Proceedings of *INFOCOM'99* March 21-25, 1999, New York City

Performance Analysis on Path Rerouting Algorithms for Handoff Control in Mobile ATM Networks

Jun Li, Roy Yates*, Dipankar Raychaudhuri
NEC USA, C&C Research Laboratories
4 Independence Way, Princeton, NJ 08540, U.S.A.
{lijun, ray}@ccrl.nj.nec.com
* Department of Electrical and Computer Engineering
Rutgers University, Piscataway, NJ 08855, U.S.A.
ryates@winlab.rutgers.edu

Abstract— **This paper studies mobile-generated traffic distributions in mobile ATM networks and evaluates the performance of path rerouting algorithms for handoff control. In mobile ATM networks, user mobility and handoff path rerouting may produce extra traffic load over network links, requiring larger network capacity to support the same QoS. In this paper, we propose a flow model for mobile ATM networks. The model represents the mobile-generated traffic as a set of stochastic flows over a set of OD (Origin-Destination) pairs. The user mobility is defined of transfer probabilities of the flows and the handoff path rerouting algorithm is modeled by a transformation between the routing functions for traffic flows. The analysis shows that user mobility may cause temporal variations as well as smoothing effects on the network traffic. Using the flow network model, typical handoff path rerouting algorithms are evaluated through both analytical and experimental approaches. The evaluation methodology can be used for either redesigning the network topology for a given path rerouting algorithm or selecting a path rerouting algorithm for a given network topology under a specific mobile service scenario.**

I. INTRODUCTION

In recent years, the demand for wireless and broadband services has been growing rapidly. Wireless ATM [1] has been considered a candidate solution for providing broadband wireless services. A wireless ATM network consists of an ATM radio access network and a "mobile ATM" core network. The "mobile ATM" network is a common network infrastructure that supports user mobility for wireless ATM as well as other mobile services, such as GSM, wireless LAN etc. [2–4].

The key mobility support function in mobile ATM networks is the path rerouting process that is required when a mobile terminal moves from one access point to another. Previous studies, conducted through both research [3–11] and standardization [12–15] activities, have mainly focused on protocol design. Performance evaluation of path rerouting algorithms has been conducted through simulation [6].

This paper intends to evaluate path rerouting algorithms for handoff control through both analytical and experimental approaches. Traditionally, in a circuit network such as telephony system, a queueing model is used to describe traffic with constant bandwidth requirements at the burst level and a Poisson arrival process at the call level. Although an ATM network uses virtual circuits, its traffic may have large variations at both burst level and call level. For example, the call lifetime can be as short as getting a small text webpage or as long as transferring a large video/audio stream. And at the burst level, the bandwidth requirement for each call may be CBR, VBR, ABR or UBR with a very diversified range.

In a mobile ATM network, traffic arrivals and departures occur in discrete bandwidth units on a call-by-call basis; however the limited capacity of wireless networks dictates that the bandwidth requirement of a wireless call are small relative to the link capacities in a fixed ATM network. Thus, as long as the number of calls (circuits) over each OD pair is large enough, the overall traffic can be approximated as a continuous function of time although the variations are caused by discrete events, such as call arrivals, departures, handoffs and call bursts. Therefore, we propose to use a *flow model*, which only describes the network traffic at the burst level using instantaneous bandwidth requirements or transmission rates.

The flow model defines the network traffic as a set of stochastic processes (*flows*) over OD (origin-destination) pairs. User mobility is represented by the probability of the flow transferring between OD (Origin-Destination) pairs. In addition, each traffic flow is associated with a routing function and the handoff path rerouting algorithms are modeled by transformations between the routing functions. Based on this model, handoff and path rerouting algorithm can be studied analytically.

This paper does not address the optimization problems in network topology or routing design. Instead, we evaluate the extra cost that a core network must pay to support mobile services using different path rerouting algorithms. The purpose of performance evaluation for handoff path rerouting algorithms is twofold. First, the results can be used for re-engineering the network topology and link capacities of the ATM core network according to the user mobility pattern and path rerouting algorithm. Second, for a given mobile ATM core network, the results can be used for selecting a path rerouting algorithm or altering user mobility patterns by relocating radio access points.

The paper is organized as follows. In section 2, we introduce the flow network model. In section 3, we give the analytical results from an example with simple traffic requirements and network topology. In section 4, through simulation, experimental results are given for a generic PNNI network topology with different user mobility patterns.

II. FLOW NETWORK MODEL FOR MOBILE ATM NETWORKS

A. Mobile ATM networks

A mobile ATM network consists of a core ATM network and wireless access points. The core ATM network has ATM switches as *nodes* and fiber optic links between switches as *edges*. The network shown in Figure 1 is a mobile ATM network. Some of the nodes (switches $A.4.5, A.4.6, A.3.4, A.3.3$) are radio access points (basestations). Each radio access point may serve several cells with multiple antennas. If there is a traf-

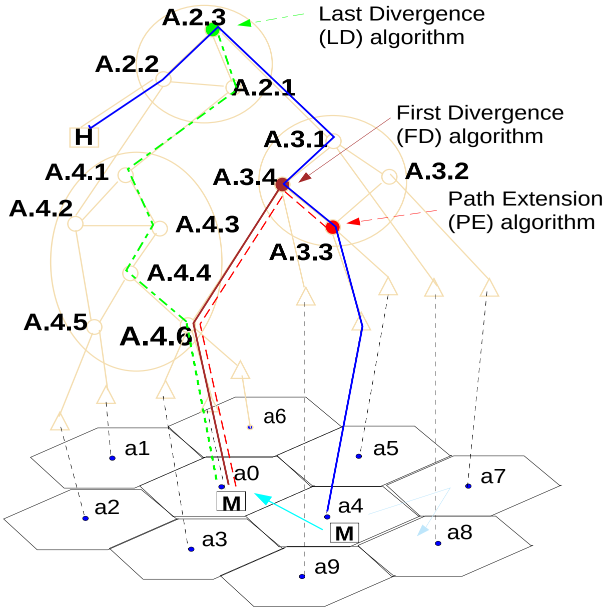

Fig. 1. A Mobile ATM Network

fic flow from a mobile user M at the access point $A.3.3$ to a fixed host H, the flow may take a route as shown in the figure. When the mobile user moves from cell $a4$ to cell $a0$, it is necessary for the network to reroute the flow to a route which goes through the new access point $A.4.6$. This requires a handoff control process and path rerouting algorithm. Figure 1 depicts three new routes. Each has a different anchor node (crossover switch or COS) to reroute the connection path, corresponding to a particular path rerouting algorithm. First is the *Last Divergence* (LD) algorithm with the COS at $A.2.3$. It aims to improve the efficiency of the network resource utilization by optimizing the path length after each handoff. Second is the *Path Extension* (PE) algorithm with the COS at $A.3.3$. It aims for low complexity by just extending the original path from the old access point to the new access point. Third is the *First Divergence* (FD) algorithm with the COS at $A.3.4$. It is a tradeoff between complexity and performance by avoiding the duplicate paths generated by PE algorithm.

