Reducing Network Congestion and Blocking Probability Through Balanced Allocation

Malwina J Luczak* Eli Upfal †

University of Oxford Brown University

Abstract

We compare the performance of a variant of the standard *Dynamic Alternative Routing (DAR)* technique commonly used in telephone and ATM networks to a path selection algorithm that is based on the balanced allocations principle [4, 18] - the *Balanced Dynamic Alternative Routing (BDAR)* algorithm. While the standard technique checks alternative routes sequentially until available bandwidth is found, the BDAR algorithm compares and chooses the best among a small number of alternatives.

We show that, at the expense of a minor increase in routing overhead, the BDAR gives a substantial improvement in network performance in terms of both network congestion and blocking probabilities.

1 Introduction

Fast, high bandwidth, circuit switching telecommunications systems such as ATM and telephone networks employ a limited path selection algorithm in order to fully utilize the network resources while minimizing routing overhead. Typically there is a dedicated bandwidth for communication between each pair of nodes in the network. This dedicated bandwidth, on one or a chain of physical links, is designed to satisfy the expected demand for communication between these stations. Only when this bandwidth is exhausted the admission control protocol tries to find an alternative route through some intermediate nodes. To minimize overhead and routing delays the protocol checks just a small number of alternative routes; if there is no free capacity on any of these alternatives, then the call or communication request is rejected. Implementations that use this technique include the Dynamic Alternate Routing (DAR) algorithm used by BT (British Telecommunications) [10], and AT&T's Dynamic Nonhierarchical Routing (DNHR) algorithm [3].

A common feature in currently implemented protocols is the sequential examination of alternative routes. Only when the algorithm decides to reject a route an alternative one is examined. The criteria for rejecting a route and the method in which the alternative route is selected have been the subject of extensive research, in particular, in the context of the BT DAR algorithm [9, 10, 13]. (See Kelly [15] for an extensive survey.)

Dynamic routing can be viewed as a special case of the "on-line" load balancing problem, where the load may be assigned to one or more servers (edges), and jobs (communication requests) can be scheduled only on specific subsets (paths) of the set of servers, as defined by the network topology. A number of recent papers demonstrated the advantage of *balanced allocations* [4, 5, 18, 19] for standard load sharing problems where jobs require only one server and can be executed by any server in the system. The goal of this work is to extend these results to the more complex setting of load sharing under constraints imposed by

*Mathematical Institute, University of Oxford, 24-29 St. Giles', Oxford OX1 3LB, UK. E-mail: luczak@maths.ox.ac.uk. Supported by a British Telecommunications studentship

†Computer Science Department, Brown University, Box 1910, Providence, RI 02912-1910, USA. E-mail: eli@cs.brown.edu. Supported in part by NSF grant CCR-9731477

network topology.

The basic idea in balanced allocations is to examine several random options and assign the job to the best of these options at the time of the assignment. The drawback of this strategy is that several alternatives are examined even when the first alternative would be satisfactory, thus increasing overhead. However, past research has shown that at least in the standard load balancing scenario a small increase in placement overhead gives a substantial improvement in load balancing under various (random) input settings.

In this paper we employ an extension of the balanced allocation principle to the problem of dynamic network routing. The goal here is to reduce system congestion and in particular to minimize the *blocking probability* - the probability that a call request is rejected. The main difficulty in applying and analyzing balanced allocation in a network setting is handling the dependencies imposed by the topology of the underlying graph. Our results show that the advantage of the load balancing principle is so significant that it holds even in the presence of a set of dependencies. We note that [6] has also applied the balanced allocation idea to a routing problem; however, the setting and technique there are different.

1.1 Model and New Results

Since the networks considered here reserve a logical link (bandwidth) for communication between each pair of stations, and use alternative routing only when that logical link is busy, it is accurate to model the problem in terms of a complete graph with n vertices (stations) and $N = \binom{n}{2}$ edges (links). The input is a sequence of call requests. The routing algorithm has to process the calls "on-line", i.e., the t-th request is either assigned a path or rejected before the algorithm receives the $t+1$-th request. Once a call is assigned a path, that path cannot be changed throughout the duration of the call. Each edge has capacity B, where B can be a constant or an increasing function of n. The goal is to assign routes to the maximum number of call requests without violating the capacity constraints on the edges. For comparison, we assume that both DAR and BDAR algorithms partition the capacity of each edge between direct and alternative routes.

The *d-Dynamic Alternative Routing (DAR)* algorithm works as follows: When a new call request arrives, it tries to route the call through the direct (one-link) path. If there is no available bandwidth on the direct path, then the algorithm chooses randomly up to d alternative routes of length two and assigns the call to the first path with available circuits. If no such path is found, then the request is lost.

The *d-Balanced Alternative Routing (BDAR)* algorithm starts by examining the direct path. If that path is blocked, then the algorithm chooses d alternative paths at random. The algorithm compares the maximum loaded edge (in a suitable sense to be defined in subsequent sections) on each of the paths and assigns the call to a path with the minimum maximum load. If there is no alternative path with free bandwidth, then the call is rejected.

