Routing in communication networks - Adaptive Random Routing *

G.P. Phillips and A.E. Krzesinski
Department of Computer Science
University of Stellenbosch
South Africa
E-mail: phil@cs.sun.ac.za

Abstract

Modern communication switching equipment is provided with programmable control units offering opportunities for traffic regulation. Traffic regulation consists of mechanisms to prevent congestion of the network and to route customers efficiently through the network. This study investigates the mechanism of *routing* in communication networks with particular emphasis on networks operating under *non-stationary* environments.

Communication networks have typically been modelled as stochastic processes. However, under non-stationary conditions these models are mathematically complex and analytical solutions have only been found for a small number of single server queueing systems.

In our approach to the problem, we construct a deterministic equation that approximates the stochastic model of a single queue. The model necessarily compromises between accuracy and speed in computation. The model is used to investigate an efficient routing mechanism at a single switch in the network. The efficiency of this routing mechanism is measured in terms of a *performance function* which in this analysis is the total holding time of all customers. The condition that minimizes this performance function is found and the Adaptive Random Routing algorithm is proposed.

*This work was funded by the Telkom Teletraffic Application Development Programme.

1 Introduction

The most comprehensive control strategy would place an entire communication network under the control of a centralised controller that uses all available information from the network. This is not possible in large communication networks for technical and administrative reasons. From a technical viewpoint there two major problems. The first is that a considerable delay in the communication of the *network state* would prevent the controller from reacting quickly. The second and most important is that the size of the network state vector, which forms the input to the controller, does not allow a tractable solution. From an administrative viewpoint, large networks cross country and organisation boundaries in which control of resources by network providers is demanded. Therefore even if centralised control were technically possible, a distributed solution would be sought.

The distributed control system consists of several controllers governing independent sets of switching centres. Each controller uses network state information to generate *control signals* that control the switching centres under its jurisdiction.

The control mechanism could include the assignment of traffic to different routes and the regulation of access to external customers. The goal of the distributed controller is to control the switching centres under its jurisdiction in such a way as to minimize a given *performance function*. The design of the controller therefore involves an optimizing problem. The solution for this problem is dependent on the explicit form of function to be minimized, which must be specified before any control algorithm can be devised. The performance function in this analysis is the total holding time of all customers in the controlled subnetwork.

The controller's output can be modified only at discrete intervals because any control algorithm that commences with a calculation will take finite time to complete. The more frequent the update cycle of control signal the more effectively the controller will adapt to changing conditions.

This paper follows the work of Filipiak [1] in which the problem of routing is approached by considering a distributed control system where each switch in the network is controlled independently. A local description of the state of the subnetwork is constructed using differential equations and an expression for the routing mechanism of this subsystem is developed.

2 Modelling queueing systems

2.1 Stochastic model of the M/M/1/∞ queue

The behaviour of a queueing system can be studied by regarding the evolution of the number of customers in queue as a stochastic process. The stochastic model of the M/M/1/∞ queue is

$$\begin{aligned}\frac{dp_0(t)}{dt} &= -\lambda(t)p_0(t) + \mu(t)p_1(t) \\ \frac{dp_n(t)}{dt} &= -(\lambda(t) + \mu(t))p_n(t) + \lambda(t)p_{n-1}(t) + \mu(t)p_{n+1}(t)\end{aligned}$$

for $n > 0$ where $p_n(t)$ is the probability that the number of customers in the queue is n at time t and $\sum_{n=0}^{\infty} p_n(t) = 1$. The average queue length at time t is defined as

$$Q(t) = \sum_{n=1}^{\infty} np_n(t) \tag{1}$$

2.2 Approximate model of the M/M/1/∞ queue

This section presents an approximate model of the M/M/1/∞ queue. This model, called the *dynamic flow model*, was proposed by Filipiak [1] and the interested reader is referred to this work for a derivation of the model. The *dynamic flow model* of the M/M/1/∞ queue is given by the equation

$$\frac{dx(t)}{dt} = -\mu G(x(t)) + \lambda(t) \tag{2}$$

where

$$G(x) = \frac{x}{x+1} \tag{3}$$

$x(t)$ is the queue length at time t and μ is constant.

