Bernstein + Sons

INFORMATION SYSTEMS CONSULTANTS

5 Brewster Lane, Bellport, New York 11713-2803
Phone: 1-516-286-1339, Fax: 1-516-286-1999
E-mail: yaya@bernstein-plus-sons.com

May 1, 1988
Revised April 26, 1997

Some Comments on Highly Dynamic Network Routing

by

Herbert J. Bernstein

Originally published as Technical Report No. 371
Computer Science Department, New York University,
May 1988, 11 pp.

Notice

ABSTRACT

Attempting to dynamically adapt network parameters to the load seen can produce unexpected results. We present a simple model network example which demonstrates unstable behavior when traffic is directed according to routing optimized for minimal delay and the load varies at a rate comparable to the routing calculation time. The instability can be avoided by using almost any alternate design which avoids the knees on the delay curves, *e.g.* a Maximum Entropy Method design. The delay penalty in this case turns out to be small. This paper is very mathematical, but the point is simple: attempting make network parameters change as quickly as possible may not be an appropriate network management strategy.

1. Introduction

In teaching about network performance it is sometimes difficult to convince students of some of the counter-intuitive facts of network routing. In this note we present a simple example of the instability which can result from too serious an attempt at optimal bifurcated routing on a

network with changing offered load. The example is given first in the form of a three node network of identical lines, and then in the form of a network providing multiple node-disjoint multi-hop paths of unequal capacity. In a large network there can be a significant benefit in avoiding poor choices of routing. When a high speed path of very few hops is available, it would seem to be sheer folly to send any traffic along slow multi-hop back-door paths. However, the addition of traffic to a route removes some of its capacity, and the next set of messages might well be better sent along an alternate path. The assignment of a set of alternate paths with an allocation of portions of the offered load among them gives rise to the bifurcated routing problem. Under reasonable constraints it is possible to find routings which optimize an appropriate payoff function, *e.g.* average end-to-end delay. (See [7] and [6].) Such calculations can be very time-consuming. A node in the network may not be able to wait for a grand plan from a central routing authority to decide where to send the next message. In that case it may be desirable to use a simple distributed shortest-path algorithm which allows each node to estimate the optimal current path for the next burst of traffic. Schwartz [11, chapter 6] gives a review of the common distributed dynamic routing techniques, and notes the necessary relationship between the diameter of a network and the number of iterations needed or convergence. Such algorithms normally do not give bifurcated routing directly, since they try for a "shortest" path. However, if the definition of length of a path is total delay along it, and if current traffic is properly accounted, then one can expect that traffic would be diverted from over-loaded paths which were once sensed as providing minimal delay, providing a somewhat oscilla-tory approximation to optimal bifurcated routing.

There are problems with the distributed algorithms. Kleinrock [10] pointed out that "... uncontrolled alternate routing in a congested net can lead to chaos. Indeed, the telephone company tends to limit (and even prohibit completely) alternate routing on unusually busy days (Mother's Day, for example)." As Schwartz notes of a common shortest path algorithm, "Although convergence to the shortest path is guaranteed, routing table entries may change during the convergence period, giving rise to possible loops during that interval," [11, p 277]. Even if one suppresses the creation of loops, there can be serious problems. When the offered load on which routing calculations are being done varies significantly on time-scales commensurate with the convergence period of the routing algorithm, one has created a feedback control system which can oscillate for very long periods. The reader is referred to standard texts on Control Theory, e.g. [1].

In this note, we present a simple example of the type of instability which can result from computing an optimal bifurcated routing for a load which changes on the time-scale of the calculation. While the example was created to clearly demonstrate the sub-optimal results of optimal routing in this case, it is not, in our opinion and observation of the Internet, unrealistic. (The Internet is a loose confederation of networks able to provide a reasonable degree of interoperability for users on connected hosts. See [5].)

As a contrast to optimal routing, we will mention routing found by the Maximum Entropy Method (MEM) [2-4, 8, 9, 12]. MEM, originally due to Jaynes, comes from the interaction of Information Theory with Statistical Mechanics. It works for underdetermined systems, producing the smoothest answer consistent with the data. In the case at hand, it would produce network flows equalized for whatever parameter we wish to smooth: traffic by links, total queue size along paths, etc. Since we need no more than that qualitative approach for our example, we will not present further detail on Maximum Entropy here.