We compare the performance of the two algorithms in *static (finite, discrete time)* and *dynamic (infinite, continuous time)* settings. In the static setting the process starts with no calls assigned to the network. A sequence of N calls is then received; the endpoints of each call are chosen independently at random from the N possible pairs (thus on average we have one call per pair of vertices).

The following theorem summarizes our results for the static case:

Theorem 1.1 *Assume that $N = \binom{n}{2}$ random calls are sequentially routed on an n node complete network with edge capacity B.*

1. *When the d-Dynamic Alternative Routing Algorithm is used, the number of lost calls is with high probability $Ne^{-\Theta(dB^2 \log B)}$. To guarantee that with high probability all N calls are routed successfully, the edge capacity must satisfy $B = \Omega(\sqrt{\frac{\log N}{d \log\log N}})$.*
2. *When the Balanced Dynamic Alternative Routing Algorithm is used, the number of lost calls is with high probability $Ne^{-d^{\Theta(B)}}$. To guarantee that with high probability all N calls are routed successfully, it is sufficient to have $B = \Omega(\frac{\log\log N}{\log d})$.*

Since we expect d to be a small constant (2-3) and B a large number (a few hundred or even thousand), the advantage of the balanced scheme is clear.

In the dynamic case we consider a stochastic process in which new requests onto each link arrive according to a Poisson process with rate λ, all arrival streams being independent. The holding period of a call is independent of all other holding periods and all arrivals, and is exponentially distributed with unit mean. The performance measure is the stationary blocking probability of the system. Our analysis of the dynamic case shows a similar gap in the performance of the two techniques.

Theorem 1.2 *In the above dynamic setting the blocking probability of the d-Dynamic Alternative Routing algorithm is* $e^{-\Omega(dB^2 \log(B/\lambda))}$*, while the blocking probability of the d-Balanced Dynamic Alternative Routing algorithm is* $e^{-d^{\Theta(B/\lambda)}}$.

2 The Finite Case

2.1 Performance of the Dynamic Alternative Routing Algorithm

Theorem 2.1 *Suppose that* $N = \binom{n}{2}$ *random calls are sequentially routed on an n node complete network using the d-Dynamic Alternative Routing algorithm. If the capacity of each edge is B, then*

1. *With high probability the number of lost calls is* $Ne^{-\Theta(dB^2 \log B)}$.

2. *To guarantee that with high probability all N calls are routed successfully, we require* $B = \Omega(\sqrt{\frac{\log N}{d \log\log N}})$.

Proof: (Sketch) To prove the upper bound assume that the capacity of edge $\{v, u\}$ is split so that $D = B/3$ circuits are reserved for direct calls and $2D$ circuits are reserved for alternative calls: D circuits for alternate calls with endpoint v, and D circuits for alternate calls with endpoint u. Let p_1 be the probability that a pair of vertices appears as the endpoints of at least D calls; then $p_1 = \Theta(\binom{N}{D}(\frac{1}{N})^D$. Thus, with high probability, for any vertex v, the number of edges adjacent to v and saturated by direct calls is $\Theta(p_1(n-1))$ and the number of calls with endpoint v that use alternative paths is $m = \Theta(p_1\frac{2N}{n}) = \Theta(p_1 n)$.

The probability that the $2m$ calls saturate a given edge with endpoints v and u is $p_2 = \Theta(\binom{2m}{D}(\frac{d}{n-1})^D)$. Thus the probability that a call to v is blocked is bounded by

$$p_1(p_2)^d \leq e^{-cdD^2 \log D}$$

for some constant $c > 0$.

To prove the lower bound we partition the N calls into three sets of $N/3$ calls each. With high probability the first set of $N/3$ calls saturates $\Omega(p_1 n)$ edges adjacent to each vertex in the graph (to bound the dependency between the events on different edges we use the fact that if $k = \Omega(n)$ then with high probability no set of k pairs of vertices receives more than $O(k + \sqrt{n})$ requests).

Using the second set of $N/3$ calls we prove that with high probability $\Omega(p_2 n)$ edges adjacent to each node are saturated by alternative paths. Using the third set of calls we show that with high probability

$$\Omega(Np_1(p_2)^d) = Ne^{-\Omega(dB^2 \log B)}$$

calls are blocked. It is easy to verify that any other partition of the edge capacity between the direct and alternative calls can only increase the constant in the exponent. □

Note that by standard results on random allocations [11, 16], if no alternative routing is used, the blocking probability is $e^{-\Omega(B \log B)}$, and edge capacity $\Omega(\frac{\log n}{\log\log n})$ is necessary to guarantee that no calls are lost.

