

Asymptotic Optimality of Shortest Path Routing Algorithms

ELI M. GAFNI, MEMBER, IEEE, AND DIMITRI P. BERTSEKAS, FELLOW, IEEE

Abstract—Many communication networks use adaptive shortest path routing. By this we mean that each network link is periodically assigned a length that depends on its congestion level during the preceding period, and all traffic generated between length updates is routed along a shortest path corresponding to the latest link lengths. We show that in certain situations, typical of networks involving a large number of small users and utilizing virtual circuits, this routing method performs optimally in an asymptotic sense. In other cases, shortest path routing can be far from optimal.

I. INTRODUCTION

MOST OF THE presently existing communication networks utilize shortest path routing as evidenced by the recent survey paper [1]. This routing method has gained popularity primarily because it is simple and adequately handles link and node failures. Relatively little is known, however, about the performance of shortest path routing under heavy traffic conditions, since most of the practical experience reported to date relates to networks that are typically lightly loaded, e.g., the ARPANET [2].

It is common to measure optimality of a routing scheme in terms of an objective function of the form

$$\sum_{(i,j)} D_{ij}(F_{ij}) \quad (1)$$

where F_{ij} denotes the arrival rate at the transmission queue of link (i, j) . Here D_{ij} is a convex monotonically increasing function such as, for example,

$$D_{ij}(F_{ij}) = \frac{F_{ij}}{C_{ij} - F_{ij}}, \quad C_{ij}: \text{capacity of } (i, j) \quad (2)$$

which corresponds to the Kleinrock independence assumption [3]. Extensive literature exists on the problem of minimizing (1) subject to known offered traffic for each origin-destination pair [4]–[12]. It makes sense to evaluate routing performance in terms of objective functions such as (1) and (2) in circumstances where the offered traffic statistics change slowly over time, and furthermore, indi-

vidual offered traffic sample functions do not exhibit frequently large and persistent deviations from their averages. A typical situation is a network accommodating a large number of relatively small users for each origin-destination (OD) pair in which a form of the law of large numbers approximately takes hold (see Lemma 1). This paper considers exclusively this type of network, and its conclusions do not apply at all to more dynamic situations characterized by the presence of a few large users that can by themselves overload the network over brief periods of time if left uncontrolled. For such cases an objective function such as (1) is not appropriate, and different methods of analysis are called for (see, e.g., [14] and [15]).

The purpose of the paper is to evaluate the performance of shortest path routing in terms of the objective function (1) when the length of each link (i, j) is periodically calculated as $D'_{ij}(F_{ij})$ —the first derivative of D_{ij} evaluated at the rate F_{ij} at queue (i, j) during the preceding period. The first derivative relation between the link lengths and objective function is motivated by the well-known optimality condition that a routing optimizes (1) if and only if it routes traffic exclusively along paths of minimum first-derivative length (see, e.g., [4] and [13]). It is known that this type of shortest path routing is strictly suboptimal, although it is believed to be close to optimal for lightly loaded networks. Furthermore, for datagram networks shortest path routing is prone to oscillations which can be severe if the length functions D'_{ij} are poorly chosen [17], [18]. Indeed, the original adaptive shortest path algorithm implemented in 1969 on the ARPANET exhibited violent oscillatory behavior which was restrained only after adding a bias to each link length at the expense of a considerable loss of adaptivity [16], [19], [20].

A key feature of a datagram network is that each packet of a user pair is not required to travel on the same path as the preceding packet. Therefore, the “holding time of each communication path” (the maximum time that a user pair will continue to use the path after it is changed due to a shortest path update) is one packet long. As a result, a datagram network reacts very quickly to a shortest path update with all traffic switching to the new shortest paths almost instantaneously.

The situation is quite different in a virtual circuit network where every conversation is assigned a fixed communication path at the time it is first established. There the holding time of the communication path (as loosely described earlier) is often large relative to the shortest path

Manuscript received September 9, 1983; revised March 3, 1986. This work was supported in part by DARPA under Contract ONR/N00014-75-C-1183. This paper was presented in part at the SIGCOMM 83 Symposium on Communications Architectures and Protocols, Austin, TX, March 1983.

E. M. Gafni is with the University of California, Los Angeles, Computer Science Department, Los Angeles, CA 90024, USA.

D. P. Bertsekas is with the Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, Room 35-210, Cambridge, MA 02139, USA.

IEEE Log Number 8610203.

updating period. As a result, the network reaction to a shortest path update is much more gradual since old conversations continue to use their established communication paths and only new conversations are assigned to the most recently calculated shortest paths.

The main result of this paper is that the performance of shortest path routing approaches the optimal achievable by any other method as

$$\frac{\text{shortest path updating period}}{\text{average holding time of the communication path}} \rightarrow 0 \quad (3)$$

and

$$n_w \rightarrow \infty \quad \gamma_w \rightarrow 0 \quad n_w \gamma_w = \text{constant} \quad (4)$$

where n_w is the average number of active conversations for the generic origin-destination (OD) pair w and γ_w is the communication rate of each conversation. Assumptions (3) and (4), together with additional Poisson-like assumptions on the offered traffic statistics, are formulated in the next section. The main result in Section III also provides bounds on the suboptimality of the shortest path method when assumptions (3) and (4) are satisfied only approximately. Roughly speaking, the theorem states that the average value of the cost (1) of the shortest path method converges to a neighborhood of the optimal cost at a natural rate which is independent of how fast the shortest paths are updated. However, the size of the neighborhood is "proportional" to the extent of violation of assumptions (3) and (4).

II. PROBLEM FORMULATION

Consider a network with a set of nodes \mathcal{N} and a set of directed links \mathcal{L} . We are given a set W of ordered node pairs referred to as origin-destination pairs. For each OD pair $w \in W$ we are given a nonempty set of directed paths P_w joining the origin node and the destination node of w . Conversations for each $w \in W$ arrive according to a Poisson process with rate $\bar{\lambda}_w/\epsilon$, where $\bar{\lambda}_w$ is given and ϵ is a positive parameter, the effect of which we wish to study. Each conversation for OD pair w is assigned upon arrival to a path $p \in P_w$, according to a rule to be described shortly, and uses this path for the entire time of its duration, which is assumed to be exponentially distributed with mean $1/\mu_w$. We assume that the Poisson arrival processes and the duration times of conversations are independent, and each path can carry unlimited conversations, so the number of active conversations for each OD pair evolves as in an $M/M/\infty$ queueing system. It follows [21, p. 101] that if $n_w(t)$ is the number of active conversations for w at time t , then its mean and variance satisfy

$$\lim_{t \rightarrow \infty} E\{n_w(t)\} = \frac{\bar{\lambda}_w}{\epsilon \mu_w} \quad \lim_{t \rightarrow \infty} \text{var}\{n_w(t)\} = \frac{\bar{\lambda}_w}{\epsilon \mu_w} \quad (5)$$

