September 26, 2001
RT0428
Computer Science 16 pages

Research Report

Study of Internet Traffic -1, Extended Poisson Distribution

Akio Koide

IBM Research, Tokyo Research Laboratory
IBM Japan, Ltd.
1623-14 Shimotsuruma, Yamato
Kanagawa 242-8502, Japan

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Contents

1 Introduction

The purpose of this report is to extend Poisson distribution for the study of real Internet traffic by use of a generating function technique.

The well-know Poisson distribution of the traffic model

$$P(n;t) = \frac{(\lambda t)^n}{n!} \exp(-\lambda t) \tag{1}$$

is derived from the three assumptions, (1) "memoryless", (2) "single request" and (3) "stationary" properties of the traffic [1, 2]. In this report, I will firstly remove the assumption of single request (separated occurrence) and next remove the stationary assumption in the traffic. Here we will see that the assumption of memoryless and stationary properties for the traffic leads to a special relation that the ratio of the variance $\langle (n - \langle n \rangle)^2 \rangle$ to the average $\langle n \rangle$ of request numbers remains constant with respect to the change of time intervals for counting requests. Then I will analyze the real access log of HTTP server with this extended Poisson distribution. We will see that the ratio of the variances to the averages increases with the increase of the counting time intervals in the real access log of HTTP servers. This strongly indicates that the memoryless assumption does not hold for real Internet traffic.

In the appendix, I will also give a very simple derivation of Poisson distribution.

2 Stationary Memoryless Stream

In this section, we derive a general form of traffic distribution in stationary state from memoryless assumption. Through this report, $P(n;t)$ represents the probability that n requests arrive during the time interval t. From this definition, $P(n;t) \geq 0$ and

$$\sum_{n=0}^{\infty} P(n;t) = 1 \tag{2}$$

for any $t > 0$.

2.1 General Solution based on Generating Function

The probability of n requests arriving in the merged interval $t_1 + t_2$ is generally given by

$$P(n;t_1+t_2) = \sum_{n'=0}^{n} P(n';t_1) P(n-n';t_2|n';t_1). \tag{3}$$

Here $P(n'';t_2|n';t_1)$ is the conditional probability that n'' requests arrive during the interval t_2 after n' requests arrive during the interval t_1. The memoryless assumption is to assume that

$$P(n'';t_2|n';t_1) = P(n'';t_2). \tag{4}$$

Thus, the memoryless assumption is alternatively written by

$$P(n;t_1+t_2)=\sum_{n'=0}^{n}P(n';t_1)P(n-n';t_2) \tag{5}$$

for any integer n and any $t_1, t_2 > 0$.

Let us derive a general solution of memoryless traffic by use of a generation function

$$W(z;t)=\sum_{n=0}^{\infty}z^n P(n;t). \tag{6}$$

Here we assume that $P(n;t)$ goes to zero sufficiently rapidly so that $W(z;t)$ is analytic function with respect to z in the range $|z| \leq 1$. Then we obtain

$$W(1;t)=1 \tag{7}$$

from Eq.(2) and

$$W(z;t_1+t_2)=W(z;t_1)W(z;t_2), \tag{8}$$

from Eq.(5). The latter equation can be written alternatively as

$$\log W(z;t_1+t_2)=\log W(z;t_1)+\log W(z;t_2). \tag{9}$$

This means that $\log W(z;t)$ is a linear function of the interval t. When $W(z;t)$ is assumed to be continuous with respect to the interval t, we can write it as

$$\log W(z;t)=u(z)t \tag{10}$$

with an appropriate function $u(z)$. Thus, finally, we obtain a general solution

$$W(z;t)=\exp(u(z)t) \tag{11}$$

for memoryless and stationary traffic in term of a generating function.

2.2 Expansion of Generating Function

The explicit form of $P(n;t)$ can be obtained by expanding $u(z)$ in series of z.

