Abstract

Wireless ATM networks can be implemented by adding mobility support functions to fixed ATM switches. This paper first describes a possible wireless ATM network and protocol architecture relying on “modular addition” of mobility support functions. This architecture sets a broad context where future WATM research can be conducted to elaborate on its various components. Within this context, this paper presents a holistic solution for one of the most important mobility support functions: the handover protocol for intra-switch handovers.

1. Introduction

The success of second generation digital cellular systems (e.g., GSM) has brought huge expectations for the third generation cellular systems based on wireless ATM [4]. There is an ever-increasing demand for wireless high bit rate services enabling simultaneous video, audio and data transmission. To help meet these expectations, international standard organizations such as the ATM Forum have been very active recently on various issues of Wireless ATM.

Because of the wide range of services supported by ATM networks, ATM technology is expected to become the dominant networking technology. It is natural to anticipate an integration of ATM transmission and wireless network technologies to meet all types of service demands effectively. An overview of wireless ATM is presented in [1], introducing ATM concepts and analyzing the requirements for Wireless ATM, in particular for the data link control and radio functions.

Wireless ATM can be implemented by enhancing the current ATM network system so that it supports mobility. On-going calls must be supported when mobile users are roaming between access points (AP) (i.e., the wireless ATM version of base transceiver stations, BTS). Handover is thus one of the most important research issues for wireless ATM. A handover scheme, which can secure a seamless and lossless ATM traffic handover, represents a crucial step toward a wireless ATM solution [6].

This paper presents an intra-switch microcellular handover scheme that draws on and improves the scheme described by the ATM Forum [9, pp. 39-49]. Other handover schemes [6, 10, 11] are also analyzed. It is assumed that the radio access part of the scheme is comparatively independent. The paper thus focuses on the network routing part to provide a robust handover mechanism that takes as many situations into consideration as possible. An intra-switch handover is a handover where the old and the new AP are connected to the same ATM switch. The proposed handover mechanism can also be extended to inter-switch handovers, that is, handovers where the old and the new APs are connected to different ATM switches.

This paper aims at supporting the WATM work of the ATM Forum and other relevant organizations as well as the WATM research community. Therefore this paper uses the same concepts and specification formats used by the ATM Forum [8, 9] and carefully describes how the proposed generic WATM protocol architecture and specific handover protocol differ from earlier works and in what respects the proposed solutions are better than earlier ones.

This paper proceeds as follows. In Chapter 2, the WATM network architecture is briefly introduced. In Chapter 3, the WATM protocol architecture is proposed. Chapter 4 first analyses the handover requirements of WATM, discusses different types of handovers, and reviews earlier handover solutions. Next, an intra-switch
2. WATM network architecture

A wireless ATM network aims at supporting integrated broadband services to mobile terminals through ATM UNI/NNI with mobility enhancements [7]. The network architecture consists of three main parts shown in Figure 1: (1) ATM switches with standard UNI/NNI and additional mobility enhancements, which form the mobility supporting ATM network; (2) ATM access points with mobility-enhanced UNI/NNI and radio interface capabilities; and (3) mobile terminals with enhanced WATM UNI.

![Figure 1. Wireless ATM network architecture](image)

The WATM UNI interface offers support for mapping user terminal ATM connections to the shared medium radio access link and for terminal mobility (handover and location management). The mobility-enhanced UNI/NNI between switches and access points contains signaling and routing protocol extensions indispensable for handover and location management.

In Figure 1, an end user with a Wireless Mobile ATM Terminal (MT) can access the Mobility Supporting ATM Network via the wireless Access Point (AP). Wireless Access Point is a logical component that, when linked with a mobility supporting ATM switch, gives that switch the ability to communicate via a wireless communication link with wireless terminals or wireless APs linked with other ATM switches.

3. WATM protocol architecture

In addition to the control and management functions of fixed ATM switches, WATM switches will also include the following mobility support functions [10, pp. 31-33]: (1) Mobile Connection Management Protocol (MCMP); (2) Mobile Handover Management Protocol (MHMP); (3) Mobile Location Management Protocol (MLMP); (4) Mobile Routing Protocol (MRP); (5) Mobile Media Access Control Protocol (MMACP); and (6) Mobile Data-Link Control Protocol (MDLCP).

We assume that these mobility support functions are relatively independent and can be added separately to fixed ATM switches. Modular addition of mobility was proposed in [6, p. 300] by adding access points and mobility control functions to fixed ATM switches so that they will have built-in, mobile specific switching capabilities. The ATM Forum [8, p. 19] also introduced an additional protocol, Access Point Control Protocol (APCP). However, none of them covered the mapping of mobility support functions to a holistic wireless ATM protocol architecture.

