Abstract

Handoff is the procedure by which a user's radio link is transferred between radio ports in the network without an interruption of the user connection. In this article, we discuss how a wireless asynchronous transfer mode (ATM) network may reroute a user connection during a handoff. The authors propose a novel procedure called "Nearest Common Node Rerouting (NCNR)." NCNR is designed to perform the rerouting of user connections due to a handoff event in a fast and efficient manner. The authors conclude by comparing NCNR to other rerouting schemes discussed in the literature.

Rerouting for Handoff in a Wireless ATM Network

BORA A. AKYOL AND DONALD C. COX

he rise of wireless communications paired with the rapid developments in asynchronous transfer mode (ATM) networking technology signals the start of a new era in telecommunications. In this era the users will not only need higher bandwidth; they will also demand mobility. A typical future wireless user may be carrying a handheld computer with audio and video conferencing capabilities and may demand high-speed data communication on the order of 10 Mb/s or higher [1]. A wireless ATM network, which is designed to provide high-speed isochronous and asynchronous communications for wireless users, is a good match for these demands [1–3].

In a network where users are mobile, handoff becomes an important function of the network. Handoff is implemented by the network to give the users freedom of motion beyond a limited wireless coverage area while they are communicating. This may be contrasted to a cordless phone where users are mobile only within a limited coverage area. The *handoff* is the procedure by which a user's radio link is transferred from one radio port to another through the network without an interruption of the user connection [4]. In this article we first summarize the handoff procedure in the wireless ATM network; we then propose a novel procedure (Nearest Common Node Rerouting or NCNR) for rerouting connections in the wireless ATM network to support a handoff event. We will conclude by comparing the proposed handoff scheme to the existing schemes in the literature alnd discussing the benefits of NCNR [3, 5–7].

Handoff Procedure

The handoff procedure is performed to ensure the integrity of a radio connection and to minimize interference to the users in the coverage area of neighboring cells [4, 8]. The wireless ATM network consists of radio ports, user terminals, and network interface equipment [2, 9]. A user terminal might have a few simultaneous connections in the wireless ATM network. When a handoff occurs these connections may need to be rerouted. In this article, we assume that a group of radio ports is connected to the same wireless ATM network interface equipment. This collection of ports is called a *zone*

[2]. The zone architecture is illustrated in Fig. 1. The zone is managed by the *zone manager* process. There are two levels in a handoff event: *network* and *radio*. The radio-level handoff is the actual transfer of the radio link between two ports; the network-level handoff supports the radio-level handoff by performing rerouting and buffering. The radio-level handoff determines some of the procedures used in network-level handoff, as shall be seen later. We also assume that the zone managers have some knowledge about the neighboring zones, and the network addresses of neighboring zones are stored in a local lookup table which is updated periodically by means of an update protocol. We refer to the user communication device as the *user terminal* and the termination point for the user connection as the *endpoint*.

The wireless ATM network architecture used in this article is discussed in detail in [2, 9]. Specifically, in [9] the authors have presented the methods by which a handoff transaction may be implemented in a wireless ATM network using either the current ATM signaling protocols [10] or a migratory wireless ATM signaling protocol; therefore, we limit our scope in this article to the rerouting of wireless ATM connections, which is a crucial part of the handoff procedure.

We also assume that these zones are interconnected by wireless ATM network switching nodes. Based on the "zone" concept, a few different situations for rerouting may be investigated:

A Handoff within a Zone (Intrazone Handoff)

In the intrazone handoff the user is moving within the zone. The only rerouting performed in this case is in the wireless ATM network interface equipment within the zone. Specifically, the zone manager is responsible for the correct update of ATM virtual circuit translation tables within the zone [2, 9]; however, this type of rerouting does not require wide-area ATM network switching; hence, it will not be discussed any further.

A Handoff between Two Zones (Interzone Handoff)

The interzone handoff occurs when the radio ports involved in the handoff belong to different zones. In this case the rerouting involves the wireless ATM network. An interzone handoff might require rerouting at one or more wireless ATM switches, depending on the location of handoff and topology of the

network. This type of rerouting will be discussed in the next section.

Rerouting for Interzone Handoff

The rerouting for an interzone handoff involves one or more wireless ATM switches. In this section we propose a novel interzone handoff procedure, Nearest Common Node Rerouting (NCNR). NCNR attempts to perform the rerouting for a handoff at the closest ATM network node that is common to both zones involved in the handoff transaction. The term "common" is used to denote a network node that is hierarchically above both of the zones in question or a parent of both zones in the network topology tree (Fig. 2).