Let $u(z)$ be expanded in

$$u(z)=-\lambda+\lambda\sum_{n=1}^{\infty}c_n z^n. \tag{12}$$

From Eq.(7), we have

$$1=\sum_{n=1}^{\infty}c_n. \tag{13}$$

Here each term λc_n can be interpreted as the rate of n requests arriving simultaneously since

$$P(n;t)/t \to \lambda c_n \ \text{ when } t \to 0. \tag{14}$$

Then we have

$$P(0;t) = \exp(-\lambda t). \tag{15}$$

and

$$P(n;t) = \sum_{k=1}^{n} c_k^{(n)} \frac{(\lambda t)^k}{k!} \exp(-\lambda t) \quad \text{for } n > 0. \tag{16}$$

Here the coefficient $c_k^{(n)}$ can be derived from the recurrence as follows.

$$c_1^{(n)} = c_n \tag{17}$$

and

$$c_{k+1}^{(n)} = \sum_{m=1}^{n-k} c_m c_k^{(n-m)} \quad \text{for } k > 0. \tag{18}$$

This can be derived from the differential equation

$$\frac{dP(n;t)}{dt} + \lambda P(n;t) = \lambda \sum_{m=1}^{n} c_m P(n-m;t), \tag{19}$$

which is obtained by expanding

$$\frac{\partial W(z;t)}{\partial t} = u(z)W(z;t), \tag{20}$$

2.3 Expectations of Traffic

The expected number of requests during the time interval t is given by

$$\langle n \rangle = \lambda t \sum_{n=1}^{\infty} n c_n \tag{21}$$

and its variance is given by

$$\langle (n - \langle n \rangle)^2 \rangle = \lambda t \sum_{n=1}^{\infty} n^2 c_n. \tag{22}$$

The ratio of the variance to the average is independent from the time interval t;

$$\frac{\langle (n - \langle n \rangle)^2 \rangle}{\langle n \rangle} = \frac{\sum_{n=1}^{\infty} n^2 c_n}{\sum_{n=1}^{\infty} n c_n} \geq \sum_{n=1}^{\infty} n c_n. \tag{23}$$

Eq(21) is derived from

$$\sum_{n=1}^{\infty} n z^{n-1} P(n;t) = \frac{\partial W(z;t)}{\partial z} = t\frac{du(z)}{dz}W(z;t).$$

Eq(22) is derived from

$$\sum_{n=1}^{\infty} n^2 z^{n-1} P(n;t) = \frac{\partial}{\partial z} z \frac{\partial W(z;t)}{\partial z} = \left(t\frac{d^2u(z)}{dz^2} + t^2 (\frac{du(z)}{dz})^2 + t\frac{du(z)}{dz} \right) W(z;t).$$

2.4 Evaluation of Extended Poisson Distribution

I showed that each term of extended Poisson distribution can be expressed by the product of exponential term and polynomial of the time interval t in the proceeding subsections. But it is not economical to evaluate $P(n;t)$ after evaluating the coefficients of polynomials.

I recommend to use directly the following recurrence formula to evaluate $P(n;t)$.

$$P(n+1;t) = \frac{t}{n+1} \sum_{m=0}^{n} (m+1) c_{m+1} P(n-m;t). \tag{24}$$

This recurrence is derived by expanding

$$\frac{\partial W(z;t)}{\partial z} = t \frac{du(z)}{dz} W(z;t)$$

with respect to z. The recurrence (24) is further simplified when specific forms of $u(z)$ are selected.

2.5 Specific Examples

Let us think about the case of

$$u(z) = \lambda \log((1-p)/(1-pz)), \tag{25}$$

which gives a generating function

$$W(z;t) = \left(\frac{1-p}{1-pz} \right)^{\lambda t}. \tag{26}$$

The parameter p controls the decreasing speed of c_n as

$$c_n = \frac{p^n}{n \log(1-p)}. \tag{27}$$

This generating function $W(z;t)$ leads to the recurrence

$$P(n+1;t) = \frac{p(n+\lambda t)}{n+1} P(n;t) \tag{28}$$

I recommend to use this recurrence with the starting value

$$P(0;t) = (1-p)^{\lambda t} \tag{29}$$

for evaluating of $P(n;t)$ although their explicitly forms are given by

$$P(n;t) = \frac{\Gamma(n+1+\lambda t) p^n}{\Gamma(1+\lambda t) n!} (1-p)^{\lambda t}. \tag{30}$$

The parameters λ and p can be determined by the average

$$\langle n \rangle = \frac{\lambda t p}{1-p} \tag{31}$$

and the variance

$$\langle (n - \langle n \rangle)^2 \rangle = \frac{\lambda t p}{(1-p)^2}. \tag{32}$$

