Connection architecture and protocols to support efficient handoffs over an ATM/B-ISDN personal communications network [☆]

Oliver T.W. Yu and Victor C.M. Leung

Department of Electrical Engineering, The University of British Columbia, 2356 Main Mall, Vancouver, B.C., Canada V6T 1Z4

Abstract. The next generation personal communication network will likely internetwork wireless networks via the ATM/B-ISDN to enable ubiquitous broadband personal communication services. Support of user terminal mobility, particularly the capability for fast and seamless handoffs, over the ATM/B-ISDN is an expected requirement that is not currently met. We propose extensions to the ATM/B-ISDN user transport and signaling network architectures and signaling protocols to meet these requirements. The new architecture employs the Mobile Virtual Circuit (MVC), a dynamic connection tree in which routes are predetermined but not set up for potential handoff connections. During a handoff, associated signaling using source-routing with a new robust adaptation feature is employed for fast resource allocation to establish the handoff connection by distributed control. We also address the new problem of packet ordering synchronization to enable a seamless handoff. The connection tree reconfigures after each handoff to enable continuous support of successive handoffs. The proposed scheme optimizes handoff delay over the ATM/B-ISDN while minimizing unnecessary resource allocation, chances of handoff failure, and call processing load in the intelligent network, and the extensions are backward compatible to current ATM/B-ISDN standards and implementations.

1. Introduction

Personal communication services (PCS) support ubiquitous personal and terminal mobility. The broadband integrated services digital network (B-ISDN) employing asynchronous transfer mode (ATM) [1–3] is a telecommunication network architecture that provides broadband and multimedia communication services. As air interface technologies are being developed to support broadband wireless access [4–6], it is expected that the next generation personal communication network (PCN) will call upon the ATM/B-ISDN to internetwork broadband wireless networks [7–9]. This motivates us to examine the issues and problems of extending PCS capability to the ATM/B-ISDN to support user terminal mobility during calls, i.e., call handoff management. This paper proposes extensions (Fig. 1) to the user transport network architecture, the signaling transport network architecture and the control signaling protocols of ATM/B-ISDN to globally support fast and seamless call handoff operations that are efficient in network resource utilization. The rationale and the alternatives for the proposed extensions are discussed below.

The architecture of a first generation PCN consists of separate narrowband mobile access networks, generally configured as regional hierarchical structures with mobile terminals (MTs) in radio cells accessing base stations (BSs) connected to mobile switching centers (MSCs), which in turn interconnect with the backbone public switched network (PSN) via gateway MSCs [10–12]. Handoff management is required when MTs cross cell boundaries while engaged in calls, and involves reassignment of radio channels and possibly rerouting of network connections. First generation PCNs restrict handoff management to cell areas associated with each individual mobile access network. As PCNs evolve to
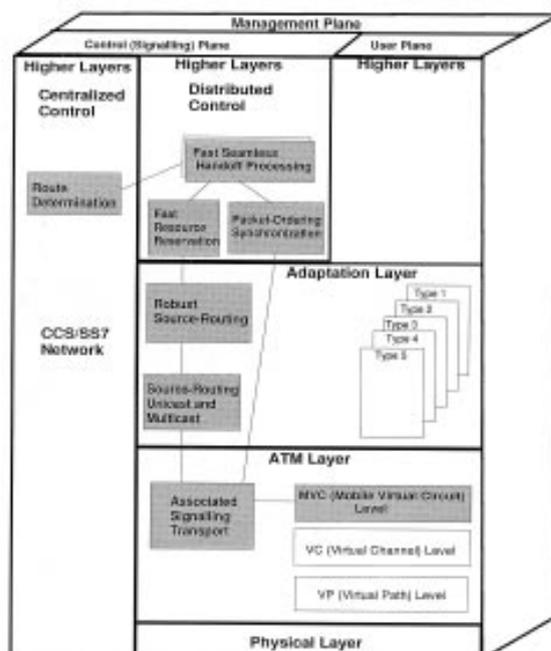

Fig. 1. B-ISDN protocol model to support mobile communications.

accommodate network-wide terminal mobility and to employ an ATM/B-ISDN backbone interconnecting broadband wireless networks, support of a mobile connection-oriented bearer service across the backbone which provides the same quality of service as a fixed connection would be essential. Such a mobile bearer service must employ a connection architecture that supports network-wide mobility management over the ATM/B-ISDN backbone to enable unrestricted terminal mobility over adjacent cells or mobile networks connected to different parts of the backbone. To minimize forced call terminations due to loss of radio contact before handoff completions, the connection architecture and associated protocols must enable fast handoff processing while utilizing network resources efficiently. This is particularly important as forced call terminations are considered much more objectionable than blocked call initiations.

In future broadband PCNs, call disruptions associated with hard handoffs over FDMA air interfaces may result in substantial and unacceptable loss of user information. TDMA air interfaces enable seamless handoffs by switching channels and time slots within a TDMA frame, while soft handoffs over CDMA air interfaces which coherently combine simultaneous transmissions over the new and old cells are inherently seamless. In both cases, it would be desirable for the mobile bearer service over an ATM/B-ISDN to support seamless handoff by incorporating protocols that preserve packet sequencing between the old and the new ATM connections. Note that a necessary condition for realizing a seamless handoff is that a new connection to the MT must be established before the old connection is terminated. Thus signaling contact with the old BS must be maintained until the new BS is ready to take over user traffic transport. This increases the potential for forced call termination unless the handoff operation can be completed before the handoff initiation signal margin is exhausted and radio contact with the BS is lost. Therefore, a mobile bearer service that supports seamless handoff must also incorporate fast handoff capabilities.

Existing ATM/B-ISDNs can support handoffs by terminating an old connection and then establishing a new connection in realtime by the usual means, i.e., employing the intelligent network (IN) [13] supported by the common channel signaling system 7 (SS7) [14,15] for route determination and resource allocation. However, while this approach economizes on resource utilization, handoffs are not seamless and potentially may be subject to long delays, due to possible heavy handoff processing load at the IN when a significant amount of PCS traffic is present. A connection tree architecture linking BSs to a common root node could simplify and speed up handoff management by restricting handoff rerouting to the links between the MT and the root node, and by identifying these links as a group, eliminating routing table updates in nodes along the new connection. This is the basis of the permanently assigned connection tree archi-

tecture [16] which connects all BSs in a mobile access network to the root node using permanent virtual channels (PVCs) in the ATM layer. Within the geographical area covered by the BSs, MTs contend for the connections. This architecture limits mobility support to within an area covered by the mobile access network, and causes network resources to be partitioned between mobile and non-mobile traffic. In contrast, we propose a new mobile virtual circuit (MVC) connection architecture consisting of dynamically configured per-call connection trees. Implemented as an extension to the ATM layer, each MVC connects the BS in the current cell serving an MT and the BSs in all the immediate neighboring cells to which the MT may potentially hand off its call, to the root node herein called the tethered-point (TP). After a handoff, the TP may shift to another node as the MVC is reconfigured. The MVC has all the desirable properties of the connection tree discussed above, while providing network-wide handoff management support across cells or mobile access networks connected to any points of the backbone network, and allowing complete sharing of network resources between mobile and non-mobile traffic. The MVC connection configuration, connection establishment, connection rerouting during handoffs and connection reconfiguration after handoffs are described in section 2.

The MVC connection establishment scheme employs the IN for route determination in configuring the MVC. While fast handoff is easily accomplished in the MVC architecture by preallocating resources over all potential handoff connections, this method is inefficient in resource utilization since most connections will remain unused. To optimize resource utilization while enabling a fast handoff operation, we propose to defer resource allocation until the handoff operation, whence establishment of the actual handoff connection is completed without IN involvement, by using a fast resource allocation scheme supported by robust source-routing over an associated signaling transport network that share the same physical and logical transmission network with the user transport network. Since the fast resource allocation scheme is based on predetermined routes, it is sensitive to network congestions and link failures. The proposed robust adaptation scheme is designed to minimize the chance of handoff failures due to these conditions. Details of these proposals are given in section 3. To enable seamless handoff over an ATM network, one particular problem that has received little attention in the literature is the synchronization of traffic between the new and the old connection after and before a handoff. We propose a packet synchronization scheme employing inband signaling to preserve packet ordering over the ATM/B-ISDN during virtual circuit rerouting. The proposed protocol for fast and seamless handoffs with packet synchronization is presented in section 4. It is a higher layer protocol in the distributed-control plane of the B-ISDN protocol model as illustrated in Fig. 1.

The flexibility offered by the MVC architecture requires that any network node be capable of operating as a TP, and that the distributed handoff signaling protocols mentioned above be supported by all network nodes. It is shown in the paper that the proposed extensions to the ATM/B-ISDN architecture and switch functions are readily implemented in a manner that is backward compatible with existing network architecture and equipment designs. The advantage of the MVC architecture is illustrated by the analysis in section 5, which shows that handoff delay performance is optimized by minimizing the average lengths of the branches of the connection tree by locating the TP as close to mobile access points as possible. Section 6 concludes the paper.

