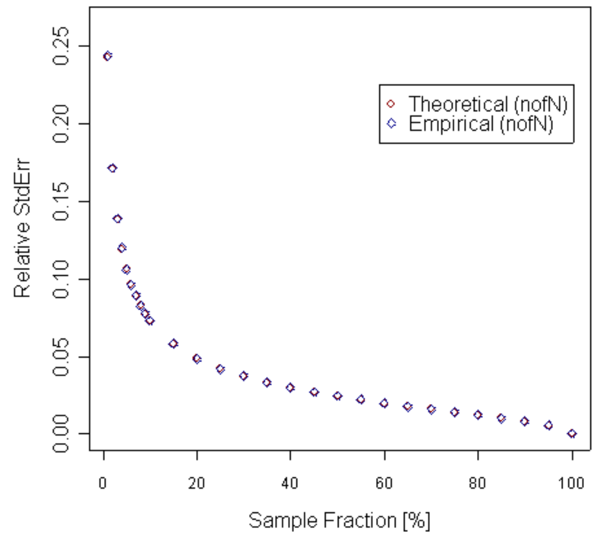


**Figure 4: Comparison Empirical Results with Theoretical Model (n-out-of-N)**

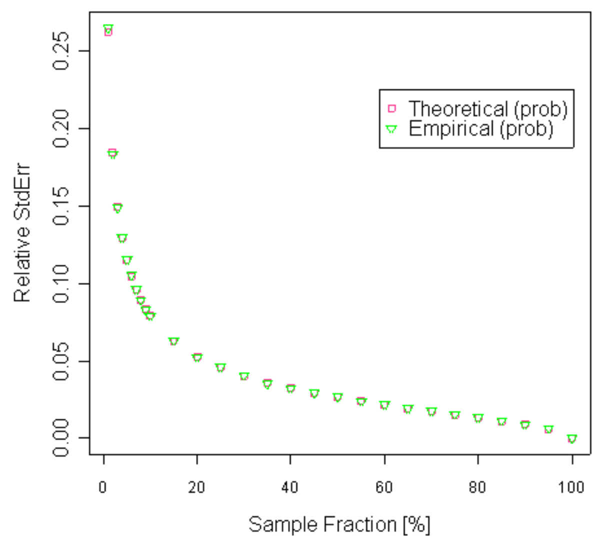


**Figure 5: Comparison Empirical Results with Theoretical Model (probabilistic)**

the estimates for different runs. We round off K=N/n, i.e. we get a higher real sample size if K is no integer.

As expected the difference between n-out-of-N and probabilistic sampling increases for larger violator proportions. For P=0.913 the n-out-of-N shows a much better performance than probabilistic sampling. The results for systematic sampling for P=0.144 cannot be matched to any model. Correlations in the trace lead to a traffic-dependent bias and therefore unpredictable accuracy for systematic methods. Above a target sampling fraction of 50% the standard error for systematic sampling is 0, because K is 1, and the real sample fraction is 100% (all packets captured). For larger proportions (P=0.913) one can see that systematic sampling is closer to n-out-of-N sampling than to probabilistic sampling.

