Dynamic Alternative Routing - Modelling and Behaviour

R.J. Gibbens [*] [†] F.P. Kelly [†] P.B. Key [‡]

Abstract

DAR, dynamic alternative routing, is a simple but effective form of dynamic routing which is decentralised and which uses only a small amount of local information. This paper examines the modelling and behaviourial aspects of such a scheme. In particular various bounds are presented which hold for any kind of dynamic routing scheme, and can be used to assess the performance of DAR. DAR can be investigated by a simple analytical model, of familar fixed-point form, and simulation studies have verified its accuracy. A simple form of dimensioning is mentioned, and methods of setting trunk reservation parameters are discussed. These are necessary both to improve performance and prevent instability. Lastly, simple ways of extending DAR are examined.

1 Introduction

DAR is a simple but effective dynamic routing strategy, which is decentralised and only uses local information. In particular, the only information required is whether trunk reservation thresholds have been exceeded on a route, and the current recommended alternative route. The information can be localised even further by limiting knowledge to outgoing links from an exchange rather than a route, and thus the scheme uses only as much information as AAR, with the additional stored information of the current best alternative.

Thus DAR stands in marked contrast to the scheme of Bell-Northern, and AT&T's DNHR. The former is centralised, time-delayed and requires detailed information about circuit occupancies and traffic arrivals, whereas the latter uses a large off-line calculation to advise on choices of alternative routes which can change hourly, coupled with a dynamic part similar in spirit to the scheme of Bell-Northern.

This paper starts by obtaining bounds which hold for any dynamic routing scheme, and the performance of DAR is compared with such bounds. A simple analytical model is then developed which enables DAR to be modelled on both large and small networks. Empirical validation of the model and a number of examples are discussed.

Any dynamic routing strategy has implications for dimensioning, and a simple way of introducing flexibility into a network is given. In addition, the setting of trunk reservation parameters is discussed, such controls being necessary to achieve high performance and prevent instability. Lastly ways of extending DAR are mentioned.

---

[*] **Supported in part by SERC grant GR/E 27089.**

[†] **Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, England (telephone: +44 223 337958).**

[‡] **British Telecom, Performance Engineering Division, Room 51 MLB5, British Telecom Research Laboratories, Martlesham Heath, Suffolk IP5 7RE, England (telephone: +44 473 645820, telex: 98376(THQRD G)).**

2 Bounds on Performance

Consider a fully connected network of J nodes, with the link between nodes i and j having capacity C_{ij}. Calls use a single circuit between i and j or can be rerouted via a tandem node k. A dynamic routing scheme is allowed to reject a call, choose a direct or alternative route via a tandem, and reroute calls in progress.

2.1 Max-Flow bound

If $\lambda_{ij}(t)$ denotes the offered traffic between i and j as a function of t and $n_{ij}(t)$ denotes the number of calls in progress, then for any dynamic routing scheme we have the following result (for related work see [2]).

Theorem 1

$$\mathsf{E}\left[\sum_{i<j} n_{ij}(t)\right] \leq f(\lambda(t))$$

where $f(\lambda(t))$ *is the maximum flow of the deterministic linear program*

$$\max \sum_{i<j}\left(x_{ij} + \sum_{k\neq i,j} x_{ikj}\right)$$

subject to

$$\begin{aligned} x_{ij} + \sum_{k\neq i,j} x_{ikj} &\leq \lambda_{ij}(t) & i<j \\ x_{ij} + \sum_{k\neq i,j} (x_{ijk} + x_{jik}) &\leq C_{ij} & i<j \\ x_{ij}, x_{ijk} \geq 0,\ x_{ijk} &= x_{kji} & \forall\, i,j,k\,. \end{aligned}$$

This theorem gives an upper bound on the carried traffic for any dynamic routing scheme, and holds irrespective of the call arrival and holding time processes.

2.2 Erlang Bounds

A second type of bound is obtained if the arrival process is assumed to be Poisson, and call durations are arbitrarily distributed with unit mean. Let

$$B = \mathsf{E}\left[\sum_{i<j} n_{ij}(t)\right] / \sum_{i<j} \lambda_{ij}\,.$$

Thus B is the overall grade of service. Now let $E(\lambda, C)$ denote Erlang's B-formula.