B. Network topology and mobility pattern

A mobile ATM network has a *topology* which can be defined as a graph $G = [X, C]$, where $X = \{1, 2, ..., N\}$ are N *nodes* in the network and $C = \{c_{ij}, i, j \in X\}$ is a *link capacity matrix*. If $c_{ij} > 0$, we say there is a link, or a C*-edge* (Capacity edge) from node i to node j.

In addition, the network topology of a mobile ATM network has a *mobility pattern*, $G_w(t) = [Y, D(t)]$, where, $Y \subset X$ is a set of wireless access points and $D(t) = \{d_{ij}(t)\}$ is the *User Transition Matrix* of the network. At time t, a mobile user departing from access point i will enter node j with probability $d_{ij}(t)$. If $d_{ij}(t) > 0$, we say there is an M*-edge* (Mobility edge) from node i to j, i.e. node j is a neighbor of node i at time t. The user transition matrix $D(t)$ will depend on the users' call lifetimes and access point dwell times.

C. Traffic flows and handoff

For a given mobile ATM network with topology G, the network traffic can be specified as bandwidth requirements over the OD (origin–destination nodes) pairs. Each requirement can be represented by a random process (flow) in the unit of bandwidth per second. A flow consists of infinite numbers of infinitesimal calls over the OD pair. Each call can be viewed as a contributor of a *mini flow*, which may take a specific route in the network. The bandwidth of a traffic flow in a mobile ATM network is randomly changed because of call arrivals and departures, call handoffs and call bursts. In general, a multimedia traffic flow may be quite complicated statistically. However, in this paper, we assume that the bandwidth requirements of traffic flows are memoryless, meaning the holding time is exponentially distributed. In this case, the traffic flow $\psi_k(t)$ on OD pair k is a Markov process which satisfies the following stochastic differential equations, for $k = 1, ..., K$:

$$\Delta\psi_k(t) = -\mu_k(t)\psi_k(t)\Delta t + \sum_{l}^{M} p_{lk}(t)\mu_l(t)\psi_l(t)\Delta t + \Delta\phi_k(t) \tag{1}$$

In equation (1), $\phi_k(t)$ is the *traffic demand* representing new bandwidth requirement (in bits/sec) over the time period Δt, $\mu_k(t)$ is the service rate of traffic flow $\psi_k(t)$ on OD pair k, and $p_{lk}(t)$ is the *handoff transfer probability*, representing the probability that a call in the traffic flow finished its service in an OD pair o_l and moves to a neighboring OD pair o_k. The probability that a departure call from o_l terminates is $p_{l0}(t)$. Clearly, $\sum_k p_{lk}(t) = 1$. With $\Psi(t) = [\psi_1(t), \ldots, \psi_K(t)]^T$ and $\Phi(t) = [\phi_1(t), \ldots, \phi_M(t)]^T$, equation (1) becomes in matrix form,

$$\dot{\Psi}(t) = A(t)\Psi(t) + \dot{\Phi}(t) \tag{2}$$

where $A(t)$ is *flow transition matrix* with $A_{ij} = -\mu_i(t)$ for $i = j$ and $A_{ij} = \mu_i(t)p_{ij}(t)$ for $i \neq j$. The flow transition matrix $A(t)$ can be derived from the user transition matrix $D(t)$ given the call lifetime and user dwell times at the wireless access points; however, space limitations preclude this development.

D. Routing functions and path rerouting algorithms

In a mobile ATM network, each traffic flow $\psi_k(t)$ has a *routing function* defined by a vector $V_k(t) = [v_{k,1}(t), \ldots, v_{k,M}(t)]^T$, where $v_{k,u}(t)$ is the probability that the link u is being used by any call in the flow $\psi_k(t)$. For example, in Figure 2, suppose ψ_1 is the flow for OD pair AD. One part of the flow takes the route ABD and rest of the flow takes the route $ACBD$. Corresponding to the order of the network links $[AB, AC, CB, CD, BD]$, one possible routing function is $V_1 = [0.8, 0.2, 0.2, 0, 1]^T$. The routing function for each traffic flow may be static, as a result of network synthesis, or dynamic, as a result of dynamic routing process. In either case, the routing function may be varied by the path rerouting process for handoff control.

In particular, we assume that the traffic demand $\phi_k(t)$ is initially routed according to the routing function $Z_k(t) = [z_{k,1}(t), \ldots, z_{k,M}(t)]^T$ where $z_{k,u}(t)$ is the fraction of the traffic demand $\phi_k(t)$ on C-edge u. In addition, we define $w_{lk,u}(t)$ as the fraction of the handoff traffic on C-edge u based on a path

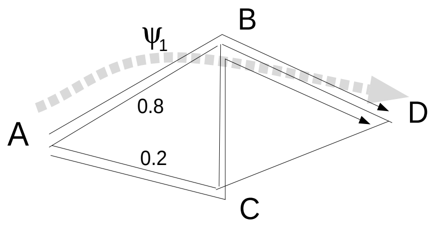

Fig. 2. Illustration of Flow Assignments by Routing Function

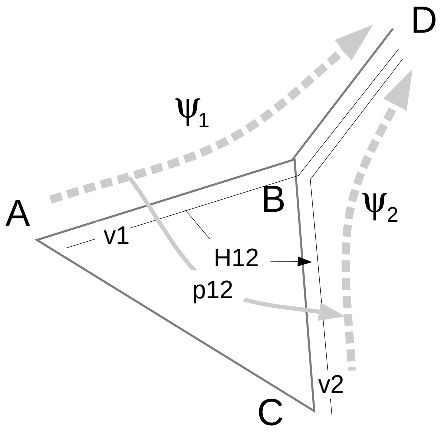

Fig. 3. Illustration of Path Rerouting Algorithms

rerouting algorithm. Suppose flow $\psi_k(t)$ on a network link u is $v_{k,u}(t)\psi_k(t)$. For all k and u, it must satisfy the traffic balance equation

$$\Delta(v_{k,u}(t)\psi_k(t)) = -\mu_k(t)v_{k,u}(t)\psi_k(t)\Delta t + \sum_{l}^{M} w_{lk,u}p_{lk}\mu_l(t)\psi_l(t)\Delta t + \Delta(z_{k,u}(t)\phi_k(t)) \quad (3)$$