The path assignment for each conversation is determined according to the following shortest path rule: at times $t = kT$, $k = 0, 1, \dots$, where $T > 0$ is given, the

length of each link (i, j) is calculated as $d_{ij}[F_{ij}(t)]$, where $F_{ij}(t)$ is the communication rate on link (i, j) given by

$$F_{ij}(t) = \sum_{w \in W} \gamma_w \sum_{\substack{p \in P_w \\ (i, j) \in p}} n_p(t). \quad (6)$$

Here $n_p(t)$ is the number of active conversations assigned on path p at time t , $\sum_{p \in P_w, (i, j) \in p} n_p(t)$ is the total number of conversations of OD pair w using (i, j) at time t , and γ_w is the communication rate per conversation of OD pair w . All conversations of OD pair w arriving at times $t \in [kT, (k+1)T)$ are assigned on a path $p \in P_w$, which is shortest relative to the link lengths $d_{ij}[F_{ij}(kT)]$. (Ties between paths are assumed resolved according to a fixed deterministic rule.)

We assume that $d_{ij}(\cdot)$ is a continuous strictly monotonically increasing function of F_{ij} satisfying $d_{ij}(F_{ij}) \geq 0$ for all $F_{ij} \geq 0$ and

$$|d_{ij}(\bar{F}) - d_{ij}(F)| \leq L|\bar{F} - F|, \quad \forall \bar{F}, F \geq 0, (i, j) \in \mathcal{L}, \quad (7)$$

where L is a given positive constant. This assumption is reasonable once the length function d_{ij} is assumed continuous. In practice, the length function is sometimes taken as discontinuous (e.g., the TYMNET [1]). We do not know whether and in what form our main result holds for this case. Note that (7) is not satisfied when d_{ij} is the first derivative of the function D_{ij} of (2) since this derivative increases without bound as F_{ij} approaches the capacity C_{ij} . As a practical matter this is not a problem since flow control will ordinarily not allow a link flow to get too close to capacity.

Regarding the communication rate γ_w , we assume that it is of the form

$$\gamma_w = \epsilon \bar{\gamma}_w \quad (8)$$

where $\bar{\gamma}_w$ is some constant. Thus we assume in effect that, even though the real communication rate of a conversation will be a random process, the rates γ_w used in the calculation of flows in (6) are obtained by averaging the real rates over a long period of time and over all conversations of OD pair w , so that the variance of γ_w is so small that γ_w can be viewed as a deterministic quantity. Note that for each OD pair w the product

$$(\text{mean arrival rate}) \cdot (\text{communication rate}) = \bar{\lambda}_w \bar{\gamma}_w$$

is independent of ϵ . We wish to study the effect of the parameters ϵ and T on various stochastic processes of interest, particularly as

$$\epsilon \rightarrow 0 \quad T \rightarrow 0.$$

Taking $\epsilon \rightarrow 0$ implies that the arrival rates tend to infinity while the communication rates tend to zero with the products staying constant and approximates a situation where many small conversations exist in the network (cf. (4)). Taking $T \rightarrow 0$ approximates a situation where updating of shortest paths is fast relative to the mean duration time of a conversation (cf. (3)).

The initial numbers $n_p(0)$ of active conversations on each path p are assumed given. These numbers, together with the earlier assumptions on the arrival processes, holding times, and the routing method, completely characterize the statistics of all processes of subsequent interest. Our main result can be proved in essentially the same form if $\{n_p(0)\}$ are random with given mean and variance (see Lemma 1 in the Appendix).

We will investigate the behavior of the processes $F(t) = \{F_{ij}(t) | (i, j) \in \mathcal{L}\}$ and

$$D[F(t)] = \sum_{(i, j) \in \mathcal{L}} D_{ij}[F_{ij}(t)]$$

where D_{ij} is some function such that

$$d_{ij}(F_{ij}) = D'_{ij}(F_{ij}) \triangleq \text{first derivative of } D_{ij} \text{ at } F_{ij}. \quad (9)$$

Note that in view of our earlier assumptions $d_{ij}(\cdot)$ uniquely defines $D_{ij}(\cdot)$ as a strictly convex monotonically increasing function up to an additive constant.

A lower bound exists to the value of $E\{D[F(t)]\}$ achievable in the long run by any rule for assigning conversations to paths. This is

$$D^* = \min_{F \in \mathcal{F}} D(F) \quad (10)$$

where \mathcal{F} is the set of all total flows $F = \{F_{ij} | (i, j) \in \mathcal{L}\}$ of the form

$$F_{ij} = \sum_{w \in W} \sum_{\substack{p \in P_w \\ (i, j) \in P}} x_p, \quad \forall (i, j) \in \mathcal{L} \quad (11)$$

where x_p are any nonnegative scalars satisfying

$$\sum_{p \in P_w} x_p = \frac{\bar{\lambda}_w \bar{\gamma}_w}{\mu_w}, \quad \forall w \in W. \quad (12)$$

In other words, \mathcal{F} is the set of all possible average total link rates resulting from the long-term average input traffic rate $\bar{\lambda}_w \bar{\gamma}_w / \mu_w$ at each OD pair w (cf. (5), (8)). Note that

and the following fact (cf. (5), (8)):

$$E\{n_w(t)\gamma_w\} \rightarrow \frac{\bar{\lambda}_w \bar{\gamma}_w}{\mu_w}, \quad \text{as } t \rightarrow \infty.$$

Our main result is that, as $\epsilon \rightarrow 0$, $T \rightarrow 0$, and $t \rightarrow \infty$, the expected cost $E\{D[F(t)]\}$, corresponding to the shortest path rule, converges to the lower bound D^* while $F(t)$ converges in mean square to the unique F^* that achieves the minimum in the deterministic optimal routing problem (10).

