ANALYSIS AND OPTIMAL DESIGN OF AGGREGATED-LEAST-BUSY-ALTERNATIVE ROUTING ON SYMMETRIC LOSS NETWORKS WITH TRUNK RESERVATIONS

Debasis Mitra  Richard J. Gibbens*  B. D. Huang

AT&T Bell Laboratories
Murray Hill, New Jersey 07974, USA

We investigate a distributed, state-dependent, dynamic routing strategy, called here Aggregated-Least-Busy-Alternative (ALBA), for circuit-switched symmetric, fully connected loss networks. We give a Fixed Point Model for ALBA. The case of no aggregation is Least-Busy-Alternative (LBA) for which we compare simulation and analytic results and find the agreement surprisingly good. We also consider two separate asymptotic scalings based on the Fixed Point Model. The first shows that there is a dichotomy in network behavior dictated by a threshold in the offered traffic rate. In the second asymptotic scaling the asymptotically optimum trunk reservation parameter is obtained; it is found to be effective in realistically sized networks.

1. INTRODUCTION

We investigate a distributed, state-dependent, dynamic routing strategy called Aggregated-Least-Busy-Alternative (ALBA), so called because the particular case in which there is no aggregation is Least Busy Alternative (LBA). The networks considered are symmetric and fully connected. The network has \( N \) nodes and a link with \( C \) circuits between every pair of nodes. The offered calls for each origin-destination pair form independent Poisson streams of rate \( \lambda \). Routes are restricted to have at most 2 links. In ALBA the route of a call is determined by certain local information on the state of the links of the possible routes at the time the call arrives; this information gets coarser with increased aggregation. In ALBA(\( K \)) the \((C+1)\) states of each link, which represent the number of occupied circuits, are lumped into \( K \) aggregates, \((s_0, s_1, \ldots, s_{K-1})\) in increasing order of levels of occupancy. The last aggregate \( s_{K-1} \) is the set of states with \( r \) or less idle circuits; we say that \( s_{K-1} \) comprises the set of reserved states and \( r \) is the trunk reservation (TR) parameter. ALBA routing is done as follows: an arriving call is attempted on the direct route and it is carried if there exists a free circuit. Otherwise, the call is attempted on an aggregated-least-busy 2-link route. An aggregated-least-busy route for origin-destination pair \( i,j \) is one which minimizes \( \max(I_{ik}, I_{kj}) \) where \( k \) is an intermediate node and, \( I_{ik} \) and \( I_{kj} \) are the aggregate states of links \((i,k)\) and \((k,j)\) respectively. We let \( M \) denote the number of possible 2-link routes between any origin-destination pair. Typically \( M = N - 2 \). Any ties for aggregated-least-busy route are broken randomly. The call is accepted on the 2-link route if it satisfies a trunk reservation criterion, namely, acceptance of the call does not leave either link in aggregate \( s_{K-1} \). Otherwise the call is blocked and lost. We further assume that the holding times of accepted calls are random and independent, exponentially distributed with mean \( 1 \). ALBA(2) and LBA represent two extremes of aggregation. The two aggregate states of ALBA(2) are the sets of unreserved and reserved states. In contrast, LBA is obtained if all but the last aggregate are composed of singletons.

Various forms of state-dependent routing have been studied. Ash [AS85] has investigated an extension to DNHR called Trunk Status Map Routing. In Canada Bell Northern Research has proposed a centralized scheme called Dynamically Controlled Routing (DCR) [CA83, GI89]. Krishnan and Ott [KR88] and, Lazarev and Starobinets [LA77] have proposed state-dependent routing schemes with roots in Markov Decision Theory. Marbukh [MA81, MA83] reports results on the analysis of LBA. See Pioro [PI89] for a broad perspective and especially for accounts of Soviet theoretical contributions. A somewhat different, distributed approach based on a random search technique is Dynamic Alternative Routing (DAR) [GI88a, GI88b, GI90].