NCNR minimizes the resources required for rerouting and conserves network bandwidth by eliminating unnecessary connections (discussed later). As mentioned previously, the users of the wireless ATM network may subscribe to services ranging from time-sensitive traffic types (audio, video) to throughput-dependent traffic types (data, file transfers, World Wide Web accesses). We note that the traffic type of the connection involved in the handoff is known by the zone managers. The two kinds of traffic types impose different constraints on the network and the handoff process. For example, time-sensitive voice traffic will not be easy to buffer due to a constant cell generation rate and strict time delay constraints; however, it can tolerate occasional loss of cells. On the other hand, data traffic will not tolerate cell loss, but may tolerate delays on the order of a few hundred milliseconds. The NCNR procedure for time-sensitive and throughput-dependent traffic will hence be different. We also note that, occasionally, a handoff will be attempted without any warning due to severe fading in the radio environment. In such a case, the upper-layer protocols will be responsible for recovery of user connection and lost cells. This case is analogous to cell loss in the fixed ATM network due to severe congestion and should be treated in a similar manner [10].
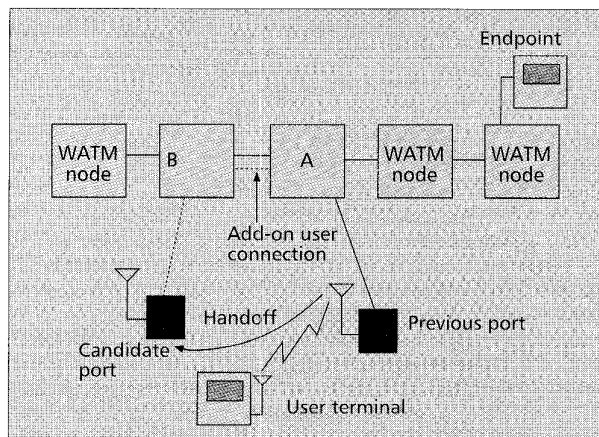
Note also that due to the nature of the fixed network, the
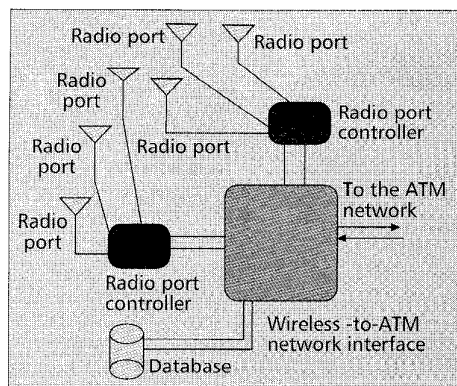
■ Figure 1. *A zone in the wireless ATM network.*
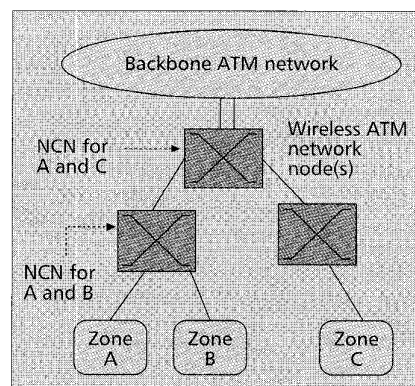
**■ Figure 2.** *Depiction of nearest common node.*

■ Figure 3. *Case 2A in a flat network.*

transmission delay and latency of the links from the nearest common node (NCN) to the zones involved in the handoff are assumed to be negligible compared to the radio transmission medium. In this section we first explain the NCNR procedure for time (delay)-sensitive traffic. We will then conclude by explaining the NCNR procedure for throughput-dependent traffic based on the NCNR for time-sensitive traffic.

NCNR for Time-Sensitive Traffic (NCNR-TS)

The NCNR for time-sensitive traffic is performed as follows:
1. A handoff session between zones A and B is started.[1] Let B be the candidate zone for the handoff and A be the present zone.
2. The zone manager of A first checks to see if a direct physical link (not involving any other network nodes) between A and B exists. There are two possible cases if this condition is satisfied:
• If A is a parent[2] of B, then A notifies B and the new connection is established without any further network involvement. This case is illustrated in Figs. 3 and 4. After the
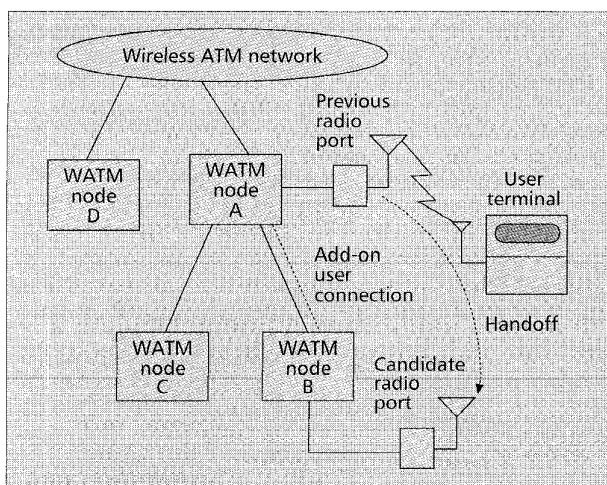
■ Figure 4. *Case 2A in a hierarchical network.*

---

[1] There might be various radio system reasons why the handoff is necessary, but we are primarily concerned with the fact that there is an interzone handoff event occurring. See [9] for details on the handoff procedure.

[2] The parent is determined by means of either the network topology in a hierarchical network or closeness in terms of hops to the endpoint. The endpoint is defined as the terminating point for the user connection in the network.

connection is established, A acts an anchor[3] for the connection. Until the stability of the handoff[4] is established, both A and B act as network connection points for the user connection. This process is explained in detail in step 6 of this procedure. Once the radio-level handoff is completed, A acts only as a wireless ATM switch in the connection path.