III. MAIN RESULT

We first introduce the following notation:

$$\begin{aligned} x_p(t) &\triangleq \epsilon \bar{\gamma}_w n_p(t), && \text{the communication rate on path } p \text{ at time } t, \\ r_w(t) &\triangleq \sum_{p \in P_w} x_p(t), && \text{the total input rate of OD pair } w \text{ at } t, \\ \bar{r}_w &\triangleq \bar{\lambda}_w \bar{\gamma}_w / \mu_w, && \text{the long-term average input rate of } w, \\ \bar{r} &\triangleq \max_w \{\bar{r}_w\}, \\ R_w &\triangleq |r_w(0) - \bar{r}_w|, && \text{the initial deviation of } r_w \text{ from its long term average,} \\ R &\triangleq \max_w \{R_w\}, \\ \mu &\triangleq \min_w \{\mu_w\}, \\ M &\triangleq \max_w \{\mu_w\}, \\ \bar{\gamma} &\triangleq \max_w \{\bar{\gamma}_w\}. \end{aligned}$$

Theorem: Positive constants c_1 and c_2 exist (which depend only on the network topology, the products $\bar{\lambda}_w \bar{\gamma}_w$, and the length functions d_{ij}) such that the total link rate vector $F(t)$ corresponding to shortest path routing satisfies, for all $t = kT$, $k = 0, 1, \dots$,

$$\begin{aligned} -c_1 R e^{-\mu t} &\leq E\{D[F(t)]\} - D^* \\ &\leq e^{-\mu t} [D[F(0)] - D^*] \\ &\quad + c_2 [a(\epsilon, T) + b(\epsilon, T) t e^{-\mu t}] \end{aligned} \quad (13)$$

where

$$a(\epsilon, T) = \bar{r} \left\{ \frac{(\epsilon \bar{\gamma} \bar{r} + \sqrt{\epsilon \bar{\gamma} (\bar{r} + R)})(e^{-\mu T} - e^{-MT})}{\bar{r}(1 - e^{-\mu T})} + 2\epsilon \bar{\gamma} + (1 - e^{-\mu T})(4\bar{r} + \epsilon \bar{\gamma}) \right\} \quad (14)$$

$$b(\epsilon, T) = R \left\{ \frac{(\epsilon \bar{\gamma} + R + 1)(e^{-\mu T} - e^{-MT}) + (1 - e^{-\mu T})[\epsilon \bar{\gamma} + (1 - e^{-\mu T})(4\bar{r} + R + \epsilon \bar{\gamma})]}{T e^{-\mu T}} \right\}. \quad (15)$$

the problem in (10) is the usual deterministic multicommodity flow problem that has been studied extensively in connection with optimal routing [4]–[13]. For any routing rule the inequality

$$D^* \leq \liminf_{t \rightarrow \infty} E\{D[F(t)]\}$$

follows from the fact

$$D[E\{F(t)\}] \leq E\{D[F(t)]\}, \quad \forall t \geq 0$$

which holds by the convexity of D , Jensen's inequality,

Furthermore,

$$\lim_{\substack{\epsilon \rightarrow 0 \\ T \rightarrow 0}} \left(\limsup_{t \rightarrow \infty} E\{D[F(t)]\} \right) = D^*.$$

If, in addition, we assume that for some $l > 0$ the length functions d_{ij} satisfy

$$l|\bar{F} - F| \leq |d_{ij}(\bar{F}) - d_{ij}(F)|,$$

$$\forall \bar{F}, F \geq 0, (i, j) \in \mathcal{L},$$

then

$$\lim_{\substack{\epsilon \rightarrow 0 \\ T \rightarrow 0 \\ t \rightarrow \infty}} E\{|F_{ij}(t) - F_{ij}^*|^2\} = 0, \quad \forall (i, j) \in \mathcal{L},$$

where F^* is the unique solution of the deterministic optimal routing problem (10).

The proof of the theorem is given in the Appendix. The idea of the proof is based on relations of shortest path routing with the flow deviation (or Frank-Wolfe) method [7] for solving problem (10) (see [13]). However, the proof here is complicated by the fact that we are dealing with a stochastic optimization problem, while the flow deviation method deals with a deterministic problem. A simpler version of the theorem that assumes that ϵ and T are so small that the path rates can be obtained as solutions of differential equations is given in [22].

The main implication of (13) is that, as $t \rightarrow \infty$, $E\{D[F(t)]\}$ comes within $c_2 a(\epsilon, T)$ of being optimal. Thus $c_2 a(\epsilon, T)$ may be viewed as the long-term deviation from optimality of shortest path routing. The key fact is that $a(\epsilon, T) \rightarrow 0$ as $\epsilon \rightarrow 0$ and $T \rightarrow 0$. The rate at which $E\{D[F(t)]\}$ approaches its long-term limit depends on the largest average holding time $1/\mu$. There are three terms here. The first term $e^{-\mu t}[D[F(0)] - D^*]$ is proportional to the initial deviation from optimality. The other two terms are proportional to the initial deviation R of the initial OD pair rates $r_w(0)$ from their long-term averages \bar{r}_w .

The three transient terms in (13) characterize the rate of convergence of the algorithm. Of these terms the slowest is the one involving $te^{-\mu t}$. Since for any $\delta > 0$ we have $te^{-\mu t} \leq (1/\delta e)e^{-(\mu-\delta)t}$, we see that even this term decays almost as fast as $e^{-\mu t}$. Thus we can conclude that at worst $E\{D[F(t)]\}$ converges to its long-term average almost like $e^{-\mu t}$ —a linear rate which is independent of ϵ and T . For specific problems the actual rate of convergence can be considerably faster, and the bound $e^{-\mu t}$ is not necessarily tight. However, $E\{D[F(t)]\}$ cannot converge to D^* much faster than $e^{-\mu t}$ since we know that the rate of change of $F(t)$ is constrained by the rate at which the number of old conversations on any path can decrease due to termination, and this rate is precisely $e^{-\mu t}$. Thus, for example, if $D_{ij}(F_{ij})$ is quadratic in F_{ij} , the rate of convergence of $E\{D[F(t)]\}$ cannot be faster than $e^{-2\mu t}$, while in the extreme case where $D_{ij}(F_{ij})$ is linear in F_{ij} , the rate of convergence cannot be faster than $e^{-\mu t}$. Therefore, there is little margin for improvement of our rate of convergence result. The conclusion is that the largest average duration $1/\mu$ of a conversation is a fundamental limiting factor in the performance of the shortest path algorithm. When $1/\mu$ is large, the algorithm tends to converge slowly to a neighborhood of the optimum. This is a manifestation of the intuitively clear fact that the routing algorithm cannot perform well if poorly routed conversations last for a long time and no provisions are made to reroute them.

ACKNOWLEDGMENT

Thanks are due to A. Segall who suggested the problem to the first author.

APPENDIX PROOF OF THE THEOREM

For brevity we use the following notation in addition to that given in the beginning of Section III:

$$\begin{aligned} n_p^k &\triangleq n_p(kT) & x_p^k &\triangleq x_p(kT) \\ r_w^k &\triangleq r_w(kT) & F_{ij}^k &\triangleq F_{ij}(kT) \\ x^k &\triangleq \{x_p^k | p \in P_w, w \in W\} & F^k &\triangleq \{F_{ij}^k | (i, j) \in \mathcal{L}\}. \end{aligned}$$

We first prove some helpful lemmas. The first lemma gives some basic facts about the transient behavior of various processes of interest. In particular, it shows that, as $\epsilon \rightarrow 0$, the processes $x_p(t)$ and $r_w(t)$ behave asymptotically as deterministic processes.