2.2 Analysis of the Balanced Alternative Routing Algorithm

Theorem 2.2 *Suppose that* $N = \binom{n}{2}$ *random calls are sequentially routed on an n node complete network using the d-Balanced Alternative Routing Algorithm. If the maximum capacity of each edge in the network is B circuits, then*

1. *With high probability the number of lost calls is* $Ne^{-d^{\Theta(B)}}$.

2. *Edge capacity* $B = O\left(\frac{\log\log N}{\log d}\right)$ *is sufficient to guarantee that with high probability all N calls are routed successfully.*

Proof: Again we assume that $D = B/3$ circuits are dedicated to direct calls, while $2D = 2B/3$ circuits are dedicated to alternatively routed calls. Consider the state of the system after the route assignment of the first t calls. We partition the load on edge $e = \{u, v\}$ into three variables:

- $X_e(t)$ is the load from direct paths.
- $Y_{e,v}(t)$ and $Y_{e,u}(t)$ are the loads from alternative paths with endpoints v and u respectively.

Clearly each path that uses edge e is included in one of these three counts. The algorithm guarantees that $X_e(N) \leq B/3$ for all e. To simplify the analysis of the $Y_{e,u}(t)$ variables we assume that all the calls use alternative paths. This assumption clearly ensures that our bounds stochastically dominate the actual values of the random variables.

For a fixed vertex v consider the set of random variables

$$\mathcal{Y}_v(t) \;=\; \{Y_{e,v}(t) \mid \text{edge } e \text{ is incident on } v\}.$$

The distribution of calls among the variables in $\mathcal{Y}_v(t)$ resembles the behavior of the "balanced allocation" systems studied in [4] with one major difference: the assignment of a call to a particular edge depends not only on the distribution of calls in the set $\mathcal{Y}_v$(t) but also on the distribution in a second system that corresponds to the other endpoint of the call. Thus, we need to analyze simultaneously the progress in time of a family of n sets of variables

$$\{\mathcal{Y}_v(t) \mid v \in V\}.$$

Each set contains $n-1$ random variables and the variables in different sets are not independent. Our analysis adapts the main argument in [4] to handle the n dependent systems.

Assume that the t-th call connects v to u and is routed through edge e that is incident on v. We define the *height* of that call in that edge to be $h_t = Y_{e,v}(t)$. Note that a call might have different heights in the two edges that carry it.

Let $L^v_{\geq i}(t) = |\{Y_{e,v}(t) \mid Y_{v,e}(t) \geq i\}|$; then $L^v_{\geq i}(t)$ counts the number of edges adjacent to v that carry calls with height at least i after the routing of the first t calls.

Define a sequence of events for $i = 1, \ldots, N$:

$$\mathcal{E}_i \;=\; \{L^v_{\geq i}(N) \leq \beta_i \;\text{ for all } v \in V\}$$

with

$$\beta_i = \begin{cases} n-1, & i = 1, \ldots, \Delta - 1; \\ \frac{n-1}{8e}, & i = \Delta; \\ \frac{e(2\beta_{i-1})^d}{(n-1)^{d-1}} = ep_{i-1}N, & i > \Delta, \end{cases}$$

where $\Delta = \lceil 8e + 1 \rceil$.

For each vertex v and index i define a sequence of binary variables $I^v_{i,t}$ for $t = 1, \ldots, N$, such that $I^v_{i,t} = 1$ if and only if

1. v is an endpoint of the t-th call;
2. $h_t \geq i + 1$;
3. $L^v_{\geq i}(t-1) \leq \beta_i$ for all $v \in V$.

The probability that v is an endpoint of the t-th call is $2/n$. If the fraction of edges with load at least i adjacent to each vertex is bounded by β_i, and $\omega_1, \ldots, \omega_{t-1}$ represent the assignments of the previous $t-1$ calls, then

$$\mathbf{Pr}(I^v_{i,t} = 1 | \omega_1, \ldots, \omega_{t-1}) \leq \frac{2}{n}\frac{(2\beta_i)^d}{(n-1)^d} = p_i.$$

Thus

$$\mathbf{Pr}(L^v_{\geq i+1} \geq k | \mathcal{E}_i) \leq \mathbf{Pr}(\sum I^v_{i,t} \geq k | \mathcal{E}_i)$$
$$\leq \frac{\mathbf{Pr}(B(N, p_i) \geq k)}{\mathbf{Pr}(\mathcal{E}_i)}.$$

The event $\mathcal{E}_\Delta = \{L^v_{\geq\Delta}(N) \leq \frac{n-1}{8e} \;\text{ for all } v \in V\}$ holds with high probability, and for $i \geq \Delta$, summing over the n systems[1],

$$\mathbf{Pr}(\neg\mathcal{E}_{i+1} | \mathcal{E}_i) \leq n\frac{\mathbf{Pr}(B(N, p_i) \geq \frac{e(2\beta_i)^d}{(n-1)^{d-1}})}{\mathbf{Pr}(\mathcal{E}_i)}$$
$$\leq n\frac{e^{-Np_i}}{\mathbf{Pr}(\mathcal{E}_i)} \leq \frac{1}{n\mathbf{Pr}(\mathcal{E}_i)}$$

provided $p_iN \geq 2\ln n$. Thus, for $p_iN \geq 2\ln n$, $\mathbf{Pr}(\neg\mathcal{E}_{i+1}) \leq \frac{1}{n} + \mathbf{Pr}(\neg\mathcal{E}_i)$.