- If B is a parent of A, then A sends a message to B relaying the handoff request. B then acts as an anchor for the handoff procedure. Until the stability of the handoff is established, both A and B may be used for information transfer from/to the terminal to/from the network (see step 6 of this procedure). Once the handoff is stable, B deletes the user connection from itself to A. The rerouting is thus completed. This case is illustrated in Figs. 5 and 6.

3. If A and B are not connected by a direct physical link, then the zone manager of A (ZMA) contacts the endpoint for the user connection by sending a *handoff start* message.[5] The handoff start[6] message contains the ATM addresses of zones A and B and the endpoint for the user connection.

4. The handoff start message traverses the network from A to the endpoint for the user connection. Upon receiving this message, the network switching nodes on this path check to see whether all three ATM addresses are routed on different egress ports of the switch (Fig. 2). When such a node is found it is designated as the NCN. The NCN sets the NCN bit in the handoff start message. The rest of the switches on this path do not perform the egress port test.[7]

5. The NCN then forwards a *reroute* message to all of the switches located between B and itself. The nodes that receive the reroute message first check for resource availability; if the resources required by the connection are available, the necessary connections are established and circuit translation tables are set up. If the resources are not available, the handoff attempt fails and the involved parties are notified.

6. When the reroute message is received by *B*, a *reroute acknowledgment* message is sent from B to A. This message completes the rerouting process. The radio-level handoff is attempted at this point by employing the procedures discussed in [2, 9]. As the radio-level handoff is started, the NCN starts to forward the user information to both A and B in a point-to-multipoint manner. This multiparty connection is necessary until the radio-level handoff is stabilized. The radio-level handoff may extend beyond one radio burst in most radio systems as the user terminal tries to select the
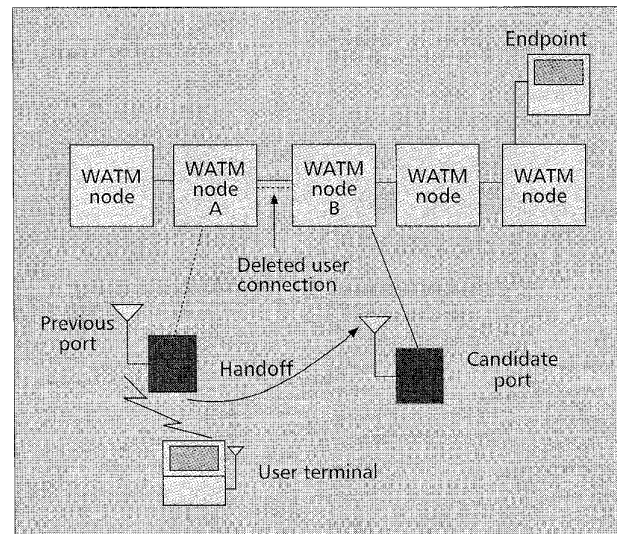
Figure 5. *Case 2B in a flat network.*

optimal link. Especially when a user is in a fading environment, a small motion of the terminal may cause the radio link to switch back and forth between the two radio ports involved in the handoff; hence, a point-to-multipoint link from the NCN to both A and B ensures the timely delivery of time-sensitive information. We note here that the user information may be discarded at the zone which is not in contact with the user terminal and transmitted from the zone which is in contact with the terminal. This zone is the zone that is currently receiving the uplink transmission from the portable. If a zone has not received an uplink transmission from the portable in a given radio transmission frame, it must assume it is not active for the next downlink radio transmission period (see [4, 8, 9] for details on the handoff procedure). In the uplink direction, the information may be transmitted through either zone involved in the handoff[8] and correctly routed to the endpoint by the NCN. Occasionally, the portable may receive duplicate information; since we are discussing time-sensitive traffic, the duplicate information may be determined by the time sequence information and discarded accordingly.[9]

7. If the radio-level handoff is successful,[10] the connection between A and the NCN is cleared by A by sending a *clear connection* message to the NCN. After the handoff is completed, any buffered time-sensitive data that has not expired will be transmitted to the current zone associated with the portable; expired data are discarded at the zones. For time-sensitive traffic, recovery of lost data may only be possible by interpolation of information, which is beyond the scope of this article.

---

[3] The word anchor is commonly used to refer to a network point that the user connection is forwarded through to the candidate zone until the handoff is completed.

[4] The stability of the handoff is used in reference to the radio link transfer which may take longer than one burst during a handoff event. During this period, the user terminal may use both network points for information transfer.

[5] If there are more than one connection for one user, then this procedure is applied to all of the connections.

[6] The signaling messages will be denoted by bold lettering in the text.

[7] Another possible implementation will be to not forward this message after the NCN is found. Both are equally viable options. By forwarding the message to the endpoint we allow the user applications to adjust to the handoff process.

[8] If a soft handoff scheme such as in code division multiple access (CDMA) is being employed, where the uplink information is being received simultaneously by two radio ports, then the NCN may be responsible for combining and sequencing the uplink information.

[9] The ATM transmission and sequencing protocols for transmission of time-sensitive traffic are being discussed in the ATM Forum as of now; hence, we will not go into specifics of transmission of time-sensitive traffic.

