

**Traffic estimation for the largest
sources on a network, using packet
sampling with limited storage**

Jonathan Jedwab, Peter Phaal, Bob Pinna
(Colorado Telecommunications Division)
Management, Mathematics and Security
Department.
HP Laboratories Bristol
HPL-92-35
March 1992

network, sampling,
estimation, truncation

In the management of networks it is often important to know the largest traffic sources and corresponding packet counts, but impractical to examine every packet due to processing and storage constraints. By sampling packets it is possible instead to estimate the actual packet counts statistically for each source.

We present an algorithm for choosing which samples to discard when limited processing and storage allows only a restricted list of sources and their packet counts to be maintained. Subject to modest assumptions on the behaviour of the network traffic, we show that the probability of miscounting sampled packets originating from the t largest sources is negligible for small t , and we bound the probability of misranking these t sources.

The method also applies to traffic estimation for the largest destinations or source-destination pairs, for example.

1 Introduction

To manage a network effectively it is often important to measure a variety of attributes of the transmitted traffic. In the long term, such measurements assist network planning and optimisation. For example, classification of traffic according to source-destination pairs may suggest an adjustment of the network topology to bring those pairs communicating the most onto the same link. An analysis of the traffic between a file server undergoing sustained overloading, and its clients, might be used to allocate selected clients to an additional file server in a balanced way.

In the short term, measurements of traffic attributes allow the prompt detection of network misbehaviour. For example, a source producing an unusually large number of broadcast or multicast packets is likely to have malfunctioned. A breakdown of traffic via source-destination pairs may indicate a disproportionate use of resources, such as interactive game-playing or large file transfers, that would be better postponed until after peak periods.

In order to measure traffic attributes exactly it is necessary to examine every packet. However this is likely to consume appreciable processing resources and may lead to data storage problems. For applications such as those mentioned it is sufficient instead to sample packets and then use statistical techniques to estimate the actual packet counts. Indeed, sampling is not only desirable for reasons of economy and timeliness, it may even be preferable on grounds of accuracy [1], [7]: if a full packet inspection is attempted but the traffic at peak periods exceeds the capability of the monitor to keep up, packet loss will systematically occur at times when measurement may be most crucial.

In this paper we analyse a packet sampling system and show how to calculate confidence intervals for the packet count estimates. In contrast to the network sampling scheme of Amer and Cassel [1] based on time, the scheme presented here is based on packet counts, which in effect removes time from the analysis. Studies of Feldmeier [3] and Minnich [11] have highlighted the difficulties of measuring heavily loaded networks without the aid of sampling.

In many applications, estimates of packet counts are required only for the t largest cases, where $t \leq 10$ say. For instance in the above examples of short term management it would suffice to measure the t largest sources or the t largest source-destination pairs. Nevertheless every incoming sample of a type not yet seen requires a separate entry in a running list of sample counts, and for a large network the maintenance of the list will quickly become costly in terms of processing and storage. We therefore propose to limit in advance the overhead of maintaining the list by restricting its length and possibly discarding some of the smaller counts during the period. Under this scheme, each entry of the list consists of an identifying type followed by a (non-zero) sample count, with the sample counts arranged in decreasing order. When a new packet is sampled, if its type already appears in the list then the corresponding sample count is incremented and the list is re-ordered. Otherwise a new entry is created and appended to the list. If this would cause the number of types in the list to exceed the storage limit then the list is first truncated by removing some entries from the bottom of the list.

We assume that the probability of a sampled packet having any particular type is constant during the period. The method of truncation overcomes the storage and processing limitations, but allows the possibility that the final sample counts for one or more of the t most abundant types may be incorrect due to earlier removal of partial counts. Subject to modest

assumptions on the behaviour of the network traffic, we show that the probability of miscounting in this way is negligible. Given that truncation of the sample counts for these t types does not occur, we bound the probability that they are incorrectly ranked in the final list (*i.e.* that they do not have the same ranks as the corresponding packet counts).

In Section 2 a sampling procedure is described which generates representative random samples and in Section 3 it is shown how to estimate packet counts from such samples. In Sections 4 and 5 the truncation algorithm is analysed according to the miscounting and misranking probabilities respectively.

Theoretical results described here are supported by empirical studies conducted at the University of Geneva and at Hewlett-Packard Laboratories, Bristol.