Let i^* be the smallest i such that $p_{i^*}N \leq 2\ln n$. Notice that $i^* \leq \ln\ln n/\ln d + O(1)$, since

$$2\beta_{i+\Delta}/(n-1) = \frac{(2e)^{(d^i-1)/(d-1)}}{(4e)^{d^i}} \leq \frac{1}{2^{d^i}}.$$

[1] We use a standard version of the Chernoff Bound [2]: $\mathbf{Pr}(B(n, p) \geq enp) \leq e^{-np}$

Thus, for $B \leq i^* = O(\log\log n/\log d)$, the blocking probability is bounded by $2^{-d^{\Theta(B)}}$, establishing (1).

To establish (2),

$$\mathbf{Pr}(\exists v L^v_{\geq i^*+1} \geq 8\ln n | \mathcal{E}_{i^*})$$

$$\leq n\frac{\mathbf{Pr}(B(N, 2\ln n/N) \geq 8\ln n)}{\mathbf{Pr}(\mathcal{E}_{i^*})} \leq \frac{1}{n\mathbf{Pr}(\mathcal{E}_{i^*})},$$

and

$$\mathbf{Pr}(L^v_{i^*+3} \geq 1 | \forall v L^v_{i^*+1} \leq 8\ln n)$$

$$\leq \frac{\mathbf{Pr}(B(N, (16\ln n/N)^d) \geq 2)}{\mathbf{Pr}(\forall v L^v_{i^*+1} \leq 8\ln n)}$$

$$\leq \frac{2\binom{N}{2}(16\ln n/N)^{2d}(1-(16\ln n/N)^d)^{N-2}}{\mathbf{Pr}(\forall v L^v_{i^*+1} \leq 8\ln n)}.$$

Thus, $\mathbf{Pr}(\exists v L^v_{i^*+3} \geq 1) \leq n(16\ln n)^{2d}N^{-2(d-1)} + \mathbf{Pr}(\exists v L^v_{i^*+1} \geq 8\ln n) = o(1)$, establishing (2). □

3 Dynamic Analysis

3.1 Performance of the Dynamic Alternative Routing Algorithm

The steady state performance of a number of variants of the Dynamic Alternative Routing algorithm has been extensively studied. A set of integral equations that characterize the behavior of the system as $N \to \infty$ has been developed in [9] and [7]. However, when network topology is taken into account, the validity of these equations has only been fully justified for bounded B [7], while we are interested in the more general case that includes edge capacity as a function of the network size. For comparison with the performance of our new method, we use a simple lower bound argument for the blocking probability under the d-Dynamic Alternative Routing regime. Note that our call allocation strategy (partitioning the capacity of each edge between direct and alternative routes) avoids the difficulty of the bistable behavior observed in [9].

Let p_1 be the stationary blocking probability for direct routes, A standard argument in the theory of loss networks (see [15] for a general survey of the area) shows that if the capacity for direct routes is D, then $p_1 = C\frac{\lambda^D}{D!}$, where $C = \sum_{i=0}^{D}\frac{\lambda^i}{i!}$. Thus $p_1 = e^{-\Theta(B\log(B/\lambda))}$.

In equilibrium, the expected number of routes with endpoint v that attempt to seize alternative paths is $\Omega(\lambda n p_1)$. Thus the mean number of alternate calls per edge is $\Omega(\lambda p_1)$, and each call is equally likely to be assigned to any edge. Let p_2 be the probability that a random edge adjacent to v is blocked by alternative routes. Then, $p_2 = \Omega\left(\binom{p_1\lambda n}{B}(\frac{1}{n})^B\right)$. Let p_3 be the probability that a call is blocked off all the alternative routes selected; then $p_3 = \Omega((p_2)^d)$ (since for $d > 1$ each call has d "chances" to find a free link).

Thus, $p_3 = e^{-\Omega(dB^2\log(B/\lambda))}$, and the total blocking probability is

$$p_1p_3 = e^{-\Omega(dB^2\log(B/\lambda))}.$$

To guarantee that in equilibrium with high probability no calls are lost in an interval of length T, we need

$$B = \Omega\left(\sqrt{\frac{\log(\lambda Tn)}{d\log\log(\lambda Tn)}}\right).$$

More formally, the performance of the DAR algorithm could be analyzed in a similar way to the performance of BDAR in the next section.

3.2 Performance of the Balanced Dynamic Alternative Routing Algorithm

Following the basic idea of Mitzenmacher's analysis of the supermarket model [18], we develop a system of deterministic differential equations that model the "average" performance of the system in the limit as the number of links tends to infinity with the offered load to each link being held constant. We compute a bound for the fixed point of that system and then use Kurtz's density-dependent jump Markov chain theory to prove that in the steady state the stochastic system approaches the fixed point of the deterministic system.