Let us think about another simple case of

$$u(z) = -\frac{\lambda(1-z)}{1-pz} = -\lambda + \frac{\lambda z(1-p)}{1-pz}, \tag{33}$$

which gives a generating function

$$W(z;t) = \exp(-\frac{\lambda t(1-z)}{1-pz}). \tag{34}$$

The parameter p controls the decreasing speed of c_n as

$$c_n = (1-p)p^{n-1} \tag{35}$$

This generating function $W(z;t)$ leads to the recurrence

$$P(n+1;t) = \frac{1}{n+1}\Big((2np + \lambda t(1-p))P(n;t) - (n-1)p^2 P(n-1;t)\Big). \tag{36}$$

The $P(n;t)$ can be easily evaluated by this recurrence and the starting values

$$P(0;t) = \exp(-\lambda t) \;\text{ and }\; P(-1;t) = 0. \tag{37}$$

The explicit form of $P(n;t)$ can be given in terms of the Laguerre polynomials by

$$P(n;t) = p^n L_n(-\frac{\lambda t(1-p)}{p})\exp(-\lambda t). \tag{38}$$

This is easily derived by noting the generating function of the Laguerre polynomials

$$\exp(-\frac{xz}{1-z}) = \sum_{n=0}^{\infty} L_n(x) z^n \tag{39}$$

The parameters λ and p can be determined by the average

$$\langle n \rangle = \frac{\lambda t}{1-p} \tag{40}$$

and the variance

$$\langle (n - \langle n \rangle)^2 \rangle = \frac{\lambda t(1+p)}{(1-p)^2}. \tag{41}$$

3 Nonstationary Memoryless Stream

3.1 General Solution based on Generating Function

Let $P(n; t_1, t_2)$ represent the probability that n requests arrive during the interval between the times t_1 and t_2 where $t_1 \leq t_2$. We set the normalization

$$\sum_{n=0}^{\infty} P(n; t_1, t_2) = 1 \tag{42}$$

and the memoryless assumption

$$P(n; t_1, t_2) = \sum_{n'=0}^{n} P(n'; t_1, t) P(n - n'; t, t_2) \tag{43}$$

for any t such that $t_1 \leq t \leq t_2$. Eq.(43) implies that the probabilities of the interval $[t_1, t]$ and $[t, t_2]$ have no correlation.

We again use a generating function

$$W(z; t_1, t_2) = \sum_{n=0}^{\infty} z^n P(n; t_1, t_2). \tag{44}$$

Then we have

$$W(1; t_1, t_2) = 1 \tag{45}$$

from Eq(42) and

$$W(z; t_1, t_2) = W(z; t_1, t) W(z; t, t_2) \tag{46}$$

from Eq.(43). This can be rewritten as

$$\log W(z; t_1, t_2) = \log W(z; t_1, t) + \log W(z; t, t_2). \tag{47}$$

Let us assume $W(z; t, t')$ is differentiable with respect to t' and there exist the limit

$$\lim_{\epsilon > 0 \to 0} \frac{\log W(z; t, t + \epsilon)}{\epsilon} = u(z; t) \tag{48}$$

where $u(z; t)$ is an appropriate function of z and t. Then, from Eq(47), we have the differential equation

$$\frac{\partial}{\partial t} \log W(z; t_1, t) = u(z; t). \tag{49}$$

By integrating to this differential equation and noting $W(z; t, t) = 1$ we obtain a general solution of memoryless and nonstationary traffic

$$W(z; t_1, t_2) = \exp(\int_{t_1}^{t_2} u(z; t))dt \tag{50}$$

in term of a generating function.