Lemma 1: For all $t \geq 0$ and $w \in W$

$$E\{r_w(t)\} = \bar{r}_w + e^{-\mu_w t}[r_w(0) - \bar{r}_w] \quad (A1)$$

$$\text{var}\{r_w(t)\} = \epsilon \bar{\gamma}_w (1 - e^{-\mu_w t}) [\bar{r}_w + e^{-\mu_w t} r_w(0)]. \quad (A2)$$

Furthermore, for each $w \in W$, if $p_k \in P_w$ is the shortest path used for routing in the interval $[kT, (k+1)T)$, we have for all $t \in [kT, (k+1)T]$

$$E\{x_p(t) | x_p^k\} = \begin{cases} e^{-\mu_w(t-kT)} x_p^k, & p \neq p_k \\ \bar{r}_w + e^{-\mu_w(t-kT)} (x_p^k - \bar{r}_w), & p = p_k \end{cases} \quad (A3)$$

$$\text{var}\{x_p(t) | x_p^k\} = \begin{cases} \epsilon \bar{\gamma}_w [1 - e^{-\mu_w(t-kT)}] e^{-\mu_w(t-kT)} x_p^k, & p \neq p_k \\ \epsilon \bar{\gamma}_w [1 - e^{-\mu_w(t-kT)}] [\bar{r}_w + e^{-\mu_w(t-kT)} x_p^k], & p = p_k \end{cases} \quad (A4)$$

Proof: Consider an $M/M/\infty$ queueing system with arrival rate Λ and service rate $1/M$. The probabilities $P_k(t)$ of k customers in the system at time t satisfy the differential equations [21, p. 59, 101]

$$\begin{aligned} \dot{P}_0 &= -\Lambda P_0 + M P_1 \\ \dot{P}_k &= -(\Lambda + kM) P_k + \Lambda P_{k-1} + (k+1) M P_{k+1}, \\ & \quad k = 1, 2, \dots \end{aligned} \quad (A5)$$

Let $N(t) = \sum_{k=1}^{\infty} k P_k(t)$ and $\sigma(t) = \sum_{k=0}^{\infty} [k - N(t)]^2 P_k(t)$ be the expected value and variance, respectively, of the number in the system. By multiplying (A5) by k and adding, we obtain by straightforward calculation the differential equation

$$\dot{N} = -MN + \Lambda. \quad (A6)$$

In addition, by multiplying (A5) by $(k - N)^2$, adding, and taking into account the fact $\dot{\sigma} = \sum_{k=0}^{\infty} (k - N)^2 \dot{P}_k$, we obtain

$$\dot{\sigma} = -2M\sigma + MN + \Lambda. \quad (A7)$$

The solutions of the linear differential equations (A6) and (A7) can be calculated by the variations of constants formula. They are

$$N(t) = \frac{\Lambda}{M} + e^{-Mt} \left[N(0) - \frac{\Lambda}{M} \right] \quad (A8)$$

$$\sigma(t) = e^{-2Mt} \sigma(0) + (1 - e^{-Mt}) \left[\frac{\Lambda}{M} + e^{-Mt} N(0) \right]. \quad (A9)$$

Applying (A8) for $M = \mu_w$ and $\Lambda = \bar{\lambda}_w/\epsilon$ and multiplying by $\epsilon\bar{\gamma}_w$ yields (A1). Applying (A9) for $M = \mu_w$, $\Lambda = \bar{\lambda}_w/\epsilon$, and $\sigma(0) = 0$ and multiplying by $\epsilon^2\bar{\gamma}_w^2$ yields (A2). A similar application of (A8) and (A9) yields (A3) and (A4). Q.E.D.

Note that from (A1) and (A2) we obtain the useful relations

$$|E\{r_w(t)\} - \bar{r}_w| \leq e^{-\mu_w t} R_w \leq e^{-\mu t} R \quad (\text{A10})$$

$$\begin{aligned} \text{var}\{r_w(t)\} &\leq \epsilon\bar{\gamma}(1 - e^{-\mu_w t})[(1 + e^{-\mu_w t})\bar{r}_w \\ &\quad + e^{-\mu_w t}[r_w(0) - \bar{r}_w]] \\ &\leq \epsilon\bar{\gamma}(\bar{r} + e^{-\mu t}R). \end{aligned} \quad (\text{A11})$$

The proof of Theorem 1 would be considerably simplified if the average holding time of a conversation were independent of the OD pair, i.e., $\mu_w = \mu = M$ for all $w \in W$. In fact, the reader may first wish to go through the proof assuming this. To cope with the case where $\mu \neq M$, we will need to introduce the following "normalized" processes

$$\tilde{x}_p(t) = \frac{x_p(t)\bar{r}_w}{r_w(t)}, \quad \forall w \in W, p \in P_w \quad (\text{A12a})$$

$$\tilde{F}_{ij}(t) = \sum_{w \in W} \sum_{p \in P_w} \tilde{x}_p(t), \quad \forall (i, j) \in \mathcal{L}. \quad (\text{A12b})$$

We denote

$$\tilde{x}_p^k \triangleq \tilde{x}_p(kT) \quad \tilde{F}_{ij}^k \triangleq \tilde{F}_{ij}(kT). \quad (\text{A12c})$$

Using the fact $\tilde{x}_p(t) \leq \bar{r}_w$, (A1), and (A11), we have

$$\begin{aligned} E\{|\tilde{x}_p(t) - x_p(t)|^2\} &= E\left\{\left|\tilde{x}_p(t)\left[1 - \frac{r_w(t)}{\bar{r}_w}\right]\right|^2\right\} \\ &\leq E\{|\bar{r}_w - r_w(t)|^2\} \\ &\leq E\left\{E\{r_w(t)\} - e^{-\mu_w t}\right. \\ &\quad \cdot [r_w(0) - \bar{r}_w] - r_w(t)\}^2 \\ &\leq \text{var}\{r_w(t)\} + e^{-2\mu_w t} R_w^2 \\ &\leq \epsilon\bar{\gamma}(\bar{r} + e^{-\mu t}R) + e^{-2\mu t} R^2. \end{aligned}$$

Since \tilde{F}_{ij} and F_{ij} are sums of \tilde{x}_p and x_p , respectively, we obtain, for some constant α_{ij} ,

$$E\{|\tilde{F}_{ij}(t) - F_{ij}(t)|^2\} \leq \alpha_{ij}[\epsilon\bar{\gamma}(\bar{r} + e^{-\mu t}R) + e^{-2\mu t}R^2]. \quad (\text{A13})$$

The next lemma provides a basic estimate.

Lemma 2: For every vector $\tilde{F} \in \mathcal{F}$ and every other total link rate vector F (not necessarily in \mathcal{F}), the following holds:

$$D(\tilde{F}) \leq D(F) + B \sum_{(i,j)} |\tilde{F}_{ij} - F_{ij}| \quad (\text{A14})$$

where B is an upper bound for $d_{ij}(\tilde{F}_{ij})$ over $(i, j) \in \mathcal{L}$ and $\tilde{F} \in \mathcal{F}$.