To characterize $w_{lk,u}(t)$, we define

$$W_{lk}(t) = [w_{lk,1}(t), \ldots, w_{lk,M}(t)]^T = H_{lk}(t)V_l(t) \quad (4)$$

where the $M \times M$ matrix $H_{lk}(t)$ is called a *path rerouting transform matrix*, which re-assigns the handoff flow, from $\psi_l(t)$ to $\psi_k(t)$, to a new route according to given path rerouting algorithm.

Let us give an example in Figure 3. The network has four nodes and two OD pairs with flows ψ_1 and ψ_2. The initial routing functions are $Z_1 = [10001]$ and $Z_2 = [00101]$, respectively, for links $[AB, AC, CB, CA, BD]$. A handoff path rerouting algorithm defined by $W_{12} = H_{12}V_1 = V_2$ and $W_{21} = H_{21}V_2 = V_1$ assigns handoff flows to the same routes as offered flow. In this case, the path rerouting transform matrix can be defined as

$$H_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad H_{21} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

In H_{12}, $[H_{12}]_{11} = 0$ means all handoff traffic from AB goes away while $[H_{12}]_{31} = 1$ means all handoff traffic from AB goes to CB.

In general, the rerouting of the handoff flows can be either static and based only on fixed network topology or dynamic and based on both network topology and traffic loads. The $H_{lk}(t)$ may be deterministic functions of time or random processes. If $H_{lk}(t)$ is deterministic, for given path rerouting algorithm (given rules of rerouting), it is not difficult, though complicated for a large network, to obtain $\{H_{lk}(t)\}$. Otherwise, it is possible to estimate $\{H_{lk}(t)\}$ through statistical measurement. With the path rerouting transform matrix $\{H_{lk}(t)\}$, we can solve for the routing functions $\{V_k(t)\}$ as follows.

$$(\psi_k(t)V_k(t))' = -\mu_k(t)\psi_k(t)V_k(t) + \sum_{l}^{K} p_{lk}(t)\mu_l(t)\psi_l(t)W_{lk}(t) + (\phi_k(t)Z_l(t))' \quad (5)$$

We call the vector $\psi_k(t)V_k(t)$ a generic *network flow* of an OD pair o_k which carries its routing information. Solving the balance equations (2) and (5), we can evaluate the traffic loads over network links.

To summarize, a mobile ATM network is modeled by a flow network model with a fixed network topology G, an access network topology (mobility pattern) $G_w(t) = (G, D(t))$, network traffic flows $\{\psi_k(t)\}$, input traffic flow $\{\Phi(t)\}$, traffic service rate $\{\mu_k(t)\}$, routing functions $\{V_k(t)\}$, and path rerouting transformations $\{H_{lk}(t)\}$. The traffic distribution over the network depends on all factors; however, our analysis will focus on the effects of the user mobility ($D(t)$) and the handoff path rerouting ($\{H_{lk}(t)\}$).

III. ANALYSIS ON TRAFFIC DISTRIBUTION

We have given the flow balance equations at two levels. One is at OD pair level and the other is at the link level. In this section, the user mobility effects on traffic flows on both levels are analyzed.

A. *Traffic flows and user mobility*

1) *Mean values and temporal variation:* Suppose the traffic demands $\{\phi_k(t)\}$ are differentiable functions and let $d\Phi(t) = \Omega(t)dt$, where $\Omega(t) = [\omega_1(t), \ldots, \omega_K(t)]^T$ represents the rates of traffic demand variation. If a steady-state can be reached, the mean values of traffic flows can be solved as

$$m_\Psi(t) = -A^{-1}(t)B(t)m_\Omega(t) \quad (6)$$

We still denote the mean value as the function of time because of hour by hour variations of mobility pattern $A(t)$ can yield different steady-state operating points.

The solution above gives the properties of traffic flows as follows. (1) The total traffic load, represented by $\sum_k m_{\psi_k}(t)$, is conserved regardless of user mobility. (2) For a balanced OD pair o_l satisfying $\sum_k p_{lk}(t)\mu_l(t)\psi_l(t) = \sum_l p_{kl}(t)\mu_k(t)\psi_k(t)$, the traffic load is independent of user mobility. A balanced OD pair has the arrival rate of handoff flow equal to the departure rate of handoff flow. Equivalently, a balanced OD pair has the flows from new calls balanced with the flows from terminated calls, i.e. $\omega_k(t) = (1 - \sum_l p_{kl}(t))\mu_k(t)\psi_k(t)$. (3) The traffic load on an unbalanced OD pair depends on user mobility. As user mobility increases, the traffic load may increase, if more

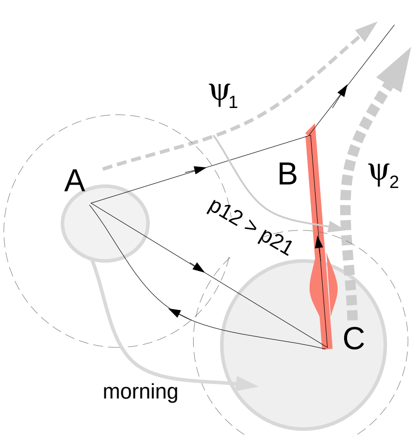

Fig. 4. Example of temporal variations

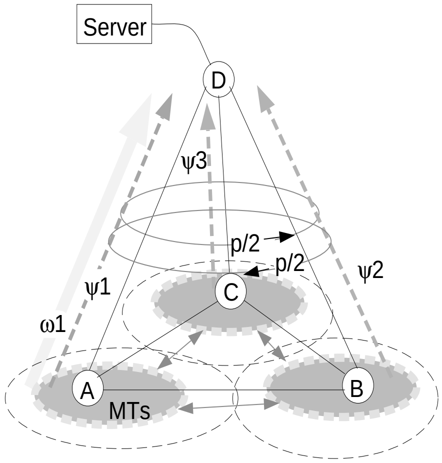

Fig. 5. Example for the Smoothing Effect

flows handoff in than flows handoff out, or decrease, if more flows handoff out than flows handoff in.