2 Selection of the sampling process

We firstly consider conditions under which the sampling process is representative in the sense that the long-term proportion of packets possessing a particular attribute, as measured by the sampling process, converges to the true proportion. It is convenient to use the framework of queueing theory, in which this property is known as ASTA (Arrivals See Time Averages); see for example Brémaud [2], Makowski *et al.* [8], Melamed and Whitt [9], [10] and Stidham and El Taha [12]. In the present context the sampling process constitutes the “arrivals”, which “see” the state of the network at precisely the moments when samples are taken. Since the sampling process is entirely under our control, we choose it to be independent of the stochastic process governing the network in order to facilitate unbiased estimation.

Following [8], let the network process and sampling process be discrete-time $\{0, 1\}$ -valued stochastic processes $U \equiv \{U_n : n \in \mathbf{N}\}$ and $A \equiv \{A_n : n \in \mathbf{N}_+\}$ respectively. $U_n = 1$ corresponds to the n th packet on the network possessing the attribute of interest, and $A_n = 1$ corresponds to the $(n - 1)$ th packet being sampled (and “seeing” the state U_{n-1}). Let

$$Y = \sum_{k=1}^n A_k U_{k-1}, \quad X = \sum_{k=1}^n A_k, \quad \theta = \sum_{k=0}^{n-1} U_k.$$

Up to the $(n - 1)$ th packet, the mean value of U seen by A is then Y/X (which we take to be 0 if $A_1 = \dots = A_n = 0$), whereas the true mean value of U is θ/n . Assume that as $n \rightarrow \infty$, the regularity conditions

$$Y/n \rightarrow \bar{y}, \quad X/n \rightarrow \bar{x} > 0, \quad \theta/n \rightarrow \bar{\theta} \tag{1}$$

hold with probability 1. Then Theorem 8 of Makowski *et al.* [8] implies that as $n \rightarrow \infty$ the ratios Y/X and θ/n converge to the same limit (ASTA holds) with probability 1, provided

$$P(A_n = 1) \rightarrow p_A \text{ as } n \rightarrow \infty, \text{ for some } p_A. \tag{2}$$

To implement the sampling process we wish to know the distribution of the number of packets T to be counted from one sample to the next, which we write as $f_t = P(T = t)$ for $t \in \mathbf{N}_+$. If we choose to satisfy (2) by taking the sampling probability $P(A_n = 1)$ to be constant for all n then T will have the geometric distribution $f_t = p_A(1 - p_A)^{t-1}$. However

there are many other possible distributions for T which will guarantee (2), and so the actual distribution may be selected according to implementation constraints. It is clearly necessary to exclude distributions for which $\{t : f_t > 0\}$ has a greatest common divisor other than 1, for otherwise (2) is false and the sampling process could synchronise with a repeated pattern in the network process to give convergence to the wrong value. Apart from these obviously unsuitable distributions, any choice of $\{f_t\}$ will imply (2) and so ensure the ASTA property, by the Renewal Theorem [4, Theorem 1 (ii), p. 330] (although convergence in (2) may be slow if the distribution of T is heavily skewed).

3 Estimation of packet counts

Suppose that in a given time period a total of n packets are counted, of which a fixed unknown number θ have the attribute of interest. Using the method of Section 2, a random sample of $X = x$ out of the n counted packets is examined for the presence of the attribute. We wish to make inferences about θ from the number of samples Y possessing the attribute, which is a random variable with the hypergeometric distribution

$$P_\theta(Y = y) = \frac{\binom{\theta}{y} \binom{n - \theta}{x - y}}{\binom{n}{x}}.$$

We assume that n is large, so this approximates to the binomial distribution

$$P_\theta(Y = y) \simeq \binom{x}{y} p^y (1 - p)^{x - y},$$

where $p = \theta/n$ is the true proportion of packets having the attribute [4, p. 172]. We also assume that x is large so that Y/x approximates to a Normal distribution with mean p and variance $\sigma^2 = p(1 - p)/x$, by the Central Limit Theorem.

