ASYMPTOTIC OPTIMALITY OF SHORTEST PATH ROUTING

Eli M. Gafni and Dimitri P. Bertsekas

Massachusetts Institute of Technology
Laboratory for Information & Decision Systems
Cambridge, Massachusetts 02139

Program Code No. 5T10
ONR Identifying No. 049-383

Approved for public release; distribution unlimited.

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.
ASYMPTOTIC OPTIMALITY OF SHORTEST PATH ROUTING

ALGORITHMS*

by

Eli M. Gafni and Dimitri P. Bertsekas**

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.

*This work was supported by DARPA under contract ONR/N00014-75-C-1183. Thanks are due to Adrian Segall who suggested the problem to the first author.

**E. M. Gafni is with UCLA, Computer Science Dept., Los Angeles, Calif. 90024. D. P. Bertsekas is with Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, Cambridge, Mass. 02139
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 handles adequately 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 customary to measure optimality of a routing scheme in terms of an objective function of the form

\[ \sum_{(i,j)} D_{ij}(F_{ij}) \]  

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

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

which corresponds to the Kleinrock independence assumption [3]. There is extensive literature 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 an objective function such as (1), (2) in circumstances where the offered traffic statistics change slowly over time and furthermore individual offered traffic sample functions do not exhibit frequently large and persistent deviations from their averages. A typical situation is a net-
work accommodating a large number of relatively small users for each origin-
destination pair in which a form of the law of large numbers approximately
takes hold (see Lemma A.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], [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 \(D^i_{ij}(F^i_{ij})\)--the first derivative of
\(D_{ij}\) evaluated at the average rate \(F^i_{ij}\) at queue \((i,j)\) during the preceding
period. The first derivative relation between link lengths and objective
function is motivated by the well known optimality condition that a rout-
ing optimizes the objective (1) if and only if it routes traffic exclusively
along paths of minimum first derivative length (see e.g. [4], [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 \(D^i_{ij}\) are chosen poorly [17], [18].
Indeed the original adaptive shortest path algorithm implemented in 1969
on the ARPANET exhibited violent oscillatory behavior which was restrained
only after using the device of adding a bias to each link length at the
expense of 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 fast 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 above) is often large relative to the shortest path 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 if

\[
\frac{\text{Shortest Path Updating Period}}{\text{Average Holding Time of the Communication Path}} \to 0 \quad (3)
\]

and

\[
n_w \to \infty, \quad \gamma_w \to 0, \quad n_w \gamma_w = \text{constant} \quad (4)
\]

where \( n_w \) is the average number of active conversations for the generic origin-destination pair \( w \), and \( \gamma_w \) is the communication rate of each conversation. Assumptions (3), (4) together with additional Poisson-like assumptions on the offered traffic statistics are formulated in the next section. The main result in Section 3 provides also bounds on the sub-
optimality of the shortest path method when the 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).
2. Problem Formulation

Consider a network with a set of nodes N and a set of directed links L. We are given a set W of ordered node pairs referred to as origin-destination (OD) pairs. For each OD pair w ∈ 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 ∈ W arrive according to a Poisson process with mean rate \( \frac{\lambda_w}{e} \) where \( \lambda_w \) is given and \( e \) 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 ∈ P_w according to a rule to be described shortly and uses this path for the entire time of its duration assumed to be exponentially distributed with mean \( \frac{1}{\mu_w} \). We assume that the Poisson arrival processes and 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/∞ 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

\[
\lim_{t \to \infty} E\{n_w(t)\} = \frac{\lambda_w}{e\mu_w}, \quad \lim_{t \to \infty} \text{var}\{n_w(t)\} = \frac{\lambda_w}{e\mu_w}. \tag{5}
\]

Path assignment for each conversation is determined according to the following shortest path rule:

At times \( t = kT \) secs, \( k = 0, 1, \ldots \), 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

\[
F_{ij}(t) = \sum_{w \in W} \gamma_w \sum_{p \in P_w, (i,j) \in p} n_p(t). \tag{6}
\]
Here $n_p(t)$ is the number of active conversations assigned on path $p$ at time $t$, $\sum_{p \in P_w} n_p(t)$ is the total number of conversations of OD pair $w$ using paths $(i,j) \in P$ at time $t$, and $\gamma_w$ 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 $d_{ij}(*)$ is a continuous strictly monotonically increasing function of $F_{ij}$ satisfying $d_{ij}(F_{ij}) \geq 0$ for all $F_{ij} \geq 0$ and

$$|d_{ij}(F)-d_{ij}(\bar{F})| \leq L|F-\bar{F}|, \forall F, \bar{F} \geq 0, (i,j) \in L,$$  \hspace{1cm} (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 discontinuous (e.g. the TYMNET [1]). We do not know whether and in what form our main result holds for this case.