A major impetus for our work here is AT&T’s recent deployment of Real Time Network Routing (RTNR) [AS91]. It is now technologically feasible to monitor the occupancy levels of alternate routes and to make routing decisions on a call-by-call basis. The role of aggregation is a central one in RTNR. By making routing decisions based only on information on the aggregate state of the links in the alternate routes, a considerable simplification of the implementation is achieved together with reductions in the signalling traffic. We show that there are

* Dr. Gibbens’ work was done while visiting AT&T Bell Laboratories. His present address: University of Cambridge, Statistical Laboratory, 16 Mill Lane, Cambridge, CB2 1SB, UK.
fundamental reasons for the performance loss due to aggregation to be small with proper design.

The role of TR is extremely important with state-dependent routing. Hence a large part of this paper is directed at the qualitative understanding of the role of TR and at the calculation of optimum TR parameters.

The report presented here is necessarily brief and [MI90a, MI90b] may be consulted for further information.

2. FIXED POINT MODEL

In this section we develop a Fixed Point Model (FPM) for ALBA. Such models have long been important in the context of state-independent routing [KA67, AK84, WH85, KE86, GI89]. The common element in the FPM is the assumption of link independence whereby each link is assumed to thin traffic independently, i.e. in a manner oblivious of the state of the rest of the network [MA83, GI90]. The new element in FPM for state-dependent routing is oblivious of the state of the rest of the network [MA83, GI90]. The new element in FPM for state-dependent routing is that the rate of the alternate-routed traffic offered to an individual link depends on the state of the link [WO90]. Aggregation reduces the number of parameters in these models since the offered traffic rates depend only on the aggregate state.

In ALBA(2) each link has 2 aggregate levels of occupancy defined in terms of parameters \( r_1, (0 \leq I < K) \) by \( A = \{ C - r_1, C - r_1 + 1, \ldots, C - r_1 + 1 + \delta_{i,K-1} \} (0 \leq I < K) \) where

\[
C = r_0 > r_1 > r_2 > \cdots > r_{K-1} > r_K = 0.
\]

Hence, \( r_{K-1} \), is the familiar trunk reservation parameter and also denoted by just \( r \). For ALBA(K), let \( \pi_i \) \( (0 \leq i \leq C) \) denote the stationary probability for the number of occupied circuits on a link and let \( P_i = \sum_{\pi_i} \pi_i \) \((0 \leq I < K)\). Let \( \Psi_i \) \((0 \leq I < K)\) denote the stationary probability that at least one link in a 2-link alternate route is in aggregate state \( I \) or higher. From the assumed independence of links, \( \Psi_i = 1 - \left( \sum_{j=0}^{I-1} P_j \right)^2 (1 \leq I < K) \) and of course \( \Psi_0 = 1 \). Given that for a particular origin-destination, the least busy 2-link alternate route is in aggregate state \( I \) \((0 \leq I < K)\), the probability that there are \((n-1)\) other such least busy 2-link routes is \((n = 1, 2, \cdots, M)\)

\[
F(n|I) = \begin{cases} M-1 \cdot \left( \Psi_i - \Psi_{i+1} \right)^{n-1} \left( \psi_{i+1}^{M-n} \right) & (0 \leq I < K) \end{cases}
\]

Hence the rate of overflow traffic from a particular origin-destination which is routed to a given 2-link route in aggregate state \( I \) is

\[
y_I = \lambda \pi C \sum_{n=1}^{M} \frac{1}{n} F(n|I) (0 \leq I < K-1).
\]

The total rate of overflow traffic routed to a link in aggregate state \( I \) from all source-destination pairs is

\[
y_I = 2M \left[ y_I \sum_{J=0}^{I} P_J + \sum_{J=I+1}^{K-1} y_J P_J \right] (0 \leq I < K-1)
\]

where, of course, \( y_{K-1} = 0 \) since it corresponds to the reserved state. The above equations define the map \( v = f(x) \).