We will form our examples by taking static cascades of independent M/M/1 queues as our model of multi-hop network paths, ignoring blocking, dependence in forwarding, and many other effects. Most importantly, we will ignore the probable failure of stochastic equilibrium by using M/M/1 queuing models even though we vary the customer arrival rate. In a network of large diameter and heavy traffic, it is not unreasonable to assume that the time scales of routing calculations are sufficiently large to consider them as leading to approximate equilibrium queue-

by-queue. It would be a good idea to do a more accurate non-equilibrium model, but that would take us beyond our simple pedagogical objective into material better suited to a research paper.

2. A Simple Three Node Example

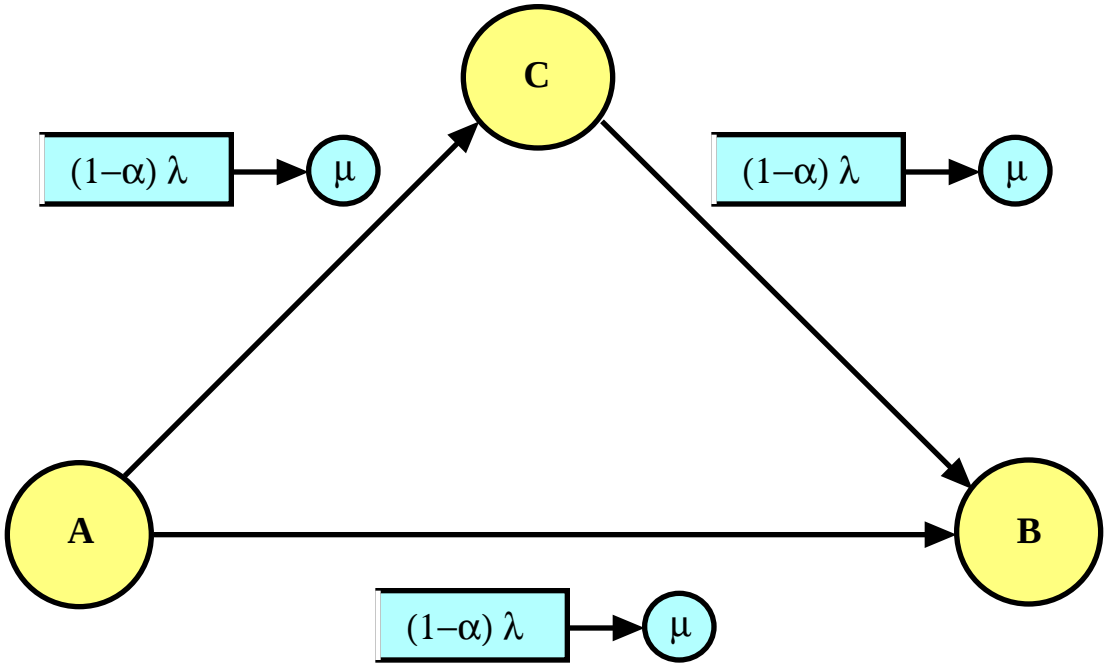

Figure 1. Three Node Network Example of Bifurcated Routing.
Traffic flows from A to B either directly or via C.

We will start with a trivial three-node network. Consider a simple example of a network consisting of three nodes, *A*, *B*, *C*, where *A* has traffic λ for *B* and may route it either directly or via *C*. We assume independent M/M/1 queues at each node, and an equal service rate μ for each queue. We assume that *A* supplies a Poisson distributed stream of packets at rate $\alpha\,\lambda$ along the direct route and $(1-\alpha)\,\lambda$ along the indirect route. Our routing decision is to choose the "best" value of α. Assume we seek to control average delay. The average delay is

$$W = \alpha\left(\frac{1}{\mu - \alpha\lambda}\right) + (1-\alpha)\left(\frac{2}{\mu - (1-\alpha)\lambda}\right)$$

This equation becomes clearer if we define

$$\rho = \frac{\lambda}{\mu}$$

use a dimensionless delay $W\mu$ (essentially queue length plus one) and define

$$\sigma = \alpha - \frac{1}{2}$$

Then

$$W\mu = \frac{3\sigma^2 - \sigma + \frac{3}{2}\left(1 - \frac{\rho}{2}\right)}{\left(1 - \frac{\rho}{2}\right)^2 - \rho^2\sigma^2}$$

and we seek to minimize $W\mu$ by varying σ. In this case, it is sufficient to look at the boundary case $\alpha = 1$ ($\sigma = .5$) and at the location of zeros of the derivative of $W\mu$ with respect to σ.