![Figure 2. Wireless ATM protocol architecture](image)

The proposed WATM protocol architecture is shown in Figure 2. Referring to [6, p. 300] and [8, p. 21], the fixed ATM protocol architecture [3, p. 115] is enhanced with the shadowed protocols to support mobility. At the radio interface, Wireless Access Layer (WAL) is added besides the physical layer. It includes wireless physical layer for cell routing (mapping to MRP), medium access control layer to share the mobile channel (mapping to MMACP), and wireless data link control layer to provide flow control and retransmission (mapping to MDLCP). In the application layer, signaling protocols are enhanced with mobility (mapping to MCMP and MLMP); User Network Interface with mobility (UNI+M), Public/Private Network-Network Interface with mobility (PNNI+M), and BISDN Inter Carrier Interface with mobility (BICI+M). Access Point Control Protocol (APCP) allows switches to interact with access points during connection set-ups and handovers (mapping to MHMP).
It is beyond the scope of this paper to discuss in detail all mobility support functions. We focus only on the handover function.

4. WATM handover

4.1. Handover requirements

To secure seamless and lossless handovers in WATM networks, the following handover requirements should be met [4, 6, 10]:

1. **QoS Guarantee.** QoS for each WATM connection should be guaranteed during handovers. This is not easy because of the nature of radio connections.

2. **Resource Consumption Efficiency.** Low buffering should be achieved during handovers to avoid latency. Bandwidth consumption should be minimized.

3. **Low Signaling Traffic.** Signaling traffic between APs and MT should be minimized to reduce the overall traffic.

4. **Scalability.** The handover protocol should be able to serve as many MT-initiated handover requests as possible. It should also be feasible for wide geographic areas.

5. **Low Latency.** The delays and delay variations during handovers should be minimized to secure the QoS of WATM connections (audio/video).

6. **Lossless Handover.** For data, ensuring low or no cell loss is as important a QoS factor as low latency is for audio and video connections.

7. **Exclusive Handover.** For a multimedia call, more than one connection (audio, video, and data) exists at the same time. All the connections for the WATM call should be handed over simultaneously.

8. **Heterogeneous Mobile Connections.** Both unicast and multicast connections should be supported.

9. **Exploitation of Radio Hints.** Signal strength detection should be used as an advance signal to trigger a handover well before MT enters fully into the new wireless cell.

10. **Maintaining the Cell Sequence.** During a handover, the cell sequence for each connection of a call should be maintained.

11. **Minimal enhancement of switch hardware.** Only minimal enhancements to switch hardware should be required to make it economically feasible to use a single platform for both wireless and fixed ATM.

4.2. Types of handover

Handovers are typically initiated by MTs, but WATM networks should also be able to trigger handovers for network management purposes [10]. This paper discusses and analyses only MT-initiated handovers that exploit radio hints.

There are two common types of handover: backward handover and forward handover. They are sometimes also called “soft handover” and “hard handover”, respectively. In a backward handover, the handover can be predicted by using radio hints and MT initiates the handover via the current AP. In a forward handover, the handover occurs when the connection to the current AP is broken. The new AP is contacted first before the handover is initiated.

From the viewpoint of the network, there are also two types of handovers: inter-switch and intra-switch handover. Here only intra-switch handover solutions will be discussed.

4.3. Related works

There are many works that have catered various aspects of the WATM handover requirements. Toh [10] proposed a handover scheme exploiting a radio hint and a partial path setup. However, the scheme assumes that MT can at any time hear only one new AP and that MT can autonomously decide whether to handover to that AP or not, thus hampering the network resource control of the switch. Mitts et al. [6] presented a lossless approach for intra-switch handovers that resembles the approach proposed by this paper. The design objectives of both solutions address all handover requirements except for multicasting and scalability. While their approach recognizes the importance of network resource control, it omits the fact that MT can hear more than one new AP simultaneously. Consequently, it can not guarantee the identification of the optimal new AP with respect to the QoS requirements. The exclusive handover of all the connections of a call can not be guaranteed either.

Yuan et al. [11] suggests a generic inter-switch handover procedure in which the current AP uses the wireless control protocol (deploying PVC or SVC virtual circuits) to contact the neighboring APs that serve as handover candidates. In this case, the AP instead of the switch determines which candidate AP will be used next. From the network perspective, AP-controlled handovers may not yield optimal results with respect to the resource consumption requirements. Moreover, in a typical cellular layout this procedure requires each AP to establish PVCs between six APs. This is likely to be viable only when PVCs or SVCs are also used to implement the forwarding of downlink cells from the old AP to the new AP (c.f., [6,
p. 307]). In the proposal, the cell sequence of all connections for MT is maintained by marshalling together all the datalink state information for all the connections (VCI, traffic class information, transmission and reception tables, outstanding cell sequence numbers and datalink cell buffers) and sending it as a single message to the new AP. But waiting for the last cell of each buffer will cause delay. For delay sensitive connections, this approach may be intolerable.