3.2 Expansion of Generating Function

Let us expand $u(z;t)$ in

$$u(z;t) = -\lambda(t) + \sum_{n=1}^{\infty} z^n u_n(t). \tag{51}$$

Since $u(1;t) = 0$ from Eq(45), we have

$$\lambda(t) = \sum_{n=1}^{\infty} u_n(t). \tag{52}$$

The function $u_n(t)$ must be nonnegative because

$$u_n(t) = \lim_{\epsilon>0\to 0} \frac{P(n;t,t+\epsilon)}{\epsilon} \quad \text{for } n > 0. \tag{53}$$

By noting that

$$\frac{\partial W(z;t_0,t)}{\partial t} = u(z,t)W(z;t_0,t), \tag{54}$$

we can derive the differential equation with respect to the time t

$$\frac{dP(n;t_0,t)}{dt} + \lambda(t)P(n;t_0,t) = \sum_{m=1}^{n} u_m(t)P(n-m;t_0,t). \tag{55}$$

Therefore, $u_n(t)$ can be interpreted as the rate of n request arriving.

3.3 Expectations of Memoryless and Nonstationary Traffic

The expected number of requests in the interval between t_1 and t_2 is given by

$$\langle n \rangle = \sum_{n=1}^{\infty} n \int_{t_1}^{t_2} u_n(t)dt \tag{56}$$

and its variance is given by

$$\langle (n - \langle n \rangle)^2 \rangle = \sum_{n=1}^{\infty} n^2 \int_{t_1}^{t_2} u_n(t)dt. \tag{57}$$

Therefore we have

$$\langle n \rangle \leq \langle (n - \langle n \rangle)^2 \rangle \tag{58}$$

and

$$\frac{\langle (n - \langle n \rangle)^2 \rangle}{\langle n \rangle} \geq \langle n \rangle. \tag{59}$$

3.4 Simplified Memoryless and Nonstationary Traffic

Now let us derive a simplified type of memoryless and non-stationary traffic.

Suppose that the relative ratios of request numbers remains unchanged;

$$u_n(t) = c_n \lambda(t) \tag{60}$$

Then, we have

$$W(z; t_1, t_2) = \exp(u(z) \int_{t_1}^{t_2} \lambda(t) dt) \tag{61}$$

where

$$u(z) = -1 + \sum_{n=1}^{\infty} c_n z^n. \tag{62}$$

Since this generating function is identical to Eq(11) by a scale transformation of time

$$t \to \int_{t_0}^{t} \lambda(s) ds, \tag{63}$$

we can express memoryless and nonstationary traffic from memoryless and stationary traffic by an appropriate time scale transformation as far as the assumption (60) approximately holds true. Also, the ratio of the variance to the average $\langle (n - \langle n \rangle)^2 \rangle / \langle n \rangle$ remains unchanged with respect to the change of time interval for counting.

In the special case that the $u_n(t) = 0$ for $n \geq 2$, we have

$$W(z; t_1, t_2) = \exp(-(1 - z) \int_{t_1}^{t_2} \lambda(t) dt). \tag{64}$$

This leads to

$$P(n; t_1, t_2) = \frac{1}{n!} \left(\int_{t_1}^{t_2} \lambda(t) dt \right)^n \exp(- \int_{t_1}^{t_2} \lambda(t) dt) \tag{65}$$

4 Merged Stream

Here we will show a merged stream of memoryless traffic is also memoryless traffic.

Let $P_1(n'; t, t')$, $P_2(n - n'; t, t')$ and $P_{12}(n; t, t')$ represent probability distributions of jobs for the stream-1, the stream-2 and the merged stream. And let $W_\alpha(z; t, t')$ be a generating function for each $P_\alpha(n; t, t')$ defined by

$$W_\alpha(z; t, t') = \sum_{n=0}^{\infty} z^n P_\alpha(n; t, t') \tag{66}$$

If the stream-1 and stream-2 has no interaction in merging, we have

$$P_{12}(n; t, t') = \sum_{n'=0}^{n} P_1(n'; t, t') P_2(n - n'; t, t'). \tag{67}$$

Then their generating functions satisfy

$$W_{12}(z;t,t') = W_1(z;t,t')W_2(z;t,t'). \tag{68}$$

This leads to the linearity of logarithm of the generating functions $W_\alpha(z;t,t)$ with respect to merging streams;

$$\log W_{12}(z;t,t') = \log W_1(z;t,t') + \log W_2(z;t,t'). \tag{69}$$

On the other hand, as shown in the previous subsection, the stationary and nonstationary Poisson Stream can be derived from the linearity of the generating functions $W_\alpha(z;t,t)$ with respect to merging time intervals. Therefore, the merged traffic of the extended Poisson distribution is also extended Poisson distribution.