The zeros of the derivative are given by

$$\sigma = \left(3 \pm 2\sqrt{2}\right)\frac{1 - \frac{\rho}{2}}{\rho}$$

The lower root is the one in the proper range ($-.5 \leq \sigma \leq .5$) as long as $\rho > .293$. Below that we use the direct route. Above that value the routing bifurcates.

Consider a situation in which the offered load switches between, say, $\rho = .3$ and $\rho = .9$, spending about half its time at each level. This might be due to the inherent characteristics of the applications, or, perhaps due to a periodic sensing of overload at the higher load and a backing off to the light load to relieve congestion. The average load is then $\rho = .6$, and we face the choice of routing for the instantaneous values or for the average. The "optimum" values of α for these values of ρ are:

ρ	α
.3	.99
.6	.70
.9	.60

(given to the nearest .01). Consider the following table of values of $W\mu$ for these values of α and for .5 (the Maximum Entropy value, see below).

	α			
ρ	.99	.70	.60	.50
.3	1.43	1.55	1.64	1.76
.6	2.46	1.94	1.99	2.14
.9	9.10	2.71	2.55	2.72

From this table, we observe that the average delay using the optimum α values for the $\rho = .3$ and $\rho = .9$ cases is $W\mu = 1.99$, which is slightly below the average 2.13 of the $W\mu$ values for $\alpha = .7$. We gained about seven percent by using highly dynamic routing. In fact, if we had used a dynamic shortest path routing, which would have taken $\alpha = 1.0$ in all these cases, we would have paid a serious delay penalty. Worse yet, suppose we had a processing time for the dynamic algorithm comparable to the cycle time of the load and switched to the values of α for $\rho = .3$ just when the load switched to $\rho = .9$ and vice-versa. Then we would have an average delay for dynamic routing of $W\mu > 5.3$. (For $\rho = .3$ we would have used $\alpha = .6$ and gotten a delay of $W\mu = 1.64$, while for $\rho = .9$ we would have used $\alpha = .99$ and gotten a delay of $W\mu = 9.10$). Of further interest is the fact that the Maximum Entropy value of $\alpha = .5$ (assuming we balance traffic by links, not paths) gives a true average delay of $W\mu = 2.24$ using the correct delay values of $W\mu = 1.76$ for $\rho = .3$ and $W\mu = 2.72$ for $\rho = .9$, a penalty of about fifteen percent for taking too low a value of α.

If we look at the Maximum Entropy solution for traffic balanced by paths, i.e. making the

average queue on the direct path equal to the average queue on the indirect path, we obtain bifurcated routing for all values of ρ:

ρ	α
0.	.97
.3	.64
.6	.62
.9	.51

with an average delay for our test case using ρ = . 6 for the alternating traffic of 2.10, a penalty of about five percent. Our conjecture is that the optimal dynamic routing solutions are, in many cases, similarly unstable under reasonable dynamic load, and that the Maximum Entropy routings will prove a more robust starting point for distributed dynamic adjustments.