Regarding the communication rate $\gamma_w$ we assume that it is of the form

$$\gamma_w = \varepsilon \bar{\gamma}_w$$

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 $\gamma_w$ 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 \( \gamma_w \) is so small that \( \gamma_w \) can be viewed as a deterministic quantity. Note that for each OD pair \( w \) the product

\[
(\text{Mean arrival rate}) \cdot (\text{Communication rate}) = \bar{\lambda}_w \bar{\nu}_w
\]

is independent of \( \varepsilon \). We wish to study the effect on various stochastic processes of interest of the parameters \( \varepsilon \) and \( T \) particularly as

\[
\varepsilon \to 0 \quad \text{and} \quad T \to 0.
\]

Taking \( \varepsilon \to 0 \) implies that arrival rates tend to infinity while communication rates tend to zero with the products staying constant, and approximates a situation where there are many small conversations in the network [cf.(4)]. Taking \( T \to 0 \) approximates a situation where updating of shortest paths is fast relative to the mean duration time of a conversation [cf.(5)].

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 A.1).

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

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

where \( D_{ij} \) is some function such that


\[ d_{ij}(F_{ij}) = D'_{ij}(F_{ij}) \triangleq \text{First derivative of } D_{ij} \text{ at } F_{ij}. \] (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.

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

\[ D^* = \min_{F \in \mathcal{F}} D(F) \] (10)

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

\[ F_{ij} = \sum_{w \in \mathcal{W}} \sum_{(i,j) \in \mathcal{E}} x_p, \quad \forall (i,j) \in \mathcal{L} \quad \text{(11)} \]

where \( x_p \) are any nonnegative scalars satisfying

\[ \sum_{p \in \mathcal{P}} x_p = \frac{\lambda_w}{\mu_w}, \quad \forall w \in \mathcal{W}. \] (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 \( \lambda_w \mu_w^{-1} \) at each OD pair \( w \) (cf. (5), (8)). Note that 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

\[ D^* \leq \liminf_{t \to \infty} E[D(F(t))] \]
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, and the fact
[cf. (5), (8)]

\[ E\{n_w(t)Y_w\} + \frac{\bar{a} Y_w}{u_w} \quad \text{as } t \to \infty \]

Our main result is that as \( \varepsilon \to 0, \ T \to 0 \) and \( t \to \infty \) 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).
3. Main Result

We first introduce some notation:

\[ x_p(t) \triangleq \epsilon_{\bar{y}_w} n_p(t) : \text{The communication rate on path } p \text{ at time } t. \]

\[ r_w(t) \triangleq \sum_{p \in P_w} x_p(t) : \text{The total input rate of OD pair } w \text{ at } t. \]

\[ \bar{r}_w = \frac{\lambda_w \bar{y}_w}{\mu_w} : \text{The long term average input rate of } w. \]

\[ \bar{r} = \max_w \{ \bar{r}_w \} \]

\[ R_w = |r_w(0) - \bar{r}_w| : \text{The initial deviation of } r_w \text{ from its long term average} \]

\[ R = \max_w \{ R_w \} \]

\[ \mu = \min_w \{ \mu_w \} \]

\[ M = \max_w \{ \mu_w \} \]

\[ \gamma = \max_w \{ \gamma_w \} \]

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

\[ -c_1 e^{-\mu t} \leq E[D(F(t))] - D^* \leq e^{-\mu t} [D(F(0)) - D^*] + c_2 [a(\epsilon, T) + b(\epsilon, T)e^{-\mu t}] \]

(13)

where

\[ a(\epsilon, T) = \bar{r} \left\{ \frac{(\epsilon_{\bar{y}_T} + \epsilon_{\bar{y}_T}(r+T))(e^{-\mu T} - e^{-MT})}{\bar{r}(1-e^{-\mu T})} + 2\epsilon_{\bar{y}} + (1-e^{-\mu T})(4r+\epsilon_{\bar{y}}) \right\} \]

(14)
\[ b(\varepsilon, T) = R\left( \frac{(\varepsilon Y + R + 1)(e^{-\mu T} - e^{-MT}) + \varepsilon Y + (1 - e^{-\mu T})(4r + R + \varepsilon Y)}{Te^{-\mu T}} \right) \]  