[10] The radio-level handoff is successful only when the stability of the new radio link is established. The determination of the stability of the new link is beyond the scope of this article.

NCNR for Throughput-Dependent Traffic (NCNR-TD)

NCNR for throughput-dependent traffic is very similar to the procedure employed for time-sensitive traffic. Throughput-dependent traffic is not sensitive to small (on the order of few hundred milliseconds) delays; however, the loss of information is not tolerated by this traffic type. A typical example of this traffic type is file transfers. We can take advantage of the delay-tolerant nature of this traffic in the rerouting process. The NCNR for throughput-dependent traffic differs from the procedure for NCNR-TS as follows:

1. As the radio-level handoff is started, the downlink user information is buffered at both A and B. No user information is transmitted in the downlink direction until the radio-level handoff is completed. Once the radio-level handoff is completed, the information is transmitted in a first-in first-out (FIFO) manner.
2. If A's buffer is non-empty before the handoff is started, then A's buffer is transmitted to the user terminal if possible; otherwise, these data are transmitted to B and go in front of all other cells buffered for transmission. This preserves the cell sequence.
3. In the uplink direction, before the radio level handoff is started the traffic is transmitted through A if possible; otherwise, it is buffered at the terminal. As the radio-level handoff is started, the user terminal starts buffering the user information. Once the handoff is stabilized the buffered information is transmitted.

These differences ensure the integrity of user data as well as the cell sequence. The cell sequence aspects of rerouting are discussed in the next section.

We also note here that multiple connections may be rerouted in the network using the virtual path concept. By assigning a virtual path identifier for connections between a user and multiple endpoints, and performing the rerouting on the virtual path instead of on a virtual circuit basis, an efficient rerouting of multiple connections may be achieved.

Preserving the Cell Sequence in NCNR

*I*n any ATM network, the cell sequence of individual ATM cells must be preserved for correct reassembly of encapsulated user information [10]. Therefore, during a handoff, the preservation of cell sequence is a primary concern of the network. In a handoff transaction, there are two parties involved: the handoff terminal (HT) and the endpoint. For time-sensitive information the cell sequence in NCNR is preserved due to the fact that the cells in the connection stream are duplicated at the NCN and transmitted to both zones involved in the handoff. The transmission delay in the network compared to the time separation of the radio bursts is negligible; hence, the cells arrive at approximately the same time at both zones. The zone that is currently active with the HT transmits the cell, and the other zone discards it. This reduces the number of active paths to one in the downlink direction, thereby preserving the cell sequence. In the uplink direction, data from only one zone are forwarded to the endpoint by the NCN. This preserves the cell sequence. As a side-note, if a soft handoff scheme is being used where the uplink direction data is received by two zones simultaneously, then NCN will be responsible for combining the two streams into one. Techniques for soft handoff are beyond the scope of this article.

For throughput-dependent traffic (World Wide Web, file
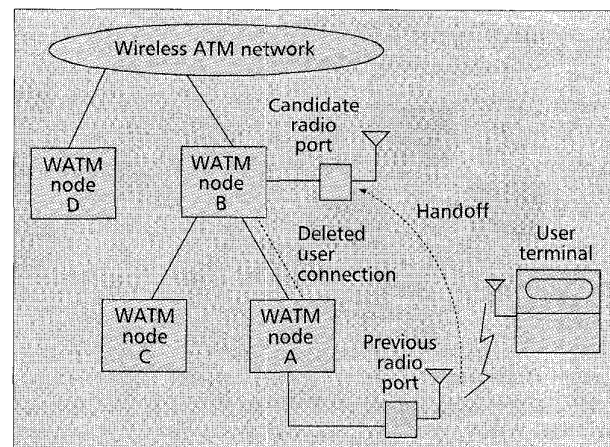
■ Figure 6. *Case 2B in a hierarchical network.*

transfers, telnet sessions), it was previously mentioned that during a handoff the involved zone managers start buffering information from the start of radio-level handoff to the completion (successful or unsuccessful). Also note that until the radio-level handoff is started, the user connection is assumed to be active and all incoming cells are transmitted to the HT through the previous port. If that is not possible, the cells are forwarded to the candidate port through the NCN for later transmission. In the reverse direction, the HT buffers the user information as soon as the radio handoff is initiated. The cells that are in transmission in the network from the HT to the endpoint are delivered normally.

By employing the procedure discussed above for throughput-dependent traffic, the cell sequence during the rerouting of a user connection due to a handoff attempt is preserved. The cell sequence is preserved because of the fact that at any given time during a handoff, there is only one active path between the two parties involved in the connection. Since there is only one active path at any given time, all the cells transmitted through the network take this path and arrive in sequence to their destination.

Occasionally, cell sequence will be broken in the wireless ATM network. Recovery from such sequencing errors is the responsibility of upper-layer protocols, such as the ATM adaptation layer (AAL) and sometimes the user application layer. For data traffic, recovery may be attempted by retransmission; however, for time-sensitive traffic the sequence error may need to be corrected by voice or video interpolation techniques.

