OPTIMAL ALLOCATION OF EFFORT IN PACKET SWITCHING NETWORKS WITH GENERALLY DISTRIBUTED TRANSMISSION TIMES

P.K. Pollett

Department of Mathematics
The University of Queensland
Queensland 4072 Australia
Ph: (07) 3365 3459 Fax: (07) 3365 1477
pkp@maths.uq.edu.au

ABSTRACT: We consider the problem of how best to assign resources in a packet switching network with generally distributed transmission times so as to minimize the average delay under a cost constraint. For such networks there are typically no analytical formulae for the delay distributions. Thus, we shall approach the optimal allocation problem using an approximation technique, namely the *residual-life approximation* [9]. This work extends previous work of author [10] and generalizes results of Kleinrock [6], who studied networks with exponentially distributed service times.

1. INTRODUCTION

In contrast to circuit switched networks, where one or more circuits are held simultaneously on several links connecting source and destination nodes, only one link is used at any given time by transmissions in a message or packet switched network; transmissions are received in their entirety at a given node before being transmitted along the next link in their path through the network. If the link is busy, packets are stored in a buffer until the link becomes available for use: hence the term *store-forward*. Thus, the total delay W is the sum of the individual delays experienced en route. We shall consider the problem of how best to allocate link capacities so as to minimize $\mathrm{E}(W)$, the expected total delay.

2. THE MODEL

Suppose that there are N switching nodes, labelled $n = 1, 2, \ldots, N$, and J communications links, labelled $j = 1, 2, \ldots, J$. We shall assume that all the links are perfectly reliable and not subject to noise, so that transmission times are determined by message length. We shall also suppose that the time taken to switch, buffer, and (if necessary) re-assemble and acknowledge, is negligible compared with the transmission times. Traffic entering the network from external sources is assumed to be Poisson, and that which originates from node m and is destined for node n is offered at rate ν_{mn}. Message lengths are assumed to be mutually independent and arbitrarily distributed with common mean μ^{-1} (bits, say). We shall assume that each link operates under the the usual first-come-first-served (FCFS) discipline and that a total effort (or capacity) of ϕ_j (bits per second) is assigned to link j. (We shall indicate later how our results extend to deal with other disciplines.)

We shall allow for two possible routing procedures, that of *fixed routing*, where there is a unique route specified for each origin-destination pair (m, n), and *random alternative routing*, where one of a number of possible paths is chosen at random. (We do not allow for *adaptive* or *dynamic routing*, where routing decisions are made on the basis of the observed traffic flow.)

For fixed routing we define $R(m, n)$ to be the (unique) collection of links used by a message emanating from node m and destined for node n. In particular, let

$$R(m, n) = \{r_{mn}(1), \ldots, r_{mn}(s_{mn})\} ,$$

where s_{mn} is the number of links used by that message and $r_{mn}(s)$ is the link used at stage s along its route (note that $r_{mn}(s)$, $s = 1, 2, \ldots, s_{mn}$, are distinct).

It is perhaps surprising that random alternative routing can be accommodated within the framework of fixed routing (see, for example, Kelly [4], Exercise 3.1.2). If there are a number of alternative routes for a given origin-destination pair (m, n), then one simply provides a finer classification for messages using these routes. We label the alternative routes as (m, n, i), $i = 1, 2, \ldots, N(m, n)$, where $N(m, n)$ is the number of alternatives for origin-destination pair (m, n), and we replace $R(m, n)$ by

$$R(m, n, i) = \{r_{mni}(1), \ldots, r_{mni}(s_{mni})\} ,$$

for $i = 1, 2, \ldots, N(m, n)$, where now $r_{mni}(s)$ is the link used at stage s along alternative route i