(15)

Furthermore

\[ \lim_{\varepsilon \to 0, T \to 0, e^T \to \infty} E\{D[F(t)]\} = D^* . \]

If in addition we assume that, for some \( \lambda > 0 \), the length functions \( d_{ij} \) satisfy

\[ \lambda |F - F'| \leq |d_{ij}(F) - d_{ij}(F')|, \quad \forall F, F' \geq 0, (i, j) \in L \]

then

\[ \lim_{\varepsilon \to 0, T \to 0, e^T \to \infty} E\{|F_{ij}(t) - F^*_{ij}|^2\} = 0, \quad \forall (i, j) \in L, \]

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 \( \varepsilon \) 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 \( t \to \infty \), \( E\{D[F(t)]\} \) comes within \( c_2 a(\varepsilon, T) \) of being optimal. Thus \( c_2 a(\varepsilon, T) \) may be viewed as the long-term deviation from optimality of shortest path routing. The rate at which \( E\{D[F(t)]\} \) approaches its long term limit depends on the largest average
holding time $\frac{1}{u}$. 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 $\bar{r}_w$. Note that $a(\varepsilon,T) \to 0$ and $b(\varepsilon,T) \to 0$ as $\varepsilon \to 0$ and $T \to 0$, so both the long term and the transient deviation from optimality are reduced as the shortest path update period is reduced and the number of conversations is increased with an attendant reduction on their communication rate that maintains the total rate of each OD pair constant.

The three transient terms in (13) characterize the rate of convergence of the algorithm. Of these terms the slowest is the one involving $t e^{-\mu t}$. Since for any $\delta > 0$ we have $t e^{-\mu t} \leq \frac{1}{\delta e} e^{-(\mu-\delta)t}$ 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 $\varepsilon$ 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.
Appendix: Proof of the Theorem

For brevity we use the following notation in addition to the one given in the beginning of Section 3:

\[ n_p = n_p(kT), \quad x_p = x_p(kT), \quad r_w = r_w(kT), \quad F_{ij} = F_{ij}(kT) \]

\[ x^k = \{x^k_p | p \in P_w, \ w \in W\}, \quad r^k = \{r^k_{ij} | (i,j) \in \mathcal{E}\}. \]

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 \( \epsilon \to 0 \) the processes \( x_p(t) \) and \( r_w(t) \) behave asymptotically as deterministic processes.

**Lemma A1**: For all \( t > 0 \) and \( w \in W \)

\[ E[r_w(t)] = \frac{-\mu_w t}{r_w + \epsilon} \left[ r_w(0) - \frac{1}{r_w} \right] \]  
(A1)

\[ \text{var}[r_w(t)] = \epsilon \frac{-\mu_w t}{r_w + \epsilon} \left[ 1 - \epsilon \frac{-\mu_w t}{r_w + \epsilon} \right]. \]  
(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 \( t \in [kT, (k+1)T] \)

\[ E[x_p(t) | x^k_p] = \begin{cases} -\mu_w(t-kT) \frac{k}{x_p} & \text{if } p \neq p_k \\ \frac{-\mu_w(t-kT)}{r_w + \epsilon} \left[ x^k_p - \frac{r_w}{x_p} \right] & \text{if } p = p_k \end{cases} \]  
(A3)
\[
\text{var}\{x_p(t) | x_p^k\} = \begin{cases} \\
   e^{-\mu_w(t-kT)}[1-e^{-\mu_w(t)kT}] \bar{x}_p^k & \text{if } p \neq P_k \\
   e^{-\mu_w(t-kT)}[1-e^{-\mu_w(t)kT}] \bar{x}_p^k + e^{-\mu_w(t)kT} \bar{x}_p^k & \text{if } p = P_k \\
\end{cases}
\]

(A4)

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

\[
\begin{align*}
\dot{P}_0 &= -\Lambda P_0 + MP_1 \\
\dot{P}_k &= -(\Lambda + kM)P_k + MP_{k-1} + (k+1)MP_{k+1}, \quad k = 1, 2, ...
\end{align*}
\]

(A5)

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

\[
\dot{N} = -MN + \Lambda.
\]

(A6)

