A NOTE ON THE DISTRIBUTION OF PACKET ARRIVALS IN HIGH-SPEED DATA NETWORKS

Vic Grout, Stuart Cunningham, Denise Oram & Rich Hebblewhite
Centre for Applied Internet Research, University of Wales
NEWI Plas Coch Campus, Mold Road, Wrexham, LL11 2AW, UK
{v.grout|s.cunningham|d.oram|r.hebblewhite}@newi.ac.uk

ABSTRACT

A simple explanation is given of an anomaly often detected in traffic measurement when packet timings are truncated to nearest whole numbers. The correct observed discrete distribution is then calculated from an underlying continuous distribution.

KEYWORDS

Traffic modeling, Packet distribution.

1. INTRODUCTION

The study of the arrival of data packets at routers, and other key network components, forms the foundation of modern data communications traffic theory. The accurate modeling and simulation of such traffic flows is an essential part of network design, resource allocation and management. Various *Quality-of-Service* (*QoS*) measures, for example, rely implicitly upon an understanding and knowledge of the distribution of these packets and, due to the high-speed, high-throughput nature of the traffic, can give poor results if inappropriately tuned. This paper addresses the standard form for the simulation of packet arrivals in a communications network, the *Poisson model*. Although various problems have been identified over time with this model (Paxson and Floyd, 1995), it remains the most useful tool in widespread use for discrete traffic modeling (Caglar, 2004). However, without appropriate application or due adaptation, it has the potential to give misleading results when applied to dense arrivals of small packets in short intervals. This paper describes the cause and effects of these errors and provides the necessary refinement. It allows packet flows obeying the Poisson distribution to be correctly identified even though logged data shows apparent deviation.

The distribution of packets arriving at a router (say) can be modeled as a sequence of events occurring over time. If these events, the arrivals of packets, are *independent* (if the Poisson model fails badly, it is generally in this respect – Paxson and Floyd, 1995) then they will have a negative exponential *probability density function* (*pdf*) given by

$$f(t) = qe^{-qt} \tag{1}$$

where t is the time between (the arrival of) two consecutive packets. In this distribution, $f(a) > f(b)$ for $a < b$ and the mean inter-arrival period is given by

$$m = \frac{1}{q}\cdot \tag{2}$$

Taking (natural) logarithms in (1) gives

$$\begin{aligned} \ln(f(t)) &= \ln(qe^{-qt}) \\ L(t) &= \ln q + \ln(e^{-qt}) = K - qt \end{aligned} \tag{3}$$

(K, constant). $L(t)$ is linear in t (Figure 1), making it convenient for data *fitness* testing. If the packet arrival distribution follows this model and we let c represent the number of packets arriving in unit time, then c follows the Poisson distribution with pdf