These properties are quite intuitive and so the proofs are not given. The third property tells us that at a given time period, the traffic requirement over an OD pair may be larger or smaller because user mobility causes temporal variations on network traffic loads. Although the total traffic load is conserved at any given time, the network capacity requirements could be larger than what for non-mobile-generated traffic, since user mobility may cause traffic requirements varying from time to time but the network topology and link capacity may not permit the corresponding changes.

Previous studies show that the dynamic routing schemes can help the network adapt to traffic load variations. Similarly, we will see the temporal variations can be relieved by dynamic path rerouting algorithms for handoff control.

Consider a simple example from Figure 4. Suppose A is in a residential area while C is in a business area. D is attached with a newspaper server. Mobile users are reading newspapers while they are commuting from home to work or from work to home. In the morning ($t = t_m$), more users move from A to C than from C to A. In the afternoon ($t = t_a$), more users move from C to A than from A to C. They are reflected by $p_{12}(t_m) > p_{21}(t_m)$ and $p_{12}(t_a) < p_{21}(t_a)$, respectively. Suppose, the new requests from A and C are symmetric and constant, i.e. $m_{\omega_1}(t) = m_{\omega_2}(t) = m_0$. The mean value of flows on OD pairs are

$$\begin{aligned} m_1(t) &= \frac{(1+p_{21}(t))(1-p_{12}(t))}{(1-p_{12}(t)p_{21}(t))} T m_0 \qquad (7) \\ m_2(t) &= \frac{(1-p_{21}(t))(1+p_{12}(t))}{1-p_{12}(t)p_{21}(t)} T m_0 \end{aligned}$$

where T is the average service time for traffic flow. When user mobility is symmetric, i.e. $p_{12}(t) = p_{21}(t)$, the traffic flows on both OD pairs are balanced, then the traffic loads are independent of user mobility, $m_1(t) = m_2(t) = T m_0$. Otherwise, the mean values of flows $\psi_1(t)$ and $\psi_2(t)$ depend on the user mobility, as reflected by the handoff probability p_{ij}. For example, if in the morning, $p_{21}(t_m) = 0.1, p_{12}(t_m) = 0.9$, then $m_2(t_m) = 1.8 T m_0$ and $m_1(t_m) = 0.2 T m_0$. Suppose traffic flow $\psi_1(t)$ always takes the route ABD, it is necessary to increase the capacity of AB to meet the traffic with 80% higher mean value. If in the afternoon, the user mobility is in the opposite direction, it requires the increment of the capacity of CB. As a result, to support user mobility, both AB and CB need more capacity!

2) *Covariance and the smoothing effect:* In a small network with few subscribers, the temporal variations caused by user mobility may be significant. However, if the number of subscribers is large, even in a short time period the traffic flow over an OD pair may be balanced by the movement of mobile users. However, even if the traffic flows are balanced, the variations in the traffic flows can still be observed through the second moments – covariances.

If the rate of traffic demand $\Omega(t)$ is a white noise, or equivalently, the traffic demand $\Phi(t)$ is a Brownian motion process, the covariance of the traffic flows satisfies

$$A(t)R_\Psi(t) + R_\Psi(t)A^T(t) + \sigma_\Omega(t)\sigma_\Omega(t)^T = 0 \qquad (8)$$

The proof refers to [16]. The covariances of the traffic flows always depend on the user mobility regardless of the balance of traffic flows. However, when traffic flows are balanced, a *smoothing effect* can be observed, in that the variances of the traffic flows decrease as the user mobility increases.

Without loss of generality, we show this smoothing effect through an example in Figure 5. The rate of traffic demands are iid processes with mean m_ω and variance σ_ω, and user mobility is symmetric among three OD pairs with all $p_{kl} = p/2$. The traffic flows ψ_1, ψ_2, ψ_3 will be same on all three OD pairs because of the symmetry and will have mean and variance

$$\begin{aligned} m_\psi &= \frac{1}{\mu(1-p)} m_\omega = T m_\omega \qquad (9) \\ \sigma_\psi^2 &= R_{kk} = -\frac{\sigma_\omega^2}{2\mu(-1+\frac{p^2}{2-p})} = T\sigma_\omega^2 \frac{2-p}{2(2+p)} \qquad (10) \end{aligned}$$

The correlation coefficient between traffic flows is

$$\rho = R_{ij}/R_{ii} = p/(2-p) \qquad (11)$$

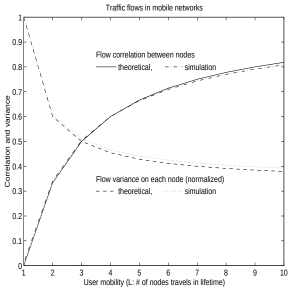

Fig. 6. Traffic Flow Distributions

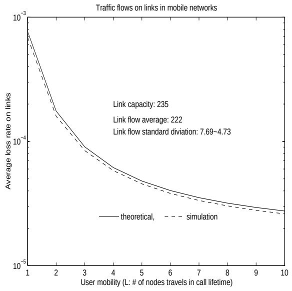

Fig. 7. Data Loss Rate

As user mobility increases, the variances of traffic flows ψ_i decrease comparing with the traffic demands $T\omega_i$, i.e. the traffic flows are being smoothed.

In Figure 6, the variance vs. user mobility is shown, where user mobility $L = 1/(1-p)$ is scaled by the average journey length of a mobile user during its average call lifetime. In addition to the variance, the correlation between traffic flows is shown. As the user mobility increases, the traffic flow correlation increases also. The simulation results (dashed lines) come from the model's simulation based on Equation (1), which verifies the correctness of Equation (8).

Since the flow variances decrease as the user mobility increases, the traffic overflow (data loss rate) will decrease on those links with limited capacity. In Figure 7, we show the loss rate vs. user mobility by assuming traffic flows obey a Gaussian distribution. In general, the loss rate is determined by the tail distributions (or effective bandwidth) of the traffic flows. It is expected that the effective bandwidth of the traffic flows will decrease if the variances of the flows decrease. Thus the smoothing effect of user mobility should reduce the loss rate. The simulation result (dashed line) is the measurement of the average of traffic overflow for a given link capacity in the model's simulation.