and s_{mni} is the number of stages. We then replace ν_{mn} by $\nu_{mni} = \nu_{mn} q_{mni}$, where q_{mni} is the probability that alternative route i is chosen. Clearly $\nu_{mn} = \sum_{i=1}^{N(m,n)} \nu_{mni}$, and so the effect is to thin the Poisson stream of messages of 'type' (m,n) into a collection of independent Poisson streams, one for each type (m,n,i). We should think of messages as being identified by their type, whether this be simply (m,n), for fixed routing, or the finer classification (m,n,i), for alternative routing. For convenience let us denote by T the set of all types, and suppose that, for each t in T, messages of type t arrive according to a Poisson stream with rate ν_t and traverse the route

$$R(t) = \{r_t(1), \ldots, r_t(s_t)\} ,$$

a collection of s_t distinct links. Having established this new nomenclature, that of *type*, the network can be perceived as a *network of queues with customers of different types* (Kelly [3]) with the queues representing the links and the customers representing the messages. Thus, in particular, if message lengths have an exponential distribution, the model is analytically tractable: in equilibrium, the links behave *independently*: indeed *as if they were isolated*, each with independent streams of Poisson offered traffic (independent among types). For example, if we let

$$\alpha_j(t,s) = \begin{cases} \nu_t, & \text{if } r_t(s) = j, \\ 0, & \text{otherwise,} \end{cases}$$

so that the arrival rate at link j is given by

$$\alpha_j = \sum_{t \in T} \sum_{s=1}^{s_t} \alpha_j(t,s)$$

and the demand (in bits per second) by $a_j = \alpha_j/\mu$, then, provided the system is stable ($a_j < \phi_j$ for each j), the expected number of messages at link j (whose transmission is incomplete) is given by

$$\mathrm{E}(n_j) = \frac{a_j}{\phi_j - a_j} \qquad (1)$$

and the expected delay by

$$\mathrm{E}(W_j) = \frac{1}{\alpha_j}\left(\frac{a_j}{\phi_j - a_j}\right) = \frac{1}{\mu\phi_j - \alpha_j} .$$

3. APPROXIMATION TECHNIQUES

In order to make satisfactory progress in cases where message lengths have an arbitrary distribution, we shall need to make one further assumption. It is similar to the celebrated independence assumption of Kleinrock [6]. We shall suppose that successive messages requesting transmission along any given link have lengths which are independent and identically distributed, and that message lengths at different links are independent. Clearly a message of a given type maintains its length as it passes through the network. However, numerous simulation results (see, for example, Kleinrock [6]) suggest that, even so, the network behaves *as if* successive message lengths at a given node are independent. This phenomenon can be explained by observing that the arrival process at a given node is the result of the superposition of a generally large number of streams, and the approximation can then be justified on the basis of limit theorems concerning the superposition of marked point processes (see Brown and Pollett [7] and the references contained therein). The assumption that independence is apparent at the links *themselves* can be justified on the basis of the corresponding results on thinning of marked point processes (see, for example, Brown [1]). Kleinrock's independence assumption differs from ours in that the message-length distribution at a given link j is assumed to be exponential with common mean μ^{-1}, a natural consequence of the usual teletraffic modelling assumption that the lengths of messages arriving from outside the network are independent and identically distributed exponential random variables. However, although the exponential assumption is usually valid in circuit switched networks, we should not expect it to be appropriate in the present context of message/packet switching, since packets are of similar length. Thus, it is more realistic to assume, as we do here, that message lengths are arbitrarily distributed. In order that this be reflected in our independence assumption, we shall allow successive messages requesting transmission along a given link j to be arbitrarily distributed. Although this distribution might be the same for each link, we shall find it no less convenient to assume that it differs from one to another. Thus, we shall assume that at link j message lengths have a distribution function $F_j(x)$ which has mean μ_j^{-1} and variance σ_j^2.