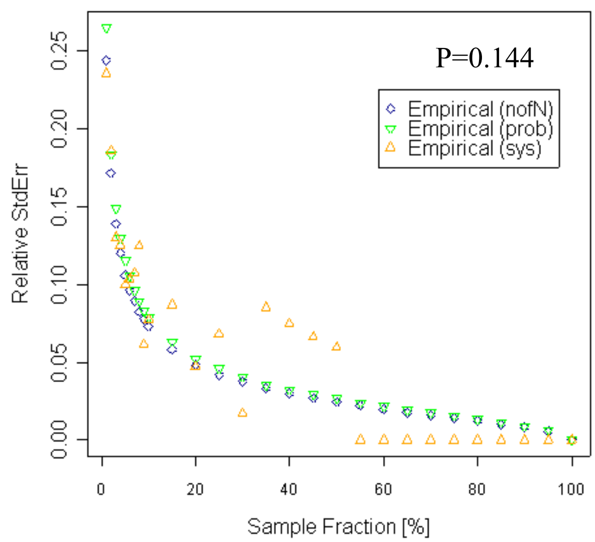


**Figure 6: Empirical Results from n-out-of-N, Probabilistic and Systematic Sampling (P=0.144)**

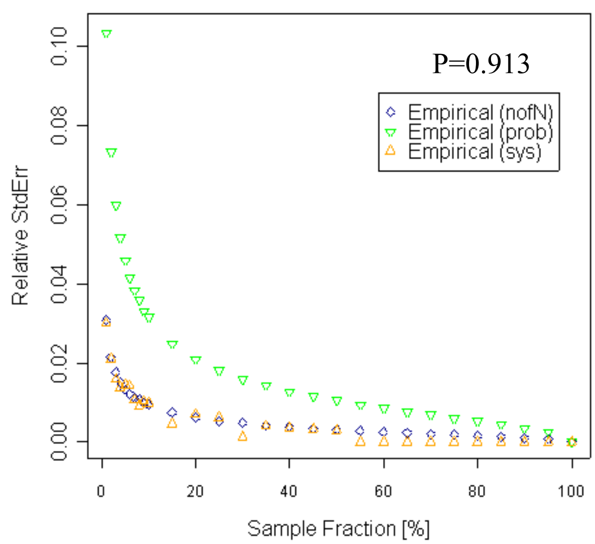


**Figure 7: Empirical Results from n-out-of-N, Probabilistic and Systematic Sampling (P=0.913)**

## V. CONCLUSION

We compared different sampling schemes for the use in non-intrusive SLA validation. From the investigated schemes n-out-of-N sampling requires the most effort and also provides the best accuracy. Probabilistic sampling requires a little bit less effort and also performs worse for traces with many SLA violators. Nevertheless, for nearly conformant traffic mixes the difference to n-out-of-N is rather marginal. Systematic sampling can be realized with very little effort. In our experiments the accuracy was at least in the same range as for the random methods. Nevertheless, the accuracy of systematic sampling is trace dependent and cannot be predicted by a generic model. Therefore those results cannot be generalized for other traces. Hash-based approaches are useful for synchronization of sampling processes in multipoint measurements and can be modeled by probabilistic sampling if sufficient randomness of the selection is ensured. Nevertheless, it requires processing of the packet content and an unbiased selection cannot be ensured for arbitrary traffic traces.

## VI. ACKNOWLEDGEMENTS

I would like to thank Bastian Schern and Guido Pohl for their help with the trace analysis software. The gaming traces were collected with IPv6 measurement software developed in the project [6QM].

## VII. REFERENCES

[6QM] 6QM Project http://www.6qm.org

[ChMC03] B.-Y. Choi, S. Moon, R. Cruz, Z.-L. Zhang, C.Diot. Practical Delay Measurements for ISPs. SPRINT ATL research report RR03-ATL-051910, May 2003,

[ChPZ02] B.-Y. Choi, J. Park, Z.-L. Zhang. Adaptive Random Sampling for Load Change Detection. Poster at ACM SIGMETRICS 2002, Marina Del Rey, California, USA, June 15-19, 2002

[ClPB93] K.C. Claffy, G. C. Polyzos, H.-W. Braun. Application of Sampling Methodologies to Network Traffic Characterization. Proceedings of ACM SIGCOMM'93, San Francisco, CA, USA, September 13 - 17, 1993

[Duff04] N. Duffield (Editor): A Framework for Packet Selection and Reporting. Internet Draft draft-ietf-psamp-framework-06, work in progress, July 2004

[DuGr00] N. Duffield, M. Grossglauser: Trajectory Sampling for Direct Traffic Observation, Proceedings of ACM SIGCOMM 2000, Stockholm, Sweden, August 28 - September 1, 2000

[DuGr00] N. Duffield, M. Grossglauser. Trajectory Sampling for Direct Traffic Observation. Proceedings of ACM SIGCOMM 2000, Stockholm, Sweden, August 28 - September 1, 2000.

[DuLT02] N. Duffield, C. Lund, and M. Thorup: Properties and Prediction of Flow Statistics from Sampled Packet Streams, ACM SIGCOMM Internet Measurement Workshop 2002, Marseille, France, November 6-8, 2002

[GrDM98] I. D. Graham, S. F. Donnelly, S. Martin, J. Martens, J. G. Cleary: Nonintrusive and Accurate Measurement of Unidirectional Delay and Delay Variation on the Internet, INET'98, Geneva, Switzerland, 21-24 July, 1998

[PSAMP] IETF PSAMP Group http://psamp.ccrle.nec.de/

[Zseb02] T. Zseby: Deployment of Sampling Methods for SLA Validation with Non-Intrusive Measurements, Proceedings of Passive and Active Measurement Workshop (PAM 2002), Fort Collins, CO, USA, March 25-26, 2002

[ZsMR04] T. Zseby, M. Molina, F. Raspall, N. Duffield. Sampling and Filtering Techniques for IP Packet Selection. IETF draft draft-ietf-psamp-sample-tech-04.txt, work in progress, February 2004

[ZsZC01] T. Zseby, S. Zander, G. Carle: Evaluation of Building Blocks for Passive One-way-delay Measurements, Proceedings of Passive and Active Measurement Workshop (PAM 2001), Amsterdam, The Netherlands, April 23-24, 2001 (PAM 2001)