Theorem 2 $B \geq g(\lambda)$ *where* $g(\lambda)$ *is the minimum of the linear program*

$$\min \frac{\sum_{i<j} \lambda_{ij} b_{ij}}{\sum_{i<j} \lambda_{ij}}$$

subject to

$$\begin{aligned} \frac{\sum_{i\in S, j\notin S} \lambda_{ij} b_{ij}}{\sum_{i\in S, j\notin S} \lambda_{ij}} &\geq E\left[\sum_{i\in S, j\notin S} \lambda_{ij}, \sum_{i\in S, j\notin S} C_{ij}\right] & \forall S \subset \{1, 2, \ldots, J\} \\ b_{ij} = b_{ji} &\geq 0 \quad i<j\,. \end{aligned}$$

This provides another bound for any dynamic routing scheme with Poisson arrivals. Such bounds, apart from their theoretical importance, enable the performance of schemes such as that described later in Section 4 to be assessed. Both types of bound are readily calculated for small networks.

3 Trunk Reservation

It is well known that individual and social optima rarely coincide. In the context of dynamic routing schemes this means that if calls are allowed too free an access to alternative routes the performance can degrade. Trunk reservation provides an effective social control in such circumstances.

To investigate trunk reservation a fixed-point model has been developed. This illustrates the important role of trunk reservation for controlling instability of dynamic routing schemes and for limiting the extent of rerouting under overloads.

3.1 Optimality

Consider a link of C circuits offered two streams of Poisson traffic λ_1 which is worth 1, and λ_2 which is worth c, and is subject to a trunk reservation parameter r. If $B_1(r)$ and $B_2(r)$ denote the blocking probabilities of the two streams, then a criterion for the choice of r is to minimise the rate of lost revenue, $\lambda_1 B_1(r) + c\lambda_2 B_2(r)$. If now λ_1 corresponds to the single link traffic in a network and λ_2 the overflow traffic carried over two-link paths then putting $c = 1/2$ corresponds to valuing two-link calls as worth half that of direct traffic. Extensions to calls carried on more than two links are possible (for further discussion and applications see [6,7]). Examples are given in Figure 1: for these examples λ_1 was chosen so that $E(\lambda_1, C) = 0 \cdot 01$.

3.2 Secondary Criteria

Overflow traffic can be unpredictable, and we might want to guarantee performance in a worst case scenario. To this end if B denotes the blocking of λ_1 when $\lambda_2 = 0$ (this blocking will then be $E(\lambda_1, C)$), and $B(\lambda_1, C, r) = B_1(r)$ when λ_2 is allowed to increase without bound, a possible choice for the trunk reservation parameter is

$$R(C) = \min\{r : B(\lambda_1, C, r) \leq KB\}$$

for some fixed constant K, illustrated in Figure 2. To construct the Figure the arrival rate λ_1 was chosen so that $B = E(\lambda_1, C) = 0 \cdot 01$.

Figure 1: Optimality criterion

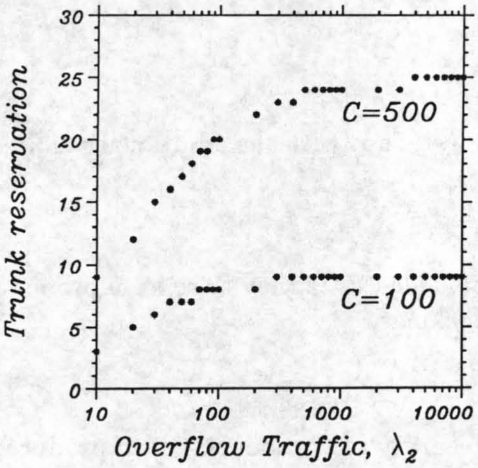
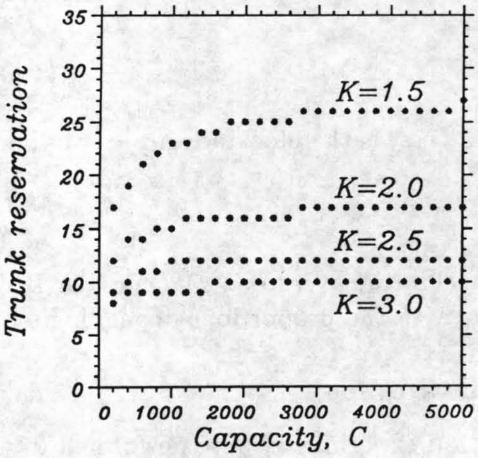