Even under the independence assumption, our model is not analytically tractable. In particular, there are no analytical formulae for the delay distributions. We shall therefore adopt one of the many approximation techniques. Consider a particular link j and let $Q_j(x)$ be the distribution function of the *queueing time*, that is, the period of time a message spends in the buffer at link j *before* its transmission begins. The *residual-life approximation*, developed by the author in [9], provides an accurate approximation for $Q_j(x)$:

$$Q_j(x) \simeq \sum_{n=0}^{\infty} \Pr(n_j = n) G_j^{(n)}(x) , \qquad (2)$$

where

$$G_j(x) = \mu_j \int_0^{\phi_j x} (1 - F_j(y))\, dy$$

and $G_j^{(n)}(x)$ denotes the n-fold convolution of $G_j(x)$. The distribution of n_j, the number of messages at link j, used in (2), is that of a corresponding *quasireversible network* (see Kelly [4]); specifically, a network of *symmetric* queues obtained by imposing a symmetry condition at each link j. In the present case, this amounts to replacing the FCFS discipline with a preemptive-resume last-come-first-served discipline at each link in the network. The term *residual-life approximation* comes from renewal theory; $G_j(x)$ is the *residual-life distribution* corresponding to the (life-time) distribution $F_j(x/\phi_j)$.

One immediate consequence of (2) is that the expected queueing time $\bar{Q}_j$ is approximately

$$\frac{1 + \mu_j^2 \phi_j^2}{2\mu_j \phi_j} \mathrm{E}(n_j)\,,$$

where $\mathrm{E}(n_j)$ is the expected number of messages at link j in the quasireversible network. Hence, the expected delay at link j is approximated as follows:

$$\mathrm{E}(W_j) \simeq \frac{1}{\mu_j \phi_j} + \frac{1 + \mu_j^2 \phi_j^2}{2\mu_j \phi_j} \mathrm{E}(n_j)\,. \qquad (3)$$

In the residual-life approximation, it is only $\mathrm{E}(n_j)$ which changes when the service discipline is altered. For the present FCFS discipline $\mathrm{E}(n_j)$ is given by (1) with $a_j = \alpha_j / \mu_j$.

Simulation results presented in [9] justify the approximation by assessing its accuracy under a variety of conditions. Even for relatively small networks with generous mixing of message streams, it is accurate, and the accuracy improves as the size and complexity of the network increases. (The approximation is very accurate in the tails of the queueing time distributions and so it allows an accurate prediction to be made of the likelihood of extreme queueing times.) For moderately large networks, the approximation becomes worse as the coefficient of variation $\mu_j \sigma_j$ of the message-length distribution deviates markedly from 1, the value which obtains in the exponential case.

4. OPTIMAL ALLOCATION OF EFFORT

We now turn our attention to the problem of how best to assign resources so that the average network delay, or equivalently the average number of messages in the network, is minimized. We shall suppose that there is some overall network budget F (dollars) which cannot be exceeded, and that the cost of operating link j is a function f_j of its capacity. Suppose that the cost of operating link j is proportional to ϕ_j, that is, $f_j(\phi_j) = f_j \phi_j$ (the units of f_j are dollars per unit of capacity (or dollar-seconds per bit)). Thus, we should choose the capacities subject to the cost constraint

$$\sum_{j=1}^{J} f_j \phi_j = F\,. \qquad (4)$$

We shall suppose that the average delay of messages at link j is adequately approximated by (3). Thus, we shall assume that

$$\mathrm{E}(W_j) = \frac{1}{\mu_j \phi_j} + \frac{1 + \mu_j^2 \sigma_j^2}{2\mu_j \pi_j} \left(\frac{\alpha_j}{\mu_j \phi_j - \alpha_j} \right)\,.$$

Using Little's Theorem, we can obtain an (approximate) expression for the mean number $\bar{m}$ of messages in the network. This is

$$\begin{aligned} \bar{m} &= \sum_{j=1}^{J} \alpha_j \left\{ \frac{1}{\mu_j \phi_j} + \frac{\alpha_j (1 + \mu_j^2 \sigma_j^2)}{2\mu_j \phi_j (\mu_j \phi_j - \alpha_j)} \right\} \\ &= \sum_{j=1}^{J} a_j \left\{ \frac{1}{\phi_j} + \frac{a_j (1 + c_j)}{2\phi_j (\phi_j - a_j)} \right\}\,, \end{aligned}$$

where $c_j = \mu_j^2 \phi_j^2$ is the squared coefficient of variation of the message-length distribution $F_j(x)$. We seek to minimize $\bar{m}$ over $\phi_1, \ldots, \phi_J$ subject to (4).