From the analysis of the means and covariances of the traffic flows, we observed that user mobility causes temporal variations in the network traffic loads and has the smoothing effect on network traffic distribution. The results tell us that if in any period (short or long), the users' movement among cells is balanced, the traffic distribution over OD pairs is smoothed due to user mobility. In other words, the network service quality may be improved due to the user mobility. However, if the users' movements are not balanced among cells in any period, the traffic load on some OD pairs may increase while the others may decrease. As the result, the network service quality may be degraded due to the user mobility.

B. *Handoff flows and path rerouting algorithms*

When a mobile user has a handover from one wireless access point to another wireless access point, all traffic flows associated with the user must migrate to routes linked to the new access point.

In the example of Figure 4, a call in flow ψ_1 at A can handoff to the flow ψ_2 at C. The traffic for the call must be rerouted to the node C. The rerouting can be done at the alternative anchor points. If the anchor point is chosen as B, the link CB will take the handoff traffic flow from ψ_1. As we pointed out, the handoff traffic may not be balanced, for example, in the morning, there are more handoffs into C than handoffs out of C. The link capacity of CB must be increased to maintain the same QoS. However, it is possible to reroute the handoff traffic at other anchor nodes. For example, reroute the handoff traffic at A, i.e. the handoff flow from ψ_1 can take route $CABD$ instead of CBD. In this case, the link capacity of CA (or AC) needs to be increased to cover the handoff traffic flow while the link capacity of CB can remain the same. If the costs of links CA and AC are lower than the costs of links CB and AB, it is worthwhile to build CA (and/or AC) instead of increasing the capacity of AB and CB.

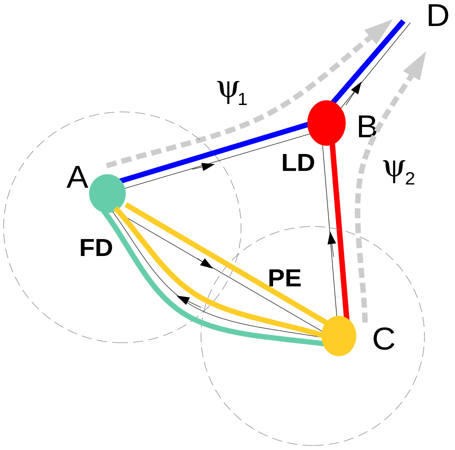

Fig. 8. Performance Analysis of Path Rerouting Algorithms

This motivates the use of alternative path rerouting algorithm for handoff traffic control. A path rerouting algorithm for handoff control specifies where and how a connection path is rerouted. We have given three typical path rerouting algorithms in section II, as illustrated by Figure 1. Theoretically, path rerouting algorithm can be defined by the transform matrix $H_{lk}(t)$ with following features for each algorithm.

- *Last Divergence (LD) Algorithm:* using the LD algorithm, statistically, all traffic flows for OD pair o_k will take the same set of routes $V_k(t)$ regardless they are from new calls or handoff calls, since handoff calls also take the optimal route. This requires the path rerouting transform matrix to satisfy $H_{lk}(t)V_l(t) = V_k(t)$.
- *Path Extension (PE):* in the PE algorithm, the path rerouting transform matrix satisfies $W_{lk} = H_{lk}(t)V_l(t) = (I + \Delta_{lk}(t))V_l(t)$, where $\Delta_{lk}(t)$ represents the handoff flows transfered to the links on a extended path. For a mobile ATM network, if the extended path are long, the performance of PE algorithm will be low. The $\Delta_{lk}(t)$ is always positive which makes the connection path longer and longer after each handoff.
- *First Divergence (FD):* the FD algorithm uses the same extended route as in PE for every handoff, however, the duplicated part is removed. Suppose the path rerouting algorithm satisfies $W_{lk}(t) = H_{lk}(t)V_l(t) = (I + \Delta_{lk}(t))V_l(t)$, where $\Delta_{lk}(t)$ represents the handoff flows transfer away from the links on the old path, and to the links on the new path.

We will outline an analysis procedure to evaluate the performance of these three typical path rerouting algorithms through a simple example. Although the example is simple, as shown in Figure 8, the methodology can be used in general situations. In Figure 8, for the LD algorithm, the COS is always at B, so the handoff traffic flow from ψ_1 takes only the optimal route CBD after handoff. For the PE algorithm, the handoff traffic flow from ψ_1 will take an extended path CA, resulting in the route becoming $CABD$. In case a call is a handoff from A to C and then back from C to A, the handoff flow of the call will take an overall extended path ACA which forms a loop. The route for the call becomes $ACABD$. For the FD algorithm, the COS is at either A or C and the handoff traffic flow from ψ_1 takes the extended path CA which makes the route become $CABD$. The FD algorithm takes the same extended path as the PE algorithm, however, it does not form a flow loop.

To identify the traffic load induced by path rerouting, we must (1) find the initial routing functions, (2) find the path rerouting matrix, (3) solve the balance equation on network links to obtain the routing functions and (4) evaluate the incremental traffic load.

First, we determine the initial routing functions, i.e. the routing functions without user mobility. In this example with five network links AB, AC, CB, CA, BD, they are

$$Z_1 = [1, 0, 0, 0, 1]^T \quad Z_2 = [0, 0, 1, 0, 1]^T \tag{12}$$

Second, we find the path rerouting transform matrices H_{12} and H_{21}. These will be different for three path rerouting algorithms.

1) *LD algorithm:* for the LD algorithm, by definition, the handoff flow in ψ_2 will take the route for flow ψ_1, i.e. $H_{21}V_2 = V_1$. Similarly, we have $H_{12}V_1 = V_2$. The path rerouting transform matrix can be represented as

$$H_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_{12} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$H_{21} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_{21} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where $\Delta_{lk} = H_{lk} - I$, represents the change in routing function due to handoff path rerouting. For example, $\delta_{lk,ij} = 1$ means the handoff flow from flow ψ_l on link i will be added to flow ψ_k onto link j.

The routing function $V_l(t)$ can be obtained by solving both balance equations for OD pairs and network links, Equation (2) and (5). In this case, we have $V_l = Z_l$, meaning the routing function does not change due to handoff.