$$Poisson(c) = \frac{l^c e^l}{c!}, \tag{4}$$

where l is the mean number of packets in unit time so that

$$l = \frac{1}{m} = q \cdot \tag{5}$$

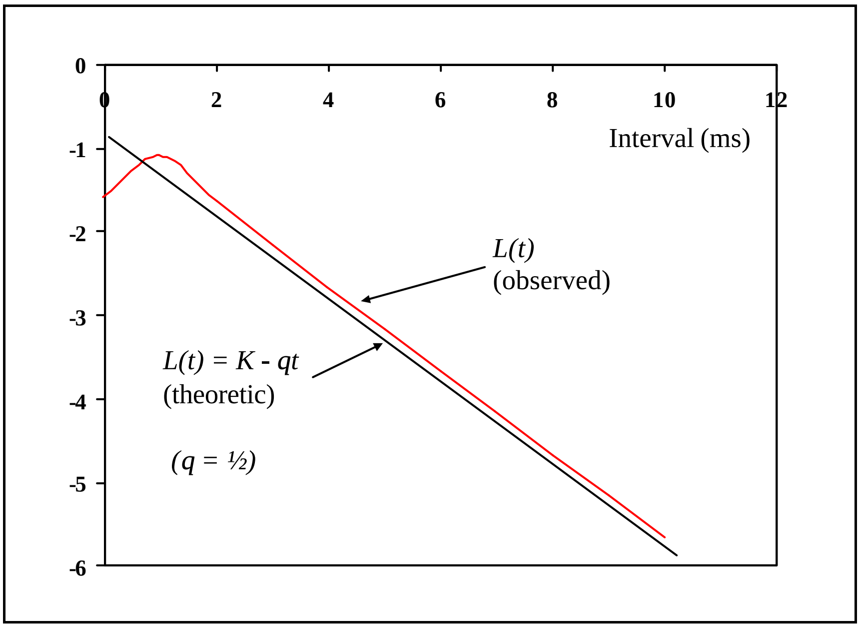

Figure 1. Comparing predicted (real) and observed (integer) ln intervals

Independent packets obeying equations (1), (3) & (4) are said to fit the Poisson model.

2. ANALYSIS AND DEVELOPMENT

However, if we record packet arrivals in a real network, even one known to have Poisson characteristics, then these are not the apparent distributions that emerge. Figure 1, for example, also shows the ln graph of observed inter-packet time probabilities to the nearest millisecond taken from a test network known to generate packets on an independently random basis. There is a clear difference for smaller intervals, precisely those in the majority for high densities of small packets, and consequently of considerable relevance.

The source of this discrepancy is simple. The *continuous* (real) distribution is being truncated to a *discrete* (integer) distribution. If the truncation were consistent (i.e. uniform), then the gap would be negligible. However, in practice, it is rarely the intervals themselves that are truncated. Instead, the arrivals of each packet are recorded to the nearest integer and the interval between two packets is calculated as the difference between the two truncated values. The truncated difference between two times and the difference between two truncated times are not necessarily equal ($[a-b]\ ^1\ [a]-[b]$ in general) and the only effect of recording to a greater (discrete) accuracy is for the discrepancy to manifest itself across smaller intervals.

The observed distribution to emerge from the arrival of independent packets is described as follows.

THEOREM:

If the real inter-arrival times (Δ) of independent packets conform to the continuous distribution

$$f(\Delta) = qe^{-q\Delta}, \tag{6}$$

then recording packet arrivals as integers will produce the observed discrete distribution (on d) of

$$g(d) = \begin{cases} 1 + \dfrac{e^{-q} - 1}{q}(d = 0) \\ \dfrac{2e^{-qd}(\cosh q - 1)}{q}(d > 0) \end{cases}. \tag{7}$$

PROOF:

Let t, then t', be the (absolute) arrival times of two successive packets. Then

$$\Delta = t' - t, \tag{8}$$

$$d = [t'] - [t] \tag{9}$$

and

$$\Delta = [\Delta] - \{\Delta\} \tag{10}$$

(where $[\Delta]$ represents the integer part and $\{\Delta\}$ the fractional part of Δ). Define $p(d|\Delta)$ to be the probability that $[t'] - [t] = d$ given that $t' - t = \Delta$. Then

$$\begin{aligned} &p([\Delta] - j|\Delta) = 0(j \geq 1) \\ &p([\Delta]|\Delta) = 1 - \{\Delta\} \\ &p([\Delta] + 1|\Delta) = \{\Delta\} \\ &p([\Delta] + j|\Delta) = 0(j \geq 2) \end{aligned} \tag{11}$$

Extending across multiple values gives

$$p(d|\Delta) = \begin{cases} 0(\Delta < d - 1) \\ \Delta - d + 1(d - 1 \leq \Delta \leq d) \\ d - \Delta + 1(d \leq \Delta \leq d + 1) \\ 0(\Delta > d + 1) \end{cases}. \tag{12}$$