A Comparison of Existing Rerouting Schemes for Wireless ATM Networks

*N*CNR enhances the current ATM signaling protocols by adding support for rerouting an active connection. In the current version of ATM signaling protocols there is no provision for rerouting a connection once it is established [10]. By using NCNR, multiple connections between a mobile portable and endpoints may be rerouted seamlessly, without re-establishing connections, while the handoff procedure is being performed. In this section we compare NCNR with some other rerouting for handoff schemes proposed for wireless ATM networks. These will be referred to as Yuan-Biswas [3], BAHAMA [5], virtual connection tree (VCT) [7], source routing mobile circuit (SRMC) [6], and Toh [11]. We will first explain these handoff schemes, then compare them to the

he BAHAMA LAN uses the virtual path indicator of the ATM cell header for routing. This simplifies the rerouting since all that needs to be changed in the cell header is the virtual path indicator provided that the virtual connection indicator is available at the new radio port.

NCNR proposed in this article. Finally, a numerical comparison of these handoff schemes will be presented.

Yuan-Biswas Rerouting Scheme

In this wireless ATM handoff scheme the rerouting of connections for handoff is based on the rerouting of connections at designated handoff switching equipment (HOS). The handoff procedure described in [3] does not specify how the HOS is determined. The example in that text uses the initial wireless ATM switch as the HOS, which in turn is very similar to cell forwarding. It is assumed that the base stations in the wireless ATM network are interconnected by permanent virtual circuits. A handoff between two ports (or base stations) attached to the *same* wireless ATM switch is handled by just updating translations tables in one switch. A handoff between two ports attached to different ATM switches is handled by forwarding the ATM cells destined for the user terminal to the user's new ATM switch. This new add-on connection is established before handoff is completed. The cell sequence is preserved since the first switch acts as a handoff server (switch).

BAHAMA Rerouting Scheme for a Wireless ATM LAN

The BAHAMA handoff scheme is proposed for a wireless ATM local area network (LAN). The BAHAMA architecture consists of a flat network of radio ports (base stations) and user terminals. The radio ports are interconnected with ATM links. The BAHAMA architecture is also based on forwarding of cells after a successful handoff. The initial radio port acts as an anchor (forwarding switch) for the handoff connection and forwards the user cells to the new radio port. After the handoff is completed, the initial radio port migrates the user connection to an optimal route in the network provided that the portable stays within the coverage area of the new radio port for an extended period of time. Since the cells are always routed through the initial switch during handoff, the cell sequence is preserved. The BAHAMA LAN uses the virtual path indicator of the ATM cell header for routing. This simplifies the rerouting since all that needs to be changed in the cell header is the virtual path indicator provided that the virtual connection indicator is available at the new radio port.

VCT: Virtual-Connection-Tree-Based Rerouting Algorithm

In the VCT handoff algorithm the concept of a virtual connection tree is utilized. A virtual connection tree is formed by a root node that is attached to the backbone ATM network. The nodes connected to that root node form the tree structure. When a mobile terminal establishes a wireless ATM connection, a connection tree is formed from the root node to the leaves in the tree effectively producing a point-to-multipoint connection; however, the mobile terminal is utilizing only one leaf node at a time, whereas the rest of the multipoint connection is not utilized. When the mobile moves within the tree, a new leaf node becomes active, and the connection is continued using one of the pre-established circuits. When a mobile steps out of the coverage area of the

VCT, the network establishes a new tree surrounding the mobile.

Source Routing Mobile Circuit Rerouting

The SRMC rerouting approach is an improvement of VCT rerouting. In this approach the rerouting functions are distributed over time. The SRMC approach uses the concept of a tethered point (TP) to serve as the root in the connection tree for handoff. When a connection is first being established, all potential network routes from the TP to the leaves due to possible handoff attempts are recognized by the network, and these connections are pre-established. Unlike the VCT algorithm, no resources are reserved. Once the handoff is initiated, only the resources for the active handoff connection are reserved. After the completion of the handoff, the TP may be migrated and new possible network routes are determined. The differences between SRMC and NCNR will be discussed later. The reader is referred to [6] for details on this scheme.

Comparison of NCNR with the Yuan-Biswas and BAHAMA Handoff Algorithms

In this section we compare NCNR with the Yuan-Biswas and BAHAMA rerouting schemes that may be employed for handoff in the wireless ATM network. The first point of interest is the cell-forwarding aspect.[11] The main reason for the use of cell forwarding in the network is the ease of cell resequencing. When the ATM switch associated with the previous port acts as an anchor for the rerouting, all the cells still go through the previously established path in the network before traversing the new add-on section established because of a handoff. This guarantees the preservation of cell sequence and also means that either the anchor ATM switch [5] (the handoff switch [3]) or the new ATM switch might need to buffer some of the cells before the handoff is successfully completed. For only one user this might not be a significant administrative load; however, the simulations in [9] predict handoff rates of 10 or more handoffs per second in a cellular environment. In such a scenario the buffering of cells due to handoff may become a burden to the network. In NCNR the buffering is mostly performed for throughput-dependent traffic and only when the radio level handoff is being performed. Moreover, for time-sensitive traffic, buffering is not feasible, and the Yuan-Biswas and BAHAMA rerouting schemes do not allow for supporting a user connection through two radio ports while the handoff stabilizes. This may ultimately cause a problem for time-sensitive traffic streams.