Similarly to the analysis of the finite case in section 2.1 we partition the load of edge $e = \{u, v\}$ at time t into three variables:

- $X_e(t)$ is the load from direct paths.
- $Y_{e,v}(t)$ and $Y_{e,u}(t)$ are the loads from alternative paths with endpoints v and u, respectively.

Clearly each path that uses edge e at time t is included in one of the three counts. W.o.l.g we assume that edge e has capacity $B/3$ for each of the three types of calls.

The algorithm guarantees that $X_e(t) \leq B/3$ for all e and t. To simplify the analysis of the $Y_{e,u}(t)$ variables we assume that all calls use alternative paths; this assumption ensures that the bounds we obtain stochastically dominate the actual values of the random variables.

Suppose that the call routed at time t has endpoints v_1 and v_2 and is assigned to edges e_1 and e_2. (Since calls arrive in a Poisson process, only one call can arrive at a time.) The *height* of that call at edge e_1 is one plus $Y_{e_1,v_1}(t^-)$. Thus, $Y_{e,v}(t)$ is the number of calls with endpoint v traversing edge e.

Let $L^v_{\geq i}(t) = |\{Y_{e,v}(t) \mid Y_{v,e}(t) \geq i\}|$; then $L^v_{\geq i}(t)$ counts the the number of edges incident on v that carry at least i calls with endpoint v at time t. In a network with n nodes let $S^n_i(v,t) = \frac{L^v_{\geq i}(t)}{n-1}$, and let $S^n_i(v,u,t)$ denote the fraction of two-link paths connecting v to u carrying at time t calls with height i or larger on *both* edges. Let $R^n_i(v,t) = \sum_{j\geq i} S^n_j(v,t)$.

The probability that a connection routed at time t has endpoints u and v is $1/\binom{n}{2}$. The probability that its path has height i or more in the edge adjacent to u is bounded by the probability that in all the d attempts to route this connection at least one of the two edges of that choice already carries a call with height at least $i-1$ at time t^-. Thus, assuming that a new path is routing at time t, the probability it uses an edge adjacent to u with hight i or more at time t. is bounded by

$$\frac{1}{\binom{n}{2}}\sum_{v\neq u}((S^n_{i-1}(v,t^-)+S^n_{i-1}(u,t^-)-S^n_{i-1}(v,u,t^-))^d.$$

(To simplify the presentation we assume that one of the choices can be the edge $e = \{u,v\}$. In that case the call is counted in both $Y_{e,v}$ and $Y_{e,u}$.) Note that when a new path with endpoint u is added to an edge adjacent to u with hight i, both $S^n_i(u,i)$ and $R^n_i(v,t)$ are increased by $\frac{1}{n-1}$.

Since the duration of a path has an exponential distribution with expectation 1, and there are i paths routed through an edge with hight i, the expected decrease in $S^n_i(v,t)$ in a short interval Δt is given by is

$$(\Delta t)i(S^n_i(v,t^-)-S^n_{i+1}(v,t^-)),$$

and the expected decrease in $R^n_i(v,t)=\sum_{j\geq i} S^n_j(v,t)$ is

$$(\Delta t)(\sum_{j\geq i} j(S^n_j(v,t^-)-S^n_{j+1}(v,t^-)) =$$

$$(\Delta t)(R^n_i(v,t^-)+(i-1)S^n_i(v,t^-)).$$

Due to the memoryless property of the exponential distribution the decrease in $(n-1)S^n_i(v,t)$ and $(n-1)R^n_i(v,t)$ have Poisson distributions with the corresponding expectations.

Let $m^n_e(t)$ be the number of directly routed calls on edge e, and let $m^n_{e,e_1,e_2}(t)$ be the number of calls alternatively routed from edge e onto edges e_1 and e_2 ($m^n_{e,e_1,e_2}=0$ if edges e_1 and e_2 do not form an alternative path to edge e). Then the vector $\mathbf{m}^n(t)$ with components $m^n_e(t)$ and $m^n_{e,e_1,e_2}(t)$ defines a Markov process that describes the state of the n-th network at time t. (In our analysis, $m^n_e(t)=0$ for all e.) It is easy to see that our Markov process is recurrent, irreducible and aperiodic. Thus, the process has a unique stationary distribution and it converges to that distribution as $t\to\infty$. Since $\mathbf{R}^n(t)=(R^n_i(v,t))_{i,v\geq 1}$ is a function of $\mathbf{m}^n(t)$, we can write $\mathbf{R}^n(t)=H(\mathbf{m}^n(t))$, and the $R_i(v,t)$'s are determined by the same process and their distributions also converge to stationary ones. Let $\Pi^n_i(v)$ denote the stationary distribution of $R^n_i(v,t)$.