If $f(\Delta)$ is the pdf of the continuous variable, Δ, then $g(d)$, the pdf of the discrete variable, d , is given generally by

$$g(d) = \int_0^\infty f(\Delta) p(d|\Delta) d\Delta \cdot \tag{13}$$

For $f(\Delta) = qe^{-q\Delta}$,

$$g(d) = \int_0^\infty qe^{-q\Delta} p(d|\Delta) d\Delta \cdot \tag{14}$$

$g(d)$ is symmetric about each point except $d = 0$. So

$$g(0) = \int_0^1 qe^{-q\Delta}(1 - \Delta) d\Delta = 1 + \frac{e^{-q} - 1}{q}, \tag{15}$$

whereas, for $d \geq 1$,

$$g(d) = \int_{d-1}^{d} q e^{-q\Delta}(\Delta - d + 1)\,d\Delta + \int_{d}^{d+1} q e^{-q\Delta}(d - \Delta + 1)\,d\Delta = \frac{2e^{-qd}(\cosh q - 1)}{q}. \tag{16}$$

ÿ

Thus, if a traffic flow, in the form of a stream of independent packets, has a distribution given by (6), then recording packet arrivals as integer values will produce the perceived distribution (7).

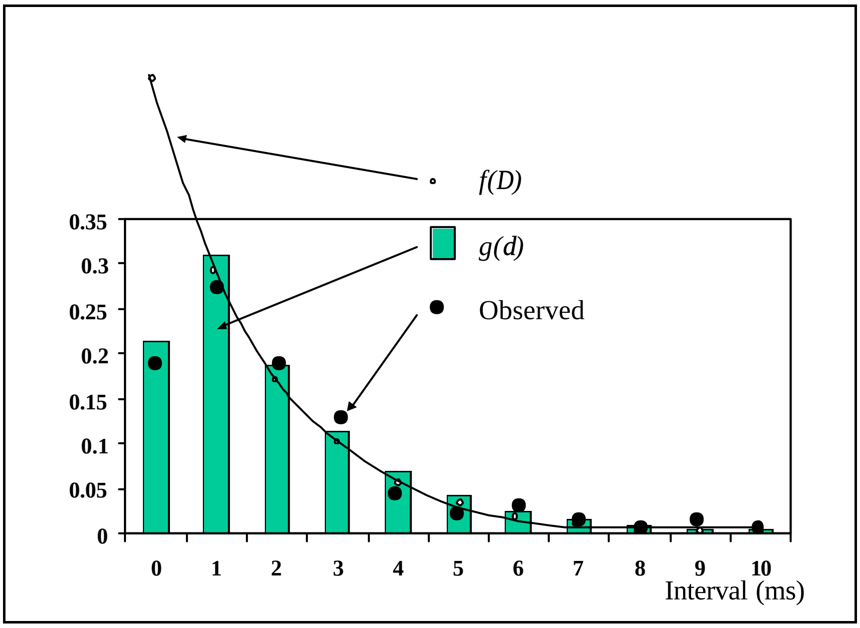

Figure 2. Comparing predicted (real), predicted (integer) and observed (integer) intervals

3. RESULTS AND CONCLUSIONS

(7) is a valid distribution since

$$\sum_{d=0}^{\infty} g(d) = 1 + \frac{e^{-q} - 1}{q} + \sum_{d=1}^{\infty} \frac{2e^{-qd}(\cosh q - 1)}{q} = 1 \tag{17}$$

as required. Figure 2 shows data from a national enterprise network plotted against the theoretical $g(d)$ distribution. (An alternative is to data-fit to the ln graph as before.) Naturally, any real-world network shows some deviation from the ideal but the general shape is consistent and the early values in particular suggest the correct form. Reversing the argument, if empirical observations, obtained from arrival time truncation, fit the distribution in (7) then the underlying flow may be assumed to be Poisson.

REFERENCES

Caglar, M. 2004. A Long-Range Dependent Workload Model for Packet Data Traffic, *Mathematics of Operations Research*, Vol. 9, No. 1, February 2004, pp92-105.

Paxson, V. and Floyd, S., 1995. Wide-Area Traffic: The Failure of Poisson Modelling. *IEEE/ACM Transactions on Networking*, Vol. 3, No. 3, June 1995, pp 226-244.