2. MVC connection architecture

The neighbouring cells of a given cell determine the potential handoff connections in the backbone network, associated with each ongoing call in this cell. For example, there are six potential handoff connections for a mobile terminal in a six-sided regular cell pattern (Fig. 2). As discussed in the introduction, our approach to enable fast handoffs over the ATM/B-ISDN backbone while minimizing resource utilization is to determine the routes for all potential handoff connections at call setup via IN control, to employ a fast resource allocation scheme for completing the establishment of the selected handoff connection, and to enable fast handoff processing during call handoff without IN involvement. The proposed approach calls for a multipoint-to-point connection or connection tree configuration to link the current and the potential handoff connections between the MT and the network. The connection tree topology can be either distinct-threaded, with individual links specific to each of the multiple endpoints (i.e., specific links)
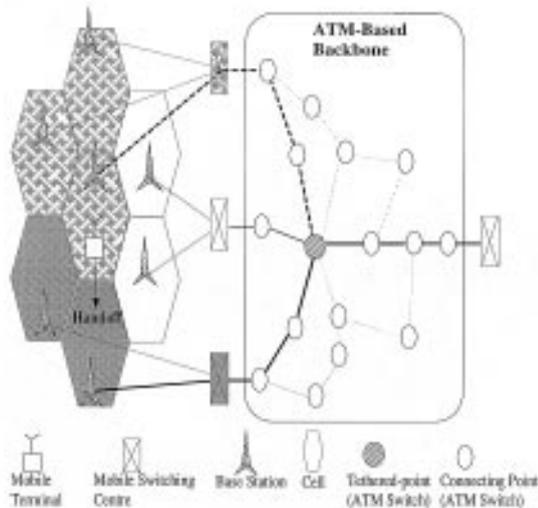
converging at a common endpoint at the other end, or common-threaded, with specific links from the multiple endpoints converging at a common intermediate node, namely the TP, within the network before progressing to the endpoint at the other end over common links. In the extreme case where the TP maps into the single endpoint, the common-threaded multipoint-to-point connection reduces to the corresponding distinct-threaded multipoint-to-point connection.

In section 5, it will be shown that the proposed fast seamless handoff scheme produces better performance with the common-threaded multipoint-to-point connections than with the corresponding distinct-threaded multipoint-to-point connections. Currently, ATM does not support common-threaded multipoint-to-point connections. In the following subsections, we propose to extend the ATM architecture to include the MVC connection, a dual common-threaded multipoint-to-point connection (Fig. 3) to support mobile connection-oriented bearer service over B-ISDN. To support network-wide terminal mobility, the potential handoff connections of the MVC are reconfigured after each call handoff to enable continuous support of successive handoffs during a call session.

2.1. MVC connection configuration

The proposed MVC connection is conceptually implemented at a new level of the ATM layer hierarchi-

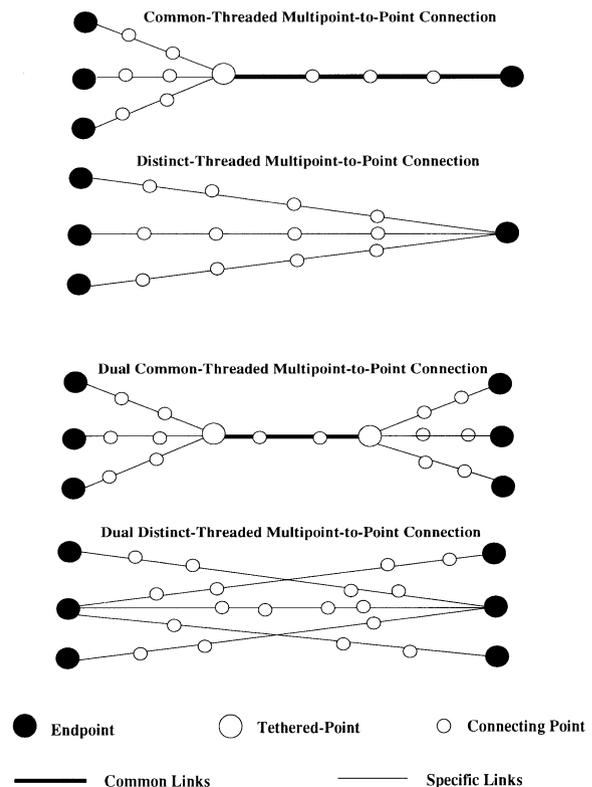Fig. 2. Location independent transport over backbone ATM network.

Fig. 3. Connection tree configurations.

cally above the VC and the VP levels (Fig. 4). A generalization of the fixed connection tree in [16], this topology dynamically changes the multipoint-to-point connection and relocates the TP to another node after each handoff. As stated before, the multiple endpoints include the current BS and the BSs in neighboring potential handoff cells. The single endpoint at the other end represents a fixed terminal, or a BS associated with a peer MT. The connecting points are ATM switches. The original connection and each of the potential handoff connections is composed of the specific connecting links and the common connecting links. The specific connecting links represent the "mobile" links which need to be reconfigured during call handoff and the common connecting links represent the "fixed" links that stay constant during call handoff. The TP represents the connecting point where the mobile links converge to a common link. Thus a MVC connection consists of multiple mobile specific link-sets jointed to a single fixed common link-set at a TP. Note that after each handoff, a new set of mobile links are chosen, and the TP is reconfigured to a possibly different connecting point in the network where the new set of mobile links converge. In the usual hexagonal cell grid, after a handoff the set of reconfigured potential handoff BSs includes the old BS, two previous potential handoff BSs, and three new BSs.

All logical links of the multipoint-to-point connection between the multiple endpoints and the TP are supported by the proposed group virtual channel (GVC). At the VC level, a logical connection between two points is defined by a series of logical link identifiers associated with the concatenation of logical links. If the multipoint-to-point connection were to be implemented at the VC level, then the logical connections between the multiple endpoints and the TP are defined by different series of logical link identifiers and the incoming logical link identifiers at the TP will be all different. On the other hand, at the proposed GVC level, all logical connections between the multiple endpoints and the TP are defined by a single logical link identifier associated with all the concatenated links that is unique at the TP (Fig. 5). Consequently, the employment of GVC allows a TP to associate a single GVC identifier (GVCI) with a group of endpoints associated with an MT. The GVCI is encoded in the standard VC identifier (VCI) field of the cell header.

The GVC is supported over a group of virtual paths (VPs) which correspond to the set of mobile specific links of the original and potential handoff connections. The VPs are fixed by network topology and bind the BSs to the TP. The VCs are multiplexed within a VP which is in turn multiplexed with others within a physical link. Unlike a VC, a VP route can be established without assigning bandwidth along the path. Since VPs are usually established permanently or for a long term basis, GVCs can be established over them on demand. Consequently, the support of the GVC over a group of VPs allows the TP to identify the MT via the GVC and to identify the mobile specific link-sets from the MT to the TP via the VPs.

Each MVC requires reservation of the following logical link identifiers: (1) the incoming and outgoing VPIs at each connecting point of current or potential handoff mobile links, up to the TP; (2) a common GVCI for all the mobile links reserved at the TP; and (3) the incoming and outgoing VCIs and VPIs reserved at each connecting point of common fixed links, starting from the TP. ATM switching nodes generally support VP crossconnection and VC/VP crossconnection, using the first bit of the 12-bit VPI field to indicate the crossconnection type. The GVCI is encoded in the 16-bit VCI field. Consequently, each node operating as a TP can maintain a local limit of 2^{16} GVCIs to identify MVCs supporting mobile calls, and each node operating as a mobile connection connecting point can maintain a local limit of 2^{11} VPI to identify mobile links. In general, MVCs multiplexed into a common VP over a mobile link must have distinct GVCIs. Therefore the GVCI space must be managed so that if the connection trees emanating from two TPs share common VPs over some mobile links, then the GVCIs assigned at these TPs do not overlap. Alternately, different VPIs could be assigned to ensure that physically overlapping connection trees are logically distinct. Even though switches may not be able to process the full range of GVCI or VPI values, by proper
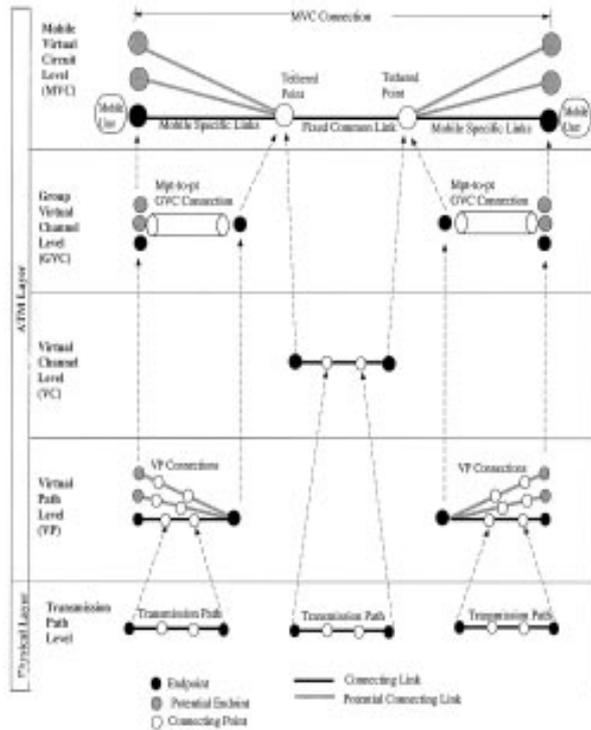

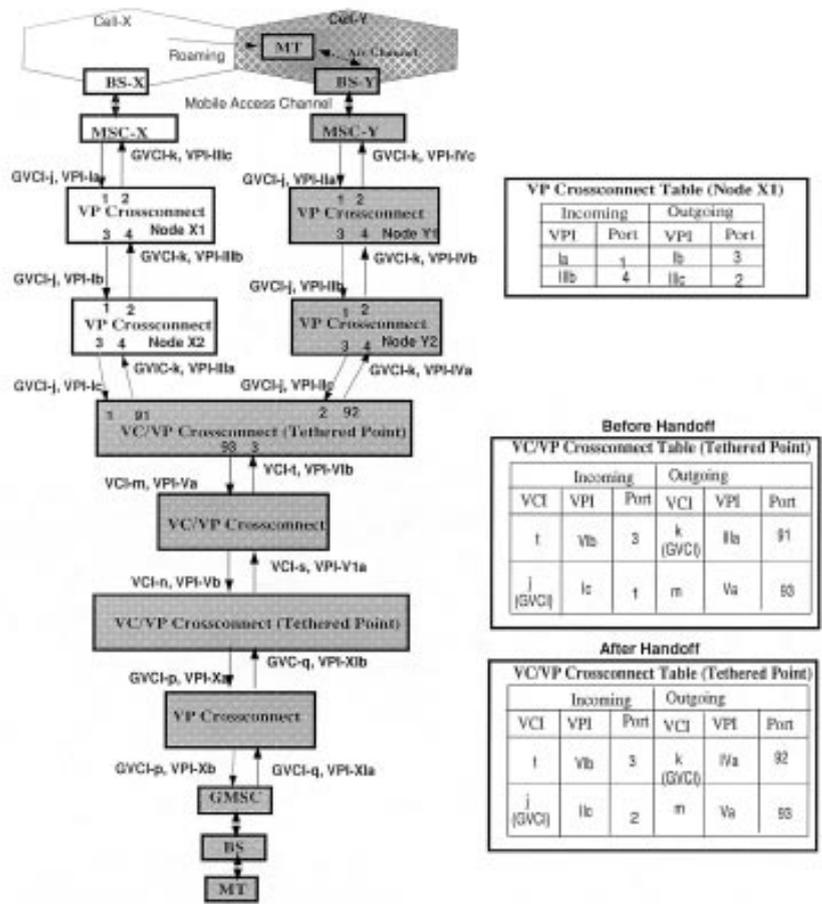Fig. 4. MVC connection architecture to support mobile communications.