5 Analysis of Real Access Log of HTTP Servers

Figure 1 shows daily behavior of HTTP requests at some US customer on August 1, 2001. Each black circle stands for the average of request counts per 2 seconds over every 2 minutes. Figure 2 gives daily behavior of the ratio of the variance to the average. The variances are also computed by statistics of request counts per 2 seconds over every 2 minutes. The average and variance largely vary with time in the day while the ratio fluctuates around constant value through the day.

From Figure 1, I selected the time range from 18:50:00 to 20:06:48 as almost stationary region of statistics.

Figure 4 shows the distribution of request count per a second from the time 18:50:00 to 20:06:48 by filled circles. The theoretical distributions, Poisson distribution and extended Poisson distributions are shown by solid and dot lines. The extended-1 is based on the model of Eq.(26) and the extended-2 is based on the model of Eq.(34). The Poisson distrbution does not fit the observed counts while the both of the extended distribution fit them well.

Figure 3 shows how the ratio of the variance to the average changes with respect to the time intervals for counting for the data from the time 18:50:00 to 20:06:48. We see that the ratio of the variances to the averages increases with the increase of the counting time intervals in the real access log of HTTP servers. This means that we can not fit the real distribution data of the request counts per different time interval with the same parameters p and λ of the extended Poisson distribution. As the ratio in Figure 2 does not show any daily behavior, we cannot explain this observation by memoryless nonstationary traffic. This strongly indicates that the memoryless assumption does not hold for real Internet traffic.

A The Elementary Derivation of Poisson Distribution

Here, we derive the well-known Poisson distribution in simple and intuitive manner,

Suppose that N jobs distribute randomly in a large time interval NT without any correlation. Then the probability $P(n;t)$ that each job falls in the interval t is given by $\lambda t/N$ where $\lambda = 1/T$. Then, we get

$$P(n;t) = \frac{N!}{n!(N-n)!}\left(\frac{\lambda t}{N}\right)^n \left(1-\frac{\lambda t}{N}\right)^{N-n} \tag{70}$$

The λ can be interpred as the rate of arriving jobs.

Let us apply Stirling's formula

$$\log M! = (M-1/2)\log M - M + (1/2)\log(2\pi) + O(1/M) \tag{71}$$

to $P(n;t)$. Here $O(1/M)$ represents the term that goes to zero in the order of $1/M$ when M becomes larger. Then $\log P(n;t)$ is simplified for large N as follows.

$$\begin{aligned}
&\log P(n;t) \\
&\quad = \log N! - \log n! - \log(N-n)! - N\log N + n\log(\lambda t) + (N-n)\log(N-\lambda t) \\
&\quad = -(1/2)\log N + (1/2)\log(N-n) - n - (N-n)(\log(N-n) - \log(N-\lambda t)) \\
&\qquad +n\log(\lambda t) - \log n! + O(1/N) \\
&\quad = -\lambda t + n\log(\lambda t) - \log n! + O(1/N).
\end{aligned}$$

Thus we obtain well-known Poisson formula of simple stream

$$P(n;t) = \frac{(\lambda t)^n}{n!}\exp(-\lambda t) \tag{72}$$

when N goes to the infinity. Eq.(72) satisfies the normalization

$$\sum_{n=0}^{\infty} P(n;t) = \exp(-\lambda t)\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!} = 1. \tag{73}$$

The expected number of jobs in the time interval t is given by

$$\langle n\rangle = \sum_{n=0}^{\infty} nP(n;T) = \lambda t \tag{74}$$

and its variance is given by

$$\langle (n-\langle n\rangle)^2\rangle = \sum_{n=0}^{\infty}(n^2 - \langle n\rangle^2)P(n;t) = \lambda t \tag{75}$$

Therefore we have

$$\langle n\rangle = \langle (n-\langle n\rangle)^2\rangle. \tag{76}$$

References

[1] B. V. Gnedenko and I. N. Kovalenko, *Introduction to Queueing Theory*, Second Edition, translated by Samuel Kotz, Birkhaeuser, ISBN 0-8176-3423-1, 1987.

[2] Donald Gross and Carl M. Harris, *Fundamentals of Queueing Theory*, Second Edition, John Wiley & Sons, ISBN 0-471-89067-7, 1985.