With the solution of $\{V_l(t)\}$, we can obtain the traffic flow on each network link u,

$$y_u(t) = \sum_{k}^{K} v_{k,u}(t)\psi_k(t) \qquad u = 1, \ldots, M \tag{13}$$

Let $Y(t) = [y_1(t), \ldots, y_M(t)]^T$. In the example, the mean and variance of traffic flows on network links are

$$\begin{aligned} m_Y &= [m_1, 0, m_2, 0, m_1 + m_2] \\ \sigma_Y^2 &= [\sigma_{\psi_1}^2, 0, \sigma_{\psi_1}^2, 0, \sigma_{\omega_1}^2 + \sigma_{\omega_2}^2] \end{aligned}$$

Combined with the statistics we previously obtained for traffic flows, we can analyze the traffic distributions on network links. For example, if mobility pattern is balanced, the only thing is affected by the user mobility is the variances of the flows on AB and CB and they decrease as user mobility increases. If mobility pattern is unbalanced, it is possible that $m_1 > m_2$, which requires AB to have higher capacity when user mobility exists.

2) *PE algorithm:* For the PE algorithm, the COS is always at the OldBS and every handoff extends the previous connection path by one more hop. This implies

$$H_{12} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$H_{21} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_{21} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Solving (5) based on the solution of Equation (6), we find that the routing functions for the PE algorithm are

$$[V_1 V_2] = \begin{pmatrix} \frac{1}{1+p_{21}} & \frac{p_{12}}{1+p_{12}} & : AB \\ \frac{p_{12}p_{21}}{1-p_{12}p_{21}} & \frac{p_{12}(1+p_{21})}{(1+p_{12})(1-p_{12}p_{21})} & : AC \\ \frac{p_{21}}{1+p_{21}} & \frac{1}{1+p_{12}} & : CB \\ \frac{p_{21}(1+p_{12})}{(1+p_{21})(1-p_{12}p_{21})} & \frac{p_{12}p_{21}}{1-p_{12}p_{21}} & : CA \\ 1 & 1 & : BD \end{pmatrix}$$

The routing functions depend on user mobility and they show how much of a given flow uses a specific network link. We can get the means and variances of link traffic flows $\{y_u(t)\}$ based on the solution of Equation (2), for traffic flows on OD pairs. In our example, the mean traffic flow on AB is $m_{AB} = v_{1,1}m_1 + v_{2,1}m_2$. By substituting m_1 and m_2 from Equation (8), we have $m_{AB} = Tm_0$, which is independent of user mobility. While the mean traffic flow on AC is

$$m_{AC} = v_{1,2}m_1 + v_{2,2}m_2 = Tm_0 \frac{p_{12}(1 + p_{21})}{(1 - p_{12}p_{21})} \tag{14}$$

which will increase greatly as user mobility increases. The cost of the PE algorithm is the traffic flows on links AC and CA. To reduce the traffic flows on extended paths, we may use the FD algorithm which removes possible duplication.

3) *FD algorithm:* For the FD algorithm, the COS is at the OldBS (or the NewBS) and a handoff may increase or decrease the path length by one. Thus,

$$H_{12} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$H_{21} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \Delta_{21} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Given H_{12} and H_{21}, the solution of (5) and (6) yields

$$[V_1 V_2] = \begin{pmatrix} 1/(1+p_{21}) & p_{12}/(1+p_{12}) & : AB \\ p_{21}/(1+p_{21}) & 0 & : AC \\ p_{21}/(1+p_{21}) & 1/(1+p_{12}) & : CB \\ 0 & p_{12}/(1+p_{12}) & : CA \\ 1 & 1 & : BD \end{pmatrix}$$

The routing functions depend on user mobility. However, unlike the PE algorithm, the traffic load on links AC and CA is not high since there is no duplicate usage of the links. The mean of traffic flow, $m_{AC} = Tm_0(v_{1,2}m_1 + v_{2,2}m_2) = Tm_0 p_{21}(1 - P_{12})/(1 - p_{12}p_{21})$, won't grow as fast as the traffic load for the PE algorithm. On the other hand, the total traffic loads on links AB and CB are same as that of the PE algorithm, which are independent of user mobility. So the FD algorithm gets the advantage of PE algorithm with a much lower cost.

C. Performance evaluation

The solution of the routing functions $V(t)$ can help us redesign the network topology or, if the network topology is fixed, evaluate which path rerouting algorithm should be chosen. As we have done in the example, we can analyze the mean and variance of the traffic flows based on Equation (13) and the solutions from flow balance Equation (2). However, we may want to have a general idea how good or bad is a path rerouting algorithm in a given network. One criterion is the sum of traffic flows on all links, or *network traffic volume*, represented

$$|Y(t)| = \sum_{u}^{M} y_u(t) \tag{15}$$

A good path rerouting algorithm should have little increment in network traffic volume due to handoff control. In the example of Figure 8, when handoff traffic flows are balanced, that is $p_{12} = p_{21} = p$, the network traffic volumes are $2(\psi_1+\psi_2)$, $(2+p/(1-p))(\psi_1+\psi_2)$ and $(2+p/(1+p))(\psi_1+\psi_2)$, for LD, PE and FD algorithms, respectively.

More generally, without the knowledge of traffic requirements $\psi_k(t)$, we may use link costs to evaluate a network topology. In particular, given g_u, the weight representing the cost of link u, we define the *route lengths of OD pairs* as

$$|V_l(t)| = \sum_{u}^{M} g_u v_{l,u}(t) \tag{16}$$

In Figure 8, the cost of link AB may be greater than the cost of link AC. In our example, assuming $g_u = 1$ for all u, the route lengths of both OD pairs are 2, $2 + p/(1-p)$ and $2 + p/(1+p)$, for LD, PE and FD algorithms, respectively. Figure 9 shows the route length versus the user mobility, as measured by $L = 1/(1-p)$ which is the average number of cells a mobile users visits during a call lifetime. It is obvious that the PE path rerouting has the route length linearly increased as user mobility increases. Using FD algorithm, only a slightly increment in route length due to user mobility. While using the LD algorithm, the route length is supposed to be a constant relative to the user mobility.

