Mathematica Balkanica

Mathematical Society of South-Eastern Europe
A quarterly published by
the Bulgarian Academy of Sciences – National Committee for Mathematics

---

Mathematica
Balkanica

New Series Vol.10, 1996, Fasc.4

A Comparison of Dynamic and Static Routing in Computer Networks

Milan Tuba

Presented by P. Kenderov

This paper gives some theoretical insight for unexpected delays in the Internet, after highly dynamic routing was introduced. By complete mathematical analysis of a simple network, we show that optimal dynamic routing, for most cases, does not offer a significant improvement over optimal static routing. That minor theoretical gain can easily be lost, and situation can actually become worse, if there is even a small error in the dynamic routing tables.

1.Introduction

Highly dynamic optimal routing has been used in the INTERNET [2], [3], [4], [5], [6]. Expectations that it will give much better results were not completely fulfilled, because unexpected delays occurred often. This paper is an attempt to give some theoretical explanation for such behavior.

We will do a complete mathematical analysis of a simple network, as proposed in [1]. It will show that dynamic routing offers an improvement over static routing, that is smaller than expected.

2.Model

Let us examine a simple three node network and even simpler offered load. Nodes are A, B and C, and all traffic is from A to B. There are two possible different routes: a direct path from A to B, and an indirect path of the length two that goes through C. Let us assume that all three lines are of the same capacity μ bits per second. Our routing problem is then reduced to making a decision about what fraction α of the total offered traffic will be sent along the indirect path of the length two. The remaining fraction $1 - \alpha$ of the total load will be send along the direct path. Let us call α a branching coefficient. Let offered load be λ bits per second and ρ will, as usually, denote utilization λ/μ. We will also assume a Poison input stream of messages, and an exponential service time on lines.

3.Limitations for Parameters

There are some limitations for the parameters that we introduced. Parameter α is a fraction (probability) so we certainly have $0 \leq \alpha \leq 1$. For this particular case, we have an even stronger condition. We may have to send some traffic along the longer route, which is more expensive, has longer wait time etc., only if the direct path is overloaded (whatever the definition of the "overload" is). It is obvious, however, that it never pays off to send more traffic along the indirect route than along the direct route. If the lines were of different capacities, costs, reliabilities etc., this would not have to be the case, but according to our assumptions, we get the limitation that reasonable interval for α is $0 \leq \alpha \leq 0.5$ (this will formally follow from the requirement that utilization for each line must be less than 1).

There may be some additional limitations for α. If the total offered load λ is less than the line capacity μ, then there are no problems. The network, however, may withstand the total offered load of $\lambda < 2\mu$ or $\rho < 2$. The reason for this is that we have two alternative paths, each of the capacity μ. It is obvious that when the total load approaches 2μ, there is no more freedom in selecting α. It has to be equal to 0.5, or one path will become overloaded, introducing infinite delays.

The new set of limitations for α can be calculated as follows. With the total load λ, line capacity μ, utilization ρ, and the branching coefficient α, the utilizations of the direct path ρ_1, and the utilization of the indirect path ρ_2 will be:

$$\rho_1 = (1-\alpha)\rho \qquad \text{and} \qquad \rho_2 = \alpha\rho \tag{0.1}$$

In order to keep the network in a stable state (to avoid infinite queues and delays), we have to avoid overloading any of the two paths. By solving $\rho_1 < 1$ and $\rho_2 < 1$, we get an additional constraints $\alpha < 1/\rho$ and $\alpha > 1 - 1/\rho$. If we check the first constraint, we see that it is completely included in the previous constraint $\alpha < 0.5$.

Then, the final set of constraints is:

(0.2) $$\mu > 0$$

(0.3) $$0 \leq \rho < 2$$

(0.4) $$\max\left(0,\ 1 - \frac{1}{\rho}\right) \leq \alpha \leq 0.5$$

The left constraint in the last expression is different from zero for $\rho > 1$.

4.Optimal Waiting Time

The waiting time (including service time) for an M/M/1 queuing system is:

(0.5) $$W_{M/M/1}(\rho) = \frac{1}{\mu(1-\rho)}$$

$W_{M/M/1}$ is a function of λ and μ, but they are connected through ρ, and μ can be considered constant.

By using Kleinrock's Independence Assumption, the total waiting time for our network is:

(0.6) $$W(\alpha,\rho) = (1-\alpha)W_{M/M/1}(\rho_1) + 2\alpha W_{M/M/1}(\rho_2)$$

By substituting (0.1) and (0.5), we get

(0.7) $$W(\alpha,\rho) = \frac{1-\alpha}{\mu[1-(1-\alpha)\rho]} + \frac{2\alpha}{\mu[1-\alpha\rho]}$$

or

(0.8) $$W(\alpha,\rho) = \frac{3\rho\alpha^2 + (1-3\rho)\alpha + 1}{\mu[1-\alpha\rho][1-(1-\alpha)\rho]}$$