Using this approximation, the maximum likelihood estimate of p is $\hat{p} = y/x$ and an approximate $100(1 - \alpha)\%$ confidence interval for p is $[y/x - z_\alpha \hat{\sigma}, y/x + z_\alpha \hat{\sigma}]$, where $1 - \alpha/2 = \Phi(z_\alpha)$ (Φ is the distribution function of a standard Normal), and

$$\hat{\sigma}^2 = \frac{\hat{p}(1 - \hat{p})}{x - 1} = \frac{y(1 - y/x)}{x(x - 1)}$$

is an unbiased estimate of σ^2 [6, p. 240]. Since $\theta = np$, the maximum likelihood estimate of θ is ny/x and an approximate $100(1 - \alpha)\%$ confidence interval for θ is

$$\left[\frac{ny}{x} - z_\alpha n \hat{\sigma}, \frac{ny}{x} + z_\alpha n \hat{\sigma} \right].$$

4 Probability of miscounting using truncation

4.1 Model

We now analyse the truncation algorithm described in Section 1. Suppose that in a given time period a large number of packets n are counted, which may be partitioned into types

$\{1, \dots, m\}$ for some large m . For example, type i might be origin from source i , or transmission between source-destination pair i . A large random sample of $X = x$ counted packets is made using the method of Section 2, each sample being examined and classified according to its type. Let θ_i be the fixed unknown number of packets having type i and let Y_i be the number of samples having type i . Similar methods to those of Section 3 show that the vector (Y_1, \dots, Y_m) is approximately distributed as a multinomial with parameters $(x; p_1, \dots, p_m)$, where $p_i = \theta_i/n$. This in turn approximates to a multivariate Normal with mean $x(p_1, \dots, p_m)$ and variance-covariance matrix xC , where $C = (c_{ij})$ is given by

$$c_{ij} = \begin{cases} p_i(1 - p_i) & \text{if } i = j \\ -p_i p_j & \text{if } i \neq j. \end{cases} \quad (3)$$

(The distribution degenerates into $m - 1$ dimensions due to the constraint $\sum p_i = 1$.)

When the truncation algorithm is used, sample counts for only $s - 1$ distinct types may be stored, and the sampling of a packet with an s th distinct type causes all but the largest $l - 1$ counts to be removed. We wish to calculate an approximate upper bound for the probability that any of the sample counts for the t most abundant types are incorrect at the end of the time period due to earlier truncations. Such a miscounting error, which arises from fluctuations in the relative ranks of the m true sample counts over the time period, should be distinguished from the usual sampling "error". Typical parameter values might be $s = 51$, $l = 26$ and $t = 5$. If l is set too large then whichever $l - 1$ counts happen to be largest at the first truncation will be unfairly difficult to displace, whereas if l is set too small then there will be a significant risk of truncating an important count.

To carry out the analysis we make the further assumption that each sample has type i with probability p_i , independently of any other sample. (This includes as a special case the continuous time model in which samples of type i arrive as a Poisson process with rate λ_i , independently of samples of other types: put $p_i = \lambda_i / \sum_j \lambda_j$.) This assumption, together with (2), implies the regularity conditions (1). Although successive packet types may be correlated, the assumption of independent sample types is reasonable provided we choose to exclude very small numbers of packets between samples (see Section 2). After relabelling if necessary we may take

$$p_1 \geq p_2 \geq \dots \geq p_m, \quad (4)$$

and we assume that

$$p_l \ll p_t. \quad (5)$$

In order to obtain numerical results we shall follow through the calculations with the $\{p_i\}$ decreasing in a fixed ratio β , so that

$$p_i \simeq (1 - \beta)\beta^{i-1} \quad (1 \leq i \leq m, \quad 0 < \beta < 1) \quad (6)$$

since m is large. The closer the ratio β is to 1, the less difference there is between successive values of p_i . This clearly makes the miscounting probability larger, but by the same token there is less difference between the true sample counts and so the consequences of miscounting samples are slight. We shall obtain numerical results for the range $0.5 \leq \beta \leq 0.95$. For most real networks we expect β to lie towards the lower end of this range, or else that the values $\{p_i\}$ decrease more rapidly than in fixed ratio. The assumption (6) is therefore likely to lead to an upper bound on the miscounting probability.

For a real number a , we use the notation $b = \lceil a \rceil$ for the unique integer satisfying $a \leq b < a + 1$, and $b = \lfloor a \rfloor$ for the unique integer satisfying $a - 1 < b \leq a$.

4.2 The first truncation

The first truncation is of great importance in the analysis. If the sample counts for types $1, \dots, t$ all lie within the top $l-1$ positions at this stage then each has a significant advantage over competitor sample counts currently set to zero, and the possibility of miscounting any of the t counts in future due to truncation may be neglected.