The dynamic flow model can be illustrated by comparing the queue lengths as computed by the queueing model (equation (1)) and the dynamic flow model (equation (2)) where both models are subject to a step input. Figure 1 shows the evolution of queue length with $\mu = 1.0$ and $\lambda(t) = 0.7\ u(t)$ where $u(t)$ is the unit step function, defined by

$$u(t) = \begin{cases} 0 & \text{for} \quad t < 0 \\ 1 & \text{for} \quad t \geq 0 \end{cases}$$

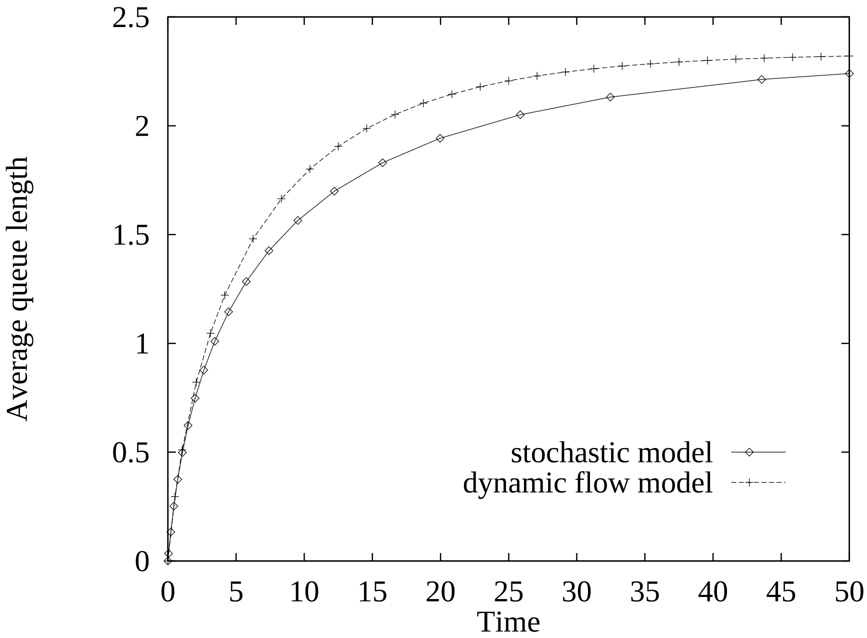


Figure 1: Evolution of queue length for the M/M/1/∞ queue

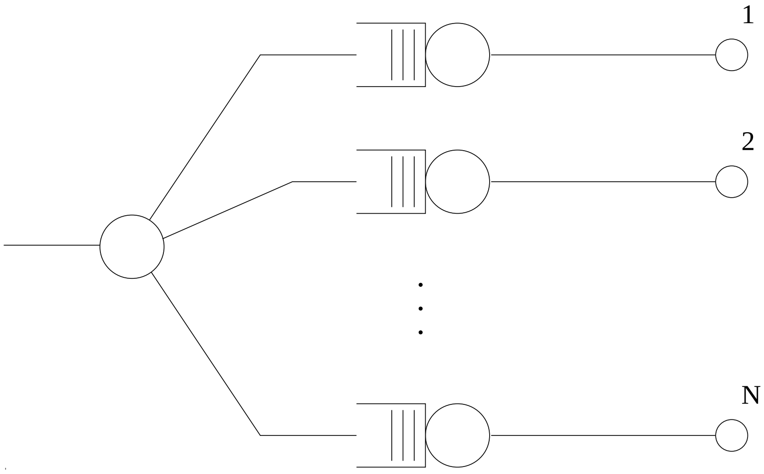
2.2.1 Modelling errors

The solution for the dynamic flow model (2) yields the queue length $x(t)$ which approximates the queue length $Q(t)$ given by the stochastic model in equation (1). Let the *dynamic flow error* be defined as

$$\epsilon = \int_0^{\infty} |Q(t) - x(t)| dt$$

If $\epsilon > 0$ the dynamic flow model is said to possess a *modelling error*. It can be shown that under steady state conditions ($t \to \infty$) the dynamic flow model yields the same result as the stochastic model. This implies that the modelling error is zero under steady state conditions. The modelling error increases as the system deviates from the steady state. The state of the system is dependent on the independent variable $\lambda(t)$ and therefore the modelling error increases as the time derivative of $\lambda(t)$ increases.