To bound the performance of the above process we introduce a *dominating* process that is easier to analyze. The dominating process, with parameter n, consists of a vector of random variables

$$\hat{\mathbf{R}}^n = \{\hat{R}^n_i(v,t) \mid i\geq i, \quad 1\leq v\leq n\}.$$

The process is governed by a sequence of increments and decrements to the variables. Increments have the same Poisson arrival rate as the original process (with the same mean $\lambda\binom{n}{2}$). The probability that an "arrival" at time t adds $\frac{1}{n-1}$ to $\hat{R}_i(u,t)$ is

$$\frac{1}{\binom{n}{2}}\sum_{v\neq u}(\hat{R}_{i-1}(u,t^-)+\hat{R}_{i-1}(v,t^-))^d.$$

The expected decrease in $\hat{R}^n_i(v,t)$ in a small interval Δt is $(\Delta t)\hat{R}^n_i(v,t)$ and the decrease in $(n-1)\hat{R}^n_i(v,t)$ has a Poisson distribution. The dominating process is a recurrent, irreducible and aperiodic Markovian process, thus, has a unique stationary distribution. Denote by $\hat{\Pi}^n_i(v)$ the stationary distribution of $\hat{R}^n_i(v,t)$.

The dominating process is defined so that whenever $\hat{R}_i^n(v,t) \geq R_i^n(v,t)$ the probability of an increase in $\hat{R}_i^n(v,t)$ is no smaller than the probability of an increase in $R_i^n(v,t)$, and the probability of a decrease in $\hat{R}_i^n(v,t)$ is no larger than the probability of a decrease in $R_i^n(v,t)$. Thus, it is easy to verify by induction on i that, starting with the same initial configuration, the vector $\hat{\mathbf{R}}^n$ stochastically dominates the vector $\mathbf{R}^n$ component by component. I.e. for any $i \geq 1$, $n \geq v \geq 1$, $t \geq 0$ and value x:

$$\mathbf{Pr}(R_i^n(v,t) \geq x) \leq \mathbf{Pr}(\hat{R}_i^n(v,t) \geq x).$$

In particular the stationary distribution of the dominating process dominates the stationary distribution of the original process.

Let $\hat{R}_i(v,t) = \lim_{n\to\infty} \hat{R}_i^n(v,t)$. We derive a set of differential equations characterizing the asymptotic change in $\hat{R}_i(v,t)$. Define $V_0(v,t) = 1$, and

$$V_i(v,t) = \lim_{n\to\infty} \frac{1}{n-1} \sum_{u\neq v} (\hat{R}_i(v,t) + \hat{R}_i(u,t))^d.$$

In a small time interval Δt we expect the change in $R_i(v,t)$ to be

$$\Delta R_i(v,t) = \frac{(\Delta t)\lambda\binom{n}{2}}{\binom{n}{2}} V_{i-1}(v,t) - (\Delta t)R_i(v,t).$$

Thus, the following system of differential equations models the behavior of $\hat{R}_i(v,t)$ under the limiting regime of the stochastic system as $n \to \infty$.

$$\begin{cases} \frac{d\hat{R}_i(v,t)}{dt} = \lambda V_{i-1}(v,t) - \hat{R}_i(v,t) & i \geq 1 \\ V_0(v,t) \ = \ 1 & \end{cases} \tag{1}$$

System (1) has a unique solution for each $d \geq 2$, and for any set of finite initial conditions (see [1] for a general result for infinite-dimensional systems):

Theorem 3.1 *The unique solution to system (1) subject to the initial conditions $\hat{R}_i(v,0) = g_i(v)$ for all i, v is given by*

$$\hat{R}_1(v,t) = \lambda + (g_1(v) - \lambda)e^{-t} \quad t \geq 0,$$

$$\hat{R}_i(v,t) = e^{-t}g_i(v) + \int_0^t e^{-(t-u)} V_{i-1}(v,u)du \quad i \geq 2.$$

We further prove

Theorem 3.2 *System (1) with $d \geq 2$ has a unique fixed point with $\sum_{v\in V}\sum_{i=1}^{\infty} \hat{R}_i(v) < \infty$ given by*

$$\hat{\Pi}_i(v) = \frac{1}{2}(2\lambda)^{\frac{d^i-1}{d-1}}$$

for all $v \in V$ and $i \geq 1$.

We can show that the dominating deterministic system converges to its fixed point. One way to prove this is by building on the proof technique in Mitzenmacher's analysis [18], which is an application of the method of Lyapunov. However, there are a few major differences between our proof and the proof in [18]. First, here we are dealing with a "doubly countably" infinite state space. That is, a state of our deterministic process can be written down as a doubly infinite matrix $\hat{R} = (\hat{R}_i(v))$, where v indexes rows and i indexes columns.