## VIII. Appendix

### A. *n-out-of-N Sampling*

Expectation and variance of hyper geometric random variable m:

$$E[m] = n_T \cdot P$$

$$V[m] = \frac{N - n_T}{N-1} \cdot n_T \cdot P \cdot (1-P)$$

Expectation of estimate $\hat{P}$:

$$E[\hat{P}] = E\left[\frac{m}{n_T}\right] = \frac{1}{n_T} \cdot E[m] = \frac{1}{n_T} \cdot n_T \cdot P = P$$

➔ $\hat{P}$ provides an unbiased estimate

Variance of estimate $\hat{P}$:

$$V[\hat{P}] = V\left[\frac{m}{n_T}\right] = \frac{1}{n_T^2} \cdot V[m]$$
$$= \frac{1}{n_T^2} \cdot \frac{N - n_T}{N-1} \cdot n_T \cdot P \cdot (1-P)$$
$$= \frac{1}{n_T} \cdot \frac{N - n_T}{N-1} \cdot P \cdot (1-P)$$

With $\frac{N - n_T}{N-1} \approx \frac{N - n_T}{N} = 1 - \frac{n_T}{N} = 1 - f_T$ the relative standard error can be derived as follows:

$$StdErr_{rel} = \frac{1}{P} \cdot \sqrt{\frac{P \cdot (1-P)}{n_T}} \cdot \sqrt{1 - \frac{n_T}{N}}$$

$$= \sqrt{\frac{(1-P) \cdot (1-f_T)}{P \cdot n_T}}$$

For small sample fraction ($f_T$< 5%), we can assume a binomial distribution and we can derive the following standard error:

$$V[\hat{P}] = V\left[\frac{m}{n_T}\right] = \frac{1}{n_T^2} \cdot V[m] = \frac{1}{n_T^2} \cdot n_T \cdot P \cdot (1-P) = \frac{P \cdot (1-P)}{n_T}$$

$$StdErr_{rel} = \frac{1}{P} \cdot \sqrt{\frac{P \cdot (1-P)}{n_T}} = \sqrt{\frac{(1-P)}{P \cdot n_T}}$$

So for small sample fraction ($f_T$< 5%), the finite population correction can be neglected.

*B. Probabilistic Sampling*

Estimate for the number of violators:

$$\hat{M} = \frac{1}{f_T} \cdot m = \frac{1}{f_T} \cdot \sum_{i=1}^{M} \omega_i$$

Expectation of the estimate $\hat{M}$ in accordance to [DuLT02]:

$$E\left[\hat{M}\right] = E\left[\frac{1}{f_T} \cdot \sum_{i=1}^{M} \omega_i\right] = \frac{1}{f_T} \cdot \sum_{i=1}^{M} E\left[\omega_i\right]$$
$$= \frac{1}{f_T} \cdot M \cdot E\left[\omega_i\right] = \frac{1}{f_T} \cdot M \cdot f_T = M$$

➔ $\hat{M}$ provides an unbiased estimate

Variance of the estimate $\hat{M}$ in accordance to [DuLT02]:

$$V\left[\hat{M}\right] = V\left[\frac{1}{f_T} \cdot \sum_{i=1}^{M} \omega_i\right] = \frac{1}{f_T^2} \cdot \sum_{i=1}^{M} V\left[\omega_i\right] = \frac{1}{f_T^2} \cdot M \cdot V\left[\omega_i\right]$$
$$= \frac{1}{f_T^2} \cdot M \cdot f_T \cdot (1 - f_T) = \frac{1}{f_T} \cdot M \cdot (1 - f_T)$$
$$= M \cdot (\frac{1}{f_T} - 1)$$

Derivation of expectation and variance for estimate $\hat{P}$:

$$E\left[\hat{P}\right] = E\left[\frac{\hat{M}}{N}\right] = \frac{1}{N} \cdot E\left[\hat{M}\right] = \frac{M}{N} = P$$

$$V\left[\hat{P}\right] = V\left[\frac{\hat{M}}{N}\right] = \frac{1}{N^2} \cdot V\left[\hat{M}\right] = \frac{M}{N^2} \cdot \left(\frac{1}{f_T} - 1\right) = \frac{P}{N} \cdot \left(\frac{1}{f_T} - 1\right)$$

Derivation of relative standard error:

$$StdErr_{rel}\left[\hat{P}\right] = \frac{\sqrt{V\left[\hat{P}\right]}}{P} = \frac{\sqrt{\frac{P}{N} \cdot \left(\frac{1}{f_T} - 1\right)}}{P}$$

$$= \sqrt{\frac{1}{N \cdot P} \cdot \left(\frac{1}{f_T} - 1\right)} = \sqrt{\frac{1}{P} \cdot \left(\frac{1}{n_T} - \frac{1}{N}\right)}$$

$$= \sqrt{\frac{N - n_T}{P \cdot n_T \cdot N}} = \sqrt{\frac{1 - \frac{n_T}{N}}{P \cdot n_T}} = \sqrt{\frac{1 - f_T}{P \cdot n_T}}$$