Our goal is to optimize the waiting time so we need a derivative. Parameter under our control is α. Differentiation gives:

$$\frac{dW(\alpha,\rho)}{d\alpha} = \frac{\rho^2\alpha^2 - 2\rho(2\rho-3)\alpha + 2\rho(\rho-2)+1}{\mu(1-\alpha\rho)^2[1-(1-\alpha)\rho]^2} \tag{0.9}$$

We are looking for the minimum for W. After somewhat tedious analysis, given in [7] (which includes proving that the second candidate for solution can not satisfy conditions, and examining the second derivative), we get

$$\alpha_{opt} = \frac{(2-\sqrt{2})\rho - 3 + 2\sqrt{2}}{\rho} \tag{0.10}$$

In the process, from the condition $\alpha_{opt} \geq 0$, we get an interesting limitation

$$\rho \geq 1 - \frac{\sqrt{2}}{2} \quad \text{or} \quad \rho \geq 0.2929 \tag{0.11}$$

Expression (0.10) gives minimum for the function $W(\alpha,\rho)$, but only when $\rho \geq 0.2929$. It is easy to see that for $\rho \leq 0.2929$, $\alpha = 0$ gives minimal value for W. This interesting result shows that for small utilizations, it does not make sense to start bifurcated routing. Only when utilization exceeds approximately 30% of the line capacity, use of the alternate route of the length two should be started.

Optimal (minimal) waiting time, when branching coefficient α is selected according to (0.10) is:

$$W_{opt} = \frac{(21-18\sqrt{2})\rho + 32\sqrt{2} - 45}{\mu\rho[(6\sqrt{2}-7)\rho - 12\sqrt{2} + 14]} \tag{0.12}$$

Here, and in the following discussion, case $\rho < 0.2929$ is not considered. It would be easy to formally include that trivial case, but it would make no changes in our analysis. For such low utilization there is no bifurcated routing, whether we use dynamic or static routing.

5. Uniformly Changing Offered Load

Previous section assumes that we know the offered load λ exactly, and that it does not change in time. This case is not really interesting (but we would be very happy to have it in practice). In reality, offered load is always changing in time, and that is what makes difference between static and dynamic routing, but also gives possibility for an error when calculating dynamic routing tables.

Let us consider more general and more realistic case when the offered load changes in time between the lower limit l and the upper limit h, where $0.3 \leq l < h < 2$ must be satisfied. To make calculations easier (or possible) we will assume that the load changes uniformly. That means that the value for the offered load spends equal amount of time inside any subinterval of the same size, included between l and h. Such distribution corresponds, for example, to constant-speed load shift from l to h, back and forth. This assumption that load changes uniformly between l and h is somewhat artificial, but not very far from what really happens in the network.

6.Optimal Dynamic Routing for Uniformly Changing Offered Load

We will now calculate the waiting time for optimal dynamic routing. We select our optimal branching probability α infinitely fast, and at any moment it follows precisely the changing load λ. The total waiting time will be expected value with regard to the distribution $g(\rho)$ of the changing load:

$$W_{opt\_dyn} = \int_l^h W_{opt}(\rho)\, g(\rho)\, d\rho \tag{0.13}$$

By substituting (0.12) and $g(\rho)$ for uniform distribution, we get

$$W_{opt\_dyn} = \frac{1}{h-l}\int_l^h \frac{(9\sqrt{2}-12)\rho - 17\sqrt{2}+24}{\mu(3\sqrt{2}-4)\rho(2-\rho)} d\rho \tag{0.14}$$

By solving this integral, we get the best we can hope for in the case of uniformly changing load. Optimal dynamic routing gives waiting time:

$$W_{opt\_dyn} = \frac{(24-17\sqrt{2})\, \ln\left(\frac{h}{l}\right) + \sqrt{2}\, \ln\left(\frac{2-l}{2-h}\right)}{2\mu\, (3\sqrt{2}-4)(h-l)} \tag{0.15}$$

7.Optimal Static Routing for Uniformly Changing Offered Load

Let us now examine static routing where the branching probability α will always have the same, fixed value. To find the optimal value for that fixed branching probability α, we do again differentiation and integration, but in the reverse order. Previously, we differentiated W to find optimal α for a particular ρ and then, using that optimal α, integrated over all possible values for ρ (with regard to distribution for ρ). Now, we will integrate over all possible values for ρ (assuming that α is fixed) to find average W and then differentiate that expression with respect to α to find the optimal fixed value for α, which minimizes W.