Also by multiplying (A5) by \( (k-N)^2 \), adding, and taking into account the fact \( \dot{\sigma} = \sum_{k=0}^{\infty} (k-N)^2 \bar{x}_p^k \) we obtain the equation

\[
\dot{\sigma} = -2M\sigma + MN + \Lambda.
\]

(A7)

The solutions of the linear differential equations (A6), (A7) can be calculated by the variations of constants formula. They are
\[ N(t) = \frac{\lambda}{M} + e^{-M \tau} [N(0) - \frac{\lambda}{M}] \]  

(A8)

\[ \sigma(t) = e^{-2M \tau} \sigma(0) + (1 - e^{-M \tau}) \left[ \frac{\lambda}{M} + e^{-M \tau} N(0) \right]. \]  

(A9)

Applying (A8) for \( M = \mu_w \), \( \lambda = \frac{\lambda_w}{\epsilon} \), and multiplying by \( e^{-\tau_w} \) yields (A1). Applying (A9) for \( M = \mu_w \), \( \lambda = \frac{\lambda_w}{\epsilon} \), \( \sigma(0) = 0 \), and multiplying by \( e^{2-2} \) yields (A2). A similar application of (A8) and (A9) yields (A3) and (A4). Q.E.D.

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

\[ |E[r_w(t)] - \bar{r}_w| \leq e^{-\mu_w \tau} R_w \leq e^{-\mu \tau} R \]  

(A10)

\[ \text{var}\{r_w(t)\} \leq e^{-\mu \tau} [(1-e^{-\mu \tau})\bar{r}_w + e^{-\mu \tau}[r_w(0) - \bar{r}_w]] \]  

(A11)

\[ \leq e^{-\mu \tau} (r + e^{-\mu \tau} R). \]

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

\[ x_p(t) = \frac{x_p(t)}{r_w(t)}, \quad \forall \ w \in W, \ p \in P_w. \]  

(A12a)

\[ \tilde{F}_{ij}(t) = \frac{1}{\epsilon} \sum_{w \in W} \sum_{p \in P_w} \tilde{x}_p(t), \quad \forall (i,j) \in L. \]  

(A12b)

We denote

\[ x_p^k \triangleq x_p(kt), \quad \tilde{F}_{ij}^k \triangleq \tilde{F}_{ij}(kt). \]  

(A12c)
Using the fact \(\tilde{x}_p(t) \leq \overline{r}_w\), and (A1), (A11) we have

\[
E\{|\tilde{x}_p(t) - x_p(t)|^2\} = E\{|\tilde{x}_p(t)[1 - \frac{r_w(t)}{\overline{r}_w}]|^2\}
\]

\[
\leq E\{|\overline{r}_w - r_w(t)|^2\}
\]

\[
\leq E\{|E\{r_w(t)\} - e^{-\mu t}\overline{r}_w - r_w(t)|^2\}
\]

\[
\leq \text{var}(r_w(t)) + e^{-2\mu t}R_w^2
\]

\[
\leq \frac{\varepsilon \gamma (\overline{r} + e^{-\mu t}R) + e^{-2\mu t}R^2}. \tag{A13}
\]

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

\[
E\{|\tilde{F}_{ij}(t) - F_{ij}(t)|^2\} \leq \alpha_{ij}[\varepsilon \gamma (\overline{r} + e^{-\mu t}R) + e^{-2\mu t}R^2]. \tag{A13}
\]

The next lemma provides a basic estimate:

**Lemma 2:** There exists \(B > 0\) such that for every vector \(F \in F\) and every other total link rate vector \(\bar{F}\) (not necessarily in \(F\)) there holds

\[
D(F) \leq D(\bar{F}) + B \sum_{(i,j) \in L} |\tilde{F}_{ij} - F_{ij}|, \tag{A14}
\]

where \(B\) is an upperbound for \(d_{ij}(F_{ij})\) over \((i,j) \in L\) and \(F \in F\).
Proof: We have by the convexity of $D$

$$D(F) \geq D(\tilde{F}) + \sum_{(i,j)} d_{ij}(F_{ij})(\tilde{F}_{ij}-F_{ij})$$

$$\geq D(\tilde{F}) - B \sum_{(i,j)} |F_{ij}-\tilde{F}_{ij}|.$$ Q.E.D.