Figure 1: Daily behavior of average count, August 01, 2000

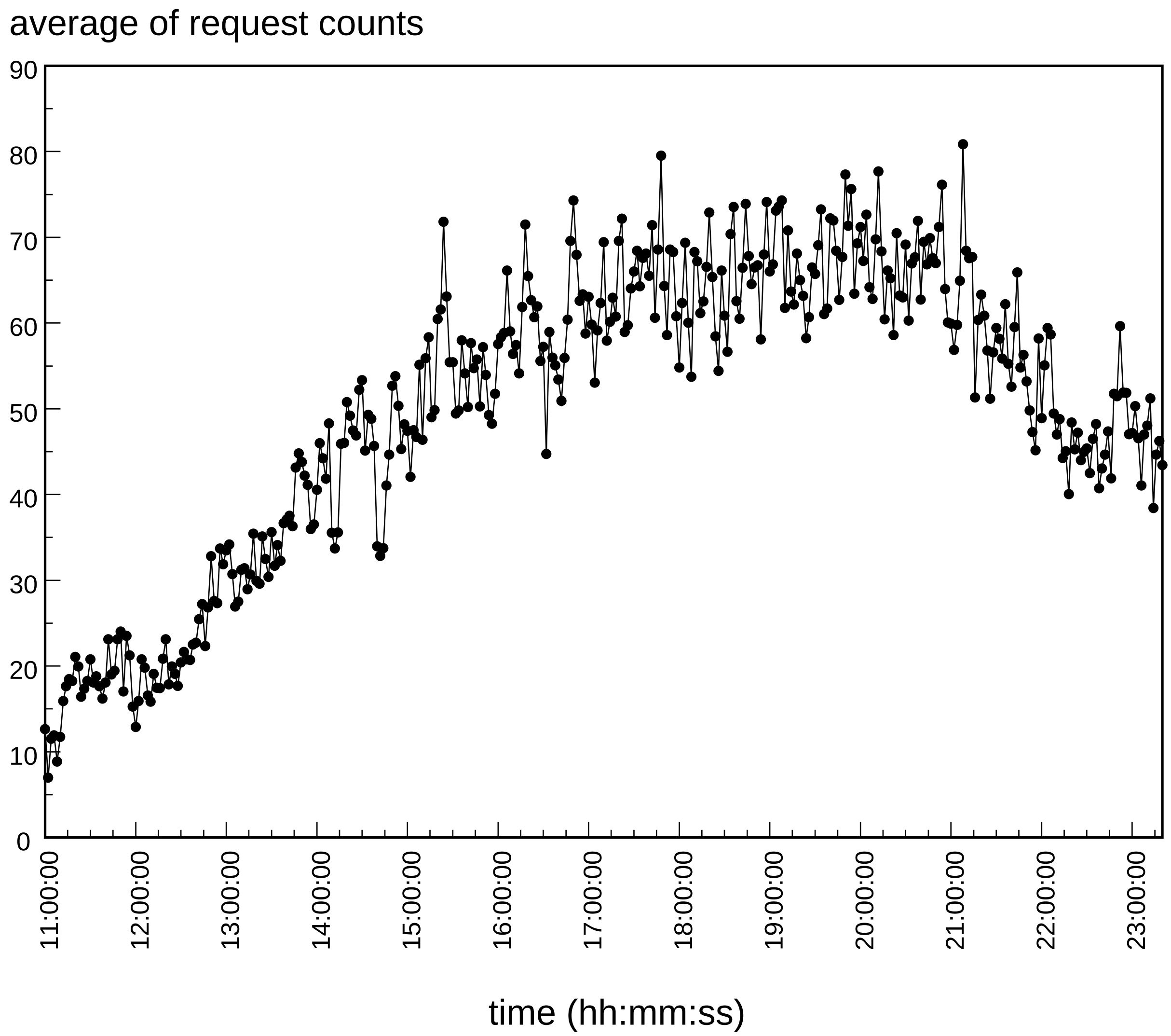

Figure 2: Daily behavior of ratio, August 01, 2000

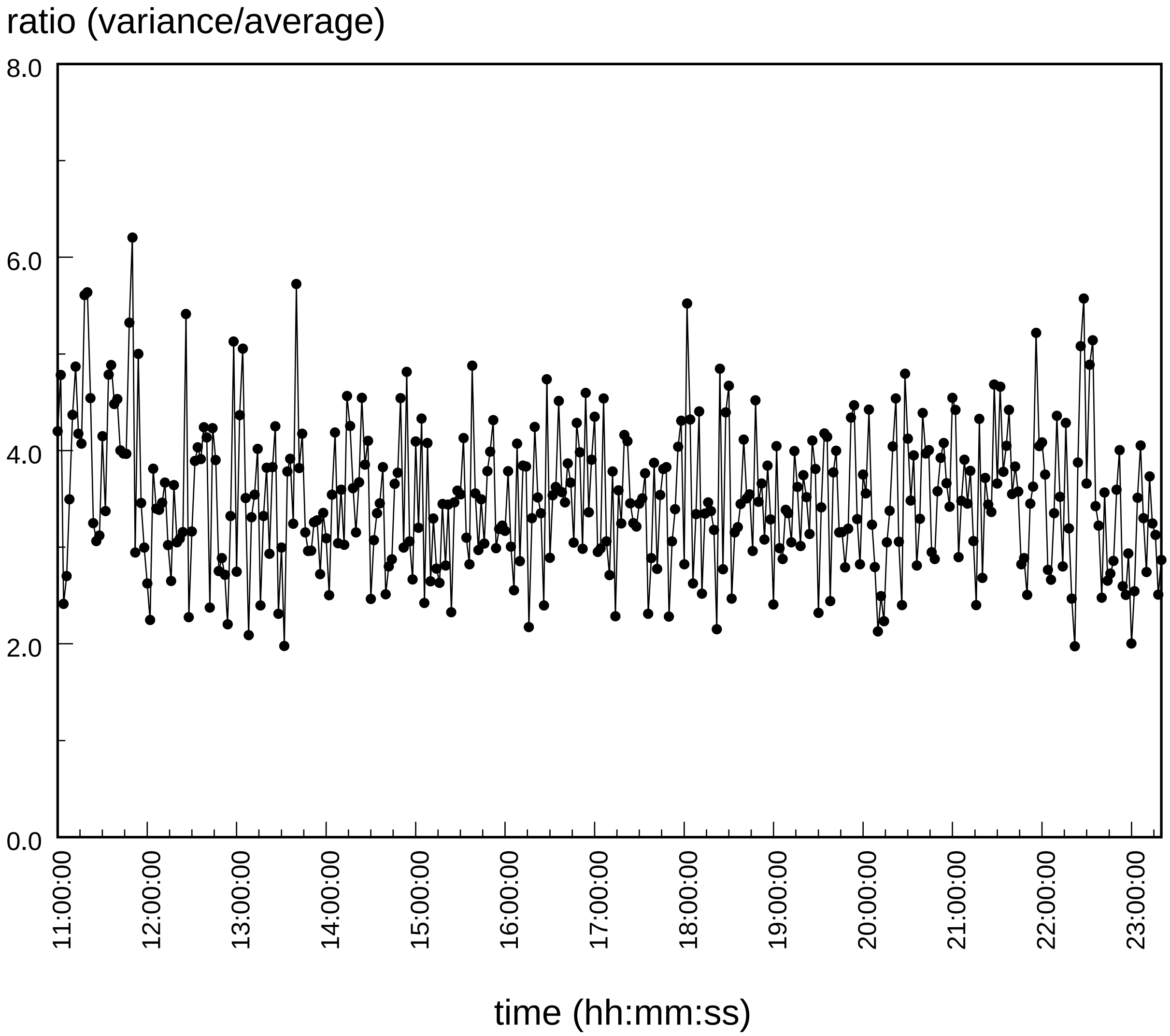

Figure 3: Averages, variances of request counts with respect to time intervals

intervals (sec)	averages	variances	ratios
1	33.768	91.496	2.710
2	67.536	258.336	3.825
4	135.071	695.413	5.148
8	270.142	1880.351	6.961
16	540.285	5078.120	9.399
32	1080.569	14621.009	13.531
64	2161.139	40453.536	18.719
128	4322.278	113315.478	26.217

Figure 4: Distribution of counts per sec, 18:50:00-20:06:48