Fig. 5. MVC mobile links rerouting during handoff.

management of the address space and reusing addresses where possible, a substantial number of MTs can be globally accommodated by the MVC connection architecture.

2.2. MVC connection establishment

In a general connection-oriented packet-switched network, connection establishment involves the following steps: (1) reservation of logical link identifier at each switch associated with the connection; (2) establishment of routing information for translating incoming logical link identifiers into outgoing logical link identifiers; (3) reservation of communication resources (buffer and physical link bandwidth) at each switch. The same procedures are employed in the ATM / B-ISDN which employ fast packet switching with IN support, where logical links are identified by virtual circuit identifiers (VCIs) or virtual path identifiers (VPIs) reserved via the IN. The new connection establishment scheme is based on reserving logical link identifiers via the IN for handoff connections, but delaying resource reservation until handoff processing (which does not involve the IN), when a fast resource reservation scheme is invoked to complete the setup of the handoff connection.

The proposed scheme for establishment of the MVC connection during initial call processing is decomposed into the following tasks, with logical link identifiers (GVCI, VCI and VPI), routing information, and resource requirements provided by the IN:

1. Establishment of the fixed common connecting links shared by the original connection and the handoff connection:

   (a) Standard VC connection establishment along the fixed common connecting links, i.e., VCIs and routing information entries in VC switch tables are reserved by the ATM switches along the fixed common connecting links.
   (b) Communication resources are reserved to satisfy connection performance requirements.

2. Establishment of the mobile specific link-sets between the multiple BSs and the TP:

   (a) GVC connection establishment between the BSs group and the TP, a single unique logical link identifier in the form of GVCI is reserved by the TP.
   (b) Standard VP connection establishment between each BS and the TP (since the VPs are usually

pre-established permanently by the network, this step may not be necessary), each BS stores its outgoing VPI accordingly.

(c) Bandwidth for the GVC connection is reserved from the VP associated with the mobile link-set specific to the original connection.

(d) For those VPs associated with the mobile link-sets specific to the potential handoff connections, bandwidth is not reserved immediately for the GVC connection; instead, a fast resource reservation scheme described later is employed to reserve bandwidth for the actual handoff connection during handoff processing.

After a handoff, the actual handoff connection becomes the original connection subject to a potential future handoff, and the TP may be mapped to a different connection point. In anticipation of a future handoff, once handoff processing is complete, the MVC needs to be reconfigured by repeating the above process. Due to the possible re-mapping of the TP to a different node, a mobile link may be changed to a common link and vice versa. The corresponding logical link identifiers therefore need to be changed accordingly. The IN is involved in this operation which is not as time critical as the handoff operation. The MVC reconfiguration process after handoff will be explained in subsection 2.4.

2.3. MVC rerouting during handoff

Mobile link-set rerouting within the MVC connection in response to MT handoff, is illustrated in Fig. 5, where an MT has roamed from cellular region X to cellular region Y and is accessing a new BS controlled by a different MSC.

Consider the initial situation when the MT was located in cellular region X. The ATM cells from the old MSC X arrives at port 1 of node X1 bearing *GVCI-j* and *VPI-Ia*. Node X1 performs VP crossconnection: it translates *VPI-Ia* at port 1 to *VPI-Ib* at port 3 with the outgoing ATM cells bearing *GVCI-j* and *VPI-Ib*. Then the ATM cells arrive at port 1 of node X2 bearing *GVCI-j* and *VPI-Ib*. Node X2 performs similar VP crossconnection. After passing through node X2, the ATM cells arrive at port 1 of the TP bearing *GVCI-j* and *VPI-Ic*. Since both the GVCI and the VCI are encoded in the same field of the ATM cell header, the TP performs standard VC crossconnection and it translates *GVCI-j* and *VPI-Ic* at port 1 to *VCI-m* and *VPI-Va* at port 93. On the other hand, the reverse traffic from the fixed link-set would arrive at port 3 of the TP bearing *VCI-t* and *VPI-VIb*. The TP performs standard VC crossconnection and it translates *VCI-t* and *VPI-VIb* at port 3 to *GVCI-k* and *VPI-IIIa* at port 91.

Now consider the handoff situation when the MT is about to enter cellular region Y. The old BS-X would signal the TP to indicate that the GVC-j connection will exit from VP-I and the new BS-Y would signal the TP to indicate that GVC-j connection will now be multiplexed instead within VP-II. The TP rerouting controller would update both the forward and reverse routes translation stored in the VC crossconnection table accordingly; i.e., the reverse group virtual circuit connection GVC-k will be moved from VP-III to VP-IV.

After successful handoff signaling, the forward traffic from MSC-Y would arrive at port 2 of the TP bearing *GVCI-j* and *VPI-IIc*. The TP performs standard VC crossconnection and it translates *GVCI-j* and *VPI-IIc* at port 2 to *VCI-m* and *VPI-Va* at port 93. On the other hand, the reverse traffic from the fixed link-set would arrive at port 3 of the TP bearing *VCI-t* and *VPI-VIb*. The TP performs standard VC crossconnection and it translates *VCI-t* and *VPI-VIb* at port 3 to *GVCI-k* and *VPI-IVa* at port 92.

2.4. MVC reconfiguration after handoff

Given that the connection tree concept is employed to support user terminal mobility over the ATM/B-ISDN and that the MT has unrestricted movement across the mobile networks, it is necessary to address the problem of supporting multiple successive handoffs over the ATM/B-ISDN as an MT roams to adjacent mobile cellular regions. One solution is to assume that the range of movement of an MT can be predetermined during call setup, in which case a static connection tree [16] is sufficient to support user terminal mobility. Essentially, this solution only supports restricted successive handoffs and forced call terminations would occur when an MT roams outside the assumed mobility range. In contrast, we allow an MT to have an unrestricted range of movement or network-wide terminal mobility. Under this non-limiting assumption, our solution is to employ the dynamic connection tree described above to provide unrestricted and continuous handoff support. The connection tree is dynamic because after each handoff, the connection tree is reconfigured to account for the new set of immediate neighbouring cellular regions into which the user terminal can potentially enter.

Since the connection tree is implemented by the proposed MVC, we now discuss how the MVC is reconfigured after each handoff to support successive handoffs. MVC reconfiguration after handoff is illustrated in Fig. 6, where an MT has roamed from cellular region 0 to cellular region 6 and is accessing a new BS controlled by a different MSC. Cellular region 0 is an immediate neighbour to cellular regions 1 to 6; while cellular region 6 is an immediate neighbour to cellular regions 0, 1, 5, 7, 8, 9. Before the handoff, the MVC is characterized as follows: TP at node D; fixed common links DE, EF, FG and GH; current ongoing connection associated with cellular region 0 has mobile specific links AB, BC and CD; and the potential handoff connection associated
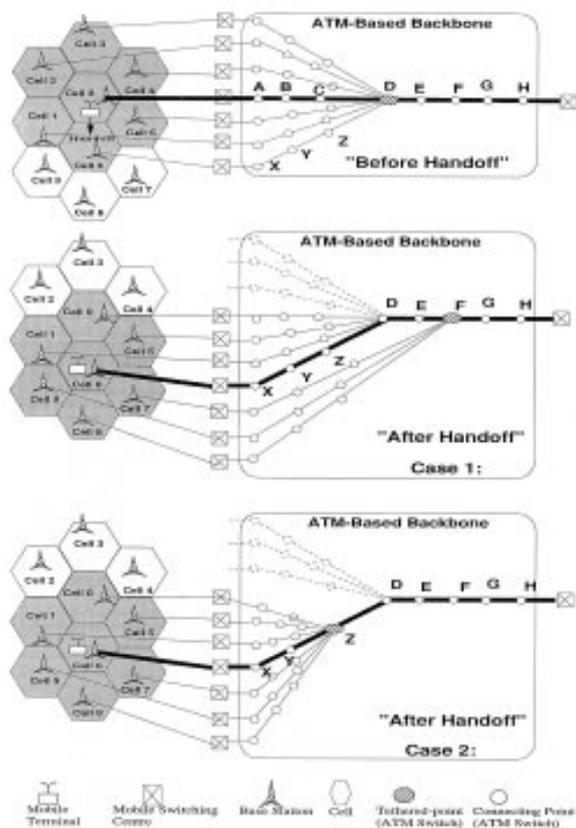
Fig. 6. MVC reconfiguration after handoff.

with cellular region 6 has mobile specific links XY, YZ and ZD.

The MVC reconfiguration after each handoff involves IN signaling control and it is described as follows:

1. The potential mobile endpoints are reconfigured to account for the new set of immediate neighbouring cellular regions into which the MT can potentially enter.