Proof: We have by the convexity of D that

$$\begin{aligned} D(F) &\geq D(\tilde{F}) + \sum_{(i,j)} d_{ij}(\tilde{F}_{ij})(F_{ij} - \tilde{F}_{ij}) \\ &\geq D(\tilde{F}) - B \sum_{(i,j)} |F_{ij} - \tilde{F}_{ij}|. \end{aligned} \quad \text{Q.E.D.}$$

Proof of Theorem 1: We first show the left side of (13). Let $\{x_p^*(t)\}$ be a set of path rates that solve the following deterministic multicommodity flow problem:

$$\begin{aligned} &\text{minimize } D(F) \\ &\text{subject to } F_{ij} = \sum_{w \in W} \sum_{\substack{p \in P_w \\ (i,j) \in p}} x_p \\ &\quad \sum_{p \in P_w} x_p = E\{r_w(t)\}, \quad \forall w \in W \\ &\quad x_p \geq 0, \quad \forall p \in P_w, w \in W. \end{aligned} \quad (\text{A15})$$

Let $F^*(t)$ be the vector of corresponding total link rates, i.e.,

$$F_{ij}^*(t) = \sum_{w \in W} \sum_{\substack{p \in P_w \\ (i,j) \in p}} x_p^*(t).$$

Define the normalized rates

$$\begin{aligned} \hat{x}_p(t) &= \frac{\bar{r}_w}{E\{r_w(t)\}} x_p^*(t) \\ \hat{F}_{ij}(t) &= \sum_{w \in W} \sum_{\substack{p \in P_w \\ (i,j) \in p}} \hat{x}_p(t). \end{aligned} \quad (\text{A16})$$

Since $\hat{F}(t) = \{\hat{F}_{ij}(t) \in \mathcal{F}\}$, we have, using (A14),

$$\begin{aligned} D^* &\leq D[\hat{F}(t)] \leq D[F^*(t)] + B \sum_{(i,j)} |\hat{F}_{ij}(t) - F_{ij}^*(t)| \\ &\leq D[E\{F(t)\}] + B \sum_{(i,j)} |\hat{F}_{ij}(t) - F_{ij}^*(t)| \\ &\leq E\{D[F(t)]\} + B \sum_{(i,j)} |\hat{F}_{ij}(t) - F_{ij}^*(t)| \end{aligned} \quad (\text{A17})$$

where the last step follows using Jensen's inequality. From (A16) we have, using the fact $\hat{x}_p(t) \leq \bar{r}_w$ and (A10),

$$|\hat{x}_p(t) - x_p^*(t)| = \left| \frac{\hat{x}_p(t)}{\bar{r}_w} [\bar{r}_w - E\{r_w(t)\}] \right| \leq Re^{-\mu t}.$$

Since $\hat{F}_{ij}(t)$ and $F_{ij}^*(t)$ consist of sums of $\hat{x}_p(t)$ and $x_p^*(t)$, respectively, we have for some constants β_{ij} ,

$$|\hat{F}_{ij}(t) - F_{ij}^*(t)| \leq \beta_{ij} Re^{-\mu t}. \quad (\text{A18})$$

Taking $c_1 = B \sum_{(i,j)} \beta_{ij}$, we obtain from (A17) and (A18)

$$D^* \leq E\{D[F(t)]\} + c_1 Re^{-\mu t},$$

and the left side of (13) is proved.

To prove the right side of (13), we first fix k and consider times $t \in [kT, (k+1)T]$. We have, using (7) and Taylor's theorem,

$$\begin{aligned} D_{ij}[F_{ij}(t)] &= D_{ij}(F_{ij}^k) + d_{ij}(F_{ij}^k)[F_{ij}(t) - F_{ij}^k] \\ &\quad + \int_0^1 \{d_{ij}[F_{ij}^k + \alpha(F_{ij}(t) - F_{ij}^k)] \\ &\quad - d_{ij}(F_{ij}^k)\} [F_{ij}(t) - F_{ij}^k] d\alpha \\ &\leq D_{ij}(F_{ij}^k) + d_{ij}(F_{ij}^k)[F_{ij}(t) - F_{ij}^k] \\ &\quad + \frac{L}{2} |F_{ij}(t) - F_{ij}^k|^2. \end{aligned}$$

By summing over all links (i, j) , we obtain

$$D[F(t)] \leq D(F^k) + \sum_{(i,j)} d_{ij}(F_{ij}^k) [F_{ij}(t) - F_{ij}^k] + \frac{L}{2} \sum_{(i,j)} |F_{ij}(t) - F_{ij}^k|^2. \quad (\text{A19})$$

We derive an upper bound for the expected value of each of the last two terms in (A19).

Denote by d_p^k the length of path p corresponding to the link flows F_{ij}^k . We have

$$d_p^k = \sum_{(i,j) \in p} d_{ij}(F_{ij}^k), \quad \forall w \in W, p \in P_w,$$

and it follows that

$$\begin{aligned} & \sum_{(i,j)} d_{ij}(F_{ij}^k) [F_{ij}(t) - F_{ij}^k] \\ &= \sum_{(i,j)} d_{ij}(F_{ij}^k) \sum_{w \in W} \sum_{\substack{p \in P_w \\ (i,j) \in p}} [x_p(t) - x_p^k] \\ &= \sum_{w \in W} \sum_{p \in P_w} d_p^k [x_p(t) - x_p^k]. \end{aligned} \quad (\text{A20})$$

Let $p_k \in P_w$ be the shortest path used for routing in $[kT, (k+1)T]$, and define

$$\bar{x}_p^k = \begin{cases} 0, & p \neq p_k \\ \bar{r}_w, & p = p_k \end{cases}. \quad (\text{A21})$$