3 Adaptive Random Routing



Consider a set of N parallel M/M/1/∞ queues. Let customers that arrive to the system be forwarded with probability $\alpha_n(t)$ to node n. Let $\lambda_n(t)$ be the time dependent rate at which customers are forwarded to node n. Thus

$$\lambda_n(t) = \alpha_n(t)\lambda(t) \tag{4}$$

and

$$\sum_{n=1}^{N} \alpha_n(t) = 1 \tag{5}$$

If the arrival process is Poisson and the routing is performed in a random manner then the input process to each parallel queue is Poisson. Using the dynamic flow model (2) the system is described by a set of differential equations

$$\frac{dx_n(t)}{dt} = -\mu_n G(x_n(t)) + \alpha_n(t)\lambda(t) \tag{6}$$

where $n = 1, \ldots, N$

We wish to route the customers such that the total holding time of all customers in the system during time $(0, T)$ is minimized. The total holding time is given by

$$\tau \quad = \quad \sum_{n=1}^{N} \int_0^T x_n(t)dt \tag{7}$$

The routing problem can now be stated as follows: find the routing probabilities $\vec{\alpha}(t) = (\alpha_1(t), \ldots, \alpha_N(t))$ that will minimize τ under the constraint (6). We shall refer to the routing strategy that minimizes τ as the *optimal routing* strategy.

As a summary of the rest of the chapter: The Pontryagin minimum principle is used to find the conditions for optimal routing given that customers are routed to K queues (where $K \leq N$). The condition under which routing should not take place to certain queues and an equation for calculating routing probabilities are given. Finally the Adaptive Random Routing algorithm is presented.

3.1 Summary of Pontryagin minimum principle

Consider a controllable system modelled by a system of non-linear differential equations, with independent variables $\vec{x}(t) = (x_1(t), \ldots, x_m(t))$. Let $\vec{x}_i$ be the initial state of the system at t_0, $\vec{x}_f$ the final state at time t_f and $\vec{u}(t) = (u_1(t), \ldots, u_r(t))$ the control vector. Define a *performance function* $J(x(t), u(t), t)$ which is continuous over all possible trajectories of system state from $\vec{x}_i$ to $\vec{x}_f$.

The *Pontryagin minimum principle* [2] provides the optimal control vector $\vec{u}^*(t)$ that moves the system state from $\vec{x}_i$ to $\vec{x}_f$ and minimizes the performance function $J(x(t), u(t), t)$. The Pontryagin minimum principle is stated as follows. Let the set of non-linear differential equations

$$
\begin{aligned}
\frac{dx_1(t)}{dt} &= f_1(x_1(t), \ldots, x_m(t), u_1(t), \ldots, u_r(t), t) \\
\frac{dx_2(t)}{dt} &= f_2(x_1(t), \ldots, x_m(t), u_1(t), \ldots, u_r(t), t) \\
&\vdots \\
\frac{dx_m(t)}{dt} &= f_m(x_1(t), \ldots, x_m(t), u_1(t), \ldots, u_r(t), t)
\end{aligned}
$$

model a physical system and

$$J(\vec{x}(t), \vec{u}_1(t), t) = \int_{t_i}^{t_f} f_0(\vec{x}(t), \vec{u}(t), t)dt$$

be the performance function. Define the Hamiltonian as

$$H(\vec{p}(t), \vec{x}(t), \vec{u}(t), t) = f_0(\vec{x}(t), \vec{u}(t), t) + \sum_{i=1}^{m} p_i(t) f_i(\vec{x}(t), \vec{u}(t), t) \tag{8}$$

where $p_k(t)$ satisfies the equation

$$\frac{dp_k}{dt} = -\frac{\partial H(\vec{p}(t), \vec{x}(t), \vec{u}(t), t)}{\partial x_k} \tag{9}$$

for $k = 1, \ldots, m$. The Pontryagin minimum principle states that for the system to be optimally controlled the Hamiltonian must be minimized over all allowable control vectors $\vec{u}(t)$.