2. One of the nodes along the new connection (i.e., nodes X, Y, Z, D, E, F, G, and H) is selected to be the reconfigured TP. The selection is based on minimizing the number of mobile specific links or maximizing the number of fixed common links of the current and the potential handoff connections. If the reconfigured TP remains at the same node (i.e., node D), no TP reconfiguration is required.

3. Encode the route for the mobile specific links from each potential mobile endpoint (i.e., those associated with cellular regions 1, 0, 5, 7, 8, 9) to the reconfigured TP in the form of source-routing information. The source-routing information would be stored in the MT, the associated BS or the MSC.

4. If TP reconfiguration is required, encode the route

for the mobile specific links from the current mobile endpoint to the reconfigured TP in the form of source-routing information.

5. Remove information storage of the predetermined routes for the mobile link-sets of all obsolete potential handoff connections (i.e., those associated with cellular regions 2, 3 and 4).

6. If TP reconfiguration is required, the reconfigured TP is requested to reserve the same current ongoing GVC identifier to support the multipoint-to-point connection between the new set of multiple endpoints and the reconfigured TP. If the particular GVC identifier is being employed by another MVC at the reconfigured TP, then a new GVC identifier must be reserved.

7. Routing table entries are established in the connecting nodes along the reconfigured route of the new current ongoing connection. The predetermined and the reconfigured routes of the new current ongoing connection include the same connecting nodes and physical links.

   (a) If the ongoing GVC is available at the reconfigured TP, the routes only differ in the point-to-point logical links between the predetermined and the reconfigured TP.

   (b) If the ongoing GVC is not available at the reconfigured TP, the predetermined and the reconfigured routes of the new current ongoing connection differ in the point-to-point logical links between the current mobile endpoint and the reconfigured TP.

8. The current ongoing connection is switched-over to the reconfigured route from the predetermined route. Since the predetermined and the reconfigured routes include the same connecting nodes and physical links, there is no need to reserve the bandwidth again and packet-ordering would still be maintained.

3. MVC associated signaling transport

3.1. Basic operation

With the proposed connection establishment support for the handoff connection, the resource reservation signaling messages must be exchanged quickly since it directly affects handoff delay and hence call continuation. Consequently, fast signaling transport capability is required. The proposed fast resource reservation scheme reserves resources along the predetermined route used by the handoff connection. The route is determined by the IN at call initiation or after the previous handoff. Repeating route determination due to congested or faulty links during handoffs would be costly in terms of
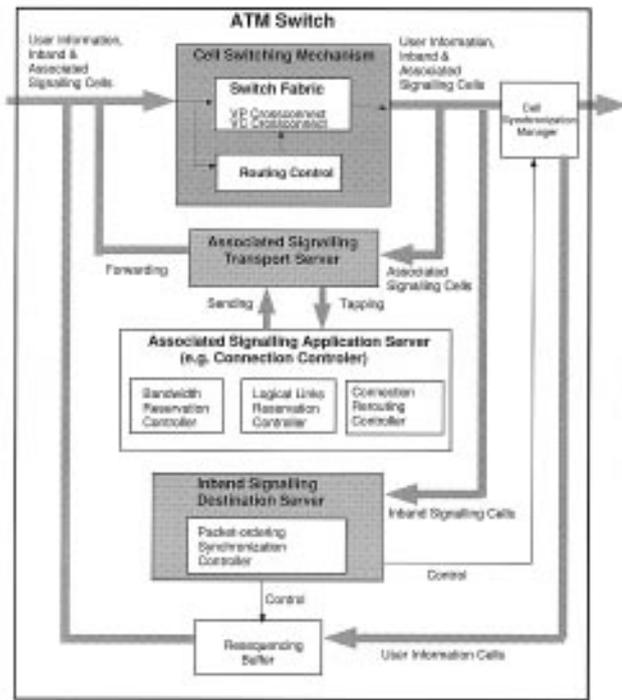
Fig. 7. ATM switch functions.

Fig. 8. Associated signaling transport.

handoff delay as IN processing must be involved. Therefore, a robust signaling transport capability to bypass problematic links would be desirable.

We propose a MVC associated signaling architecture to support fast and robust signaling transport. With the associated signaling mode, the user and signaling networks would share the physical, virtual circuit (VC) and virtual path (VP) transmission facilities from the ATM network. The associated signaling services are supported over logical signaling links established by preassigned permanent VCs and VPs within the ATM network.

The proposed signaling network architecture provides source-routing unicast and multicast services (Fig. 8) and protocols in the signalling ATM adaptation layer (Fig. 1) which is conceptually above the ATM layer (i.e. ATM layer provides network services to the adaptation layer) where associated signaling services are provided. The source-routing services are further adapted to enhance robustness.

The source-routing information identifies the series of preassigned destination VCI/VPI/port for the subsequent signaling transport stages. For k transport stages associated with the source, connecting and destination nodes, the source-routing information is subdivided into k subfields. Subfield i contains the next forwarding VCI/VPI/port at stage i. With the source-routing unicast, a signaling cell is forwarded to the next node according to the corresponding subfield information. With the source-routing multicast, a signaling cell is copied and the copy is passed to the local application protocol entity before it is forwarded to the next node. Assuming a 12-bit port identifier, a 40-bit $(16 + 12 + 12)$ address is
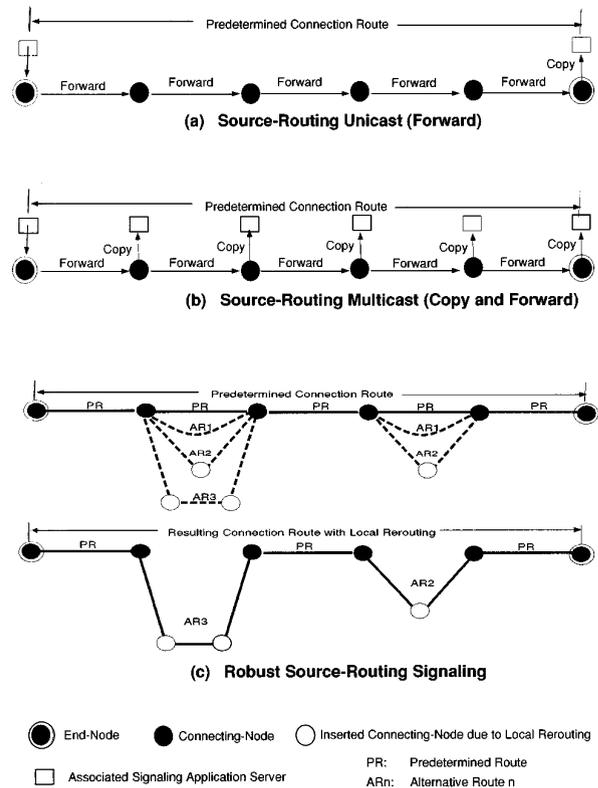
required to identify a node along the source-routed path. Once the source-routing information is known (e.g. obtained from the AIN during call processing), the signaling initiator can start transmitting signaling messages without having to establish a signaling connection explicitly; thus enabling fast signaling transport.

The associated signaling cell (Fig. 9) consists of the ATM cell header with the ATM cell payload containing the signaling adaptation layer header and payload. The signaling adaptation layer header consists of the following fields: (1) protocol type indicating either source-routing unicast or source-routing multicast; (2) number of source-routing information entries; (3) next entry pointer; and (4) source-routing information entries.

As illustrated in Table 1, the associated signaling cell is identified via the cell header by preassigned VCI/VPI and by the 2-bit payload type (PT) indicator being set to some non-zero code (e.g. "01"). The PT code to identify signaling cells (including operation and maintenance traffic) has not been finalized by the ATM standard; however, the general agreement is to identify user cells by setting the PT bits to "00" [2].

There are two approaches for adding the proposed associated signaling transport service functions to an ATM network. One is to integrate the service functions into the cell routing mechanism in the ATM switch. A more flexible alternative that can be implemented by retrofitting existing ATM switches is to append the service functions in an independent unit external to the cell rout-

Table 1
ATM cell header fields encoding and ATM switch functions.

ATM cell header fields	ATM switch functions			
	VP crossconnect (user or inband signaling cells)	VC/VP crossconnect (user or inband signaling cells)	Inband signaling destination endpoint	Associated signaling transport
VPI (12 bits)	Most significant bit of VPI = 0	Most significant bit of VPI = 1	Preassigned VPI. e.g., VPI = All 1's	Preassigned VPI
VCI (16 bits)	VCI assigned per call	VCI assigned per call	VCI assigned per call	Preassigned VCI
PT (2 bits)	PT = 00 (user) or PT = 01 (signaling)	PT = 00 (user) or PT = 01 (signaling)	PT = 01	PT = 01

ing mechanism (as illustrated in Fig. 7). The associated signaling transport server acts as a self-contained unit to process associated signaling cells arriving from one or more outputs of the switch fabric. The processed associated signaling cell returns to one or more inputs of the switch fabric for forwarding over the appropriate outgoing link. While endowed with flexibility, the stand-alone approach does impose extra cell routing delay since an associated signaling cell is routed through each cell routing mechanism twice.

Encoded as source-routing information in the signaling adaptation layer, the predetermined route for fast resource reservation during handoff processing would be stored in the handoff signaling application initiator and there are three possible cases: (1) the MT; (2) each of the potential handoff BSs; (3) each of the potential handoff MSCs. While the first case would dictate a MT with intelligence for signaling, the third case may cause heavy signaling activities of the MSC. Therefore, the second case of configuring the BS to store the self-routing information and to act as the handoff signaling initiator would be a good compromise.