Proof of Theorem 1:

We first show the left side of (13). Let \( x^*(t) \) be a set of path rates that solve the deterministic multicommodity flow problem

$$\text{minimize } D(F)$$

subject to

$$F_{ij} = \sum_{w \in \mathcal{W}} \sum_{p \in \mathcal{P}_w} x_p$$

$$\sum_{p \in \mathcal{P}_w} x_p = E(r_w(t)), x_p \geq 0.$$ (A15)

Let \( F^*(t) \) be the vector of corresponding total link rates, i.e.
Define the "normalized" rates

\[ \hat{x}_p(t) = \frac{\bar{r}_w}{E\{r_w(t)\}} x^*_p(t) \]  

(A16)

\[ \hat{F}_{ij}(t) = \sum_{w \in W} \sum_{p \in P_w} \hat{x}_p(t) \]  

\[ (i,j) \in \mathcal{P} \]

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

\[ D^* \leq D[F(t)] \leq D[F^*(t)] + B \sum_{(i,j)} |\hat{F}_{ij}(t) - F^*_{ij}(t)| \]  

(A17)

\[ \leq D[E\{F(t)\}] + B \sum_{(i,j)} |\hat{F}_{ij}(t) - F^*_{ij}(t)| \]

\[ \leq E\{D[F(t)]\} + B \sum_{(i,j)} |\hat{F}_{ij}(t) - F^*_{ij}(t)| , \]

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)

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

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 \( \beta_{ij} \)

\[ |\hat{F}_{ij}(t) - F^*_{ij}(t)| \leq \beta_{ij} R e^{-\mu t} \]  

(A18)

Taking \( c_1 = B \sum_{(i,j)} \beta_{ij} \) we obtain from (A17) and (A18)
D^* \leq E\{D(F(t))\} + c_1 R e^{-\lambda 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:

\[
D_{ij}[F_{ij}(t)] = D_{ij}(F^k_{ij}) + d_{ij}(F^k_{ij})[F_{ij}(t) - F^k_{ij}] \\
+ \int_0^1 \{d_{ij}[F^k_{ij} + \alpha(F_{ij}(t) - F^k_{ij})] - d_{ij}(F^k_{ij})\}[F_{ij}(t) - F^k_{ij}]d\alpha
\]

\[
\leq D_{ij}(F^k_{ij}) + d_{ij}(F^k_{ij})[F_{ij}(t) - F^k_{ij}] + \frac{L}{2} |F_{ij}(t) - F^k_{ij}|^2
\]

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

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

(A19)

We derive an upper bound for the expected value of each of the last two terms above.

Denote by \( d^k_p \) the length of path \( p \) corresponding to the link flows \( F^k_{ij} \).

We have

\[
d^k_p = \sum_{(i,j) \in p} d_{ij}(F^k_{ij}), \quad \forall \omega \in W, \ p \in P_w
\]

and it follows that
\[
\sum_{(i,j)} d_{ij}(F_{ij}^k)(F_{ij}(t) - F_{ij}^k) = \sum_{(i,j)} d_{ij}(F_{ij}^k) \sum_{w \in W} \sum_{p \in P_w} [x_p(t) - x_p^k]
\]

\[
= \sum_{w \in W} \sum_{p \in P_w} d_p^k [x_p(t) - x_p^k]
\]

(A20)

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

\[
\frac{d_x^k}{x_p} = \begin{cases} 
0 & \text{if } p \neq p_k \\
\frac{d_x^k}{x_p} & \text{if } p = p_k
\end{cases}
\]

(A21)

Taking conditional expectation in (A20) and using (A3)

\[
E\{\sum_{(i,j)} d_{ij}(F_{ij}^k)(F_{ij}(t) - F_{ij}^k)\mid x^k\} = \sum_{w \in W} \sum_{p \in P_w} d_p^k [E(x_p(t)\mid x_p^k) - x_p^k]
\]

(A22)

\[
= \sum_{w \in W} [1 - e^{-\mu_w(t-kT)}] \sum_{p \in P_w} d_p^k (x_p^k - x_p^k)
\]

\[
= \sum_{w \in W} [1 - e^{-\mu_w(t-kT)}] [\sum_{p \in P_w} d_p^k ] (x_p^k - x_p^k) +
\]

\[
\sum_{p \in P_w} d_p^k (x_p^k - x_p^k)
\]

where \(x_p^k\) is given by (A12). Since \(\sum_{p \in P_w} x_p^k = \sum_{p \in P_w} x_p^k = x_w\) and, for each \(w\), \(p_k\) is the shortest path we obtain using (A21)

\[
\sum_{p \in P_w} d_p^k x_p^n < \sum_{p \in P_w} d_p^k \tilde{x}_p^k
\]
so (A22) can be strengthened to yield