Figure 2: Secondary criterion



3.3 Remarks

The above provide simple and effective ways of setting trunk reservation parameters. The former method tends to give larger values. In general the overall performance of schemes (such as that to be described in the next Section) is relatively insensitive to the trunk reservation parameter provided that it is not too small, and moreover the above methods seem a good way of deriving suitable values.

4 The DAR Scheme

The DAR scheme operates as follows [4]. Suppose there are n nodes in the network with link (i, j) having a capacity C_{ij}. Each link is assigned a trunk reservation parameter and each source destination pair stores the identity of its current tandem k for use in two-link alternative routes. Fresh offered traffic between nodes i and j is first offered to the direct link and is always routed along that link if there is a free circuit. Otherwise, the call attempts the two-link alternative route via tandem node k with trunk reservation applied to both links. If the call fails to be routed via k, this call is lost and, further, the identity of the tandem node is reselected (at random perhaps) from the set $\{ 1, ..., n \} \setminus \{ i, j \}$. Note especially that the tandem node k is not reselected if the call is successfully routed on either the direct link or the two-link alternative path.

The simplicity of the DAR scheme allows it to be readily modelled and easily modified to cope with particular circumstances. We illustrate these points in the next two Sections.

5 Modelling DAR

An important consequence of the simplicity of the DAR scheme is the ease with which a mathematical model of its long run stationary behaviour may be constructed. For example, observe that if the reselection of a tandem node is uniform, then DAR equalizes the blocking rates over the separate two-link paths for each source destination pair. In general, a fixed-point model may be set up for the long run routing proportions and the loss probabilities in the network under DAR.

If $p_t(r, s)$ denotes the proportion of calls between r and s which are offered to tandem node t, (taken over a long time interval), and $y_t(r, s)$ is the mean length of sequences of such alternatively routed calls, where a sequence is terminated by the first call to be blocked on the tandem route, then under uniform reselection

$$y_k(r, s) : y_t(r, s) = p_k(r, s) : p_t(r, s) .$$

Under the assumption of link independence, we have

$$y_t(r, s) \propto \frac{1}{L_t(r, s)}$$

where $L_t(r, s)$ is the blocking on the two-link path $r - t - s$. Thus with the additional requirement that

$$\sum_t p_t(r, s) = 1$$

we have that the overflow streams under DAR can be modelled *as if* they arise from proportional routing, with the proportions depending on the link blockings.

5.1 Examples

Simulation experiments were performed to assess the accuracy of the model, which produced excellent agreement. More specifically, 5 node and 14 node networks were considered, and the overall grade of service and the traffic-weighted overall stream grade of service distribution compared.

5 node network: The network was simulated under normal load conditions and with one stream and two streams at 10% overload. For this case the stream grade of service of every stream from the analytical model lay within the 95% confidence limits of the simulation. In addition, the calculated proportions of calls which are blocked on the direct link that are offered to tandem nodes were plotted in 2-space with 95% confidence regions from the simulation, and again the agreement was good.

14 node network: The network was a subset of main network traffic, and the network dimensioned in an analogous way, with target grades of service for the design date traffic, 10% and 20% overload. The network was then simulated at general overloads of −5%, 0%, 5% and 10%. Again the agreement between the analytic model and the simulation was very good. It is worth noting that fixed point models such as the above are generally least accurate at low blockings; however, in these examples the analytic results and simulation agreed well even at low blockings, when the overall grade of service was 0·03%.