3.2 Condition for optimal routing

Applying the Pontryagin minimum principle to the routing system above we have from equation (7)

$$f_0(\vec{x}(t), \vec{\alpha}(t), t) = \sum_{n=1}^{N} x_n(t)dt \tag{10}$$

while from equation (6)

$$f_n(\vec{x}(t), \vec{\alpha}(t), t) = -\mu_n G(x_n(t)) + \alpha_n(t)\lambda(t)$$

for $n > 0$ where

$$\begin{aligned} \vec{x}(t) &= (x_1(t), \ldots, x_N(t)) \\ \vec{\alpha}(t) &= (\alpha_1(t), \ldots, \alpha_N(t)) \end{aligned}$$

Therefore the Hamiltonian is

$$H(\vec{p}(t), \vec{x}(t), \vec{\lambda}(t), t) = \sum_{n=1}^{N} x_n(t) + \sum_{n=1}^{N} p_n(t) \left(-\mu_n G(x_n(t)) + \alpha_n(t)\lambda(t)\right) \tag{11}$$

which is minimized over all $\vec{\alpha}(t)$ subject to the constraint (5). The only term in the Hamiltonian that depends on $\alpha_n(t)$ is

$$\sum_{n=1}^{N} p_n(t)\alpha_n(t)\lambda(t) \tag{12}$$

Therefore we need only minimize equation (12) subject to constraint (5). Let $\widehat{p}(t)$ be the minimum of $p_n(t)$, that is

$$\widehat{p}(t) = \min_{1 \le n \le N} p_n(t)$$

To minimize equation (12), the smallest $p_n(t)$ must be maximally weighted. Therefore

$$\alpha_n(t) > 0 \tag{13}$$

only if $p_n(t) = \widehat{p}(t)$. If $\widehat{p}(t)$ is attained by only one variable $p_n(t)$ then $\alpha_n(t) = 1$. On the other hand if $\widehat{p}(t)$ is attained by more that one variable $p_n(t)$ then control variables $\alpha_n(t)$ are not specified by minimizing the Hamiltonian.

If during some time interval customers are forwarded to K queues, that is,

$$\alpha_n(t) > 0$$

for $n = 1, \ldots, K$ and

$$\alpha_{K+1}(t) = \ldots = \alpha_N(t) = 0 \tag{14}$$

and queues are arranged in increasing order of $p_n(t)$ then from (13)

$$p_1(t) = p_2(t) = \ldots = p_K(t) = \widehat{p}(t) \tag{15}$$

Applying equation (9) to equation (11) and substituting into (15) we obtain

$$\mu_1 \frac{dG(x_1)}{dx_1} = \mu_2 \frac{dG(x_2)}{dx_2} = \ldots = \mu_K \frac{dG(x_K)}{dx_K} \tag{16}$$

This relation which is necessary for minimizing the performance function given by equation (10), is called the *property of optimal flow*. This condition gives the optimal routing strategy when forwarding customers to K queues.

3.3 Switching conditions

The routing system described above assumes that customers can reach their destinations from any node $1, \ldots, N$. Therefore routing to several nodes can be terminated without blocking the customer from its required destination. Equation (15) suggests that customers should not be forwarded to queues $K+1, \ldots, N$. This section establishes the condition for which certain routes should be closed. Note that the termination of routes is not permanent but dependent on the density of incoming traffic.