3. A More General Case -- Inhomogeneous Rates, Many Paths

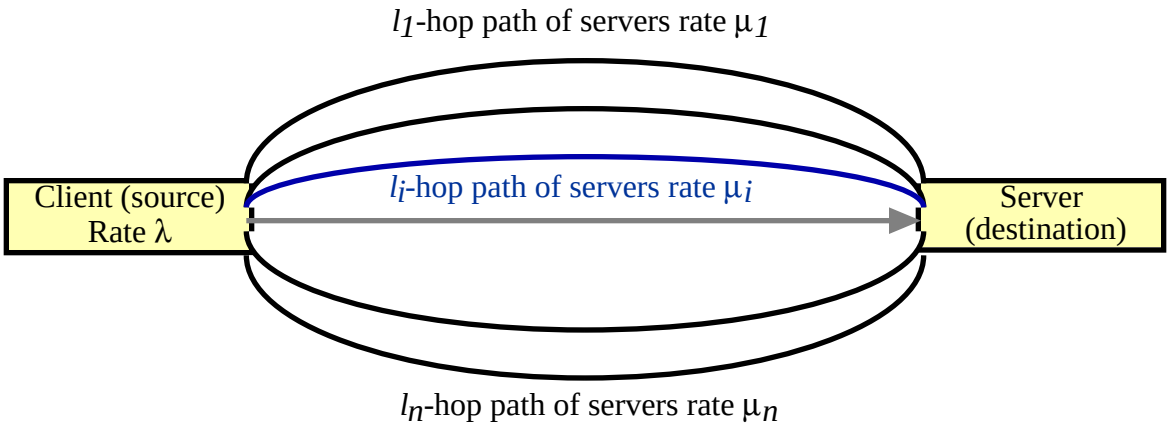

Figure 2. Inhomogeneous node-disjoint paths of multi-hop M/M/1 queues. Traffic $\alpha_i \lambda$ goes onto the *i*th path and sees l_i hops of servers at rate μ_i.

It may seem we are extracting too much from the three node case. It is more realistic to consider a network offering multiple paths with differing numbers of hops of differing capacity. For reliability, it is desirable that separate routes share as few nodes as possible. We will restrict our attention to node-disjoint paths in which only the origin and destination are common to distinct paths. We also require that on any single path the service rates on all of the links forming the path be the same, even though links on different paths may have different service rates. We contend that this is not a severe restriction, since very high capacity links in series with much lower capacity links will be dominated by the bottleneck formed by the slower links, and can be effectively ignored for our purposes. In Section 4, below, we will give a conservative approximation to the case of mixed rates on a single path.

Thus consider the extension of this analysis to a network with a single client offering traffic load λ trying to reach a server via *n* node-disjoint paths, where each path, *i*, acts as a cascade of l_i independent M/M/1 queues of service rate μ_i. Suppose the client allocates his traffic to path *i* with weight α_i, providing Poisson distributed traffic at rate $\alpha_i \lambda$ to the path. Then the independent M/M/1 queuing model gives an expected delay of

$$W = \sum \frac{\alpha_i l_i}{\mu_i - \alpha_i \lambda}, \quad \text{for } 0 \le \alpha_i \le 1, \ \sum \alpha_i = 1$$

Define

$$d_i = \frac{\alpha_i l_i}{\mu_i - \alpha_i \lambda}$$

and compute the partial derivatives which will be needed in finding minima of *W*.

$$\frac{\partial d_i}{\partial \alpha_i} = \frac{\mu_i l_i}{(\mu_i - \alpha_i \lambda)^2}$$

for $i = 1, ..., n-1$ and

$$\frac{\partial d_n}{\partial \alpha_i} = -\frac{\mu_n l_n}{(\mu_n - \alpha_n \lambda)^2}$$

by taking $\alpha_n = 1 - \alpha_1 - \alpha_2 - ... - \alpha_{n-1}$, so that the critical points of *W* as a function of $\alpha_1, ... \alpha_{n-1}$, occur when

$$0 = \frac{\partial W}{\partial \alpha_i} = \frac{\mu_i l_i}{(\mu_i - \alpha_i \lambda)^2} - \frac{\mu_n l_n}{(\mu_n - \alpha_n \lambda)^2}$$

i.e. when

$$\frac{\mu_i l_i}{(\mu_i - \alpha_i \lambda)^2} = \tau^2$$

for some τ independent of *i*, $i = 1, ..., n$. We will solve these equations for α_i, but first note that

$$\mu_i - \alpha_i \lambda = \varepsilon_i \frac{(\mu_i l_i)^{\frac{1}{2}}}{\tau}$$

with $\varepsilon_i = \pm 1$. The negative value of ε_i is “unphysical”, since that would require an overload of the first queue on path *i* by the distributed traffic. Thus we accept only the positive roots and obtain

$$\alpha_i = \frac{1}{\lambda}\left(\mu_i - \frac{(\mu_i l_i)^{\frac{1}{2}}}{\tau}\right)$$