To proceed in the reverse direction from \( v \) to \( x \), \( x = g(v) \), it is necessary to solve the birth and death equations with transition rates \( q(i, i') \) given thus: \( q(i, i-1) = i \) for \( 0 < i \leq C \), and if \( i = 0 \) then \( q(i, i+1) = \lambda + v_I \) \((0 \leq I < K-1)\).

In summary, the fixed-point equations are \( v = f(x) \), \( x = g(v) \).

Define \( B_1 = \pi_C \) and \( B_2 = P_{K-1} = \sum_{\pi_i} \pi_i \), the probability of blocking on the direct path and of the link being in the reserved state, respectively. Then the network loss probability, \( L \), is

\[
L = B_1 \left[ 1 - (1 - B_2)^2 \right]^M.
\]

In the special case of ALBA(2) it may be shown that, writing \( v \) for \( v_0 \),

\[
v = \frac{2\lambda (B_1 - L)}{1 - B_2}.
\]

Note that the only difference between a classical single link model and the FPM for ALBA(2) is that in the former the overflow rate \( v \) is given a priori, while in the latter it is obtained by solving equation (1).

3. NUMERICAL RESULTS

3.1 Numerical Results for LBA

The LBA routing strategy gives rise to some sensitive behavior in \( x \) and \( v \). For instance, the state-dependent overflow rates \( v_I \) make a large transition over a small number of states \( i \) (see Figure 3.1). Observe that except for a few large values of \( i \), \( \pi_i \) is negligibly small. The numerical procedures which we have devised to calculate \( x \) take advantage of these features. Observe the truncated Gaussian form for \( x \).

We have simulated the LBA routing strategy and the results are compared in Figure 3.2 to those obtained from the FPM. The confidence interval of the simulation data points are about the widths of the points shown in the figure. The agreement will be observed to be good.

Figure 3.3 shows \( L^* \) vs \( \lambda \) for various values of \( M \) where \( L^* \) is the network loss probability minimized with respect to the trunk reservation parameter \( r \) separately for each \( \lambda \).

3.2 Numerical Results for ALBA(K), K Small

We report on the comparative performance of ALBA(2) and LBA. For offered traffic above a threshold value (whose existence is examined in the next section), the difference in \( B_1, B_2 \) and \( L \) is negligible, and this is true
for a broad range of values of $M$. For offered traffic below the threshold there are significant differences in $B_1$ and $B_2$. However, in spite of this, if $M$ is large then $L$ is effectively 0 for both ALBA(2) and LBA. See Table 3.1 in which the threshold is about 107.5. When $M$ is not large we have identified a narrow range of offered traffic rates in the neighbourhood of the threshold where non-negligible differences in $L$ exist. It is in this important region that a small number of properly designed aggregate states, say $K=4$ or 5, makes ALBA(K) perform better than ALBA(2) and negligibly differently from LBA. See Table 3.2; the data presented is of particular interest in light of design goals which aim to have $L$ less than 0.1%. The case of $M$ not large is important since this may in effect be the case locally in large non-homogeneous networks.

4. ASYMPTOTICS

4.1 Dichotomous Behavior for Large $M$

The FPM is the base for investigating behavior in asymptotic regimes in which certain parameters are made large. The regime investigated in this section starts with the FPM for ALBA(2) and lets $M$ be large while $\lambda$, $C$ and $r$ are held fixed. The hope is that state-dependent routing succeeds in making the entire network one sharable resource and that there are large gains accruing from large $M$.