Another problem with the cell-forwarding model is that it inherently assumes a flat (or ring) network model where all neighboring ATM switches (or zones) will have direct connections in between. The reason for this assumption is that in a flat network cell forwarding does indeed minimize the number of ATM switches involved in a handoff. However, in a hierarchically organized network, cell-forwarding-based rerouting automatically involves the nodes referred to as the NCNs in this article (Fig. 2). Once the NCN is involved in cell forwarding, it is more advantageous to use the NCNR scheme proposed herein. This is primarily because of two points:

1. When the NCN is involved in the handoff, the network bandwidth is actually minimized when the connection is simply rerouted to the new ATM switch (zone) of the wire-

---

[11] *The YUAN scheme has provisions for dynamic rerouting of connections that are similar to NCNR; however, presently the details on the dynamic rerouting are not available.*

t remains to be determined whether NCNR or SRMC will have the best performance in terms of speed of handoff. However, in terms of efficiency we believe that NCNR will indeed outperform SRMC for the reasons stated in this section.

less network. This prevents the waste of band-width for the portion of the forwarding connection between the NCN and the previous ATM switch (zone).[12]

2. Utilization of NCNR means that the data for all the connections involved in the handoff are buffered at the zones (see the third section). Buffering of data is hence performed at the edge of the network, requiring no intervention from ATM switching equipment.[13] NCNR also preserves cell sequence, as explained in the fourth section.

Another possible problem with cell forwarding is the fact that for a fast-moving user in a system that has small radio coverage areas, there will be a network trail left behind the user consisting of cell forwarding among multiple zones. The NCNR always performs the rerouting at the NCN, thereby minimizing the amount of bandwidth used and the amount of rerouting.

The BAHAMA scheme uses virtual path indicators of the ATM cell header for routing purposes in the wireless LAN. This approach, while suitable for a LAN, will not scale well for a wider-area network. The NCNR is developed with wide area networks (WANs) in mind, and does not have such a problem.

Comparison of NCNR with the VCT-Based Handoff Algorithm

In this section we compare NCNR to the VCT-based handoff proposed in [7]. VCT-based handoff is similar to NCNR. There are three primary differences:

- NCNR takes the concept proposed by VCT one step further by deleting the need for multiple virtual circuits to be reserved at a given time to support a single connection. The zone concept utilized in this article is effectively equivalent to the connection tree concept given in [7]. The zone manager node is the root, the radio ports attached to the zone manager node the leaves. Since the radio ports and the zone manager maintain constant communication there is no need for multiple virtual circuits [2].
- NCNR guarantees the preservation of cell sequence. VCT-based handoff does not implement cell sequence preservation. The mobile is responsible for cell sequencing.
- NCNR recognizes the differing constraints associated with time-sensitive and throughput-dependent traffic types and implements the handoff procedure to accommodate both types of traffic in an efficient manner. VCT-based routing, described in [7], does not address this issue.

Because of these reasons, NCNR may be superior to VCT-based handoff.

Comparison of NCNR with the SRMC Algorithm

The SRMC algorithm improves the VCT handoff algorithm by addressing most of the concerns expressed in the previous sections. The SRMC does not reserve bandwidth in the connection tree until the actual rerouting is performed. This clearly is an advantage over the VCT algorithm. When compared to the NCNR, however, we can still point out these unaddressed issues:

---

[12] _The BAHAMA scheme allows for reorganizing a connection based on cell forwarding once the user becomes stationary within the coverage area of a radio port._

[13] _Note here that the radio controller equipment may coexist with the switching equipment._

- SRMC, by predetermining all possible handoff paths from a root node, attempts to avoid the actual connection establishment during handoff. It does not, however, avoid the resource allocation that must be performed before the actual rerouting is completed. This resource allocation process involves sending a message from the root node (TP) in the connection tree to all the nodes on the active path. There are also messages sent from the leaf node to the root node to notify the TP of an ongoing handoff attempt. In summary, a pair of messages are sent for notification and for resource allocation from one side of the connection tree to the other. Since the root node has to be involved in all handoff attempts, even an attempt between neighboring nodes is managed through the root node. When we compare this with NCNR we see that the worst case in NCNR degenerates to the SRMC algorithm in terms of the number of messages sent between the NCN and nodes involved in the handoff. When the nodes involved in the handoff are neighbors, the NCNR outperforms the SRMC since the messages go up only one level in the hierarchy. Therefore, in terms of the number of messages sent between the nodes in the network hierarchy, NCNR is at worst comparable to, and for neighbors better than, SRMC.
- SRMC uses the centralized intelligent network (IN) concept for predetermining the possible routes that may be involved in a handoff. When a user does not perform a handoff, all the overhead of calculating these possible routes is wasted. It is also not clear whether the most resource-intensive part of a handoff rerouting is resource allocation or finding a possible route for rerouting. NCNR, on the other hand, is a fully distributed algorithm. It performs the work only when it is necessary, avoiding wasted computational overhead.
- SRMC does not address the constraints associated with different traffic types (see the previous discussion).
- SRMC inherently assumes a hierarchical topology. The scheme is not effective in a flat network.