The approximation procedure can break down, for example, if the overflow is large and needs to be spread over a number of alternatives. The most extreme example of this is where there is no direct traffic on a link. DAR may still perform well in such circumstances, as simulation results show for the international access example of the next Section.

6 Extensions

DAR can easily be modified to cope with different situations. Some examples follow.

1. A set of tandem nodes can be used and cycled through with successive overflow calls. By this means the overflow traffic will be purposely shared across a number of tandem nodes.
2. The tandem nodes might also be selected according to some distribution other than uniform, resulting in some tandem nodes being given preference or some excluded from two-link routes. Two possible applications of this are first, in the case of link or node failure where routes would be debarred from being attempted (a network management option), and secondly to bias tandem routing towards local rerouting.
3. The trunk reservation mechanism can be extended, so that a second larger parameter is assigned for each link to signal the re-selection of the tandem node but not the rejection of the call.
4. Overflow paths do not have to be two-link, and in general networks DAR can operate by utilising the "sticky-random" principle, where a route is stuck with if it is successful, but when a call fails a new route is chosen from some feasible set. The scheme can also operate without a designated first choice route (i.e. with a "dummy" first choice route). This extension means that DAR can be used in networks which are not fully connected.

For other examples of extensions see [4,10].

An example of the use of DAR in a non-fully connected network is the model of the international access network shown in Figure 3 with nominal offered traffics:

	A	B	C
S1	50	4000	200
S2	50	4000	200

and with circuits dimensioned to a nominal 1% gos. Thus for this network, traffic to destination A goes via one ISC, whilst that to B has access to all 4 ISC's and that to C has access to 2 ISC's, and the traffic which is carried over several routes is under the control of DAR. A number of runs were performed about the base traffic and circuits, two examples are with the stream from S1 to B increased by 50%, and with a link from an ISC to B reduced by 50%.

All paths have the same length in this network, and access links to the ISC's carry a mix of traffic with differing number of alternative routes. Trunk reservation on the DMSU to ISC link can be used to protect the grade of service of links which do not have alternative routes available to them. Various routing options were tried, including

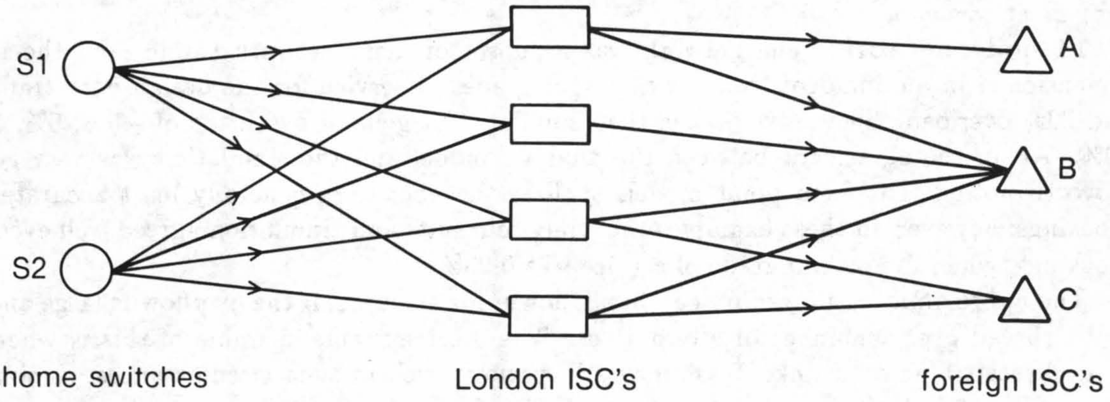


Figure 3: International access network

1. Proportional routing with the proportions chosen *optimally* for each scenario by means of a hill-climbing technique
2. DAR with a dummy first choice
3. DAR with a fixed first choice (thus routing to destination C operates as AAR)
4. DAR with a fixed first choice and trunk reservation to protect first offered traffic.