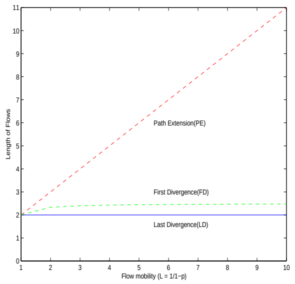

Fig. 9. The OD Pair Length

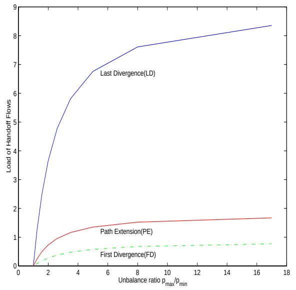

Fig. 10. The Handoff New Path Length

When link costs g_u are not uniform, for example $g_{AB} = g_{CB} = 5, g_{AC} = g_{CA} = 1, g_{BD} = 1$, the lengths of traffic flows for LD, PE and FD algorithms will be 6, $6 + p/1 - p$ and $6 + p/1 + p$. The relative increasing rate of the lengths for the PE and FD algorithms will be smaller compared with the original length of traffic flows.

In the case of unbalanced traffic, the performance cannot be evaluated only by the route length of OD pairs. The traffic flows ψ_1 and ψ_2 may reach their peak values at different times. If the new path which takes handoff traffic flow is long, the extra traffic load caused by user mobility will be large. For example, if the peak value of m_1 is $1.8Tm_0$, the extra cost of using LD algorithm is $0.8g_{AB} = 4$ while only $0.8g_{AC} = 0.8$ for FD algorithm. Thus in the case of unbalanced traffic, we can evaluate the performance of path rerouting algorithms using the *handoff new path route length*

$$|U_k(t)| = |\sum_{l} H_{lk}(t)V_l(t) - Z_k(t)| \tag{17}$$

for the route length of OD pair k. In our example, suppose at time t_m, $p_{12} = p_{max}$ and $p_{21} = p_{min}$, at time t_a, $p_{12} = p_{min}$ and $p_{21} = p_{max}$. And $g_{AB} = g_{CB} = 5, g_{AC} = g_{CA} = 1, g_{BD} = 1$. The handoff new path route length is shown in Figure 10 as a function of the imbalance ratio p_{max}/p_{min}.

Through simple examples, we have obtained analytical results on the performance of typical path rerouting algorithms. They are reflected by the route lengths of OD pairs and the route length of handoff new paths. We have observed that the PE algorithm is not a good choice if the user mobility is high. And the LD algorithm may not be a good choice if traffic is unbalanced. For both concerns, it seems the FD algorithm is a good choice since it has the smallest handoff new path route length and only a slightly greater overall route lengths in OD pairs. We now verify the analytical observations through simulations of a larger network configuration.

IV. PERFORMANCE ANALYSIS THROUGH SIMULATIONS

In the previous section, we have seen that the route lengths of OD pairs and the route lengths of handoff new paths can measure the performance of path rerouting algorithms. The route lengths of OD pairs reflect overall traffic load in the network

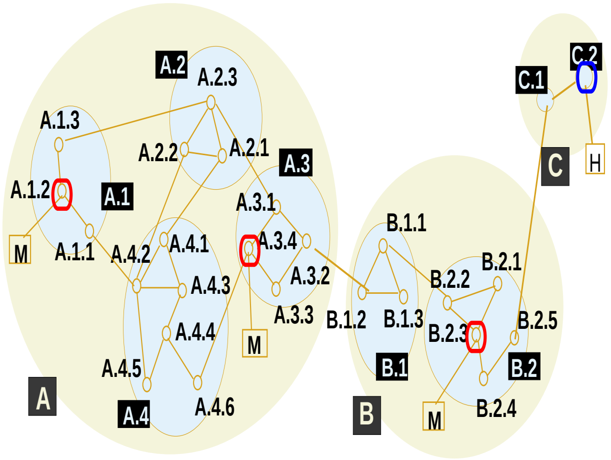

Fig. 11. The PNNI Network in Simulation

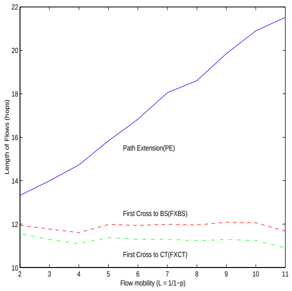

Fig. 12. Average OD Pair Length

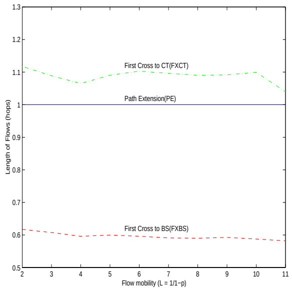

Fig. 13. Average New Path Length

Simulation with fixed mobility neighborhood

and the route lengths of handoff new paths reflect the extra traffic load when handoff traffic flows are unbalanced.

In the simulations, we are going to measure the connection path length of individual calls to estimate the route lengths of OD pairs and measure the new path length of individual call handoff to estimate the route lengths of handoff new paths.

The network used in the simulation is as shown in Figure 11. It is an example in ATM Forum PNNI specifications [17] for PNNI hierarchical structure. In this network, there are 27 switches and three hierarchical levels. We use peer group C as a fixed backbone network and peer group A and B as the mobile ATM access network. Every switch in group A and B is considered a basestation. Mobile terminals can be attached through any switch in A, B and move (handoff) from one to another.

In the first simulation, we assume the user mobility pattern matches the network topology, i.e. any two nodes within A, B are neighboring access points (having an $M - edge$) if there is a link ($C - edge$) between them.

The simulation is conducted as follows. Three mobile terminals starting calls at $A.1.2$, $A.3.4$ and $B.2.3$ to a remote server at $C.2$ in the fixed network. Each call starts with an exponential distributed lifetime t_f with average L and performs t_f handoffs then terminates. At the access point where the call is terminated, a new call is started. The procedure continues until each mobile terminal performs 10,000 handoffs. A mobile user will randomly handoff to one of the neighboring cells with equal probability.