Let us define $\delta_i(x_i)$ as

$$\delta_i(x_i) = \mu_i \frac{dG(x_i)}{dx_i}$$

for $i = 1, \ldots, N$. It can be shown that $dG(x)/dx$ decreases as x increases, for $G(\cdot)$ given by equation (3). Therefore if the queues are numbered in order of decreasing capacity, that is,

$$\mu_1 \geq \mu_2 \geq \ldots \geq \mu_N$$

then from equation (16)

$$x_1(t) \geq x_2(t) \geq \ldots \geq x_N(t)$$

Let the system start in operation from the empty state $x_i(0) = 0$ for $i = 1, \ldots, N$, with all initial flow directed at queue 1. As x_1 increases $dG(x_1)/dx_1$ decreases. At some point $\delta_1(x_1)$ will reach $\delta_2(0)$. According to the *property of optimal flow*, routing commences to queue 2. As x_1 and x_2 increase $\delta_1(x_1)$ and $\delta_2(x_2)$ decrease together until $\delta_1(x_1) = \delta_3(x_3)$ at which point routing commences to queue 3. This process is reversed when the total inflow begins to decrease, that is, routing to queue i is terminated when $\delta_i(x_i)$ first ceases to be equal to $\delta_{i-1}(x_{i-1})$.

We shall call the point at which routing commences (or terminates) to queue i, the *switching condition* of the queue.

A closed form of this switching condition can be found if $G(\cdot)$ fulfils the following property.

PROPERTY 1 *This property is said to hold for a function $G(x)$ if there exists a function $V(x)$ such that $V(x)$ is the inverse of $dG(x)/dx$, that is, $x = V(dG(x)/dx)$ and*

$$V\left(\alpha \frac{dG(x)}{dx}\right) = \gamma_1(\alpha) + \gamma_2(\alpha)x \tag{17}$$

for any real number α.

It can be shown that this property holds for $G(\cdot)$ in the dynamic flow model of the M/M/1/∞ queue.

Equation (16) can be rewritten as

$$\mu_i \frac{dG(x_i)}{dx_i} = \mu_{i+1} \frac{dG(x_{i+1})}{dx_{i+1}} \tag{18}$$

for $i = 1, \ldots, K-1$. By applying equation (17) to both sides of equation (18), we obtain

$$x_{i+1} = \frac{\gamma_2(\mu_i)}{\gamma_2(\mu_{i+1})} x_i + \frac{\gamma_1(\mu_i) - \gamma_1(\mu_{i+1})}{\gamma_2(\mu_{i+1})} \tag{19}$$

The switching condition for queue $i+1$ is $x_i(t) = \widehat{x}_i$ where

$$\mu_i \frac{dG(x_i)}{dx_i}\bigg|_{\widehat{x}_i} = \mu_{i+1} \frac{dG(x_{i+1})}{dx_{i+1}}\bigg|_0$$

Substitution of $x_{i+1} = 0$ into equation (19) yields

$$\widehat{x}_i = \frac{\gamma_1(\mu_{i+1}) - \gamma_1(\mu_i)}{\gamma_2(\mu_i)} \tag{20}$$

In the case when queues have equal capacities, routing must commence (or terminate) at the same time and each of these queues is to receive an equal share of the traffic.

3.4 Calculation of routing probabilities

This section investigates a method for calculating the routing probabilities to K queues. Differentiating equation (16) we obtain

$$\mu_i \frac{d^2G(x_i)}{dx_i^2} \frac{dx_i}{dt} = \mu_{i+1} \frac{d^2G(x_{i+1})}{dx_{i+1}^2} \frac{dx_{i+1}}{dt} \tag{21}$$

for $i = 1, \ldots, K-1$.

Substituting equation (6) into equation (21) yields

$$\begin{aligned} -\mu_i^2 \frac{d^2G(x_i)}{dx_i^2} G(x_i(t)) + \mu_i \frac{d^2G(x_i)}{dx_i^2} \alpha_i(t)\lambda(t) = \\ -\mu_{i+1}^2 \frac{d^2G(x_{i+1})}{dx_{i+1}^2} G(x_{i+1}(t)) + \mu_{i+1} \frac{d^2G(x_{i+1})}{dx_{i+1}^2} \alpha_{i+1}(t)\lambda(t) \end{aligned} \tag{22}$$

The K unknowns $\alpha_i(t)$ can be solved using equations (22), (14) and (5).