By taking the conditional expectation in (A20) and using (A3), we find that

$$\begin{aligned} & E \left\{ \sum_{(i,j)} d_{ij}(F_{ij}^k) [F_{ij}(t) - F_{ij}^k] | x^k \right\} \\ &= \sum_{w \in W} \sum_{p \in P_w} d_p^k [E \{ x_p(t) | x_p^k \} - x_p^k] \\ &= \sum_{w \in W} [1 - e^{-\mu_w(t-kT)}] \sum_{p \in P_w} d_p^k (\bar{x}_p^k - x_p^k) \\ &= \sum_{w \in W} [1 - e^{-\mu_w(t-kT)}] \left[\sum_{p \in P_w} d_p^k (\bar{x}_p^k - \tilde{x}_p^k) \right. \\ & \quad \left. + \sum_{p \in P_w} d_p^k (\tilde{x}_p^k - x_p^k) \right] \end{aligned} \quad (\text{A22})$$

where \tilde{x}_p^k is given by (A12). Since $\sum_{p \in P_w} \bar{x}_p^k = \sum_{p \in P_w} \tilde{x}_p^k = \bar{r}_w$ and, for each w , p_k is the shortest path, we obtain the following by using (A21):

$$\sum_{p \in P_w} d_p^k \bar{x}_p^k \leq \sum_{p \in P_w} d_p^k \tilde{x}_p^k$$

so that (A22) can be strengthened to yield the following:

$$\begin{aligned} & E \left\{ \sum_{(i,j)} d_{ij}(F_{ij}^k) [F_{ij}(t) - F_{ij}^k] | x^k \right\} \\ & \leq [1 - e^{-\mu(t-kT)}] \sum_{w \in W} \sum_{p \in P_w} d_p^k (\bar{x}_p^k - \tilde{x}_p^k) \\ & \quad + \sum_{w \in W} [1 - e^{-\mu_w(t-kT)}] \sum_{p \in P_w} d_p^k (\tilde{x}_p^k - x_p^k) \\ & = [1 - e^{-\mu(t-kT)}] \sum_{w \in W} \sum_{p \in P_w} d_p^k (\bar{x}_p^k - \tilde{x}_p^k) \\ & \quad + \sum_{w \in W} [e^{-\mu(t-kT)} - e^{-\mu_w(t-kT)}] \sum_{p \in P_w} d_p^k (\tilde{x}_p^k - x_p^k). \end{aligned} \quad (\text{A23})$$

We proceed to bound each of the two terms in the right side of (A23).

Let $\{x_p^* | w \in W, p \in P_w\}$ be any set of path flows minimizing $D(F)$ over \mathcal{F} , i.e., any $x_p^* \geq 0$ such that

$$F_{ij}^* = \sum_{w \in W} \sum_{\substack{p \in P_w \\ (i,j) \in p}} x_p^*, \quad \forall (i,j) \in \mathcal{L}.$$

Since for each w the shortest path is p_k and $\sum_{p \in P_w} x_p^* = \sum_{p \in P_w} \bar{x}_p^k = \bar{r}_w$, we have

$$\sum_{p \in P_w} d_p^k (\bar{x}_p^k - x_p^k) \leq \sum_{p \in P_w} d_p^k (x_p^* - x_p^k) \quad (\text{A24})$$

while similarly as earlier we have (cf. (A20))

$$\sum_{w \in W} \sum_{p \in P_w} d_p^k (x_p^* - x_p^k) = \sum_{(i,j)} d_{ij}(F_{ij}^k) (F_{ij}^* - F_{ij}^k). \quad (\text{A25})$$

Since D is convex, we obtain

$$\begin{aligned} \sum_{(i,j)} d_{ij}(F_{ij}^k) (F_{ij}^* - F_{ij}^k) &\leq D(F^*) - D(F^k) \\ &= D^* - D(F^k). \end{aligned} \quad (\text{A26})$$

By combining (A24)–(A26), we see that

$$\sum_{w \in W} \sum_{p \in P_w} d_p^k (\bar{x}_p^k - x_p^k) \leq D^* - D(F^k), \quad (\text{A27})$$

which provides a bound for the first term on the right in (A23).

To obtain a bound for the second term on the right of (A23), we write

$$\begin{aligned} \sum_{p \in P_w} d_p^k (\tilde{x}_p^k - x_p^k) &= \sum_{p \in P_w} (d_p^k - \tilde{d}_p^k) (\tilde{x}_p^k - x_p^k) \\ & \quad + \sum_{p \in P_w} \tilde{d}_p^k (\tilde{x}_p^k - x_p^k) \end{aligned} \quad (\text{A28})$$

where \tilde{d}_p^k is the length of path p if each flow x_p^k is replaced by \tilde{x}_p^k , i.e.,

$$\tilde{d}_p^k = \sum_{(i,j) \in p} d_{ij}(\tilde{F}_{ij}^k).$$

By using (7) and (A13), we can easily see that for some constant $\xi > 0$,

$$\begin{aligned} & E \left\{ \sum_{p \in P_w} (d_p^k - \tilde{d}_p^k) (\tilde{x}_p^k - x_p^k) \right\} \\ & \leq LE \left\{ \sum_{p \in P_w} \sum_{(i,j) \in p} |\tilde{F}_{ij}^k - F_{ij}^k|^2 \right\} \\ & \leq \xi [\epsilon \bar{r} + e^{-\mu k T R} + e^{-2\mu k T R^2}]. \end{aligned}$$

By using (A12), we have

$$\begin{aligned} \sum_{p \in P_w} \tilde{d}_p^k (\tilde{x}_p^k - x_p^k) &= \frac{\bar{r}_w - r_w^k}{r_w^k} \sum_{p \in P_w} \tilde{d}_p^k x_p^k \\ &\leq \frac{|\bar{r}_w - r_w^k| B}{r_w^k} \sum_{p \in P_w} x_p^k = B |\bar{r}_w - r_w^k| \end{aligned}$$

where B is the constant defined in Lemma 2.

We have

$$\begin{aligned} E \{ |\bar{r}_w - r_w^k| \} &\leq E \{ |\bar{r}_w - E \{ r_w^k \}| \} + E \{ |E \{ r_w^k \} - r_w^k| \} \\ &\leq E \{ |\bar{r}_w - E \{ r_w^k \}| \} + \sqrt{\text{var} \{ r_w^k \}} \end{aligned}$$

where the last step follows from Jensen's inequality. Therefore,

using (A10) and (A11), we obtain

$$\begin{aligned} E\{|r_w^k - r_w^k|\} &\leq e^{-\mu t} R + \sqrt{\epsilon \bar{\gamma}(\bar{r} + e^{-\mu t} R)}, \\ &\leq e^{-\mu t} R + \sqrt{\epsilon \bar{\gamma}(\bar{r} + R)}, \end{aligned}$$

and

$$E\left\{\sum_{p \in P_w} \tilde{d}_p^k(\tilde{x}_p^k - x_p^k)\right\} \leq B[e^{-\mu k T} R + \sqrt{\epsilon \bar{\gamma}(\bar{r} + R)}].$$

By taking expectation over x^k in (A28) and using the foregoing inequalities, we obtain the following for some constant $\xi > 0$:

$$\begin{aligned} E\left\{\sum_{w \in W} [e^{-\mu(t-kT)} - e^{-\mu_w(t-kT)}] \sum_{p \in P_w} d_p^k(\tilde{x}_p^k - x_p^k)\right\} \\ \leq \xi [e^{-\mu(t-kT)} - e^{-M(t-kT)}] [\epsilon \bar{\gamma}(\bar{r} + e^{-\mu k T}) \\ + e^{-2\mu k T} R^2 + e^{-\mu k T} R + \sqrt{\epsilon \bar{\gamma}(\bar{r} + R)}]. \quad (\text{A28}) \end{aligned}$$