4.4. The handover solution

Backward and forward handovers are mutually complementary. They will be introduced in sequence. The proposed handover phases are described using examples. Message sequence charts are used to clarify the handover procedure. In the examples, MT has one ongoing call, which includes two active connections. To simplify the case, only one connection of the call will be described in each phase.

Backward handover

During the normal operation, MT maintains a list of APs (other than the current one) it can hear. The list is prioritized by the signal strength of each AP. When the radio link between MT and the old AP (AP1) becomes weak, radio hints show that MT needs to initiate a handover. MT requests a handover by sending HO_REQUEST (1) message to AP1. This message includes the list of APs MT can hear. The message is further forwarded to the switch. When the switch receives the HO_REQUEST message, it sends an RR_STATUS_ENQUIRY (2) message to all APs on the list (or some of them depending on the status of the network) to find out which APs can support the connections. The APs check their resource situation and answer the inquiry with an RR_STATUS (3) message. The switch selects the best AP (in this case AP2) which is able to provide QoS guarantees for all the connections of the call. This design supports the exclusive handover of calls. The handover will be made and RR_ALLOC (4) message will be sent to AP2.

Contrary to the solution described in [9, p. 39], in this solution resource allocation is done after the switch has made the decision. Usually there will be at least two APs on the list. It is not necessary to reserve resources in each AP for the connection when they receive the RR_STATUS_ENQUIRY because only one AP will be finally chosen. By reserving resources upon receiving RR_ALLOC, the resource levels shown in APs will always be the real ones regardless of the number of simultaneous handovers handled by the switch. The switch needs to send only one RR_ALLOC(4) message to AP2 instead of sending many RR_DEALLOC messages to other APs. The solution also reduces the message processing time of other APs. The efficiency of the whole network will thus be improved.

When AP2 has been determined, the switch informs MT that the handover will be made to AP2 by sending it a HO_RESPONSE (5) message via AP1. At the same time, the old downlink route is switched from AP1 to AP2 (6). AP2 starts to buffer downlink cells. When the switching has been done, the switch via the Virtual Channel (VC) of the old connection sends a Down_ready (7) inband mark. Down_ready indicates the end of downlink data stream. In the “Sudden Withdrawal” situation, there is a potential risk involved in buffering downlink cells in AP2. If MT suddenly withdraws from entering the new cell of the cellular network, this withdrawal will cause time-outs and abolish the new route. In this case, downlink cells buffered in AP2 need to be forwarded to the next AP for buffering. To maintain the cell arrival sequence, the forwarded cells should be transmitted first when MT has set up a new radio link.

If MT must disconnect its radio link from AP1 before all cells buffered in AP1 have been successfully sent to MT, the switch enables cell forwarding from AP1 to AP2 by sending AP1 the FORWARD (8) message which includes the ID of AP2. The forwarding of cells is always enabled in this solution and AP1 is responsible for it. The switch thus does not need to be informed whether the downlink cell delivery was successful or not. It should be noted that the ATM Forum [9, p. 40] does not mention the forwarding of cells at all.

When MT receives HO_RESPONSE carrying the identity of AP2, it can initiate the radio handover. MT knows that the downlink connection has already been re-
routed in the switch and a handover to AP2 is inevitable. However, both uplink and buffered downlink cells are still transmitted normally over the radio interface. There are two choices as to when MT should initiate the radio handover: (1) MT can wait for AP1 to send all buffered cells. This is a good choice for connections without critical timing requirements; (2) MT can start the handover immediately to minimize jitter introduced into the downlink cell stream. Cells buffered in AP1 will be forwarded to AP2. This can be achieved by adding a buffer to save forwarded cells in APs. In this way, the cell sequence is always maintained. Both choices have their pros and cons. When calls involve several connections each with different QoS, things become more complex. We argue that the highest QoS connection should be used as the decision criteria. Although some of the connections might be “promoted” to a higher level QoS, it is better to provide more rather than less. The most important issue is that all the connections of a call can be handed over exclusively while QoS for the call as a whole can still be guaranteed.