First, we need to define

$$F(\lambda, C, r) = \frac{C!}{\lambda^r + 1} \frac{1}{(C-r-1)!}$$

$$G(\lambda, C, r) = \frac{C!}{\lambda^C} \sum_{n=C-r}^{\infty} \frac{\lambda^n}{n!}.$$
Proposition 1. In ALBA(2) let $\lambda$, $C$ and $r$ be fixed while $M \to \infty$. The condition

$$1 < \frac{2}{F(\lambda, C, r)}$$

is necessary and sufficient for the existence of a unique solution $v$ to (1) with the property that $v = O(M^{1/2})$. In this solution $B_1 = O(1)$, $1-B_2 = O(M^{-1/2})$ and $L = O(1)$. Specifically,

$$L = \frac{1}{G(\lambda, C, r)} \left[ 1 - F(\lambda, C, r)/2 \right] \left[ 1 + O(M^{-1/2}) \right].$$

The above inequality defines the condition for the offered traffic $\lambda$ to exceed a threshold value $\lambda_T = \lambda_T(C, r)$.

Proposition 2. For the scaling in the previous Proposition if

$$\frac{2}{F(\lambda, C, r)} < 1$$

then one or more solutions to (1) exist in which $v = O(1)$, $B_1 = O(1)$, $1-B_2 = O(1)$ and $L$ is exponentially small in $M$, i.e., $L = O(c^{-M})$ where $c > 1$.

Our results show that the network behavior is dichotomous. If the offered traffic rate is below a threshold $\lambda_T$ determined by $C$ and $r$ then the network loss probability is exponentially small in $M$. In sharp contrast, if the offered traffic rate exceeds the threshold then the network loss probability remains essentially constant. The threshold thus serves to delineate the "engineered" designs for which state-dependent routing is most effective.

The above pair of Propositions have given information on network behavior for large $M$. In particular, for $M = \infty$:

$$L \to \begin{cases} 0 & \text{if } 1 > 2/F \\ \frac{1}{G} \left[ 1 - \frac{F}{2} \right] & \text{if } 1 < 2/F. \end{cases}$$

4.2 Optimal Trunk Reservation for Large Networks

The second asymptotic regime which we consider is in connection with the large network limit in which $N = \infty$. Specifically, we have from (2),

$$L_{N=\infty} = \max \left[ 0, \frac{1}{G} \left[ 1 - \frac{F}{2} \right] \right].$$

This expression for the loss probability in this limit was obtained by Marbukh [MA81] for both LBA and ALBA(2). Gibbens and Kelly [GI90] have also given an informal derivation.

We next impose on (3) the asymptotic scaling

$$C = \lambda - \alpha \lambda^{1/2} \text{ as } \lambda \to \infty, \quad \alpha = O(1).$$

This scaling has a long history in the literature of both loss and queueing networks [JA74, MC81, MI88, RE89]. We call it the "moderate traffic" regime to distinguish it from the "light" and "heavy" traffic regimes in which, respectively, $p < 1$ and $p > 1$, where $\lambda/C = p$. This regime is of the greatest practical interest. Further, assume that

$$C - r - 1 = \lambda - \beta \lambda^{1/2}, \quad \beta = O(1).$$

Note that $\alpha \leq \beta$ and that equivalently, $r + 1 = (\beta - \alpha) \lambda^{1/2}$. It turns out that in this scaling, $2/F(\lambda, C, r) \sim 2e^{-2(\alpha^2 - \beta^2)/2}$. We have obtained the following results [Mi90b].

Proposition 3.

$$L_{N=\infty} = 0 \quad \text{if } 2e^{-2(\alpha^2 - \beta^2)/2} < 1$$

$$= \frac{1}{2\lambda^{1/2}} \Psi(\beta; \alpha) \left[ 1 + O(\lambda^{-1/2}) \right] \quad \text{otherwise},$$

where $\Psi(\beta; \alpha) = (2e^{-\alpha^2/2} - e^{-\beta^2/2})\int_{\alpha}^{\beta} e^{-y^2/2} dy$.

To recapitulate, this expression for the network loss probability has been obtained as a consequence of two limiting processes: first, $N \to \infty$ while $\lambda$, $C$ and $r$ are held
fixed, and, second, $\lambda, C$ and $r$ are simultaneously made large. In [MI90b] a quite separate asymptotic scaling is considered in which $N, \lambda, C$ and $r$ are all simultaneously made large. This asymptotic scaling yields, to leading order, a result identical to Proposition 3.