It remains to be determined whether NCNR or SRMC will have the best performance in terms of speed of handoff. However, in terms of efficiency we believe that NCNR will indeed outperform SRMC for the reasons stated in this section.

Other Related Work: Toh's Hybrid Handover Protocol

In addition to the rerouting for handoff algorithms, the "hybrid handover" scheme has been proposed in [11] for a wireless LAN environment. This scheme is similar to the scheme discussed in this article, but it is implemented for a wireless LAN. This scheme has been implemented using the Cambridge Fairisle ATM Switch in a LAN environment, and the results in terms of handoff rerouting rates that may be accommodated in a LAN environment are certainly encouraging. There are two main differences between the hybrid handover scheme and NCNR:

- NCNR is proposed for a wireless WAN and works with current ATM signaling specifications using the overlay signaling proposed in [9].
- NCNR does not require any buffering to be performed in the network switching nodes. The buffering of data during a

Measure	NCNR		Yuan-Biswas	BAHAMA	VCT	SRMC	Toh
	Direct Link	No Direct Link					
N_h	7	9	12	x	2 (est.)	x	10
N_r	2	4	4	2 (est.)	$2N$	Centralized	6 (*)
N_n	2	$4 + D$	$2 + D$	$2 + D$ (est.)	$N + D$	$N + D$	$3 + D$
N_c	1	1	1	1	N	N	1
B	1	1	1	1	N	1	1
W/L	W	W	W	L	W	W	L
Rerouting scheme	DY	DY	CF/DY	CF	T	T	DY
Robustness estimated by the author	5	5	2	3	4	2	3

x: not described in literature
N: number of leaves in the tree structure in virtual-tree-based schemes
D: number dependent on network topology
(*) N_r includes partial teardown; equals 4 without partial teardown

■ **Table 1.** *Comparing rerouting algorithms.*

handoff is performed at the zones for throughput-dependent information only, avoiding a buffer management problem as mentioned in the previous sections. Buffering cannot be used for time-sensitive information, as noted earlier.
• NCNR provides support for different types of traffic, and provides point-to-multipoint support for handling time-sensitive traffic streams in a handoff event. No distinction of different traffic types is made in the hybrid handover procedure.

Numerical Comparison of Rerouting Algorithms

Quantifying network management algorithms presents a challenge. There are two possible approaches. One approach is extensive simulation of a WAN either on a hardware level or by using network simulators. A second approach is to characterize the algorithm and simplify the measurement approach. In this article we choose the second approach for characterizing and quantifying various handoff and associated rerouting algorithms. In general, the following may be of interest for comparing handoff and rerouting algorithms:

Signaling Bandwidth Used in Handoff — The signaling bandwidth used in a handoff event may limit how much bandwidth is available for the user connections; therefore, for the procedures that achieve the same functionality,[14] the amount of bandwidth used for signaling is a valid performance measure. Unfortunately, this aspect of the other handoff protocols is not discussed in the literature and hence cannot form a part of our comparison.

Number of Signaling Messages Exchanged During Handoff (N_h) — The number of messages involved in the handoff procedure increases the complexity of the signaling software and may slow the handoff process.; therefore, the smaller the number of messages, the easier the handoff procedure is to implement, and the faster it should run.

---

[14] *In order to be able to compare different handoff procedures, we need to define a common functional description that is to be expected from a handoff and rerouting protocol. If, for example, we were to design a circuit, we may compare two designs that meet the functional description of the circuit. The same concept holds true for network protocols. A common functional description is needed as a reference point.*

Number of Signaling Messages Exchanged for Rerouting During a Handoff (N_r) — The number of signaling messages due to the rerouting may slow the handoff process and increase the complexity of the signaling software; therefore, the smaller the number of rerouting messages, the easier the handoff procedure is to implement and the faster it should run.

Number of Network Nodes Involved in the Rerouting (N_n) — This criterion determines the time it takes to perform the rerouting and the bandwidth required for signaling.

Number of User Connections Established for Rerouting (N_c) — This parameter is significant for VCT-based handoff procedures and determines the amount of overhead the network has to perform in order to achieve handoff.

User Bandwidth Allocated for Handoff (B) — This is expressed as a multiple of the basic bandwidth allocated for the user connection. In some VCT-based handoff procedures, the amount of user bandwidth allocated for a user connection may be a multiple of the actual needed bandwidth.

The Time It Takes to Execute a Handoff — This measure is greatly influenced by the underlying radio technology of the network; hence, two procedures that run on different radio networks may not be compared directly. (The hybrid handover algorithm proposed by Toh has been implemented by using the Cambridge Fairisle switch, and handoff times on the order of tens of milliseconds have been reported [11].)

Whether the Protocol Is Proposed for WANs (W) or LANs (L)

Rerouting Scheme Used — Cell forwarding (CF), dynamic such as in NCNR (DY), or tree such as in VCT or SRMC (T).