Equation (22) is integrated over an interval delimited by the points at which the routing probabilities $\alpha_i(t)$ are updated. The routing probabilities remain constant in this interval therefore $\alpha_i(t)$ can be replaced by α_i.

If we assume that equation (22) can be written as

$$\begin{aligned} -\mu_i^2 \frac{d^2G(x_i)}{dx_i^2} G(x_i) + \mu_i \frac{d^2G(x_i)}{dx_i^2} \alpha_i\lambda = \\ -\mu_{i+1}^2 \frac{d^2G(x_{i+1})}{dx_{i+1}^2} G(x_{i+1}) + \mu_{i+1} \frac{d^2G(x_{i+1})}{dx_{i+1}^2} \alpha_{i+1}\lambda \end{aligned} \tag{23}$$

then the values of α_i can be calculated explicitly. The solution of equation (23) has the form

$$\alpha_i = \frac{1}{\lambda}\left(\frac{\gamma_2(\mu_1)}{\gamma_2(\mu_i)}\frac{\lambda + B_K}{A_K} - \sum_{j=1}^{i-1}\frac{\gamma_2(\mu_j)}{\gamma_2(\mu_i)}\mu_j G(\widehat{x}_j)\right) \tag{24}$$

where

$$\begin{aligned} A_K &= \sum_{i=1}^{K}\frac{\gamma_2(\mu_1)}{\gamma_2(\mu_i)} \\ B_K &= \sum_{i=1}^{K}\sum_{j=1}^{i-1}\frac{\gamma_2(\mu_j)}{\gamma_2(\mu_i)}\mu_j G(\widehat{x}_j) \end{aligned}$$

3.5 Adaptive Random Routing algorithm

The Adaptive Random Routing algorithm is presented below. This algorithm calculates routing *probabilities* while an independent mechanism is required to route individual customers to different nodes in a random way so as to satisfy the routing probabilities.

```
Adaptive_random_routing_algorithm
begin
    Calculate x̂_i using equation (20);
    Assign all traffic to queue 1, (α_1 = 1, α_2 = ... = α_N = 0);
    Repeat forever
    begin
        Integrate equations (6);
        If x_i(t) = x̂_i then
        begin
            If x_i(t) is increasing then
                commence routing to queue i + 1
            else
                terminate routing to queue i + 1, (α_{i+1} = 0);
            Calculate routing probabilities using equation (24);
        end
    end
end
```

3.6 Conclusion

This paper presents the Adaptive Random Routing algorithm proposed by Filipiak [1]. The system consists of N parallel M/M/1/∞ queues. Customers arriving to the system are forwarded with probability $\alpha_n(t)$ to node n. Control of the routing is achieved using routing probabilities as control signals for switching equipment. The algorithm periodically calculates routing probabilities in such a way as to minimize the total holding time of all customers in the subnetwork. The switching equipment is responsible for routing individual customers in a random manner such that the routing probabilities are satisfied.

Several assumptions are made by the Adaptive Random Routing algorithm:

- Communication channels are modelled by M/M/1/∞ queues with constant service rates.
- The arrival process to the system is Poisson.
- Customers may be routed to any of N nodes.

The Adaptive Random Routing algorithm has two disadvantages. The first is that the waiting line buffers are assumed infinite. A further study is required to determine the effects of limited buffer size in the communication channels. The second disadvantage is that a practical routing mechanism cannot assume that all customers can reach their destination nodes via any of N intermediate nodes since a circular route may develop.

The distributed nature of the Adaptive Random Routing algorithm together with its undemanding requirement for network state information makes it suitable for large networks where centralized control is impractical.

References

[1] Janusz Filipiak. *Modelling and Control of Dynamic Flows in Communication Networks.* Springer-Verlag, 1988. ISBN 3-540-18292-6.

[2] Donald E Kirk. *Optimal Control Theory.* Prentice-Hall, Inc., 1970.