**Proposition 4.** Define $\lambda^+(C) = C - \sqrt{\log 4} C^{1/2}$. If $\lambda < \lambda^+(C)$ then there exist TR parameters $r$ such that $L_{N_{\infty}} = 0$. The smallest such TR parameter $r^* = r^+(\lambda, C)$ is

$$r^* + 1 = C - \lambda - \sqrt{(C - \lambda)^2 - \lambda \log 4}.$$ 

If $\lambda > \lambda^+$, then $L_{N_{\infty}} > 0$. In this case an important design problem is the selection of $r$ which minimizes $L_{N_{\infty}}$ for given $\lambda$ and $C$. The solution to this problem is given by the following proposition.

**Proposition 5.** If $\lambda > \lambda^+$ in the above asymptotic regime, then the trunk reservation parameter $r^*$ which asymptotically minimizes the loss probability is given thus: $(r^* + 1) = (\beta^* - \alpha) \lambda^{1/2}$ for $\beta^*$ the unique solution in $(\alpha, \infty)$ of $\Psi(\beta^*; \alpha) = \beta^*$ where $\Psi(\beta; \alpha)$ is given in Proposition 3.

It may also be shown that when $\lambda > \lambda^+$ the minimized loss probability, $L_{N_{\infty}}$, is given by

$$L_{N_{\infty}} = \frac{\beta^*}{2C^{1/2}}.$$ 

In Figure 4.1 the curve marked $\gamma = \infty$ corresponds to the solution $\beta^*$ which has been computed for a range of values of $\alpha$. A striking feature is the almost linear behavior of $\beta^*$ with respect to $\alpha$.

Figure 4.2 gives results for $C = 120$. The optimum trunk reservation parameters have been computed in two different ways: first, via the asymptotic analysis and, secondly, by numerically minimizing $L_{N_{\infty}}$ in (3) with respect to $r$. The two approaches are found to be in excellent agreement. We have found such agreement to also hold for smaller trunk groups, for instance, $C = 30$.

We conclude this section by remarking on a bridge between the results just obtained and asymptotic results [RE89] for the classical single link model with two offered traffic streams: the first of high priority calls at rate $\lambda$ and the second of low priority calls at rate $\nu$ and which is subject to trunk reservation. In addition to the parameters $(\lambda, \nu, C, r)$ we assume that each carried call of the high priority stream earns $w(w > 1)$ as much as a low priority call. In appropriate units the net revenue is thus: $W = w\lambda(1-B_1) + \nu(1-B_2)$. In the asymptotic regime suppose that $\nu = \gamma \lambda^{1/2}$. Then the following Proposition holds [MI90b].

**Proposition 6.** The $r^*$ obtained earlier which asymptotically minimizes the loss probability subject to ALBA routing is identical to the asymptotically optimal revenue maximizing trunk reservation parameter for the single link model when $w = 2$ and $\gamma$ is large.

This result is consistent with a central characteristic of ALBA routing, namely, when the offered traffic exceeds the threshold the overflow traffic is large.

Figure 4.1 shows solutions $\beta^*$ for the single link model with $w = 2$ and various values of $\gamma$. Particularly striking features are the almost linear dependence on $\alpha$ and the relative insensitivity of $\beta^*$ with respect to $\gamma$.

### 5. CONCLUSIONS

This paper has offered a variety of results for the analysis and design of aggregated state-dependent routing strategies in symmetric loss networks. Basic to these results are approximate Fixed Point Models which have been compared with simulation results and which have been examined in two separate asymptotic frameworks.

![Figure 4.1: Asymptotically optimal trunk reservation parameter.](image1)

![Figure 4.2: Comparison of $r^*$. $C = 120$.](image2)
ACKNOWLEDGEMENTS

The authors acknowledge with gratitude the careful reading of the manuscript by G. R. Ash whose comments were helpful.

REFERENCES