Let the first truncation occur when the R th sample is made, and let the sample counts at $R = r$ be $Y_i(r)$ ($1 \leq i \leq m$). The sample count $Y_i(r)$ will be truncated if

$$\text{rank}(Y_i(r)) \geq l.$$

This condition is difficult to evaluate because it depends on the joint distribution of the $\{Y_i(r)\}$. The calculation is greatly simplified if this is approximated by the condition

$$Y_i(r) \leq \lceil \mathbb{E}Y_i(r) \rceil = \lceil rp_l \rceil,$$

(using $\lceil \mathbb{E}Y_i(r) \rceil$ rather than $\mathbb{E}Y_i(r)$ since $Y_i(r)$ can only take integer values), because the distribution of $Y_i(r)$ is simply a binomial with parameters (r, p_i) .

We therefore approximate the miscounting probability by

$$\begin{aligned} P_C &= P(Y_i(R) \leq \lceil Rp_l \rceil \text{ for at least one } i \text{ in } 1 \leq i \leq t) \\ &\leq \sum_{i=1}^t P(Y_i(R) \leq \lceil Rp_l \rceil) \\ &= \sum_{i=1}^t \sum_{r=s}^{\infty} P(Y_i(r) \leq \lceil rp_l \rceil) \cdot P(R = r) \end{aligned}$$

using the law of total probability.

Suppose for the moment that a function $g_i(r)$ can be found for which

$$P(Y_i(r) \leq \lceil rp_l \rceil) \leq g_i(r) \quad \text{for all } 1 \leq i \leq t. \quad (7)$$

Then we can write

$$\begin{aligned} P_C &\leq \sum_{i=1}^t \sum_{r=s}^{\infty} g_i(r) [P(R \geq r) - P(R \geq r+1)] \\ &= \sum_{i=1}^t \left[\sum_{r=s}^{\infty} [g_i(r) - g_i(r-1)] P(R \geq r) + g_i(s-1) P(R \geq s) \right]. \quad (8) \end{aligned}$$

We show in the Appendix how to bound the right hand side in terms of s, l, t and β , and give an explicit form for $g_i(r)$.

4.3 Numerical results

Figure 1 shows numerical results obtained for the parameter values

- $s = 51, l = 26, t = 10$
- $s = 51, l = 26, t = 5$

in the range $0.5 \leq \beta \leq 0.95$. Both of these parameter sets give rise to a “sawtooth” shaped graph with discontinuities at the same values of β , for reasons explained in the Appendix. We expect the miscounting probability to increase smoothly with β , so we might reasonably replace each rising and vertically falling section of the graphs by a straight line (*i.e.* connect the bottom of adjacent “teeth”).

The relation (5) may be said to hold provided $p_l < 0.1p_t$, say. Using (6), this is equivalent to $\beta < 0.1^{1/(l-t)}$, which for the parameter sets above gives $\beta < 0.86$ and $\beta < 0.89$ respectively. Outside these ranges the upper bound should be interpreted with caution.

As mentioned earlier, in most real networks we expect values of β nearer to 0.5 than 0.95, or else that the decrease in probabilities $\{p_i\}$ is more rapid than in fixed ratio. We conclude that for practical networks involving parameters such as those above, the proposed truncation algorithm carries negligible probability of miscounting.

5 Probability of misranking using truncation

In this section the misranking probability of the truncation algorithm is analysed under the assumptions of Section 4.1. We have seen that for suitable parameter values the possibility of miscounting samples having types $1, \dots, t$ may be neglected. Therefore we assume that no such miscounting occurs, and now find an approximate upper bound for the probability that these t sample counts are incorrectly ranked in the final list.

We use the model (6), and again consider values of β in the range $0.5 \leq \beta \leq 0.95$. The sample counts for types i and j , where $i < j$, will be incorrectly ranked with respect to each other if $Y_i < Y_j$. Since $p_i = \beta^{i-j}p_j$, to a first approximation we may assume that if misranking of the sample count for type i occurs with respect to any other type $j > i$ then it is for $j = i + 1$, when $Y_i < Y_{i+1}$. In this case the misranking probability of the truncation algorithm is approximately

$$\begin{aligned} P_R &= P(Y_i < Y_{i+1} \text{ for at least one } i \text{ in } 1 \leq i \leq t) \\ &\leq \sum_{i=1}^t P(Y_i < Y_{i+1}). \end{aligned} \quad (9)$$

Recall that the vector (Y_1, \dots, Y_m) is approximately distributed as a multivariate Normal with mean $x(p_1, \dots, p_m)$ and variance-covariance matrix xC , where $C = (c_{ij})$ is given by (3). Therefore [5] $Y_i - Y_{i+1}$ is distributed as a univariate Normal with mean

$$x(p_i - p_{i+1}) \simeq x(1 - \beta)^2 \beta^{i-1}$$

and variance

$$x [p_i(1 - p_i) + p_{i+1}(1 - p_{i+1}) + 2p_i p_{i+1}] \simeq x [(1 - \beta^2)\beta^{i-1} - (1 - \beta)^4 \beta^{2i-2}].$$