We can solve for τ by

$$1 = \sum \alpha_i = \frac{1}{\lambda} \sum \mu_i - \frac{1}{\lambda \tau} \sum (\mu_i l_i)^{\frac{1}{2}}$$

$$\tau = \frac{\sum (\mu_i l_i)^{\frac{1}{2}}}{\sum \mu_i - \lambda}$$

so that

$$\alpha_i = \frac{1}{\lambda}\left(\mu_i - \left(\sum \mu_j - \lambda\right)\frac{(\mu_i l_i)^{\frac{1}{2}}}{\sum (\mu_j l_j)^{\frac{1}{2}}}\right)$$

While all the resulting values of αi are certainly critical points of *W*, they may not be valid minima. We can eliminate any concern about convexity of the problem by noting that

$$\frac{\partial^2 W}{\partial \alpha_i \partial \alpha_j} = \frac{2\mu_i \lambda l_i \delta_{i,j}}{(\mu_i - \alpha_i \lambda)^3} + \frac{2\mu_n \lambda l_n}{(\mu_n - \alpha_n \lambda)^3}$$

which, as the sum of a diagonal matrix with positive terms and a scalar times the matrix of all ones, is positive definite as long as we have $\mu_i > \alpha_i \lambda$. (This is only to be expected since this routing problem is one of a much wider class of convex minimization problems). The real question is whether the minima are within the region of interest, $0 \leq \alpha_i \leq 1$. We may drive some α_i negative with too small a value of λ. This corresponds to the $\rho < .293$ cases in our simple example above. In that case, we must reduce the allowed range of *i* by dropping appropriate paths.

To select the paths to be dropped, order the paths so that

$$\frac{\mu_i}{l_i}$$

is monotone non-increasing with *i, i.e.* so that the path with the fastest hop-corrected service rate comes first and the slowest path comes last. Compare λ to

$$\sum \mu_i - \left(\frac{\mu_n}{l_n}\right)^{\frac{1}{2}} \sum (\mu_i l_i)^{\frac{1}{2}}$$

If λ is smaller, drop path *n* and recompute on the reduced network, since in that case α_n will be negative. To see this

$$0 > \alpha_n$$

if and only if

$$0 > \mu_n - \left(\sum \mu_j - \lambda\right)\frac{(\mu_n l_n)^{\frac{1}{2}}}{\sum (\mu_j l_j)^{\frac{1}{2}}}$$

if and only if

$$\frac{\mu_n \sum \left(\mu_j l_j\right)^{\frac{1}{2}}}{\left(\mu_n l_n\right)^{\frac{1}{2}}} < \sum \mu_j - \lambda$$

from which the bound on λfollows.

We can actually drop more such lines at the same time, since the effect of taking out lines with negative α is to reduce load on other lines, but we cannot assume that the calculation need not be repeated for the reduced set, since we have no assurance that more α will not go negative with this reduced load.

Once we enter a regime in which the critical points are indeed the minima, we can compute the minimal *W* from

$$d_i = \frac{l_i}{\lambda}\left(\left(\frac{\mu_i}{l_i}\right)^{\frac{1}{2}} \frac{\sum \left(\mu_j l_j\right)^{\frac{1}{2}}}{\sum \mu_j - \lambda} - 1\right)$$

$$W = \sum d_i$$

4. Unequal Service Rates on a Given Path

In the previous section, we did not use the fact that the number of hops was an integer, just that it was nonnegative. Thus we may perform the same analysis with fractional numbers of hops. This allows us to make a conservative correction for paths consisting of links of different rates of service. We certainly cannot use a rate any higher than the rate of the slowest link on the path, for once we hit the knee on that link the entire path will block. However, estimating all links at that lowest capacity gives unduly pessimistic estimates of the response of the path. Let $\mu_{i,j}$ be the service rates of the l_i links on path *i*, reordered so that $\mu_{i,1} \le \mu_{i,j}$, $j = 2, ..., l_i$. Then define