\[
E\{ \sum_{(i,j)} d_{ij}(F_{ij}^k[F_{ij}(t)-F_{ij}])|x^k} \leq [1-e^{-\mu(t-kT)}] \sum_{weW} \sum_{peP_w} d^k_{p}(x^k_p-x^k_p)
\]

\[
+ \sum_{weW} [1-e^{-\mu(t-kT)}] \sum_{peP_w} d^k_{p}(x^k_p-x^k_p)
\]

\[
= [1-e^{-\mu(t-kT)}] \sum_{weW} \sum_{peP_w} d^k_{p}(x^k_p-x^k_p)
\]

\[
+ \sum_{weW} [e^{-\mu(t-kT)} - e^{-\mu_w(t-kT)}] \sum_{peP_w} d^k_{p}(x^k_p-x^k_p).
\]

We proceed to bound each of the two terms in the right side above.

Let \(\{x^*_p|weW, peP_w\}\) be any set of path flows minimizing \(D(F)\) over \(F\)
i.e., any \(x^*_p \geq 0\) such that

\[
F^*_{ij} = \sum_{weW} \sum_{peP_w} x^*_p, \forall (i,j) \in L.
\]

\[(i,j) \in P\]

Since for each \(w\) the shortest path is \(p_k\) and \(\sum_{peP_w} x^*_p = \sum_{peP_w} \overline{x}^k_p = \overline{r}_w\)

we have

\[
\sum_{peP_w} d^k_{p}(x^k_p-x^k_p) \leq \sum_{peP_w} d^k_{p}(x^*_p-x^k_p)
\]

(A24)

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

\[
\sum_{weW} \sum_{peP_w} d^k_{p}(x^*_p-x^k_p) = \sum_{(i,j)} d_{ij}(F^k_{ij}(F^*_k-F^k_{ij})).
\]

(A25)
Since D is convex we obtain
\[ \sum_{(i, j)} \frac{d_{ij}(F^*_{ij}) - d_{ij}(F^k_{ij})}{d_{ij}(F^k_{ij})} \leq D(F^*) - D(F^k) = D^* - D(F^k). \] (A26)

Combining (A24)-(A26) we see that
\[ \sum_{w \in W} \sum_{p \in P_w} d_k^p(x^k_{\text{p}} - x^k_{\text{p}}) \leq D^* - D(F^k). \] (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
\[ \sum_{p \in P_w} d_k^p(x^k_{\text{p}} - x^k_{\text{p}}) = \sum_{p \in P_w} (d_k^p - d^p_k)(x^k_{\text{p}} - x^k_{\text{p}}) + \sum_{p \in P_w} d_k^p(x^k_{\text{p}} - x^k_{\text{p}}) \] (A28)

where \( d_p^k \) is the length of path \( p \) if each flow \( x^k_p \) is replaced by \( \tilde{x}^k_p \), i.e.
\[ \tilde{d}_p^k = \sum_{(i, j) \in P} d_{ij}(F^k_{ij}). \]

Using (7) and (A13) it is easily seen that for some constant \( \xi > 0 \)
\[ E\{ \sum_{p \in P_w} (d_k^p - d^p_k)(x^k_{\text{p}} - x^k_{\text{p}}) \} \leq L E\{ \sum_{p \in P_w} \sum_{(i, j) \in P} |F_{ij} - F^k_{ij}|^2 \} \]
\[ \leq \xi[\epsilon \gamma (\tau + e^{-\mu k T R}) + e^{-2\mu k T R^2}]. \]

Using (A12) we have
\[ \sum_{p \in P_w} \tilde{d}_p^k(x^k_{\text{p}} - x^k_{\text{p}}) = \frac{\tilde{r}_w^k - r_w^k}{r_w^k} \sum_{p \in P_w} d_p^k x^k_{\text{p}} \]
\[ \leq \frac{|\tilde{r}_w^k - r_w^k| B}{r_w^k} \sum_{p \in P_w} x^k_{\text{p}} = B|\tilde{r}_w^k - r_w^k|. \]
where $B$ is the constant defined in Lemma 2.