Hence from (9)

$$P_R \leq \sum_{i=1}^t \left[1 - \Phi \left(\frac{\sqrt{x}(1-\beta)^2 \beta^{i-1}}{\sqrt{(1-\beta^2)\beta^{i-1} - (1-\beta)^4 \beta^{2i-2}}} \right) \right] \quad (10)$$

Figure 2 shows numerical results for several values of x . For a specific value of β , this graph indicates the minimum number of samples needed to ensure the misranking probability falls below a desired level. We can express this in terms of the network load ρ as follows. Consider a 10Mbit/s ethernet network with a mean packet size of 100 bytes, from which 0.5% of packets are sampled over one minute to give x samples. The network load is then $\rho = x/3750$, as indicated by the bracketed labels in Figure 2. For any value of β , the misranking probability decreases as the network load ρ increases.

If the ordering of the sample counts for types $1, \dots, t$ amongst themselves is unimportant, a corresponding approximate bound for the misranking probability is given by taking only the summation term on the right hand side of (10) for which $i = t$.

6 Conclusion

We have presented a packet sampling system for estimating attributes of network traffic. Consideration of the maximum likelihood estimate leads to confidence intervals on the packet count estimates. The proposed truncation algorithm limits in advance the length of the list of packet counts of various types. Provided the probability of a sampled packet having a particular type is constant and independent of other samples we can bound the probability of miscounting and misranking the largest packet types under this algorithm.

7 Appendix: Bounding P_C

We derive here an upper bound for the right hand side of (8) in terms of s, l, t and β , and give an explicit form for the function $g_i(r)$.

7.1 Bounding $P(R \geq r)$

We firstly bound the probability $P(R \geq r)$. The first truncation occurs at the arrival of the s th distinct sample type. Suppose we fix a subset of types $\{1, \dots, j\}$ where $j < s$ is to be determined, and only consider samples whose type does not belong to the subset. Then the first truncation cannot occur before $s - j$ such samples are received, say at the D th sample (and truncation will occur later unless all of the $s - j$ types are distinct). Then $D = d$ if $s - j - 1$ of the first $d - 1$ samples, and also the d th sample, have types greater than j . Since sample types are independent, this means that D has a negative binomial distribution with parameters $(s - j, \gamma)$, where

$$\gamma = P(\text{sample type} > j) = \sum_{i=j+1}^m p_i, \quad (11)$$

and

$$P(D = d) = \binom{d-1}{s-j-1} \gamma^{s-j} (1-\gamma)^{d-s+j} \quad (d = s - j, s - j + 1, \dots). \quad (12)$$

Furthermore if $D = r$ then $R \geq r$, so

$$P(R \geq r) \geq P(D \geq r). \quad (13)$$

We may substitute (13) in (8) to give an upper bound on P_C , provided $g_i(r) - g_i(r-1) \leq 0$ i.e. provided $g_i(r)$ is a decreasing function of r for all $1 \leq i \leq t$. In this case we obtain

$$\begin{aligned} P_C &\leq \sum_{i=1}^t \left[\sum_{r=s}^{\infty} [g_i(r) - g_i(r-1)] P(D \geq r) + g_i(s-1) \right] \\ &= \sum_{i=1}^t \left[\sum_{r=s}^{\infty} g_i(r) P(D = r) + g_i(s-1) [1 - P(D \geq s)] \right] \\ &= \sum_{i=1}^t \left[\sum_{r=s}^{\infty} g_i(r) P(D = r) + g_i(s-1) \sum_{r=s-j}^{s-1} P(D = r) \right]. \end{aligned} \quad (14)$$

This bound depends on the value of j , which we may choose freely in the range $0 \leq j < s$. We choose that value of j which maximises ED , so that large values of $P(D = r)$ coincide with small values of the decreasing function $g_i(r)$. Now $ED = (s-j)/\gamma$, and using the distribution (6) for $\{p_i\}$ in (11) we have

$$\gamma \simeq \beta^j, \quad (15)$$

giving $ED \simeq (s-j)\beta^{-j}$. By differentiation, the maximum of this function occurs at $j = s + 1/\log \beta$. Since we require an integer value for j we take the rounded value

$$j = \lfloor s + 1/\log \beta + 1/2 \rfloor \quad (16)$$

(which lies in the range $0 \leq j < s$ provided $e^{-2} < \beta \leq e^{-2/(2s+1)}$, and this holds for values of s and β in the ranges we shall consider). We may now use (12), (15) and (16) to calculate the terms involving D in (14).