Figure 4. Rerouting the downlink from AP1 to AP2 and releasing the radio link with AP1

If there is enough time for AP1 to send MT all cells buffered for a given VC (AP1 knows this when the Down_ready is the next downlink cell to be transmitted), AP1 indicates the last cell sent to MT with a No_more_traffic (9) flag. Else, a forwarding procedure, which forwards cells buffered in AP1 to AP2, is needed. This procedure will be explained in Section “Forward handover.” Here we assume that MT can wait for AP1 to send all buffered cells.

AP1 transmits any remaining uplink cells and sends the switch an inband mark called Up_ready (10 in Figure 4) after the last cell (10). After receiving Up_ready, the switch knows that it is safe to allow AP2 to start transmission so that the cell order of the connection is maintained. Figure 4 describes the rerouting of the downlink from AP1 to AP2 and the releasing of the old radio link.

Next, MT establishes a radio link with AP2 and releases the radio link with AP1. Contrary to the solution described in [9, p. 40], MT should establish the link with AP2 before tearing down the link with AP1. In this way, AP1 can still signal another handover if MT cannot establish the new radio link, thus ensuring a lossless handover. MT sends a CONN_ACTIVATE message (11) to AP2 to activate all VCs. Then MT releases the radio link (12). A DR_flag for each VC in this message indicates whether MT did receive all downlink cells of the connection via AP1 (that is, whether MT received a No_more_traffic flag from AP1 for a given connection). In this case, DR_flag is set to indicate that all cells were received.

The CONN_ACTIVATE message includes any radio interface identifiers to be used and indicates that MT is ready to receive cells using VC. At this point, AP2 can start sending downlink traffic to MT. In the example, AP2 has buffered downlink traffic and because no cells need to be forwarded, it can immediately start to transmit the buffered downlink cells to MT.

Meanwhile, the Up_ready mark has reached the switch. This indicates that there are no more uplink cells coming from AP1 and uplink cells can now be sent via AP2 while maintaining the cell sequence. The uplink connection routing is updated in the switch (13), CONN_SWITCHED (14) is sent to AP2 and
CONN_RELEASE (15) to AP1. When AP2 receives the
CONN_SWITCHED message, it sends a
CONN_ACTIVE message to MT (16). This message tells
MT that it can start transmitting uplink cells and the
handover has been completed. Figure 5 describes new
radio link establishment and uplink switching.

The sequence of messages needed in the backward
handover is illustrated with a message sequence chart
(Figure 6). The chart depicts a situation where MT has
two active connections and both of them are handed over
to AP2. The chart also shows the buffers needed in
different phases. An arrow above a buffer indicates that
the buffer starts to fill whereas the emptying of a buffer is
indicated with an arrow placed below the buffer. A
switching event is illustrated with X in the figure.

Forward handover

Forward handover is initiated when the radio link
between MT and AP1 is lost (1). Downlink cells are
buffered in AP1 and uplink cells are buffered in MT. AP1
transmits any remaining uplink cells to the switch. When
the last uplink cell has been transmitted from the buffer,
AP1 sends an inband mark called Up_ready to the switch
(2). This mark serves as the end point of uplink cells in
the user data cell stream, thus indicating that there are no
more uplink cells coming from AP1.

When MT has established a radio link with AP2, it
sends the switch a HO_REQUEST (3) message via AP2.
The message contains a list of the APs (in a prioritized
order) to which MT could handover. When the switch
receives the HO_REQUEST message, it sends a
RR_STATUS_ENQUIRY (4) message to each AP in the
list to determine which AP could best support the
connection. Now each AP checks the resource situation
and answers the inquiry via a RR_STATUS (5) message.
Figure 7 shows the initiating of forward handover.

After receiving all the RR_STATUS messages, the
switch has to decide to which AP the handover will be
directed (Figure 8). In this example, the switch decides to
route the connection to AP2 (6). After the decision, the
switch sends a RRALLOC (7) message to AP2 so that it
can allocate the resource for this call. The switch also
sends a FORWARD (8) message to AP1 requesting AP1
to start sending buffered downlink cells to AP2.

Contrary to the solution described in [9, p. 43], we
argue that MT must not send any further contact
messages to any APs before the switch has decided to
which AP the handover will be directed. The purpose of
maintaining an AP list in MT is to provide the switch with information needed to pick up the best AP. So the AP chosen by the switch may not be the AP with which MT already built the radio link. If the switch chooses another AP, all the communication efforts with the current AP will be wasted. The use of the AP list thus increases the probability of a successful handover. On the other hand, the use of the list may cause some delay.