Besides associated signaling transport, inband signaling transport is required to transport packet-ordering synchronization data during connection rerouting (as described in subsection 4.2) to preserve the timing relationship between the user data and the synchronization data. The inband signaling cell (Fig. 9) consists of the ATM cell header with the ATM cell payload containing the inband signaling application header and payload. As illustrated in Table 1, the TP identifies the inband synchronization marker cells via the cell header by preassigned VPI, and identifies the MVC connection and the MT by the incoming GVCI encoded in the VCI field. Once the TP detects cells associated with a GVC that requires resequencing, it will request the cell synchronization manager to direct the user information cells to the resequencing buffer. When synchronization period ends, the TP will direct the resequenced cells to the inputs of the switch fabric to continue their transits to the destinations.

3.2. Robust adaptation

According to the proposed associated signaling transport mechanism, the signaling initiator obtains, e.g., from a route server in the IN, a predetermined route to the signaling responders via a set of connecting nodes that the signaling messages will follow. The source-routing information is contained in the payload of the ATM cell. Each connecting node along the predetermined route relays the message to the next connecting node according to the specified routing information. However, if message relaying fails, a connecting node routes the messages to the next connecting node according to the robust source-routing scheme (Fig. 8) described below. Signaling message relaying for call routing in an associated signaling network architecture can fail due to the following reasons: (1) failure of signaling or user transport resources (links or nodes are shared by both signaling and user transports); (2) congestion in user transport resources (e.g. bandwidth congestion).

The robust source-routing scheme require each connecting node to maintain local routing information to access neighbouring nodes (i.e., immediate neighbour 1 hop away or regional level neighbours several hops away). The local routing information can be either statically configured by the IN or dynamically updated via the IN or the neighbours along the source-routing route. The strategy consists of the following rules: (1) choose the local alternative route to the downstream neighbour along the predetermined route with the least number of connecting links; (2) if two alternative routes are of the same number of connecting links, choose the one with less loading or higher capacity, otherwise, choose arbitrarily; (3) reject any alternative route that contains connecting nodes that also belong to the predetermined routes or alternative routes associated with the preceding node-pairs, or the predetermined routes associated with the subsequent node-pairs. The first rule minimizes the number of additional connecting links and nodes relative to the predetermined route. The second rule optimizes congestion delay. The third rule avoids back-

Inband Signalling Cell
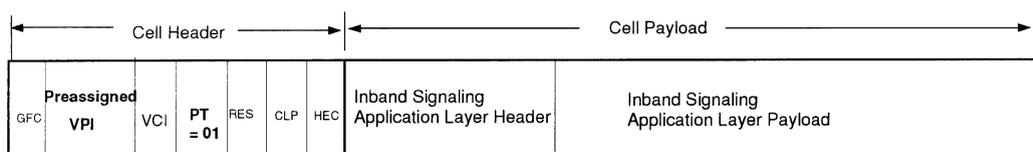
Preassigned VPI to indicate inband signalling destination endpoint

Associated Signalling Cell
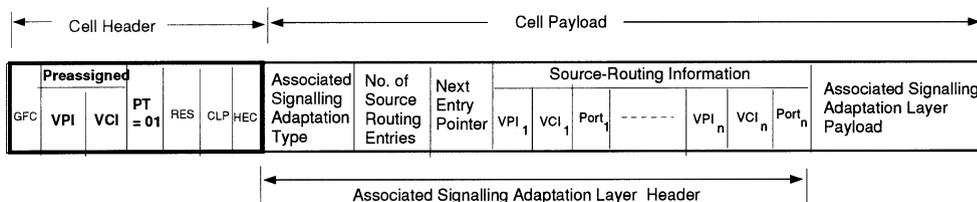
Associated Signalling Adaptation Type:
(1)   Unicast (Forwarding only)
(2)   Multicast (Copy-and-Forward)

Fig. 9. Signaling cell format.

hauling or looping where a connecting link is traversed twice for a connection.

After a connecting node has determined the local alternative route to the neighbouring connecting node that a message will follow, it inserts the sequence of alternative node addresses (VCIs/VPIs/Ports) into the original source-routing information encoded in the subsequent signaling cells. Each connecting node along the alternative route relays and employs the above local alternative routing scheme.

4. Fast seamless handoff processing

To support a reliable mobile connection-oriented bearer service, a fast seamless handoff scheme is required. For a handoff processing scheme to be qualified as seamless, it must minimize call disruptions and call degradations due to transient traffic lagging-gap or packet-disordering. The scheme must also be fast to minimize forced call terminations.

While forced call termination due to lack of network resources to accommodate a handoff is addressed by proper dimensioning of network resources, a handoff processing scheme which expedites call handoffs will minimize forced terminations due to the MT losing radio contact before handoff processing is complete. The proposed handoff scheme expedites handoffs by minimizing delays for handoff connection establishment as discussed above.

Call disruption is caused by the delays in synchronizing handoff connections in different parts of the network. The proposed handoff scheme eliminates call disruption by means of in-channel sequence terminating and sequence beginning markers to ensure no loss of user traffic. However, mismatch of packet or cell transit time between the old and the new connections during the handoff instance could result in transient traffic lagging-gap or packet-disordering. For synchronous traffic, if the inaugural cell of the new connection arrives later than the expected time at the destination, a transient traffic lagging-gap results. For both synchronous and asynchronous traffic, if the inaugural cell of the new connection overtakes cells of the old connection at the destination, a transient packet-disordering traffic results. Resequencing of packet-disordering traffic is achieved by buffering incoming and outgoing traffic from the new connection at the new BS and the TP, respectively.

The handoff processing scheme is based on the following assumptions: (1) the MT supports single outgoing and incoming air channels only; (2) the TP and BS support the resequencing synchronization protocol, and synchronization markers are transported via the user information network with the same priority as the user information; (3) signaling information between the TP and BSs is transported via the proposed associated signaling transport mechanism.

The proposed call handoff processing scheme for duplex MVC with MT-initiated handoff is shown in Fig. 10 and described as follows:
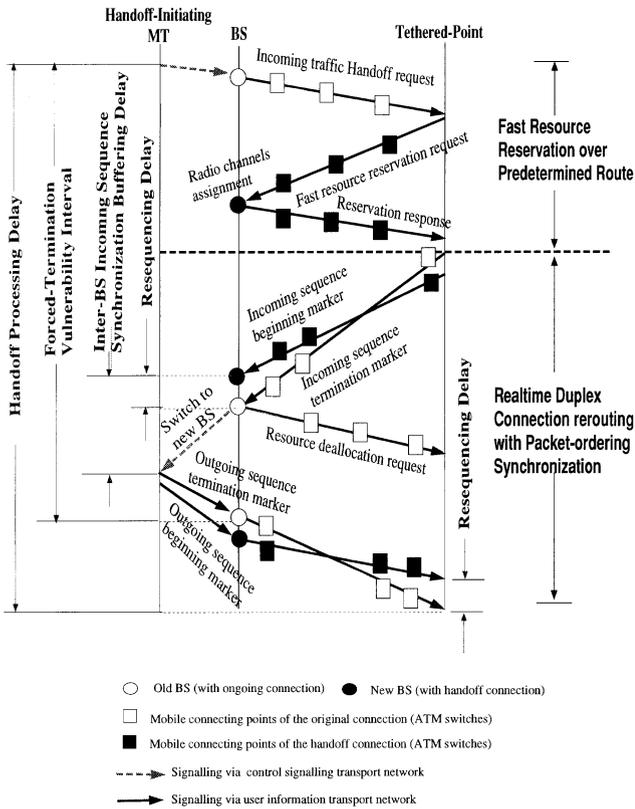
Fig. 10. Fast seamless handoff processing with MVC connection.

Fig. 11. Fast resource reservation without local alternative routing.

1. The MT signals the TP via the old BS to switch MT's incoming traffic to the mobile links of the handoff connection. The TP is equipped with the source-routing information of the mobile links to the new BS during call initiation.

2. The TP signals the new BS to assign a duplex radio channel for the handoff connection.

3. The TP signals the mobile connecting points of the handoff connection (including the new BS) for fast resource allocation as elaborated in subsection 4.1.

4. Packet-ordering synchronization during connections switch-over is carried out by the TP, the old BS, the new BS and the MT, as described in subsection 4.2.

4.1. Fast resource reservation scheme for mobile links

The fast resource reservation scheme for mobile links during handoff processing involves the TP ATM switch, the BS and the connecting ATM switches between them. The scheme is a two-phase commit transaction procedure [18] with reservation controls executed in parallel by the co-resident reservation controllers of the connecting switches. The reservation controller is an associated signaling application server supported by the associated signaling transport mechanism described in section 3.
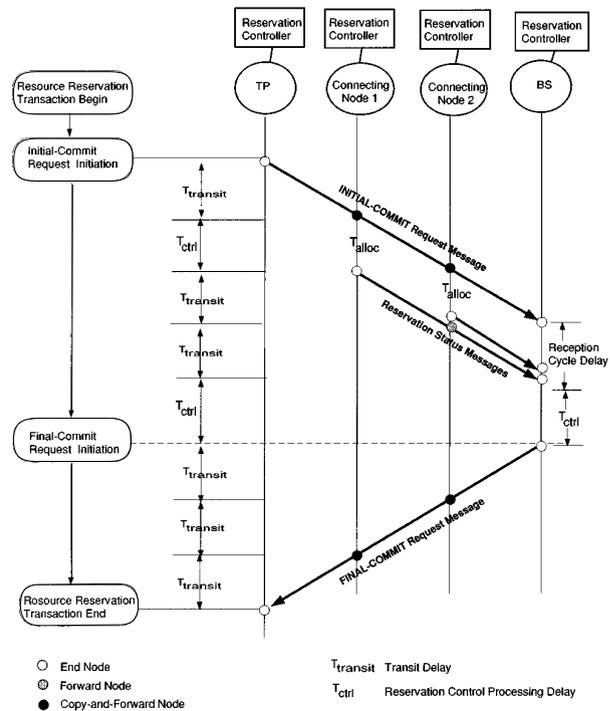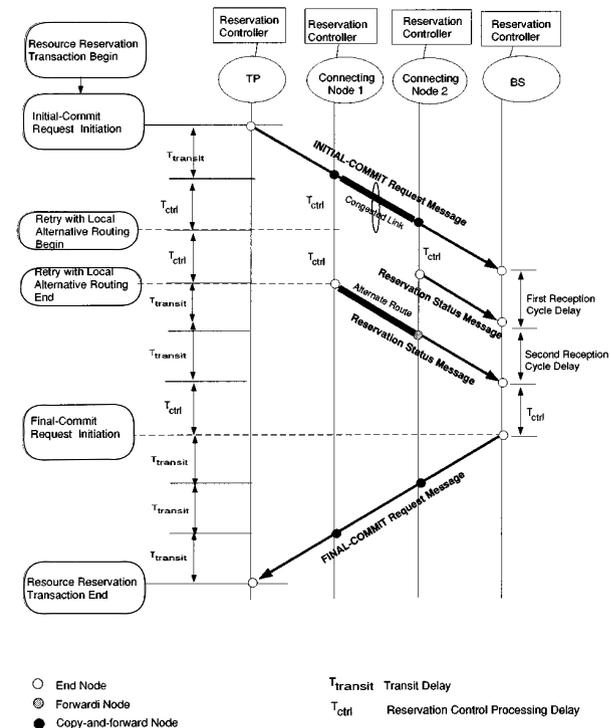

Fig. 12. Fast resource reservation with local alternative routing.