7.2 Finding a decreasing function $g_i(r)$

It remains to find a decreasing function $g_i(r)$ satisfying (7). Write $v = \lceil rp_l \rceil - 1$, so that

$$rp_l - 1 \leq v < rp_l. \quad (17)$$

Since $Y_i(r)$ is a binomial with parameters (r, p_i) ,

$$P(Y_i(r) \leq \lceil rp_l \rceil) = \sum_{u=0}^v \binom{r}{u} p_i^u (1-p_i)^{r-u} + \binom{r}{v+1} p_i^{v+1} (1-p_i)^{r-v-1}.$$

Now for

$$0 \leq u \leq v-1 \quad (18)$$

the ratio of the $(u+1)$ th term of the summation to the u th is

$$a = \frac{(r-u)p_i}{(u+1)(1-p_i)}.$$

From (17) and (18) we have $r - u > r(1 - p_l)$ and $u + 1 < rp_l$, so $a > (1 - p_l)p_i/p_l \simeq p_i/p_l$. Therefore using (4) and (5), the ratio $a \gg 1$ for $1 \leq i \leq t$ and we may neglect all terms of the summation except for $u = v$:

$$\begin{aligned} P(Y_i(r) \leq [rp_l]) &\simeq \binom{r}{v} p_i^v (1 - p_i)^{r-v} + \binom{r}{v+1} p_i^{v+1} (1 - p_i)^{r-v-1} \\ &= \binom{r}{v} p_i^v (1 - p_i)^{r-v-1} \left[1 - p_i + \frac{(r-v)p_i}{v+1} \right]. \end{aligned}$$

Using (17) we have $(r - v)/(v + 1) < r/(v + 1) \leq 1/p_l$ and therefore we may satisfy (7) by choosing

$$g_i(r) = \binom{r}{v} p_i^v (1 - p_i)^{r-v-1} (1 + p_i/p_l).$$

All the terms in (14) can now be calculated.

Furthermore $g_i(r)$ is a decreasing function of r , independently of (6), because

$$b = \frac{g_i(r+1)}{g_i(r)} = \frac{(r+1)(1-p_i)}{r+1-v},$$

and from (17) we have $v < (r+1)p_l$ and so $(r+1)/(r+1-v) < 1/(1-p_l)$. Therefore

$$b < \frac{1-p_i}{1-p_l} < 1 \quad \text{for all } 1 \leq i \leq t$$

due to the labelling (4).

7.3 Calculation of numerical results

To calculate numerically the approximate upper bound for P_C given by (14) we truncate the infinite summation in r as soon as the decreasing function $g_i(r)$ falls below some tolerance τ . To determine the value of τ for specific values of s , l and t , we find the truncated bound to four significant figures for representative values of β in the range $0.5 \leq \beta \leq 0.95$ and for $\tau = 10^{-j}$, $j = 1, 2, \dots$. We fix τ by taking the smallest value of j for which the truncated bound does not change for any β or i when j is replaced by $j + 1$.

The discontinuities in the graphs of Figure 1 occur at the values of β for which the rounded value of j determined by (16) changes.

8 References

- [1] P.D. Amer and L.N. Cassel. Management of sampled real-time network measurements. In *Proceedings of the 14th Conference on Local Computer Networks, October 1989, Minneapolis*, pages 62–68. IEEE, 1989.
- [2] P. Brémaud. Characteristics of queueing systems observed at events and the connection between stochastic intensity and Palm probability. *Queueing Systems*, 5:99–112, 1989.