$$l_i^{'} = 1 + \mu_{i,1} \sum_{j=2}^{l_i} \frac{1}{\mu_{i,j}}$$

as the pseudo-hop count to use of links all of rate $\mu_{i,1}$. The difference between the delay estimated on this path with the pseudo-hop count and the real delay is

$$\frac{1 + \mu_{i,1} \sum_{j=2}^{l_i} \frac{1}{\mu_{i,j}}}{\mu_{i,1} - \alpha_i \lambda} - \sum_{j=1}^{l_i} \frac{1}{\mu_{i,j} - \alpha_i \lambda} = \sum_{j=1}^{l_i} \frac{\left(1 - \frac{\mu_{i,1}}{\mu_{i,j}}\right)\alpha_i \lambda}{\left(\mu_{i,1} - \alpha_i \lambda\right)\left(\mu_{i,j} - \alpha_i \lambda\right)} \ge 0$$

with equality at $\alpha_i \lambda = 0$ and for equal service rates.

5. Instability in the General Case

We could extend this example to chains of M/G/1 queues, or even to more general models of the node-disjoint paths, but qualitatively we expect the same basic behavior. If we do our route planning for a light load case, we will tend to favor the "shortest" paths. If the load then forces those paths onto their delay curve knees, that routing will be significantly worse than a route plan which off-loaded some portion of the excess load onto longer paths earlier. It is tempting to

think that we can solve this problem by responding to the load change quickly enough. The calculation of an optimal bifurcated route for the node-disjoint M/M/1 cascaded path model is simple, requiring only accurate data on hop counts and service rates. The difficulty lies in gathering the data, not in using it. If we rely on multi-hop distributed reporting of effective service rates and connectivity, by the time we have it available for use, it may well be out of date. At the very least, if we must compute "optimal" routing, we should do so not for the current load, but for a load which we can reasonably expect not to exceed often until the next routing update. Accumulation of variances of loads and delays would make such estimates feasible.

Acknowledgement

We wish to thank Frances C. Bernstein, Milan Tuba and the students in G22.2263 for their helpful comments.

References

1. S. Barnett, *Introduction to Mathematical Control Theory,* Clarendon Press, Oxford University Press, Oxford (1975). 264 pp.

2. H. J. Bernstein, *Tutorial on Maximum Entropy*, Philadelphia, Pennsylvania (6 February 1986). Lecture at the Institute for Cancer Research, Fox Chase Center.

3. H. J. Bernstein, *Digital Communications,* New Paltz, New York (9 October 1986). Lecture presented in 1986 New Horizons in Physics and Engineering Lecture Series on Telecommunications at the College of New Paltz, State University of New York.

4. G. Bricogne, “Maximum Entropy and the Foundations of Direct Methods,” *Acta Crystallographica,* A40 (4), pp. 410-445 (1984).

5. V. Cerf, “The Catenet Model for Internetworking,” *DARPA/IPTO,* IEN-48 (July, 1978).

6. L. Fratta, M. Gerla, and L. Kleinrock, “The Flow Deviation Method: An Approach to Store and Forward Communication Network Design,” *Networks,* 3, pp. 97-133 (1973).

7. M. Gerla, “The Design of Store and Forward Networks for Computer Communications,” *Ph.D. Thesis,* Dept. of Computer Science, UCLA (1973).

8. E. T. Jaynes, “Information Theory and Statistical Mechanics,” *Physical Review,* 106 (4), pp. 620-630 (15 May 1957).

9. E. T. Jaynes, “Information Theory and Statistical Mechanics. II,” *Physical Review,* 108 (2), pp. 171-190 (15 October 1957).

10. L. Kleinrock, “Computer Networks” in *Computer Science,* ed. A. F. Cardenas, L. Presser & M. A. Marin, pp. 241-284, Wiley-Interscience, New York (1972).

11. M. Schwartz, *Telecommunications Networks,* Addison-Wesley Publishing Company, Reading, Massachusetts (1987).

12. C. E. Shannon, “A Mathematical Theory of Communication,” *The Bell System Technical Journal,* XXVII (3), pp. 379-423, 623-656 (July 1948).