The results showed that under a variety of offered traffic conditions DAR with a fixed first choice performed as well as optimal proportional routing. In this example all routes use two links, and trunk reservation is less critical than in the example described in Section 4. If trunk reservation is used then the performance is slightly worse overall although the gos of the small streams is improved. DAR with a dummy first-choice performs slightly worse.

7 Dimensioning for Network Flexibility

Dynamic routing schemes attempt to take advantage of spare capacity that exists in networks to cope with various types of overload and failure. Such spare capacity arises from forecasting errors, changes in traffic patterns, the timing of link upgrades as well as circuit group modularity and, perhaps, even transient drops in offered loads. Thus, this spare capacity tends to be spread around the network in a fairly haphazard fashion. We outline a method, derived from the early work of [1,5], of assigning additional spare capacity so that it forms a rich collection of two-link paths. The network so constructed arises as a feasible solution for the following link overload problem.

Suppose that the network is in addition required to carry a given proportion p of the nominal offered traffic λ_{ij} for any single source destination pair. Only one stream is in overload but it could be any of the traffic streams in the network. For the particularly simple cost structure where link costs are determined just by link capacities there is a simple divide and conquer algorithm for constructing a feasible network which carries the additional traffic over at most two-link routes and where for *deterministic* streams the cost is minimal:

1. Form the maximal spanning tree T of the requirements graph (the network graph with values $p\lambda_{ij}$).
2. Decompose T into a sum of k sub-trees, $T_1, \ldots, T_k$ with uniform requirements r_i.

3. Synthesise each uniform tree by putting $r_i/(n-1)$ on each link of the complete graph through the nodes of T_i, where n is the number of nodes.

4. Superpose the resulting solutions.

With N nodes the complexity of the above algorithm is $O(N^2)$. If the solutions are quantities c_{ij} and if link capacities satisfy

$$E\left(\lambda_{ij}+c_{ij}, C_{ij}\right) \leq B$$

then the overall blocking between pairs of nodes is not more than

$$\frac{B+p(2B)}{1+p}.$$

Moreover, this algorithm involves a decomposition of the traffic matrix which may be viewed as a data analysis tool with which to investigate features of the traffic pattern relating to dimensioning.

If the costs are more complicated, this simple algorithm does not produce a minimum cost solution. Instead the problem can be formulated as a linear programming problem of size $O(N^2)$. This is of too high an order for large networks; however, other approaches can be used such as constraint generation, or a neighbourhood approach to improve upon the solution generated by the above algorithm. For a fuller discussion of dimensioning issues see [8].

Acknowledgement

Acknowledgement is made to the Director of Research and Technology of British Telecom to make use of part of the information contained in this paper and to colleagues in the Performance Engineering Division, British Telecom for simulation studies.

References

[1] Ford L R and Fulkerson D R. *Flows in Networks.* Princeton University Press, 1962.

[2] Franks R L and Rishel R W. Optimum Network Call-Carrying Capacity. *Bell System Technical Journal*, 52(7), September 1973.

[3] Gibbens R J. Some Aspects Of Dynamic Routing In Circuit-Switched Telecommunication Networks. January 1986. Rayleigh Prize Essay, University of Cambridge.

[4] Gibbens R J and Kelly F P. Telecommunications Network and Method. British patent application 8531138, 18 December 1985.

[5] Gomory R E and Hu T C. Multi-Terminal Network Flows. *J Soc Indust Appl Math*, 9(4), 1961.

[6] Kelly F P. Routing and capacity allocation in networks with trunk reservation. 1987.

[7] Kelly F P. Communications Network and Method. British patent application 8607576, 26 March 1986.

[8] Key P B. Network Synthesis and Dynamic Routing Methods. 1987.

[9] Key P B and Whitehead M J. Cost-Effective Use Of Networks Employing Dynamic Alternative Routing. *International Teletraffic Congress 12 ,Torino, Italy*, June 1988. Submitted.

[10] Stacey R R and Songhurst D J. Dynamic Alternative Routing in the British Telecom Trunk Network. *International Switching Symposium*, March 1987. Phoenix.