To prove that the system converges to its fixed point, we define for each vertex v a potential function $\Phi(v,t) = \sum_{i=0}^{\infty} w_i|\hat{R}_i(v,t) - \hat{R}_i(v)|$. We show that there exists a constant $\delta > 0$ such that, unless at the fixed point, for each vertex v,

$$d\Phi(v,t)/dt \leq -\delta\Phi(v,t) + \frac{\delta}{2}\lim_{n\to\infty}\frac{1}{n-1}\sum_{u\neq v}\Phi(u,t).$$

We find δ by applying a technique analogous to the one used by Mitzenmacher in [18]. If we start the process with an empty system, or any other symmetric configuration, that is, $\hat{R}_i(v,0) = \hat{R}_i(u,0)$ for all $i \geq 0$ and all pairs of vertices u and v, then the form of the differential equations implies that $\hat{R}_i(v,t) = \hat{R}_i(u,t)$ for all $t \geq 0$, and thus the $\Phi(v,t)$ are all equal for all $t \geq 0$. Therefore, $d\Phi(v,t)/dt \leq -\frac{\delta}{2}\Phi(v,t)$ for all v, which yields exponential convergence of the system to its fixed point. For "non-symmetric" initial conditions, we still have, for all v and t,

$$d\Phi(v,t)/dt \leq -\delta\Phi(v,t) + \frac{\delta}{2}\sup_u \Phi(u,t)$$

$$= -\frac{\delta}{2}\Phi(v,t) + \frac{\delta}{2}(\sup_u \Phi(u,t) - \Phi(v,t)).$$

Now at any time t that we are away from the fixed point there exists a vertex v^* such that $\Phi(v^*,t) >$

$2(\sup_u \Phi(u,t) - \Phi(v^*,t))$. While v^* satisfies this condition, $\Phi(v^*,t)$ and all $\Phi(v,t) \geq \Phi(v^*,t)$ converge to 0 exponentially, with all the other $\Phi(v,t)$ remaining no larger than $\Phi(v^*,t)$. We divide the time interval $[0,\infty)$ into subintervals $[T_i, T_{i+1})$ of non-zero length (with $T_0 = 0$) such that $v^* = v_i$ in $[T_i, T_{i+1})$. Let $C_i = \Phi(v_i, T_i)$. Then $C_{i+1} \leq C_i e^{-(\delta/4)(T_{i+1}-T_i)}$, and the sequence C_i converges to zero exponentially, which proves the convergence of the system to its fixed point.

Alternatively (and more simply), one can use the integral form of the solution to system (1) to deduce that, for any set of bounded initial conditions, there exists a sequence of positive numbers $C_1, C_2, \ldots$ such that, for all i, v,

$$|\hat{R}_i(v,t) - \frac{1}{2}(2\lambda)^{\frac{d^i-1}{d-1}}| \leq C_i t^{i-1} e^{-t} \quad t \geq 0.$$

This again implies the convergence of system (1) to its fixed point.

Using techniques similar to those in [21], we can establish yet another interesting property of system (1):

Lemma 3.1 *Suppose for some u, v, $\hat{R}_i(v,0) \geq \hat{R}_i(u,0)$ for $i = 1, \ldots$. Then $\hat{R}_i(v,t) \geq \hat{R}_i(u,t)$ for $i = 1, \ldots$ for all $t \geq 0$.*

Lemma 3.1 gives that if we introduce a "dummy" vertex v^* with the property that $\hat{R}_i(v^*,0) = \sup_{v\geq 1} \hat{R}_i(v,0)$, for all i, v, then $\hat{R}_i(v^*,t) = \sup_{v\geq 1} \hat{R}_i(v,t)$ at all times $t \geq 0$. This implies that one can redefine $V_i(v,t)$ as $V_i(v,t) = (2\hat{R}_i(v^*,t))^d$. The fixed point of the modified system is the same as in the original, and all the analysis goes through as before. (In fact, it is sufficient to show convergence for the subsystem corresponding to v^*, and then we know that all the other vertices behave no worse than v^*.) This alteration allows one to avoid the technical difficulty associated with ensuring that the limit in $V_i(v,t) = \lim_{n\to\infty} \frac{1}{n-1} \sum_{u\neq v} (\hat{R}_i(v,t) + \hat{R}_i(u,t))^d$ is well-defined for all v, t.

Finally using Kurtz's convergence theory we conclude that the finite stochastic system converges to its steady state distribution, which approaches the fixed point of the deterministic continuous-time system as $n \to \infty$. Note that any solution $R_i(v,t)$ to the equations (1) satisfies $R_i(v,t) \leq R_1(v,0)$ for all $i, v \geq 1$ and for all $t \geq 0$. Hence we can work in the space X of sequences $x_i(v)$ that satisfy $K \geq x_i(v) \geq 0$ for all $i, v \geq 1$, and $x_1(v) \geq x_2(v) \geq \ldots \geq 0$ for all $v \geq 0$, for some positive bound K. In our proof, we define the following metric for the state space of our process: given two states $x = (x_i(v))$ and $y = (y_i(v))$, the distance between them is

$$|x - y| = sup_{v\geq 1, i\geq 1}\{\frac{1}{iv}|x_i(v) - y_i(v)|\}.$$

With this metric at hand, we show that the infinite system of equations satisfies a Lipschitz condition. We then adapt the technique of [8] and [18] to prove a generalization of Kurtz's theorem that is suitable for our setup.