Robustness to Instabilities in Handoff — The instability in a handoff event refers to the ping-pong effect that sometimes occurs when the user terminal is in an area where the radio link to both radio ports may fade. In such a situation, the terminal may switch back and forth between the two radio ports a few times until one of the links stabilize. Hence, a handoff

procedure should be robust with respect to this effect. We will subjectively rate the respective handoff algorithms on a scale of 1 to 5, with 5 being the most robust in our comparison, based on our perception of the algorithms. The most robust algorithms take the ping-pong effect into account and are designed accordingly to minimize additional network traffic due to the ping-pong effect, the least robust algorithms are prone to generating additional network traffic due to the ping-pong effect. The algorithms rated 3 are in the middle of the scale in both aspects.

We can use the criteria above to compare the various handoff and rerouting procedures discussed in this article. The results are given in Table 1 This table may be used as a first step in choosing a handoff and rerouting algorithm for a given wireless network.

Conclusion

In this article we proposed a novel rerouting scheme for a handoff transaction in a wireless ATM network. This rerouting scheme is referred to as Nearest Common Node Rerouting (NCNR). We have utilized the zone concept to illustrate the NCNR algorithm. NCNR is based on finding the ATM node that is a root of or common to both of the zones involved in the handoff transaction. The rerouting is then performed starting from this common node. After describing NCNR we have summarized five other schemes for rerouting for handoff from the literature. A comparison of all six schemes, including NCNR, was also presented. Another contribution of this article is the recognition of different constraints associated with the handoff of different traffic types. From the comparisons, NCNR is seen as a promising scheme for rerouting a wireless ATM connection for handoff.

References

[1] D. Raychaudhuri and N. Wilson, " ATM Based Transport Architecture for Multiservices Wireless Personal Communication Networks," *IEEE JSAC*, Oct. 1994, pp. 1401–14.
[2] B. A. Akyol and Donald C. Cox, "Handling Mobility in a Wireless ATM Network," *Proc. INFOCOM '96* San Francisco, CA, March 24–28, 1996,.
[3] R. Yuan *et al.*, "Mobility Support in a Wireless ATM Network," *Proc. 5th Wksp. on Third Generation Wireless Info. Networks*, 1995, Kluwer, pp. 335–45.
[4] D. C. Cox, "Wireless Network Access for Personal Communications," *IEEE Commun. Mag.*, vol. 30, no. 12, Dec. 1992, pp. 96–115.
[5] K. Y. Eng *et al.*, "A Wireless Broadband Ad-Hoc ATM Local-Area Network," *Wireless Networks*, vol. 1, 1995, pp. 161–74.
[6] O. T. W. Yu and V. C. M. Leung, "B-ISDN Architectures and Protocols to Support Wireless Personal Communications Internetworking," *Proc. PIMRC '95*, Sept. 27–29, 1995, Toronto, Canada.
[7] A. S. Acampora and M. Naghshineh, "An Architecture and Methodology for Mobile-Executed Cell Hand-off in Wireless ATM Networks," *Proc. Int'l. Zurich Seminar on Digital Commun.*, Zurich, Switzerland, March 8–11, 1994.
[8] D. C. Cox, "A Radio System Proposal for Widespread Low-Power Tetherless Communications," *IEEE Trans. Commun.*, vol. 39, no. 2, Feb. 1991, pp. 324–35.
[9] B. Akyol and D. C. Cox, "Signaling Alternatives in a Wireless ATM Network," to be published , *IEEE JSAC* Special Issue on Wireless ATM Networks.
[10] ATM Forum, "ATM Forum User Network Interface Specification," v. 3.1, May 1994.
[11] C. Toh, "The Design and Implementation of a Hybrid Handover Protocol for Multimedia Wireless LANs," *Proc. Mobicom '95*, Nov. 13–15, 1995, Berkeley, CA..

Additional Reading

[1] K. Keeton *et al.*, "Providing Connection-Oriented Network Services to Mobile Hosts," *Proc. USENIX Symp. on Mobile and Location-Independent Comp.*, Aug. 2–3, 1993, Cambridge, MA..

Biographies

Bora A. Akyol received the B.S. degree in electrical engineering from Bilkent University, Ankara, Turkey, in 1992, and the M.S. degree in electrical engineering from Stanford University, Stanford, California, in 1993. He is currently a Ph.D. candidate working on wireless asynchronous transfer mode networking at Stanford University. His research interests include network signaling, media access control, and resource allocation for wireless communication networks.

DONALD C. COX [F] is the Harald Trap Friis Professor of Engineering at Stanford University and director of the Center for Telecommunications. He did research at Bell Laboratories in the late 1960s and early '70s on wireless mobile systems that still provides basic input to the design of cellular and personal communications systems. From the late '70s to 1993 he led and was actively involved in pioneering wireless research, first at Bell Laboratories and then at Bell Communications Research (Bellcore), that started and fueled the current explosion in wireless personal communications. He was manager of all radio research at Bellcore for ten years, and was instrumental in evolving this research intohe Bellcore WACS specification which was combined with PHS into the PACS standard for low-tier PCS in the United States. He has also done research on earth-space and transhorizon radio systems and propagation. He received a B.S. in electrical engineering in 1959, an M.S. in electrical engineering in 1960, and an honorary Doctor of Science in 1983 from the University of Nebraska, as well as a Ph.D. in electrical engineering in 1968 from Stanford University. He is a fellow of the AAAS and RCA.