The proposed fast resource reservation scheme is shown in Figs. 11 and 12, and it is described as follows:

1. The reservation controller of the TP signals an initial-commit reservation request message to the new BS via an associated signaling channel along the connecting switches of the mobile links by using source-routing.

2. Upon receiving the initial-commit reservation message, each connecting switch: (1) forwards the signaling message to the next connecting switch along the route designated by the incoming source-routing information; and (2) taps the signaling message for processing by the reservation controller.

3. Upon receiving the signaling message, the reservation controller initiates a statistical or deterministic reservation control algorithm. If the controller accepts the request, appropriate resources are reserved. Upon completion of control processing, each controller sends either a rejection or acceptance message to the reservation controller of the BS by means of source-routing over the associated signaling channels. The source-routing information is encoded in the reservation request message.

4. The reservation controller of the BS signals a final-commit request message to the TP via the associated signaling channels along the connecting switches of the mobile links by using source-routing.

5. Upon receiving the final-commit reservation message, each connecting switch must either commit or release the previously reserved resources.

Fig. 11 illustrates the timing diagram of the fast resource reservation scheme without activating the robust adaptation of the associated signaling transport. Fig. 12 illustrates the timing diagrams of the fast resource reservation scheme when the robust adaptation of the associated signaling transport is used to bypass congested or faulty links.

4.2. Packet-ordering synchronization scheme

The proposed packet-ordering synchronization scheme during connection switch-over is shown in Fig. 10 and described as follows:

1. The TP signals the old BS that the MT's incoming traffic via the original connection is now terminated by issuing an incoming sequence termination marker via the inband signaling transport mechanism proposed in section 3.

2. The TP signals the new BS that the MT's incoming traffic via the handoff connection is now activated by issuing an incoming sequence beginning marker via the inband signaling transport mechanism.

3. Upon receiving the incoming sequence termination marker, the old BS signals the MT to switch to the new BS.

4. Upon receiving the incoming sequence beginning marker, the new BS buffers the MT's incoming traffic.

5. Before switching to the new BS, the MT prepares to switch to the new BS; but first, the MT signals the TP via the old BS that the MT's outgoing traffic via the original connection is now terminated by issuing an outgoing sequence termination marker via the inband signaling transport mechanism.

6. The MT dissociates from the old BS.

7. The MT signals the new BS to activate the assigned radio channels of the duplex handoff connections. The new BS then presents the incoming buffered data to the MT.

8. The MT signals the TP via the new BS that the MT's outgoing traffic via the handoff connection is now activated by issuing an outgoing sequence beginning marker via the inband signaling transport mechanism.

9. The TP synchronizes the MT's outgoing traffic via the old and the new mobile links by employing the outgoing sequence termination and beginning markers, and buffering the outgoing traffic on the new link as necessary.

5. Performance analysis

The MVC facilitates flexible configuration of the connection tree involving only mobile endpoints in neighbouring cells. Given the mobile endpoints and the destination endpoint at the other end, the number of possible connection trees is determined by the TP availability ratio (i.e. proportion of network nodes that can function as TPs) supported by the network. We present analysis to show that it is desirable to maximize the TP availability ratio in order to minimize the handoff processing delay. This is possible if every network node can function as a TP, and our previous discussions show that this is not a particularly onerous requirement since the existing switch architecture can be augmented with add-on signaling components. Under such conditions, the connection tree can be configured using the following algorithm:

1. Determine the optimal route for the current connection between the current mobile endpoint and the destination endpoint. The decision is based on minimizing the number of total links of the current connection.

2. Select one of the nodes along the determined route of the current connection to be the TP. The decision is based on minimizing the number of mobile specific links or maximizing the number of fixed common

links of the current and the potential handoff connections.

3. Determine the route for the mobile specific links from each potential mobile endpoint to the TP.

In general, the handoff processing delay includes the following components: (1) handoff signaling delay; (2) the handoff connection establishment delay; and (3) connection switch-over synchronization delay. Since the MVC predetermines all handoff routes, the resource reservation delay is the only component of the handoff connection establishment delay. In the following subsections, we determine how the mobile links ratio (proportion of mobile links to total links) is affected by the TP availability ratio, and their effects on the resource reservation delay and hence the handoff processing delay, and illustrate the results by numerical examples.

5.1. Effect of TP availability

In subsequent subsections, we will show that the handoff processing delay is proportional to the mobile links ratio. Therefore, the selection of MVC during call setup should aim to minimize the number of mobile links. For a given set of endpoints, the selection is limited by the number of available MVCs which is limited by the TP availability ratio. A simulation program is written to see how the mobile link ratio depends on the TP availability ratio. The program performs a random experiment by generating a random network of 100 nodes based on a network connectivity parameter and the TP availability ratio, and determines the set of mobile links ratios for all possible MVC connection trees that may be configured for this network. Each MVC connection tree is generated based on randomly selected mobile endpoints associated with a proximity parameter and a randomly selected destination endpoint; and the mobile links ratio of each generated connection tree is recorded. For each specified combination of network connectivity and TP availability ratio, the experiment is repeated 100 times and an overall average mobile links ratio is obtained by averaging over all elements in the mobile links ratio sets resulting from all repetitions of the experiment.

Fig. 14 shows the effect of the TP availability ratio on the mobile links ratio for different network connectivities. For a given TP availability ratio, increasing the network connectivity tends to decrease the mobile links ratio. Regardless of network connectivity, the mobile links ratio is inversely proportional to the TP availability ratio. As the TP availability ratio approaches 1, the mobile links ratio approaches a minimum value and becomes less dependent on the TP availability ratio.

5.2. Resource reservation delay

Under the proposed fast resource reservation scheme,

resource reservation applies along the predetermined route of the handoff connection between the BS and the TP. At the BS reservation controller, the reception of the commit status messages sent by the controllers of the mobile connecting switches and the TP is modelled by using the concept of reception cycle (Figs. 11 and 12). The reception cycle starts after the BS controller has received the initial-commit reservation request message. Each cycle duration is bounded by the average reservation control processing delay. During each cycle, none or one or more commit status messages from the reservation controllers of the connecting nodes will arrive at the BS reservation controller. The cycle ends when all the commit status messages have arrived at the BS reservation controller.

Let T_{CTRL} be the average reservation control processing delay experienced by each reservation controller, $T_{transit}$ be the average transit delay between two connecting switches along the associated signaling channel, and M_H be the number of mobile connecting links or the number of mobile connecting and TP switches of the handoff connection. Assume that T_{CTRL} is the same for all controllers, the length of each reception cycle is characterized by T_{CTRL}. Upon the end of the first reception cycle, up to M_H commit status messages may be sent by the reservation controllers of the mobile connecting and TP switches. For those reservation controllers that do not respond, they would engage in resource reservation along alternative routes to their next immediate neighbouring controllers via the proposed robust source-routing scheme described in subsection 3.2. We now derive a recursive relation for $P_r(M_H)$, the probability distribution that r reception cycles are required to service all M_H reservation controllers of the handoff connection; i.e., $P_1(M_H)$ is the probability that the BS controller receives all M_H commit status messages in the first reception cycle.