Next, the uplink and downlink connections are simultaneously switched to AP2. Immediately after switching, the switch sends AP1 an inband mark called \textit{Down\_ready} (9). This mark indicates that the cell before \textit{Down\_ready} was the last downlink cell sent to AP1. After sending \textit{Down\_ready}, the switch sends a \textit{HO\_RESPONSE} (10) message to MT (transparently through AP2) as an answer to the \textit{HO\_REQUEST} message. Immediately after receiving \textit{HO\_RESPONSE}, MT sends AP2 the \textit{CONN\_ACTIVATE} message (11) asking it to activate connections.

Figure 8. Switching both uplink and downlink

AP1 forwards all buffered downlink cells to AP2. The \textit{Down\_ready} mark is also buffered and forwarded like a normal user data cell. When AP2 receives \textit{Down\_ready}, it knows it has received all the forwarded cells. In order to maintain cell sequencing, AP2 buffers the forwarded cells separately from the cells received directly from the switch and sends them to MT first.

As soon as AP2 starts receiving the cells forwarded from AP1, it can begin to send them to MT (12) because the radio downlink has been opened by \textit{CONN\_ACTIVATE} already.

After the switching instant, the switch sends a \textit{CONN\_SWITCHED} (13) message to AP2 to indicate that the connection is successfully routed to AP2. This switch also sends a \textit{CONN\_RELEASE} (14) message to AP1 to indicate that AP1 can release all the resources possibly reserved for MT. When AP2 receives the \textit{CONN\_SWITCHED} message, it sends a \textit{CONN\_ACTIVE} (15) message to MT. This message indicates that MT can now send the buffered uplink cells to AP2.

Figure 9. AP2 sends buffered cells to MT and starts normal operation

Figure 10. The message sequence chart for the forward handover protocol
The sequence of messages needed for the forward handover is shown in Figure 10.

5. Conclusions and future research

The current developments of Wireless ATM are mainly based on ATM as the backbone network with enhanced mobility. Mobility functions are ‘added’ into the ATM switches and/or base stations so that ATM networks can support both fixed and mobile ATM users.

This paper has contributed to the WATM research area by (1) outlining ATM network and WATM protocol architectures with full range of mobility support functions, and (2) focusing on an efficient, simple, and cost effective handover mechanism supporting both forward and backward handovers. Specifically, we have drawn on the work of the ATM Forum and presented the following improvements to their handover solution:

1. The resource allocation scheme of APs has been changed so that resources are reserved only in the AP that the switch has chosen for establishing the connection(s) with MT.
2. Cell forwarding is always enabled and APs are responsible for it so that unexpected disconnections of radio links can be handled efficiently.
3. The timing of the initiation of a radio handover is determined by the highest QoS connection of a call so that the QoS for the call as a whole can be guaranteed.
4. MT establishes the link with the new AP before tearing down the link with the old one so that the old AP can be used to initiate a new handover if MT fails in establishing the new link.
5. In the case of a forward handover, MT minimizes its communication with APs before the switch informs it which AP should be used.

The following issues have not been discussed in detail because they are covered by Mitts et al. [6] in ways that can also be applied to our solution:

- various alternative methods for implementing the forwarding of cells;
- the reasons for buffering the uplink cells in MT instead of AP; and
- the performance of the handover solution when various errors (e.g., the temporary unreachability of MT due to deep fade) occur.

Network-initiated and inter-switch handovers have been outside the scope of this paper, thus representing an area where our solution needs to be elaborated on. QoS renegotiation and provisioning will also be our next research target.

Two requirements of the handover solution have been researched relatively little, that is, scalability and support for multicast connections. For example, our solution and those of the ATM Forum [9], Mitts et al. [6] and Toh [10] do not support handovers of multicast connections. Yuan et al. [11], while not explicitly mentioning multicast connections, present an approach that groups together multicast and unicast connections for MT and then enacts an exclusive handover (see also Section 4.3). They do not describe this aspect of the approach in great detail, but it might nevertheless provide a direction for further research in this area.

There are WATM research projects that relate to our work from the viewpoint of scalability, thus yielding new insights for the further development of the handover solution. For example, Lobley [5] proposed a solution in which mobility management functions are supported by the intelligent network (IN). The use of IN for handover control would remove the processing burden of handover control from WATM switches because the solution makes the switches of the fixed network responsible for only mobile call-handling functions. The scalability of handovers would thus be improved and the switches could also handle more calls. Cáceres and Padmanabhan [2] suggested that it may be beneficial for the handover mechanism to differentiate between traffic types by prioritizing the traffic so that delay-sensitive traffic streams (e.g. audio) would be handled first.

Further research is needed to investigate (1) how these and other developments affect the handover solution proposed in this paper and (2) how the solution can be tested and further improved and elaborated on. Simulations will be used to verify which strategy yields the best performance. A solid research agenda must thus be established in the area of WATM handover.

6. References