- [3] D.C. Feldmeier. Statistical monitors for local area networks. In *Proceedings of the 11th Conference on Local Computer Networks, October 1986, Minneapolis*, pages 142–146. IEEE, 1986.
- [4] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume I. John Wiley & Sons, New York, 3rd edition, 1968. Revised printing.
- [5] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume II. John Wiley & Sons, New York, 2nd edition, 1971.
- [6] W. L. Hays. *Statistics*. Holt, Rinehart and Winston, Inc., New York, 4th edition, 1988.
- [7] L.L. Lapin. *Statistics for Modern Business Decisions*. The Harbrace Series in Business and Economics. Harcourt Brace Jovanovich, New York, 1973.
- [8] A. Makowski, B. Melamed, and W. Whitt. On averages seen by arrivals in discrete time. In *Proceedings of the 28th IEEE Conference on Decision and Control, December 1989, Tampa, Florida*, pages 1084–1086.
- [9] B. Melamed and W. Whitt. On arrivals that see time averages. *Operations Research*, **38**:156–172, 1990.
- [10] B. Melamed and W. Whitt. On arrivals that see time averages: a martingale approach. *Journal of Applied Probability*, **27**:376–384, 1990.
- [11] N.M. Minnich. A packet capture system for LAN software development. In *Proceedings of the 11th Conference on Local Computer Networks, October 1986, Minneapolis*, pages 68–76. IEEE, 1986.
- [12] S. Stidham, Jr. and M. El Taha. Sample-path analysis of processes with imbedded point processes. *Queueing Systems*, **5**:131–166, 1989.