Let $f(k, m)$ be the probability that k controllers do not respond out of m non-responding controllers from the previous reception cycle. Subsequently, each of the non-responding controllers may require a various number of reception cycles to send the commit status message. Assume that reservation on the predetermined route or each alternative route between two neighbouring connecting switches is a Bernoulli event with failure probability Q and success probability $1 - Q$, then

$$f(k, m) = \begin{bmatrix} m \\ k \end{bmatrix} Q^k (1 - Q)^{m-k}. \qquad (1)$$

Given k non-responding controllers in the first reception cycle, the conditional probability to succeed in the rth reception cycle is $P_{r-1}(k)$. Therefore,

$$P_r(M_H) = \sum_{k=1}^{M_H} f(k, M_H) P_{r-1}(k). \qquad (2)$$

Consequently, the expected number of reception cycles is

$$\bar{r} = \sum_{r=1}^{\infty} r P_r(M_H) . \qquad (3)$$

Employing eq. (1) for induction on eq. (2), then

$$P_r(M_H) = (1 - Q^r)^{M_H} - (1 - Q^{r-1})^{M_H}$$

$$= \sum_{k=0}^{M_H} \begin{bmatrix} M_H \\ k \end{bmatrix} (-Q^r)^k - \sum_{k=0}^{M_H} \begin{bmatrix} M_H \\ k \end{bmatrix} (-Q^{r-1})^k .$$

$$(4)$$

Substituting eq. (4) into eq. (3), then

$$\bar{r} = \sum_{k=1}^{M_H} \begin{bmatrix} M_H \\ k \end{bmatrix} (-1)^{k+1} \left(\frac{1}{1 - Q^k} \right) . \qquad (5)$$

Let $T_{RESERVE}$ be the resource reservation delay, then from Figs. 11 and 12, we can see that

$$T_{RESERVE} = T_{CTRL} + \bar{T}_{CTRL} + 2 M_H T_{transit}$$

$$= (\bar{r} + 1) T_{CTRL} + 2 M_H T_{transit} . \qquad (6)$$

From Fig. 13, we can see that for link congestion and failure probability $Q \leqslant 0.1$ (which is already quite high for normal network operation), $\bar{r}$ is upper-bounded by 3 as M_H increases; i.e.,

$$(T_{RESERVE})_{max} = 4 T_{CTRL} + 2 M_H T_{transit} \quad \text{for } Q \leqslant 0.1 . \qquad (7)$$

It can seen that the delay of the proposed resource reservation scheme has two components, the reservation control processing delay and the signaling transit delay. With regards to the number of mobile links along the handoff connection, the signaling transit delay component is directly proportional to it while the reservation control processing delay is independent of it when the link congestion and failure probability is less than 0.1. Therefore, the proposed reservation scheme minimizes
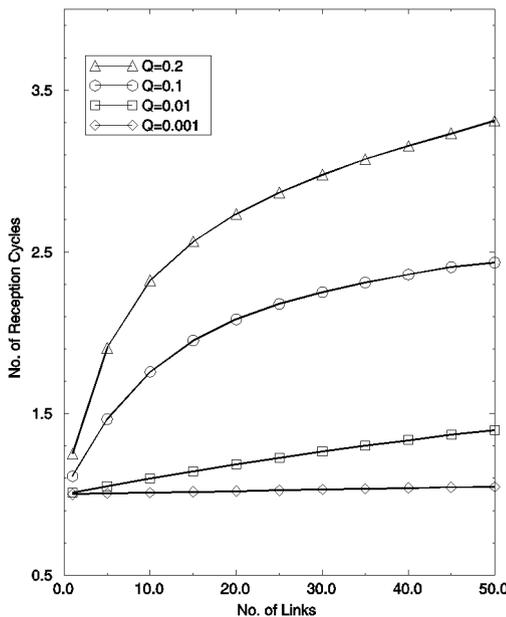

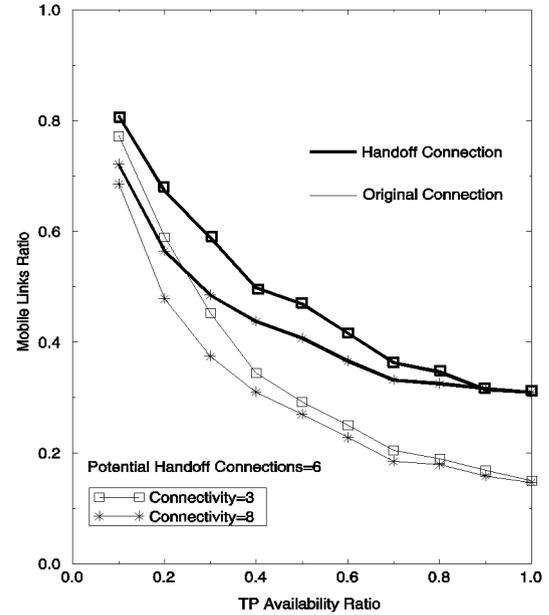Fig. 13. Reception cycles of fast resource reservation scheme.

Fig. 14. Effect of TP availability.

the dependency of the control processing delay on the number of mobile connecting switches of the handoff connection.

5.3. Handoff processing delay

The performance parameters characterizing our seamless call handoff scheme include: (1) the handoff processing delay, (2) the forced-termination vulnerability interval, (3) the incoming sequence synchronization buffering delay, and (4) the resequencing delay. The forced-termination vulnerability interval is the interval during handoff processing that is vulnerable to forced-termination due to the premature dropping of the radio channel. Incoming sequence buffering is employed to eliminate call disruption due to the inability of the handoff-initiating MT to associate with the original and the handoff connection simultaneously. Resequencing is employed to eliminate packet sequence disordering due to the instantaneous mismatch of packet transit time between the original and the handoff connection at the handoff epoch. In general, minimization of the above time intervals or delays is desired.

Let τ_{BT}^S be the signaling transit delay between the BS and the TP, τ_{BM}^S be the signaling transit delay between the BS and the MT, τ_{BT}^U be the user information transit delay of the synchronization marker between the BS and the TP, τ_{BM}^U be the user information transit delay between the BS and the MT, $T_{transit}$ be the average signaling or user information transit delay via a mobile connecting link of the MVC, $T_{RESERVE}$ be the resource reservation or allocation delay, M_0 be the number of mobile connecting links of the original connection, and M_H be the number of mobile connecting links of the handoff connection.

With reference to Fig. 10, let T_{RSQ} be the resequencing delay, then the average forced-termination vulnerability interval is given by

$$T_{VUL} = \tau_{BT}^S + T_{RESERVE} + \tau_{BT}^U + 2\tau_{BM}^S + \tau_{BM}^U + T_{\text{RSQ}}$$
$$= 2M_O T_{transit} + T_{RESERVE} + 2\tau_{BM}^S + \tau_{BM}^U + T_{RSQ}. \tag{8}$$

The average handoff processing time is determined as

$$T_{HO} = \tau_{BT}^S + T_{RESERVE} + 2\tau_{BT}^U + 2\tau_{BM}^S + \tau_{BM}^U + 2T_{RSQ}$$
$$= (2M_O + M_H)T_{transit} + T_{RESERVE} + 2\tau_{BM}^S$$
$$+ \tau_{BM}^U + 2T_{RSQ}. \tag{9}$$

The inter-BS incoming sequence synchronization buffer delay on the handoff-initiating side is given by

$$T_{BUF} = \tau_{BM}^S + T_{RSQ}. \tag{10}$$

Switching from the original connection to the handoff connection, the resulting difference or mismatch in cell transmit times would generate either a lagging gap (positive difference) or packet-disordering (negative difference). The seamless handoff scheme eliminates packet-disordering by introducing resequencing delay T_{RSQ}. The cell transit time consists of the deterministic delays (e.g. transmission, propagation and switching delays) and the stochastic queueing delays. The queueing mechanisms at the output ports of the connecting nodes form a network of queues. Consequently, the difference in cell transit times has a deterministic delay difference component and a stochastic queueing delay difference component. However, the transit delay for high priority control packets in high-speed networks is dominated by the propagation delay T_P. Under this situation, T_{RSQ} becomes a deterministic quantity and

$$T_{RSQ} = (M_O - M_H)T_P. \tag{11}$$

In subsection 5.2, we have shown that

$$(T_{RESERVE})_{max} = 4T_{CTRL} + 2M_H T_{transit} \quad \text{for } Q \leqslant 0.1, \tag{12}$$

where Q is the per-link reservation success probability. Substituting eqs. (11) and (12) into eqs. (8)–(10), and replacing $T_{transit}$ by T_P, then

$$T_{VUL} \leqslant 3M_O T_P + M_H T_P + 4T_{CTRL} + 2\tau_{BM}^S + \tau_{BM}^U, \tag{13}$$

$$T_{HO} \leqslant 3M_O T_P + 2M_H T_P + 4T_{CTRL} + 2\tau_{BM}^S + \tau_{BM}^U, \tag{14}$$

$$T_{BUF} \leqslant (M_O - M_H)T_P + \tau_{BM}^S. \tag{15}$$

Therefore, T_{BUF} can be minimized by selecting M_O and M_H as close as possible. On the other hand, T_{VUL} and T_{HO} can be minimized by decreasing the number of mobile connecting links; or by selecting the tether-point to be as close as possible to the mobile endpoints. In the extreme case where the tether-point of the MVC connec-tion is at the destination endpoint (i.e., maximum mobile connecting links no common connecting links), upper bounds of the delays result.

5.4. Numerical example

Due to signal reception fading in the mobile cellular environment, the handoff process must be initiated at a margin well before the received signal between the MT and the current BS falls below the receiver threshold level. Increasing the initiation margin would decrease the probability of forced call terminations but it would also increase the probability of unnecessary handoffs. To avoid forced call termination due to handoff failure, the handoff processing delay must not exceed the initiation margin.

For a given established initiation margin, the intelligent decision algorithm for handoff initiation tries to minimize the number of unnecessary handoffs by averaging received signal values and employing hysteresis. The intelligent decision process would incur detection delay which should be maximized to minimize the probability of unnecessary handoffs. However, increasing the detection delay would also decrease the allowable time for handoff processing. Thus, it is important to minimize the actual handoff processing delay to allow optimal minimization of unnecessary handoffs while maintaining the required probability of forced call terminations.