Average waiting time for a fixed α will be:

$$W_{avg} = \int_l^h W(\rho)\, g(\rho)\, d\rho \tag{0.16}$$

or, after we substitute (0.7) and $g(\rho)$ for uniform distribution

$$W_{avg} = \frac{1}{h-l}\int_l^h \left[\frac{1-\alpha}{\mu[1-(1-\alpha)\rho]} + \frac{2\alpha}{\mu(1-\alpha\rho)}\right] d\rho \tag{0.17}$$

By solving this, we get

$$W_{avg} = \frac{1}{\mu(h-l)} \ln \frac{(1-\alpha l)^2[1-(1-\alpha)l]}{(1-\alpha h)^2[1-(1-\alpha)h]} \tag{0.18}$$

Now, we differentiate this expression with regard to α:

$$\frac{dW_{avg}}{d\alpha} = \frac{3\alpha l - 2l + \alpha^2 hl - 4\alpha hl + 2hl + 1 + 3\alpha h - 2h}{\mu\,(1-h+\alpha h)\,(-1+\alpha h)\,(1-l+\alpha l)\,(-1+\alpha l)} \tag{0.19}$$

Again, after even more tedious analysis, given in [7], the solution is:

$$\alpha_{opt\_stat} = \frac{1}{2hl}\,[R + 4lh - 3l - 3h] \tag{0.20}$$

where

$$R = \sqrt{(8h^2 - 16h + 9)l^2 + 2(7-8h)hl + 9h^2} \tag{0.21}$$

By substituting this α_{opt_stat} into the function for the average waiting time (0.18), we get an expression for the optimal waiting time for static routing:

$$(0.22) \qquad W_{opt\_stat} = \frac{\ln\left(\frac{l^3[R+h(4l-5)-3l]^2[R+h(2l-1)-3l]}{h^3[R+h(4l-3)-5l]^2[R+h(2l-3)-l]}\right)}{\mu(h-l)}$$

This case represents pure static routing if the boundaries l and h are fixed and never change. In practice, we use a quasi-static routing where the boundaries l and h do change over time, but much slower than the offered load ρ. We adjust l and h, and corresponding α_{opt_stat}, but we do it once every hour or so. For shorter periods of time routing is static, while dynamic routing chases changing offered load continuously.

8.Comparison

Now, we will compare optimal dynamic routing and optimal static routing. Formula that is used to calculate improvement is

$$(0.23) \qquad \text{Improvement} = \left(\frac{W_{opt\_stat}}{W_{opt\_dyn}} - 1\right) * 100\%$$

The following Table 0.1 shows improvement in percents (reduction of delays) when optimal static routing is replaced by optimal dynamic routing, for different intervals $[l, h]$, where offered load ρ is uniformly changing.

l,h	0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9
0.3	0.6	1.3	1.8	2.2	2.5	2.7	2.8	2.9
0.5		0.2	0.6	1.0	1.3	1.7	2.0	2.3
0.7			0.1	0.4	0.7	1.1	1.5	2.0
0.9				0.1	0.3	0.6	1.0	1.6
1.1					0.1	0.3	0.7	1.3
1.3						0.1	0.4	1.1
1.5							0.1	0.7
1.7								0.4

Table 0.1: Dynamic vs. Static routing, improvement in percents

Rows in the table give corresponding improvement for particular l, columns for h. Since $l < h$, only the upper right triangle of the table is used, diagonal excluded. First impression is surprisingly small improvement that dynamic routing introduces. It allows us to make claim that too zealous optimization is harmful. Even without any errors in calculating routing tables, best improvement we can hope for, the upper limit, is given in Table 0.1. Average improvement is about 1%, maximal improvement is less than 3%. It is not surprising that maximal improvement is achieved when interval $[l, h]$ is wide. Traffic then varies a lot, and if we can follow that wide variations, improvement will be more significant.

When we look at this modest improvement, we should keep in mind that we are dealing with a very simple model with only three nodes and one source. In a larger network, it is possible that improvement would be better, but chances for an error in the dynamic routing tables would also be better. The combined effect would probably be the same.

The conclusion is that optimal dynamic routing gives modest improvement over optimal static routing. That small improvement can easily be annihilated, and actually dynamic routing can give larger delays than static, if there are any errors in the dynamic routing tables. Such errors always exist, because it takes significant time to calculate new routing tables, both to accumulate data in any node, and to exchange data among nodes. By the time the calculation is finished, load may be sufficiently different to make the tables obsolete, and the routing far from optimal.

References

[1] H. Bernstein. Some Comments on Highly Dynamic Network Routing, *New York University Technical Report*,**371**, 1988, 11.

[2] R. Blokzijl. RIPE Terms of Reference, *RFC*,**1181**, 1990.

[3] H-W. Brown. Models of Policy Based Routing, *RFC*,**1104**, 1989.

[4] D. Clark. Policy Routing in Internet Protocols, *RFC*, **1102**, 1989.

[5] C. Hedrick. Routing Information Protocol, *RFC*, **1058**, 1988.

[6] D. L. Mills. Exterior Gateway Protocol formal specification, *RFC*,**904**, 1984.

[7] M. Tuba. New Approach to Topology and Routing Problems in Network Design, *Ph. D. thesis*, University of Belgrade, 1993, 138 (in Serbian)

The basic ideas for this paper originated while author was at Courant Institute of Mathematical Sciences, New York University, New York. The paper was founded by Ministry of Science, Republic of Serbia. Author can be contacted at:

Milan Tuba
Matematički fakultet
Studentski trg 16
11000 Beograd
YUGOSLAVIA
e-mail: tuba@eunet.yu

Received 30.03.1994