By combining (A23), (A27), and (A28) and taking expectation over x^k , we obtain the following for some constant β_1 :

$$\begin{aligned} E\left\{\sum_{(i,j)} d_{ij}(F_{ij}^k) [F_{ij}(t) - F_{ij}^k]\right\} \\ \leq [1 - e^{-\mu(t-kT)}] [D^* - E\{D(F^k)\}] \\ + \beta_1 [e^{-\mu(t-kT)} - e^{-M(t-kT)}] [\epsilon \bar{\gamma}(\bar{r} + e^{-\mu k T} R) \\ + e^{-2\mu k T} R^2 + e^{-\mu k T} R + \sqrt{\epsilon \bar{\gamma}(\bar{r} + R)}], \quad (\text{A29}) \end{aligned}$$

which provides the desired bound on the expected value of the next to last term in (A19).

We now bound the expected value of the last term in (A19). Since F_{ij}^k and $F_{ij}(t)$ are sums of path flows x_p^k and $x_p(t)$, respectively, we have that a constant θ exists such that

$$\sum_{(i,j)} |F_{ij}(t) - F_{ij}^k| \leq \theta \sum_{w \in W} \sum_{p \in P_w} |x_p(t) - x_p^k|. \quad (\text{A30})$$

We have

$$\begin{aligned} E\{|x_p(t) - x_p^k|^2 | x_p^k\} &= \text{var}\{x_p(t) | x_p^k\} \\ &+ [x_p^k - E\{x_p(t) | x_p^k\}]^2, \end{aligned}$$

and using Lemma 1, we obtain

$$\begin{aligned} \sum_{p \in P_w} E\{|x_p(t) - x_p^k|^2 | x_p^k\} \\ = \epsilon \bar{\gamma}_w [1 - e^{-\mu_w(t-kT)}] [\bar{r}_w + e^{-\mu_w(t-kT)} r_w^k] \\ + [1 - e^{-\mu(t-kT)}]^2 \left[(\bar{r}_w - x_p^k)^2 + \sum_{\substack{p \in P_w \\ p \neq p_k}} (x_p^k)^2 \right] \\ \leq [1 - e^{-\mu(t-kT)}] [\epsilon \bar{\gamma}_w (\bar{r}_w + r_w^k) \\ + (1 - e^{-\mu T}) (\bar{r}_w + r_w^k)^2]. \end{aligned}$$

By taking expectation over x^k and using (A10) and (A11), we

obtain

$$\begin{aligned} \sum_{p \in P_w} E\{|x_p(t) - x_p^k|^2\} \\ \leq [1 - e^{-\mu(t-kT)}] \{ \epsilon \bar{\gamma}(\bar{r}_w + E\{r_w^k\}) \\ + (1 - e^{-\mu T}) [\bar{r}_w^2 + 2\bar{r}_w E\{r_w^k\} \\ + (E\{r_w^k\})^2 + \text{var}\{r_w^k\}] \} \\ \leq [1 - e^{-\mu(t-kT)}] \{ \epsilon \bar{\gamma}(2\bar{r} + e^{-\mu k T} R) \\ + (1 + e^{-\mu T}) [(2\bar{r} + e^{-\mu k T} R)^2 + \epsilon \bar{\gamma}(\bar{r} + e^{-\mu k T} R)] \}. \quad (\text{A31}) \end{aligned}$$

We now combine (A19) and (A29)–(A31) to obtain the following for all $t \in [kT, (k+1)T]$ and some positive constant β_2 :

$$\begin{aligned} E\{D[F(t)]\} - D^* &\leq e^{-\mu(t-kT)} [E\{D(F^k)\} - D^*] \\ &+ \beta_1 [e^{-\mu(t-kT)} - e^{-M(t-kT)}] \\ &\cdot [\epsilon \bar{\gamma}(\bar{r} + e^{-\mu k T} R) + e^{-2\mu k T} R^2 \\ &+ e^{-\mu k T} R + \sqrt{\epsilon \bar{\gamma}(\bar{r} + R)}] \\ &+ \beta_2 [1 - e^{-\mu(t-kT)}] \\ &\cdot \{ \epsilon \bar{\gamma}(2\bar{r} + e^{-\mu k T} R) + (1 - e^{-\mu T}) \\ &\cdot [(2\bar{r} + e^{-\mu k T} R)^2 \\ &+ \epsilon \bar{\gamma}(\bar{r} + e^{-\mu k T} R)] \}. \quad (\text{A32}) \end{aligned}$$

By applying this inequality for $t = (k+1)T$, setting $c_2 = \max\{\beta_1, \beta_2\}$, and collecting terms, we obtain

$$\begin{aligned} E\{D(F^{k+1})\} - D^* &\leq e^{-\mu T} [E\{D(F^k)\} - D^*] \\ &+ c_2 [\bar{a}(\epsilon, T) + \bar{b}(\epsilon, T) e^{-\mu k T}] \quad (\text{A32}) \end{aligned}$$

where

$$\begin{aligned} \bar{a}(\epsilon, T) &= \bar{r} \left\{ (e^{-\mu T} - e^{-M T}) \left(\epsilon \bar{\gamma} + \frac{\sqrt{\epsilon \bar{\gamma}(\bar{r} + R)}}{\bar{r}} \right) \right. \\ &\left. + (1 - e^{-\mu T}) [2\epsilon \bar{\gamma} + (1 - e^{-\mu T})(4\bar{r} + \epsilon \bar{\gamma})] \right\} \quad (\text{A33}) \end{aligned}$$

$$\begin{aligned} \bar{b}(\epsilon, T) &= R \{ (e^{-\mu T} - e^{-M T}) (\epsilon \bar{\gamma} + R + 1) + (1 - e^{-\mu T}) \\ &+ [\epsilon \bar{\gamma} + (1 - e^{-\mu T})(4\bar{r} + R + \epsilon \bar{\gamma})] \}. \quad (\text{A34}) \end{aligned}$$

By applying (A32) repeatedly for k equal to zero up to $(k-1)$, we obtain

$$\begin{aligned} E\{D(F^k)\} - D^* &\leq e^{-\mu k T} [D(F^0) - D^*] \\ &+ c_2 \left[\frac{\bar{a}(\epsilon, T)}{1 - e^{-\mu T}} + \frac{\bar{b}(\epsilon, T)}{T e^{-\mu T}} k T e^{-\mu k T} \right], \end{aligned}$$

which is the desired right side of relation (13) (compare (14) and (15) with (A33) and (A34)).