Figure 1: Upper bound on probability of miscounting

bound on probability

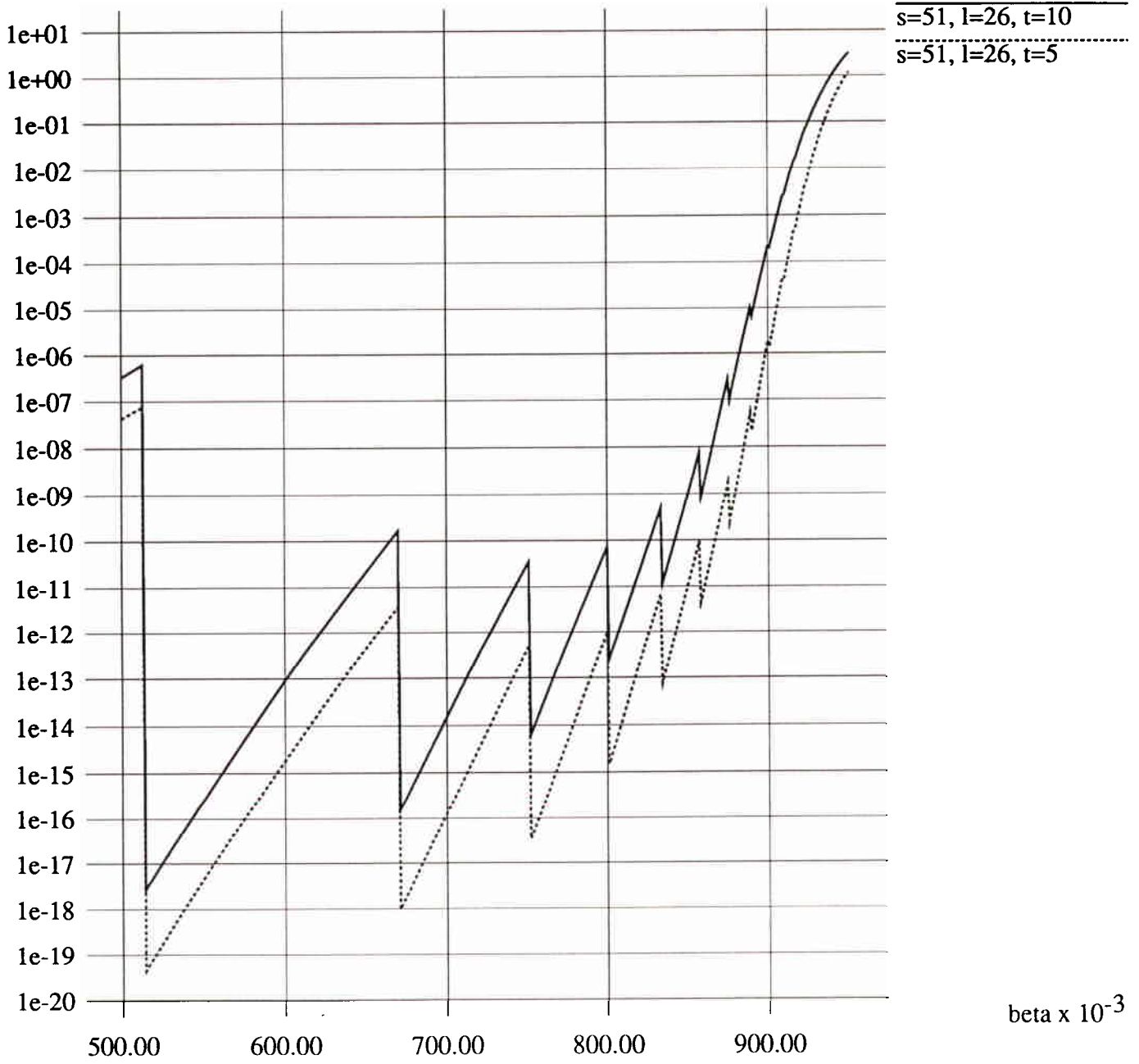
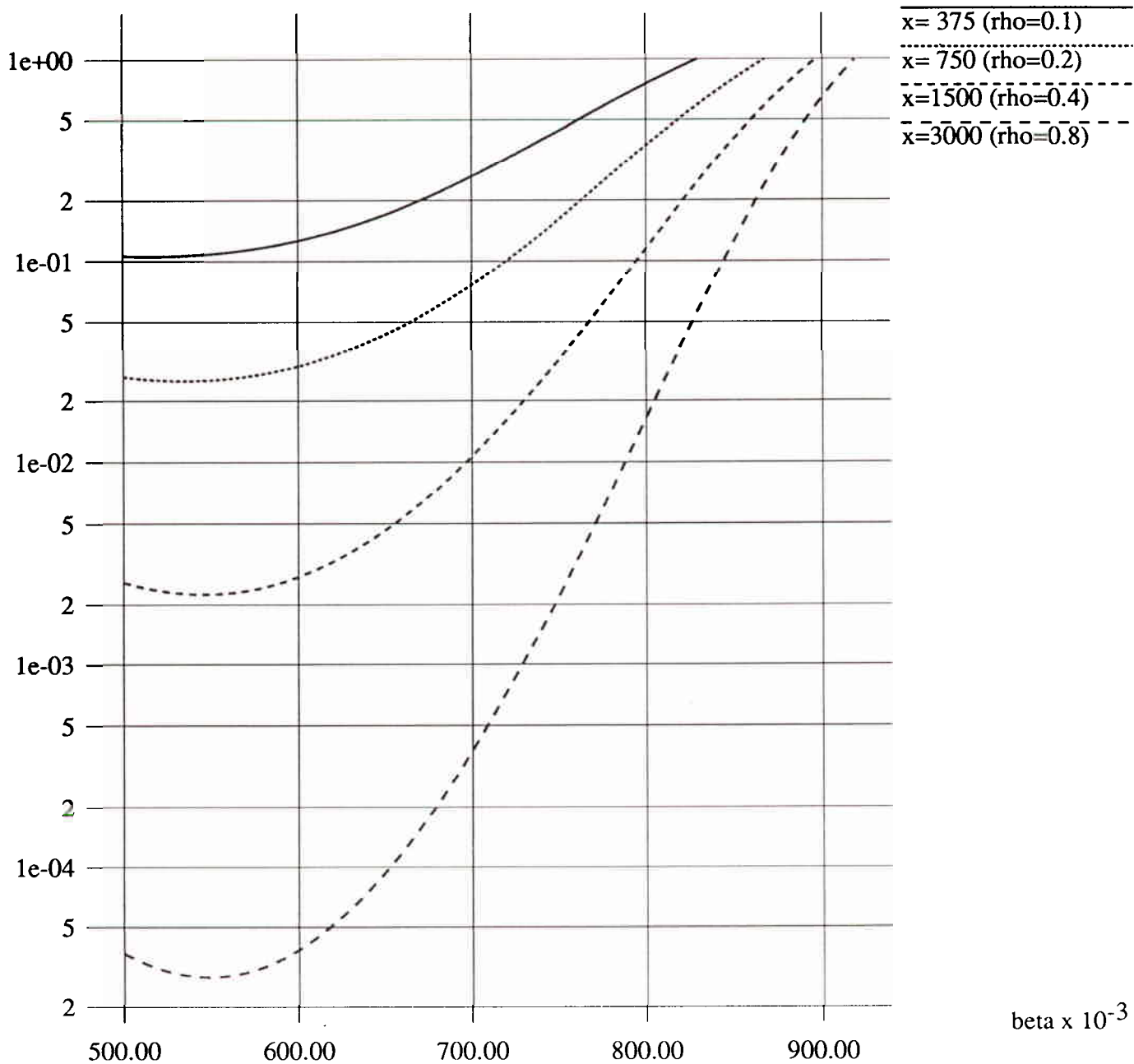


Figure 2: Upper bound on probability of misranking

bound on probability



$\beta \times 10^{-3}$