Handoff across an ATM/B-ISDN personal communications network would incur handoff processing delays across the fixed ATM network and the mobile subnet. In this paper, we have shown how the handoff processing delay across the fixed ATM network could be minimized by the proposed connection architecture and protocols; thus minimizing the total handoff processing delay. Let T_{margin} be the initiation margin, $(T_{handoff})_{allowable}$ be the allowable handoff processing delay across an ATM/B-ISDN personal communications network, $(T_{ATM})_{max}$ and $(T_{subnet})_{max}$ be the actual maximum or worst-case handoff processing delay across the fixed ATM network and the mobile subnet respectively, $T_{detection}$ be the handoff detection delay. Since $(T_{handoff})_{allowable} = (T_{ATM})_{max} + (T_{subnet})_{max}$ to avoid forced call terminations, then

$$T_{margin} = (T_{ATM})_{max} + (T_{subnet})_{max} + T_{detection}. \tag{16}$$

With our proposed handoff scheme across the fixed ATM network, we can see from eq. (14) that

$$(T_{ATM})_{max} = 3M_O T_P + 2M_H T_P + 4T_{CTRL}, \tag{17}$$

where M_O and M_H are the numbers of mobile connecting links of the original connection and the handoff connection respectively, T_P is the per-link propagation delay, and T_{CTRL} is the average reservation control processing delay experienced by each reservation controller. For simplification, assume $M_H = M_O$, then

$$(T_{ATM})_{max} = 5(\text{propagation delay between mobile}$$

$$\text{endpoint and TP}) + 4T_{CTRL}. \qquad (18)$$

In [17], the control processing delay for their proposed bandwidth reservation control algorithm (run on a VAX 8600) is found to be 3.02 ms for the connecting controller and 3.18 ms for the destination controller. In [18], the propagation delay for a 10 km metropolitan path is assumed to be on the order of 0.05 ms; for a 5000 km coast-to-coast path around 25 ms. Therefore, $(T_{ATM})_{max} = 12.49$ ms for a metropolitan path and $(T_{ATM})_{max} = 137.24$ ms for a coast-to-coast path.

To give some examples of initiation margins established under standard mobile cellular conditions, we will reference to page 101 of [19]. It is assumed that -100 dB m is a threshold level at the cell boundary at which a handoff must be taken, and the handoff level is set at -95 dB m. The MT is assumed to be roaming at a constant velocity of 50 miles/hour along a straight path between two transmitting antennas of neighbouring cells. Then based on the propagation path loss figure in page 102 of [19], the T_{margin} should be set at 180 sec for a suburban area with a mobile path loss of 38.4 dB/dec, and at 36 sec for New York City with a mobile path loss of 48 dB/dec. Thus, we can see that with the proposed handoff scheme over an ATM/B-ISDN personal communications network, $(T_{ATM})_{max}$ consumes less than 0.076% and 0.38% of the above established T_{margin} of 180 sec and 36 sec respectively for a coast-to-coast path.

Assuming $(T_{subnet})_{max}$ is of the same order of magnitude as $(T_{ATM})_{max}$, then we can see that $T_{detection}$ can be optimized almost up to the limit of the established T_{margin}. According to [20] with a hysteresis margin of 10 dB, when $T_{detection}$ is allowed to operate at 36 sec, the probability of unnecessary handoff is around 10^{-4} in macrocells and around 10^{-9} for microcells. When $T_{detection}$ is allowed to operate at 180 sec, the probability of unnecessary handoff is negligible for both macrocells and microcells. Therefore, for a given intelligent decision algorithm and a given established initiation margin, optimizing the handoff delay across networks would minimize the probability of unnecessary handoff while maintaining the probability of forced call terminations.

6. Conclusions

We have proposed extensions to the user and signaling transport network architectures and the signaling protocols of ATM/B-ISDN to support broadband wireless PCS by enabling fast and seamless handoffs between radio cells connected to any nodes of the backbone network while minimizing network resource utilization. An integrated framework for implementing connection-oriented mobile bearer services over ATM/B-ISDN, based upon a new MVC connection architecture in the ATM layer of the user transport network, has been presented. The MVC connection has a common-threaded multipoint-to-point topology with mobile links converging at the TP. Since a TP is only required to support standard VC crossconnection and standard rerouting update, the MVC connection architecture allows any standard network node to operate as a TP when additional signaling capabilities are incorporated. Consequently, the multipoint-to-point configuration can be optimized by choosing the TP as close as possible to the mobile access points, thus reducing the complexity of the mobile links. The MVC connection architecture also allows reconfiguration of the multipoint-to-point topology after call handoff to ensure support of network-wide handoff management. The link switching mechanism and the process of identifying the mobile user are simplified by employing the proposed group virtual channel (GVC) supported by a group of virtual paths. We have proposed a fast, seamless handoff signaling scheme using the MVC connection, based on identifying the routes of all possible handoff connections in the MVC before a handoff, and using the proposed fast resource reservation mechanism to activate the handoff connection during handoff processing. The MVC connection and the associated fast handoff signaling scheme minimizes the number of network links involved during handoff processing. Analysis has been presented to show that the above attribute minimizes the reservation and hence the handoff delay. Furthermore, to facilitate robust realtime transport of handoff signaling messages, we have proposed the MVC associated signaling architecture employing source-routing with robustness adaptation to bypass network congestion and link failures. We have also proposed a packet synchronization scheme to preserve packet ordering for seamless handoffs.

References

[1] M. Kawarasaki and B. Jabbari, B-ISDN architecture and protocol, IEEE J. Select. Areas Commun. (Dec. 1991) 1405–1415.

[2] H. Breuer, ATM-layer OAM: Prinicples and open issues, IEEE Commun. Mag. 29 (1991) 75–78.

[3] J. Anderson and M.D. Nguyen, ATM-layer OAM implementation issues, IEEE Commun. Mag. 29 (1991) 79–81.

[4] D.L. Schilling et al., Broadband CDMA for personal communications systems, IEEE Commun. Mag. 29 (1991) 86–93.

[5] R.L. Pickholtz, D.L. Schilling and L.B. Milstein, Spread spectrum for mobile communications, IEEE Trans. V.T. 40 (May 1991).

[6] J.N.H. Grond, H.P.J. Hecker and C. Shepherd, Broadband on the move, *IEEE Int. Conf. on Integ. Broadband Services and Networks* (Oct. 1990) pp. 206–210.

[7] B. Materna, B. Vaughan and C. Britney, Evolution from LAN and MAN access networks towards the integrated network, *IEEE Global Commun. Conf. Globecom'89* (Nov. 1989) pp. 1455–1461.

[8] A.D. Malyan, L.J. Ng, V.C.M. Leung and R.W. Donaldson, Network architecture and signalling for wireless personal communications, IEEE J. Select. Areas Commun. 11 (1993) 830–841.

[9] V.C.M. Leung et al., Call control and traffic transport for connection-oriented high speed wireless personal communications over metropolitan area networks, IEEE J. Select. Areas Commun. 12 (1994) 1376–1388.

[10] S. Suzuki and K. Funakawa, Signalling protocol architecture for digital mobile system, IEEE Veh. Tech. 89 (1989) 729–734.

[11] B. Jabbari, Intelligent network concepts in mobile communications, IEEE Commun. (Feb. 1992) 64–69.

[12] J. Homa and S. Harris, Intelligent network requirements for personal communications services, IEEE Commun. (Feb. 1992) 70–76.

[13] R.K. Berman and J.H. Brewster, Perspectives on the AIN Architecture, IEEE Commun. 31 (Feb. 1992) 27–33.

[14] P.J. Kuhn, C.D. Pack and R.A. Skoog, Common channel signalling networks: Past, present, future, IEEE J. Select. Areas Commun. 12 (1994) 383–394.

[15] T. Porta and M. Veeraraghavan, Description of a functional signaling architecture for broadband networks, *IEEE Globecom* (Dec. 1993) pp. 1012–1016.

[16] A.S. Acampora and M. Naghshineh, An architecture and methodology for mobile-executed handoff in cellular ATM network, IEEE J. Select. Areas Commun. 12 (1994) 1365–1375.

[17] D. Ferrari and D. Verma, A scheme for real-time channel establishment in wide-area networks, IEEE J. Select. Areas Commun. 8 (1990) 368–379.

[18] I. Cidon, I. Gopal and A. Segall, Connection establishment in high-speed networks, IEEE/ACM Trans. Networking 1 (1993) 469–481.

[19] W.C.Y. Lee, *Mobile Cellular Telecommunications Systems* (McGraw-Hill, 1990) Ch. 4.

[20] A. Murase, I. Symington and E. Green, Handoff criterion for macro and microcellular systems, IEEE Veh. Tech. 91 (1991) 524–530.

Oliver T.W. Yu received the B.A.Sc. (1981) and M.A.Sc. (1991) degrees in electrical engineering from the University of British Columbia, where he is currently pursuing a Ph.D. degree. He had worked in the telecommunication industry for over ten years, employed, respectively, by Microtel Pacific Research Ltd., Bell Northern Research Ltd. and Hughes Aircraft of Canada Ltd. He had also been employed as a software engineering manager in Soltech Industries Ltd. His past research and development works included communication protocol design and verification testing, broadband digital cross connect systems, interworking of heterogeneous networks, intelligent network architectures and aviation communication systems. He authors various technical papers and holds a U.S. patent (no. 4,672,370) relating to the Videotex display system. His current research interests are in the areas of broadband ISDN and ATM networks, with applications in network management, mobile, personal and multimedia communications.

Victor C.M. Leung received the B.A.Sc. (Hons.) degree in electrical engineering from the University of British Columbia (U.B.C.) in 1977, and was awarded the APEBC Gold Medal as the head of the graduating class in the Faculty of Applied Science. He attended graduate school at U.B.C. on a Natural Sciences and Engineering Research Council Postgraduate Scholarship and obtained the Ph.D. degree in electrical engineering in 1981. From 1981 to 1987, he was a Senior Member of Technical Staff at Microtel Pacific Research Ltd. (now known as MPR Teltech Ltd.), specializing in the planning, design and analysis of satellite communication systems. He also held a part-time position as Visiting Assistant Professor at Simon Fraser University in 1986 and 1987. In 1988, he was a Lecturer in the Department of Electronics at the Chinese University of Hong Kong. He joined the Department of Electrical Engineering at U.B.C. in 1989, where he is an Associate Professor and a member of the Centre for Integrated Computer Systems Research. He is also a project leader in the Canadian Institute for Telecommunications Research, a Network of Centres of Excellence funded by the Canadian Government. His research interests are in the areas of architectural and protocol design and performance analysis for computer and telecommunication networks, with applications in satellite, mobile, personal communications and high speed networks.