Since

$$\lim_{\substack{\epsilon \rightarrow 0 \\ T \rightarrow 0}} \frac{\bar{a}(\epsilon, T)}{1 - e^{-\mu T}} = 0 \quad \lim_{\substack{\epsilon \rightarrow 0 \\ T \rightarrow 0}} \frac{\bar{b}(\epsilon, T)}{T e^{-\mu T}} < \infty,$$

we see that $E\{D[F(kT)]\} \rightarrow D^*$ as $\epsilon \rightarrow 0$, $T \rightarrow 0$, and $kT \rightarrow \infty$. It follows from (A31) that $E\{D[F(t)]\} \rightarrow D^*$ as $\epsilon \rightarrow 0$, $T \rightarrow 0$, and $t \rightarrow \infty$.

To show the last part of the theorem, we use Taylor's theorem and the hypothesis $l|\bar{F} - F| \leq |d_{ij}(\bar{F}) - d_{ij}(F)|$ to write, for any vector $F \in \mathcal{F}$,

$$\begin{aligned} D(F) &= D(F^*) + \sum_{(i,j)} d_{ij}(F_{ij}^*)(F_{ij} - F_{ij}^*) \\ &\quad + \sum_{(i,j)} \int_0^1 \{d_{ij}[F_{ij}^* + \alpha(F_{ij} - F_{ij}^*)] \\ &\quad - d_{ij}(F_{ij}^*)\} (F_{ij} - F_{ij}^*) d\alpha \\ &\geq D(F^*) + \sum_{(i,j)} d_{ij}(F_{ij}^*)(F_{ij} - F_{ij}^*) \\ &\quad + \frac{l}{2} \sum_{(i,j)} |F_{ij} - F_{ij}^*|^2. \end{aligned}$$

Since F^* minimizes D over \mathcal{F} , we have the optimality condition $\sum_{(i,j)} d_{ij}(F_{ij}^*)(F_{ij} - F_{ij}^*) \geq 0$, and it follows that

$$D(F) \geq D^* + \frac{l}{2} \sum_{(i,j)} |F_{ij} - F_{ij}^*|^2, \quad \forall F \in \mathcal{F}.$$

Therefore, using Lemma 2, we have

$$\begin{aligned} D^* + \frac{l}{2} \sum_{(i,j)} E\{|\tilde{F}_{ij}(t) - F_{ij}^*|^2\} \\ \leq E\{D[\tilde{F}(t)]\} \\ \leq E\{D[F(t)]\} + B \sum_{(i,j)} E\{|\tilde{F}_{ij}(t) - F_{ij}(t)|\}. \end{aligned}$$

Since $E\{|\tilde{F}_{ij}(t) - F_{ij}(t)|\} \rightarrow 0$ (cf. (A13)) and $E\{D[F(t)]\} \rightarrow D^*$, as $\epsilon \rightarrow 0$, $T \rightarrow 0$, and $t \rightarrow \infty$, we obtain that $\tilde{F}_{ij}(t)$ converges in mean square to F_{ij}^* . Since $\{\tilde{F}_{ij}(t) - F_{ij}(t)\}$ also converges to zero in mean square (cf. (A13)) we obtain that $F(t)$ converges to F^* in mean square. Q.E.D.

REFERENCES

- [1] M. Schwartz and T. E. Stern, "Routing techniques used in computer communication networks," *IEEE Trans. Commun.*, vol. COM-28, pp. 539-552, 1980.
- [2] J. M. McQuillan, I. Richer and E. C. Rosen, "The new routing algorithm for the ARPANET," *IEEE Trans. Commun.*, vol. COM-28, pp. 711-719, 1980.
- [3] L. Kleinrock, *Communication Nets: Stochastic Message Flow and Delay*. New York: McGraw-Hill, 1964.
- [4] R. G. Gallager, "A minimum delay routing algorithm using distributed computation," *IEEE Trans. Commun.*, vol. COM-25, pp. 73-85, 1977.
- [5] A. Segall, "Optimal routing for virtual line-switched data networks," *IEEE Trans. Commun.*, vol. COM-26, 1979.
- [6] T. E. Stern, "A class of decentralized routing algorithms using relaxation," *IEEE Trans. Commun.*, vol. COM-25, pp. 1092-1102, 1977.
- [7] L. Fratta, M. Gerla, and L. Kleinrock, "The flow deviation method: An approach to store-and-forward communication network design," *Networks*, vol. 3, pp. 97-133, 1973.
- [8] D. G. Cantor and M. Gerla, "Optimal routing in a packet switched computer network," *IEEE Trans. Comput.*, vol. C-23, pp. 1062-1069, 1974.
- [9] D. P. Bertsekas and R. G. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
- [10] D. P. Bertsekas and E. M. Gafni, "Projection methods for variational inequalities with application to the traffic assignment problem," *Math. Programming Study*, vol. 17, pp. 139-159, 1982.
- [11] —, "Projected Newton methods and optimization of multicommodity flows," Mass. Inst. Technol., Cambridge, LIDS Rep. P-1140, Aug. 1981; *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 1090-1096, 1983.
- [12] D. P. Bertsekas, E. M. Gafni, and R. G. Gallager, "Second derivative algorithms for minimum delay distributed routing in networks," *IEEE Trans. Commun.*, vol. COM-32, pp. 911-919, 1984.
- [13] D. P. Bertsekas, "Optimal routing and flow control methods for communication networks," in *Analysis and Optimization of Systems*, A. Bensoussan and J. L. Lions, Eds. New York: Springer-Verlag, 1982, pp. 615-643.
- [14] F. H. Moss and A. Segall, "An optimal control approach to dynamic routing in communication networks," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 329-339, 1982.
- [15] B. Hajek and R. G. Ogier, "Optimal dynamic routing in communication networks with continuous traffic," *Networks*, vol. 14, pp. 457-487, 1984.
- [16] H. Schwartz, *Computer Communication Network Design and Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1977.
- [17] D. P. Bertsekas, "Dynamic behavior of shortest path routing algorithms for communication networks," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 60-74, 1982.
- [18] —, "Dynamic models of shortest path routing algorithms for communication networks with multiple destinations," in *Proc. 1979 IEEE Conf. Decision and Control*, 1979, pp. 127-133.
- [19] J. M. McQuillan, G. Falk, and I. Richer, "A review of the development and performance of the ARPANET routing algorithm," *IEEE Trans. Commun.*, vol. COM-26, Dec. 1978.
- [20] L. Kleinrock, *Queueing Systems, Vol. II*. New York: Wiley, 1976.
- [21] —, *Queueing Systems, Vol. I*. New York: Wiley, 1975.
- [22] E. M. Gafni and D. P. Bertsekas, "Path assignment for virtual circuit routing," in *Proc. SIGCOMM 83 Symp. Communications Architectures and Protocols*, Mar. 1983, pp. 21-25.