To this end, we introduce a lagrange multiplier λ^{-2}; our problem then becomes one of minimizing

$$L(\phi_1, \ldots, \phi_J; \lambda^{-2}) = \bar{m} + \frac{1}{\lambda^2} \left(\sum_{j=1}^{J} f_j \phi_j - F \right)\,.$$

Setting $\partial L / \partial \phi_j = 0$ for fixed j yields a quartic polynomial equation in ϕ_j, namely

$$\begin{aligned} 2 f_j \phi_j^4 - 4 a_j f_j \phi_j^3 + 2 a_j (a_j f_j - \lambda^2) \phi_j^2 & \\ - 2\epsilon_j a_j^2 \lambda^2 \phi_j + \epsilon_j a_j^3 \lambda^2 = 0, & \qquad (5) \end{aligned}$$

where $\epsilon_j = c_j - 1$, and our immediate task is to find solutions such that $\phi_j > a_j$ (recall that this latter condition is a requirement for stability). The task is simplified by observing that the transformation

$$\phi_j f_j / F \to \phi_j,\ a_j f_j / F \to a_j,\ \lambda^2 / F \to \lambda^2, \quad (6)$$

reduces the problem to one with unit costs $f_j = F = 1$, whence the polynomial equation (5) becomes

$$\begin{aligned} 2\phi_j^4 - 4 a_j \phi_j^3 + 2 a_j (a_j - \lambda^2) \phi_j^2 & \\ - 2\epsilon_j a_j^2 \lambda^2 \phi_j + \epsilon_j a_j^3 \lambda^2 = 0, & \qquad (7) \end{aligned}$$

and the constraint becomes

$$\phi_1 + \phi_2 + \cdots + \phi_J = 1 \,. \tag{8}$$

If transmission times are exponentially distributed ($\epsilon_j = 0$ for each j) it is easy to verify that (7) has a unique solution on (a_j, ∞) given by

$$\phi_j = a_j + |\lambda| a_j^{1/2} \,.$$

Upon application of the constraint (8) we arrive at the optimal capacity assignment

$$\phi_j = a_j + \left(1 - \sum_{k=1}^{J} a_k\right) \frac{a_j^{1/2}}{\sum_{k=1}^{J} a_k^{1/2}} \,,$$

for unit costs. In the case of general costs this becomes

$$\phi_j = a_j + \frac{1}{f_j}\left(F - \sum_{k=1}^{J} f_k a_k\right) \frac{(f_j a_j)^{1/2}}{\sum_{k=1}^{J} (f_k a_k)^{1/2}} \,,$$

after applying the transformation (6). This is a result obtained by Kleinrock [6] (see also Kelly [4]): the allocation proceeds by first assigning enough capacity to meet the demand a_j, at each link j, and then allocating a proportion of the affordable excess capacity,

$$\frac{1}{f_j}\left(F - \sum_{k=1}^{J} f_k a_k\right)$$

(that which could be afforded to link j), in proportion to the square root of the cost $f_j a_j$ of meeting that demand. In the case where some or all of the ϵ_j, $j = 1, 2, \ldots, J$, deviate from zero, (7) is difficult to solve analytically. We shall adopt a perturbation technique, assuming that the lagrange multiplier and the optimal allocation take the following forms:

$$\begin{aligned} \lambda &= \lambda_0 + \sum_{k=1}^{J} \lambda_{1k}\epsilon_k + O(\epsilon^2) \,, \\ \phi_j &= \phi_{0j} + \sum_{k=1}^{J} \phi_{1jk}\epsilon_k + O(\epsilon^2) \,, \\ &\qquad\qquad j = 1, \ldots, J, \end{aligned} \tag{9}$$

where by $O(\epsilon^2)$ we mean terms of order $\epsilon_i\epsilon_k$. The zero-th order terms come from Kleinrock's solution: specifically,

$$\phi_{0j} = a_j + \lambda_0 a_j^{1/2}, \quad j = 1, \ldots, J,$$

where

$$\lambda_0 = \frac{1 - \sum_{k=1}^{J} a_k}{\sum_{k=1}^{J} a_k^{1/2}} \,.$$

On substituting (9) into (7) we obtain an expression for ϕ_{1jk} in terms of λ_{1k}, which in turn is calculated using the constraint (8) and by setting $\epsilon_k = \delta_{kj}$ (the Kronecker delta). We find that the optimal allocation, to first order, is

$$\begin{aligned} \phi_j = a_j + \lambda_0 a_j^{1/2} &- \frac{a_j^{1/2}}{\sum_{k=1}^{J} a_k^{1/2}} \sum_{k \neq j} b_k \epsilon_k \\ &+ \left(1 - \frac{a_j^{1/2}}{\sum_{k=1}^{J} a_k^{1/2}}\right) b_j \epsilon_j \,, \end{aligned} \tag{10}$$

where

$$b_k = \frac{1}{4}\lambda_0 a_k^{3/2} \frac{a_k + 2\lambda_0 a_k^{1/2}}{(a_k + \lambda_0 a_k^{1/2})^2} \,.$$

For most practical applications, higher-order solutions are required. To achieve this we can simplify matters by using a single perturbation $\epsilon = \max_{1 \le j \le J} |\epsilon_j|$. For each j we then define a quantity $\beta_j = \epsilon_j / \epsilon$ and write ϕ_j and λ as power series in ϵ:

$$\begin{aligned} \lambda &= \sum_{n=0}^{\infty} \lambda_n \epsilon^n \\ \phi_j &= \sum_{n=0}^{\infty} \phi_{nj} \epsilon^n \,, \quad j = 1, \ldots, J. \end{aligned} \tag{11}$$

Substituting as before into (7), and using (8), gives rise to an iterative scheme, details of which can be found in Pollett [8]. The first-order approximation is useful, none-the-less, in dealing with networks whose message-length distributions are all ‘close’ to exponential in the sense that their coefficients of variation do not differ significantly from 1. It is also useful in providing some insight into how the allocation varies as ϵ_j, for fixed j, varies. Let ϕ_j, $j = 1, 2, \ldots, J$, be the new optimal allocation obtained after incrementing ϵ_j by a small quantity $\delta > 0$. We find that to first order in δ

$$\phi_j' - \phi_j = \left(1 - \frac{a_j^{1/2}}{\sum_{k=1}^{J} a_k^{1/2}}\right) b_j \delta > 0$$

and, for $i \neq j$,

$$\phi_i' - \phi_i = -\frac{a_i^{1/2}}{\sum_{k=1}^{J} a_k^{1/2}} (\phi_j' - \phi_j) < 0 \,.$$

Thus, if the coefficient of variation of the message-length distribution at a given link j is increased (respectively decreased) by a small quantity δ, then there is an increase (respectively decrease) in the optimal allocation at link j which is proportional to δ. All other links experience a complementary decrease (respectively increase) in their allocations and the resulting deficit is reallocated in proportion to the square root of the demand.

In Pollett [8] empirical estimates were obtained for the radii of convergence of the power series (11) for the optimal allocation. In all cases considered there, the closest pole to the origin was on the negative real axis outside the physical limits for ϵ_i, which are of course $-1 \leq \epsilon_j < \infty$. The perturbation technique is therefore useful for networks whose message-length distributions are, for example, Erlang (Gamma) $(-1 < \epsilon_j < 0)$ or, for example, Hyperexponential $(0 < \epsilon_j < \infty)$ with a not too large a coefficient of variation.