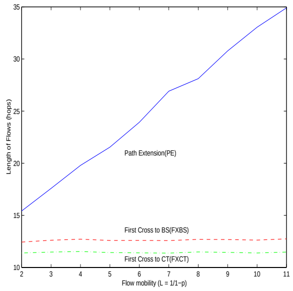

Fig. 14. Average Path Length

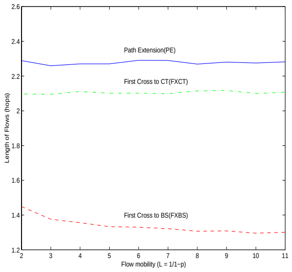

Fig. 15. Average New Path Length

Simulation with random mobility neighborhood

In the simulation, we use three path rerouting algorithms, Path Extension (PE), First-Cross-to-CT (FXCT, as an approximation of LD algorithm, which uses first crossover switch toward the CT as the COS) and First-Cross-to-OldBS (FXBS, as an approximation of FD algorithm, which uses first crossover switch toward the OldBS as the COS).

Figure IV show the route length for the three path rerouting algorithms. It is obvious that PE has linear increasing in route length. Higher user mobility implies more handoffs per call lifetime. It is not surprising that FXBS has only a little extra path length over the FXCT algorithm and is almost constant as mobility increases. This is consistent with our analytical example in Figure 9. Using the FXBS algorithm, the network needs about 10% more capacity than using FXCT algorithm, however, the capacity increment does not change much as the user mobility increases.

Figure 13 shows the route lengths of handoff new paths for three algorithms. They do not change much with the user mobility. However, they will cause the handoff traffic flow to increase at different rate when unbalanced ratio increases. FXCT algorithm has a new path length about 0.5 hops more than what of the FXBS algorithm. When handoff flows are not balanced, the capacity increment on the new path for FXCT algorithm can be much larger than FXBS algorithm. Since the new path takes the handoff flows, if handoff traffic flows double the original traffic load on the new paths, the FXCT algorithm needs one more hop ($2 * 0.5$) on average than the FXBS algorithm, which is also about 10% of total average path length. So depending upon the mobility pattern, one algorithm may consume less of the network resources.

In the second simulation, the neighborhood of the access network is randomly generated among the nodes in peer groups A, B. On the average, each node has 4 neighbors and the neighbors are randomly selected. If a node x is closer to y in terms of PNNI hierarchy, the probability that x is a neighbor of y is higher.

From Figure 14, we can see the route lengths for FXCT and FXBS are about the same as the first example. The FXBS needs about 10% more hops than FXCT algorithm. However, in Figure 15, we observe the difference between the route length for handoff new paths for FXCT and FXBS is bigger about 0.7 hops. Therefore when handoff flows are unbalanced, the handoff flows may cause more capacity requirements for the FXCT algorithm than the FXBS.

V. CONCLUSION

We have investigated the traffic distribution in mobile ATM networks through a dynamic flow model. We observed that user mobility will produce two effects over network traffic. One is the temporal variations in network traffic loads which can cost network resources when the handoff traffic is unbalanced and the other is the smoothing effect of the network traffic distribution which can relieve traffic load, especially when handoff traffic is balanced. Typical handoff path rerouting algorithms for handoff control are evaluated. By measuring the route lengths of OD pairs and the route length of handoff new paths, we conclude that the FD (or FXBS) algorithm has advantages over other algorithms in most situations, especially when the link capacity in the access network costs less than that in the core network. The performance analysis can also help us redesign the network topology by adding more links or increasing capacity in the network based on the user mobility pattern. Furthermore, the results may suggest a rearrangement of access points to balance handoff traffic flows. Future work will include analytical models for PNNI hierarchical networks and the simulation in a large scale network setup.

REFERENCES

[1] D. Raychaudhuri and N. D. Wilson. ATM-based transport architecture for multiservices wireless personal communication networks. *IEEE Journal on Selected Areas in Communications*, 12(8):1401–1414, December 1994.

[2] D.Raychaudhuri, R.Yuan, A.Iwata, and H.Suzuki. Rationale and framework for wireless ATM specification. ATM Forum/95-1646/PLEN, 1995.

[3] R. Yuan, S.K.Biswas, L.J.French, J.Li, and D.Raychaudhuri. A signaling and control architecture for mobility support in wireless ATM networks. *ACM/Baltzer Mobile Networks and Applications*, 1(3), December 1996.

[4] A. Acharya, J.Li, and D. Raychaudhuri. Mobility management in wireless ATM networks. *IEEE Communication Magazine*, 35(11):100–109, November 1997.

[5] A. Acampora and M. Naghshineh. An architecture and methodology for mobile-executed handoff in cellular ATM networks. *IEEE Journal on Selected Areas in Communications*, 12(8):1365–1375, December 1994.

[6] C. K. Toh. Crossover switch discovery for wireless ATM LANs. *ACM/Baltzer Mobile Networks and Nomadic Applications*, 1(2), December 1996.

[7] M. Veeraraghavan, M.Karol, and K.Eng. Mobility and connection management in a wireless ATM LAN. *IEEE Journal on Selected Areas in Communications*, 15(1):50–68, January 1997.

[8] H. Mitts, H. Hansen, J. Immonen, and S. Veikkolainen. Lossless handover for wireless ATM. *ACM/Baltzer Mobile Networks and Applications*, 1(3), December 1996.

[9] B.Akyol and D. Cox. Rerouting for handoff in a wireless ATM network. *IEEE Personal Communications*, 3(5), October 1996.

[10] S. Seshan, H. Balakrishnan, and R.H. Katz. Handoffs in cellular wireless networks: the daedalus implementation and experience. *IEEE Wireless Personal Communications*, 4(2):141–162, March 1997.

[11] J.Li, R.Yates, and D. Raychaudhuri. Unified handoff control protocol for dynamic path rerouting in mobile ATM networks. In *Proceedings of the Ninth IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, Boston, MA, September 1998.

[12] J. Li, A. Acharya, and D. Raychaudhuri. Signaling syntax extensions for handoff control in mobile ATM. ATM Forum/96-1625, December 1996.

[13] A.Acharya, J. Li, and D. Raychaudhuri. Primitives for location management and handoff control in mobile ATM networks. ATM Forum/96-1121, August 1996.

[14] B. Rajagopalan H.Mitts, K.Rauhala and G.Bautz. Proposed handover signaling architecture for release 1.0 WATM baseline. ATM Forum/97-0845, September 1997.

[15] The ATM Forum. Baseline text for wireless ATM specifications. ATM Forum/BTD-WATM-01.16, Feburary 1998.

[16] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag Press, 1988.

[17] The ATM Forum. Private Network-Network Interface Specification (PNNI) Version 2.0. ATM Forum/BTD-PNNI 02.00, September 1997.