The above analysis works for $\lambda < 1/2$. For a larger (but bounded as $n \to \infty$) injection rate we partition the incoming calls randomly between $\lceil 4\lambda \rceil$ "systems" such that the injection rate to each system is strictly less than $1/2$. We run the balanced allocation routing algorithm independently in each system, though the calls are actually using the same edges. Each system gets an equal share of edge capacity, which is $\Omega(B/\lambda)$. Thus, the blocking probability in the combined system is bounded above by

$$\lceil 4\lambda \rceil (\frac{1}{2})^{d^{(B/4\lambda+1)}} = 2^{-d^{\Theta(B/\lambda)}}.$$

Setting $B = O(\log\log(\lambda TN)/\log d)$ guarantees that with high probability all calls within an interval of length T are routed successfully.

References

[1] Abraham, R., Marsden, J., and Raitu, T. (1983). *Manifolds, Tensor Analysis, and Applications*. Addison-Wesley.

[2] Alon, N., Spencer, J., and Erdös, P. (1992). *The Probabilistic Method*. Wiley, New York.

[3] Ash, G.R., Cardwell, R.H., and Murry, R.P. (1981). Design and Optimization of Networks with Dynamic Routing. *Bell Systems Technical Journal* **60** 1787-1820.

[4] Azar, Y., Broder, A., Karlin, A., and Upfal, E. (1994). Balanced Allocations. In *Proceedings of*

the 26th ACM Symposium on the Theory of Computing, 593-602.

[5] Broder A.Z., Frieze A.M., Lund C., Phillips S., Reingold N. (1995). Balanced Allocations for Tree-Like Inputs. *Information Processing Letters*, 55 329-332.

[6] Cole, R., Maggs, B.M, Meyer auf der Heide, F., Mitzenmacher, M., Richa, A., Schroder, K., Sitaraman, R., and Vocking, B. (1998). Randomized Protocols for Low-Congestion Circuit Routing in Multistage Interconnection Networks. In *Proceedings of the 30th ACM Symposium on the Theory of Computing*, 378-388.

[7] Crametz, J.P. and Hunt, P.J. (1991). A Limit Result Respecting Graph Structure for a Fully Connected Loss Network with Alternative Routing. *The Annals of Applied Probability* **1** 436-444.

[8] Ethier, S.N. and Kurtz, T.G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

[9] Gibbens, R.J., Hunt, P.J., and Kelly, F.P. (1990). Bistability in Communication Networks. In *Disorder in Physical Systems* (G.R. Grimmet and D.J.A. Welsh, eds.) 113-128. Oxford University Press.

[10] Gibbens, R.J., Kelly, F.P., and Key, P.B. (1995). Dynamic Alternative Routing. In *Routing in Communications Networks* (M.E. Steenstrup ed.) 13-47. Prentice Hall.

[11] Gonnet, G.H. (1981). Expected Length of the Longest Probe Sequence in Hash Code Searching. *Journal of the ACM* **28** 289-304.

[12] Hunt, P.J. (1990). Limit Theorems for Stochastic Loss Networks. Ph.D. thesis, Cambridge.

[13] Hunt, P.J. and Laws, C.N. (1993). Asymptotically Optimal Loss Network Control. *Mathematics of Operations Research* **18**(4) 880-900.

[14] Kelly, F.P. (1995). Modelling Communication Networks, Present and Future. *Proceedings of the Royal Society of London. Series A* **444** 1-20.

[15] Kelly, F.P. (1991). Loss Networks. *The Annals of Applied Probability* **1** 319-378.

[16] Kolchin, V.F., Sevastyanov, B.A., and Chistyakov, V.P. (1978). *Random Allocations*. Wiley, New York.

[17] Mees, A.I. (1986). Simple is the Best for Dynamic Routing of Telecommunication. *Nature* **323** 108.

[18] Mitzenmacher, M. (1996). The Power of Two Choices in Randomised Load Balancing. Ph.D. dissertation, Berkeley.

[19] Mitzenmacher, M. (1997). On the Analysis of Randomized Load Balancing Schemes. In *Proceedings of the Ninth Annual Symposium on Parallel Algorithms and Architectures.*

[20] Mitzenmacher, M. (1996). Load Balancing and Density Dependent Jump Markov Processes. In *Proceedings of the 37th IEEE Symposium on Foundations of Computer Science*, 213-222.

[21] Vvedenskaya, N.D., Dobrushin, R.L., and Karpelevich, F.I. (1996). Queueing System with Selection of the Shortest of Two Queues: An Asymptotic Approach. *Problems of Information Transmission* **32** (1) 15-27.