We have assumed that the capacity does not depend on the state of the link (as a consequence of the FCFS discipline), and, that the cost of operating a link is a linear function of its capacity. Let us briefly consider some other possibilities. Let $\phi_j(n)$ be the effort assigned to link j where there are n messages present. If, for example,

$$\phi_j(n) = \frac{n}{n+\eta-1}\phi_j\,,$$

where η is a positive constant, the zero-th order allocation, optimal under (4), is precisely the same as before (the case $\eta = 1$). For values of η greater than 1 the capacity increases as the number of messages at link j increases and levels off at a constant value ϕ_j as the number becomes large. If we allow η to depend on j we get a similar allocation but with the factor

$$\frac{(f_j a_j)^{1/2}}{\sum_{k=1}^{J}(f_k a_k)^{1/2}}$$

replaced by

$$\frac{(f_j \eta_j a_j)^{1/2}}{\sum_{k=1}^{J}(f_k \eta_k a_k)^{1/2}}\,.$$

See Kelly [4] for further details. The higher order analysis is very nearly the same as before. The factor $1+c_j$ is replaced by $\eta_j(1+c_j)$; for the sake of brevity, we shall omit the details.

As another example, suppose that the capacity function is linear, that is $\phi_j(n) = \phi_j n$, and that message lengths are exponentially distributed. In this case, the total number of messages in the system has a Poisson distribution with mean $\sum_{j=1}^{J} a_j/\phi_j$, and it is elementary to show that the optimal allocation subject to (4) is given by

$$\phi_j = \frac{(f_j a_j)^{1/2}}{f_j \sum_{k=1}^{J}(f_k a_k)^{1/2}} F, \quad j = 1, \ldots, J.$$

It is interesting to note that we get a *proportional* allocation, $\phi_j/\phi_k = a_j/a_k$, in this case if (4) is replaced by

$$\sum_{j=1}^{J} \log \phi_j = 1\,.$$

More generally, we might use the constraint

$$\sum_{j=1}^{J} f_j \log(g_j \phi_j) = F$$

to account for 'decreasing costs', when costs become less with each increase in capacity. Under this constraint, the optimal allocation is $\phi_j = \lambda a_j/f_j$, where

$$\log \lambda = \frac{F - \sum_{k=1}^{J} f_k \log(g_k a_k/f_k)}{\sum_{k=1}^{J} f_k}\,.$$

5. REFERENCES

[1] Brown, T.C, *Some Distributional Approximations for Random Measures*, Ph.D. thesis (University of Cambridge, 1979).

[2] Kleinrock, L., *Queueing Systems Vol. II: Computer Applications* (Wiley, New York, 1976).

[3] Kelly, F.P., Networks of queues with customers of different types, *J. Appl. Probab.* **12** (1975) 542–554.

[4] Kelly, F.P., *Reversibility and Stochastic Networks* (Wiley, Chichester, 1979).

[5] Kelly, F.P., Networks of queues, *Adv. Appl. Probab.* **8** (1976) 416–432.

[6] Kleinrock, L., *Communication Nets* (McGraw-Hill, New York, 1964).

[7] Brown, T.C. and Pollett, P.K., Some distributional approximations in Markovian queueing networks, *Adv. Appl. Probab.* **14** (1982) 654–671.

[8] Pollett, P.K., *Distributional Approximations for Networks of Queues*, Ph.D. thesis (University of Cambridge, 1982).

[9] Pollett, P.K., Residual life approximations in general queueing networks, *Elektron. Informationsverarb. u. Kybernet.* **20** (1984) 41–54.

[10] Pollett, P.K., Analysis of response times and optimal allocation of resources in message and packet switched networks, *Asia-Pacific J. Operat. Res.* **3** (1986) 134–149.