We have

$$E[|\tilde{r}_w - r^k_w|] \leq E[|\tilde{r}_w - E(r^k_w)|] + E[|E(r^k_w) - r^k_w|]$$

$$\leq E[|\tilde{r}_w - E(r^k_w)|] + \sqrt{\text{var}(r^k_w)}$$

where the last step follows using Jensen's inequality. Therefore using (A10) and (A11) we obtain

$$E[|\tilde{r}_w - r^k_w|] \leq e^{-\mu t} R + \sqrt{\varepsilon Y (\tilde{r} + e^{-\mu t} R)}$$

$$\leq e^{-\mu t} R + \sqrt{\varepsilon Y (\tilde{r} + R)},$$

and

$$E\left[ \sum_{p \in P} d^k_p (x^k_p - x^k_p) \right] \leq B [e^{-\mu t} R + \sqrt{\varepsilon Y (\tilde{r} + R)}].$$

Taking expectation over $x^k$ in (A28) and using the inequalities above we obtain for some constant $\zeta > 0$

$$E\left[ \sum_{w \in W} \left[ e^{-\mu (t-kT)} - e^{-\mu_w (t-kT)} \right] \sum_{p \in P} d^k_p (x^k_p - x^k_p) \right]$$

$$\leq \zeta [e^{-\mu (t-kT)} - e^{-M(t-kT)}] \left[ \frac{e^{-\varepsilon Y (\tilde{r} + e^{-\mu t} R)}}{\tilde{r} + e^{-\mu t} R} + e^{-2\mu t R} + e^{-\mu t R} + \sqrt{\varepsilon Y (\tilde{r} + R)} \right]$$

Combining (A23), (A27), (A28), and taking expectation over $x^k$ we obtain for some constant $\beta_1$
\[ E\{ \sum_{(i,j)} d_{ij}(F_{ij}^k)(F_{ij}(t) - F_{ij}^k) \} \leq [1 - e^{-\mu(t-kT)}][D^* - E\{D(F^k)\}] \] (A29)

\[ + \beta_1 [e^{-\mu(t-kT)} - e^{-M(t-kT)}][e^{-\gamma(t+kT)} R + e^{-\mu kT} R^2 + e^{-\mu kT} R + \sqrt{\varepsilon_Y} (R + R)] \]

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 there exists a constant \( \theta \) such that

\[ \sum_{(i,j)} |F_{ij}(t) - F_{ij}^k|^2 \leq \theta \sum_{w \in W} \sum_{p \in P_w} |x_p(t) - x_p^k|^2. \] (A30)

We have

\[ E\{|x_p(t) - x_p^k|^2| x_p^k\} = \text{var}(x_p(t)|x_p^k) + [x_p^k - E\{x_p(t)|x_p^k\}]^2 \]

and using Lemma 1 we obtain

\[ \sum_{p \in P_w} E\{|x_p(t) - x_p^k|^2| x_p^k\} = \epsilon_{Y_w}[1 - e^{-\mu w(t-kT)}]^2 \left[ (x_p^k)^2 + e^{-\mu w(t-kT)} x_p^k \right] \]

\[ + [1 - e^{-\mu(t-kT)}][e^{-\gamma(t+kT)} R + e^{-\mu kT} R^2 + e^{-\mu kT} R + \sqrt{\varepsilon_Y} (R + R)] \]

Taking expectation over \( x^k \) and using (A10), (A11) we obtain
\[
\sum_{p \in \mathcal{P}_w} E\{ |x_p(t) - x^k_p|^2 \} \leq [1-e^{-\mu(t-kT)}] \{c_Y(r_w^k + E\{r^k_w\}) \\
+ (1-e^{-\mu T}) \{r^2_w + 2r_w E\{r^k_w\} + (E\{r^k_w\})^2 + \text{var}\{r^k_w\} \}
\]  
\[
\leq [1-e^{-\mu(t-kT)}] \{c_Y(2r + e^{-\mu kT} R) \\
+ (1-e^{-\mu T}) [(2r + e^{-\mu kT} R)^2 + c_Y(r + e^{-\mu kT} R)] \}
\]  

We now combine (A19), (A29)-(A31) to obtain for all \( t \in [kT, (k+1)T] \) and some positive constant \( \beta_2 \)

\[
E\{D(F(t))\} - D^* \leq e^{-\mu(t-kT)} [E\{D(F^k)\} - D^*] \\
+ \beta_1 [e^{-\mu(t-kT)} - e^{-M(t-kT)}] [c_Y(r + e^{-\mu kT} R) + e^{-2\mu kT} R^2 + e^{-\mu kT} R + \sqrt{c_Y(r+R)}] \\
+ \beta_2 [1-e^{-\mu(t-kT)}] \{c_Y(2r + e^{-\mu kT} R) + (1-e^{-\mu T}) [(2r + e^{-\mu kT} R)^2 + c_Y(r + e^{-\mu kT} R)] \}
\]  

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

\[
E\{D(F^{k+1})\} - D^* \leq e^{-\mu T} [E\{D(F^k)\} - D^*] \\
+ c_2 [\bar{a}(\varepsilon, T) + \bar{b}(\varepsilon, T) e^{-\mu kT}]
\]  

where

\[
\bar{a}(\varepsilon, T) = \bar{f}\{(e^{-\mu T} - e^{-MT})(c_Y + \sqrt{c_Y(r+R)} \over r) + (1-e^{-\mu T})[2c_Y + (1-e^{-\mu T})(4r+c_Y)] \}
\]  

(A33)

\[
\bar{b}(\varepsilon, T) = \bar{f}\{(e^{-\mu T} - e^{-MT})(c_Y + R + 1) + c_Y + (1-e^{-\mu T})(4r+R+c_Y) \}.
\]  

(A34)
Applying (A32) repeatedly for \( k \) equal to zero up to \((k-1)\) we obtain

\[
E\{D(F^k)\} - D^* \leq e^{-\mu kT}[D(F^0) - D^*] + \epsilon \left[ \frac{a(\epsilon, T)}{1-e^{-\mu T}} + \frac{b(\epsilon, T)}{Te^{-\mu T}} \right] kTe^{-\mu T}
\]

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

Since

\[
\lim_{\epsilon \to 0} \frac{a(\epsilon, T)}{1-e^{-\mu T}} = \lim_{\epsilon \to 0} \frac{b(\epsilon, T)}{Te^{-\mu T}} = 0
\]

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

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

\[
D(F) = D(F^*) + \sum_{(i,j)} d_{ij}(F^*) (F_{ij} - F_{ij}^*)
\]

\[
+ \sum_{(i,j)} 0 \{d_{ij}(F_{ij}^* + \alpha(F_{ij} - F_{ij}^*)) - d_{ij}(F_{ij}^*)\} (F_{ij} - F_{ij}^*)d\alpha
\]

\[
\geq D(F^*) + \sum_{(i,j)} d_{ij}(F_{ij}^*) (F_{ij} - F_{ij}^*) + \frac{\lambda}{2} \sum_{(i,j)} |F_{ij} - F_{ij}^*|^2 .
\]

Since \( F^* \) minimizes \( D \) over \( \mathcal{F} \) we have the optimality condition

\[
\sum_{(i,j)} d_{ij}(F_{ij}^*) (F_{ij} - F_{ij}^*) \geq 0 \quad \text{and it follows that}
\]

\[
D(F) \geq D^* + \frac{\lambda}{2} \sum_{(i,j)} |F_{ij} - F_{ij}^*|^2 , \quad \forall F \in \mathcal{F}.
\]
Therefore using also Lemma 2 we have

\[ D^* + \frac{k}{2} \sum_{(i,j)} E[|F_{ij}(t) - F_{ij}^*|^2] \leq E[D(F(t))] \]

\[ \leq E[D(F(t))] + B \sum_{(i,j)} E[|F_{ij}(t) - F_{ij}(t)|] \]

Since \( E[|F_{ij}(t) - F_{ij}(t)|] \to 0 \) [cf. (A13)] and \( E[D(F(t))] \to D^* \) as \( \varepsilon \to 0 \),

\( T \to 0 \) and \( t \to \infty \) we obtain that \( F_{ij}(t) \) converges in mean square to \( F_{ij}^* \).

Since \( F_{ij}(t) - F_{ij}(t) \) also converges to zero in mean square [cf. (A13)]

we obtain that \( F(t) \) converges to \( F^* \) in mean square. Q.E.D.
References


Best Available Copy
Distribution List

Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314

Assistant Chief for Technology
Office of Naval Research, Code 200
Arlington, Virginia 22217

Office of Naval Research
Information Systems Program
Code 437
Arlington, Virginia 22217

Office of Naval Research
Branch Office, Boston
495 Summer Street
Boston, Massachusetts 02210

Office of Naval Research
Branch Office, Chicago
536 South Clark Street
Chicago, Illinois 60605

Office of Naval Research
Branch Office, Pasadena
1030 East Greet Street
Pasadena, California 91106

Naval Research Laboratory
Technical Information Division, Code 2627
Washington, D.C. 20375

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, D.C. 20380