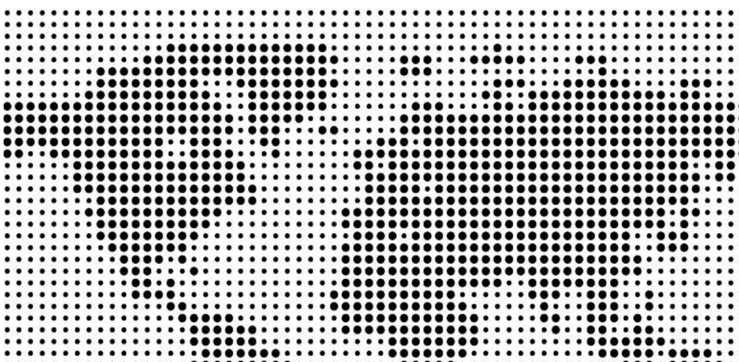
QUALCOMM

UMB Network Architecture



Qualcomm, Incorporated
December 2007

Table of Contents

Table of Figures

Executive Summary

Ultra Mobile Broadband radio-access technology uses advanced innovations in wireless communications to deliver industry-leading high-speed data throughputs, low latencies and high quality of service (QoS) for an enhanced mobile-broadband experience. It is optimally designed from the ground up, at both the physical air-interface and the upper layers, to support bandwidth-intensive mobile services and concurrent VoIP and data services with greater flexibility.

UMB enables efficient wireless transfer of IP packets at very high data rates, while providing seamless mobility and best QoS, even at the cell edges, without lowering frequency re-use. The UMB network architecture is designed for deployments with maximum frequency re-use.

UMB systems benefit from a highly innovative flat network architecture that simplifies the core network and network interfaces, making it easy to scale the network. With UMB, networks can be easily scaled to serve a myriad of base stations for different deployment scenarios, such as environments with varying coverage and capacity requirements. The distributed network architecture allows operators to spread the processing load across the network elements, thus simplifying the overall design. Leveraging standard IP components, operators should be able to scale their networks based on system usage, at reduced time and costs.

One of the key principles for UMB architecture is seamless mobility. A major emphasis is placed on the design of network architecture to facilitate seamless handoffs both within the UMB network and across different technologies. Innovative concepts enable fast switching between base stations while minimizing overhead and offering simpler network interfaces. New tunneling mechanisms provide signaling exchange at the data-link layer (layer 2) and IP layer (layer 3) to enable faster mobility across the base stations.

This paper discusses key features of UMB network architecture, and provides insight into various architecture design choices.

[1] Introduction

The main objectives of the UMB network design are:

- Seamless mobility
- QoS for a wide range of applications, including latency-sensitive traffic
- Efficient frequency re-use deployment
- Interoperability of equipment from different vendors, by simplifying the interface between the network elements
- Ease of network management and a reduced number of network elements
- Network scalability and flexibility in service deployment

All of the above objectives are accomplished simultaneously in the UMB design. The main features of UMB network architecture can be broadly summarized as follows:

- A flat network architecture that simplifies the core network design and eliminates the need for centralized base station controllers (BSCs)
- A converged-access network (CAN) design that enables seamless mobility
- A multi-route feature that enables fast switching between base stations and provides requisite support for latency-sensitive applications
- Layer 2 and layer 3 tunneling mechanisms to simplify the network interfaces

Decoupling of layer 3 handoffs with layer 1 handoffs, ensuring fast and efficient mobility across base stations

[2] UMB: Flat Network Architecture

The UMB solution, based on flattened network architecture, is a marked departure from the traditional hierarchical architectures defined with multiple layers of control and interconnect platforms. In a traditional hierarchical architecture, the access terminal (AT) maintains a single air-interface protocol stack for communication with multiple base stations (BSs). A centralized control entity called a BSC keeps the protocol state coordinated between multiple BSs. The BSC coordinates fast-changing state information across multiple BSs and maintains a common state, such as layer 2 buffers, across multiple BSs.

A UMB network using a flat architecture does not require a centralized entity such as a BSC (Figure 1). In this scheme, there is no need to coordinate the connection state across the UMB's BS equivalent, the evolved base station (eBS). This flat-architecture eBS combines the functions of a traditional BS, a BSC, and some functions of the packet-data serving node (PDSN) into a single node, making the deployment of the networks simpler. As the number of elements required to build a network are reduced, the networks become more reliable, more flexible, easier to deploy and less costly to operate. There is a significant reduction in inter-eBS interaction, leading to a vastly simpler inter-eBS interface, which thus facilitates multivendor interoperability.

The flat-architecture BS can be connected directly to the Internet. Traditional hierarchical mobile-broadband radio access networks serve user traffic in multiple types of nodes. For example, in legacy networks, the BS, BSC, PDSN and mobile IP home agent all cooperate to serve user traffic. Combining these functions into fewer nodes reduces latency, decreases capital and maintenance costs, and reduces the complexity of interactions between the nodes to deliver end-to-end QoS.

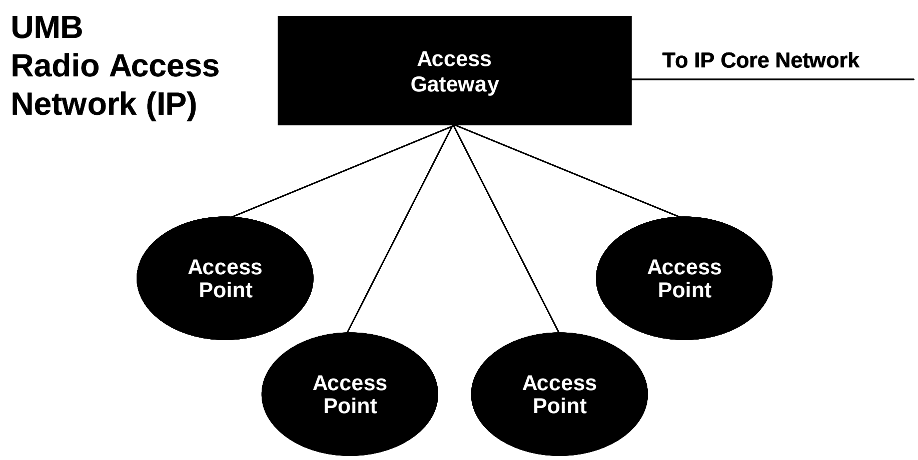


Figure 1. The UMB flat network architecture, depicting access points connected on their backhauls to a single IP attachment point.

[3] UMB and the Converged Access Network

One of the core principles of the UMB network architecture is a CAN that seamlessly integrates UMB with existing 3G core networks. In a converged network, an operator can gain additional capacity via UMB wherever necessary, while using existing 3G for mobile-broadband access in the majority of the network. An example of a UMB CAN with appropriate interfaces in interoperation with a CDMA2000® 1x EV-DO network is depicted in Figure 2.

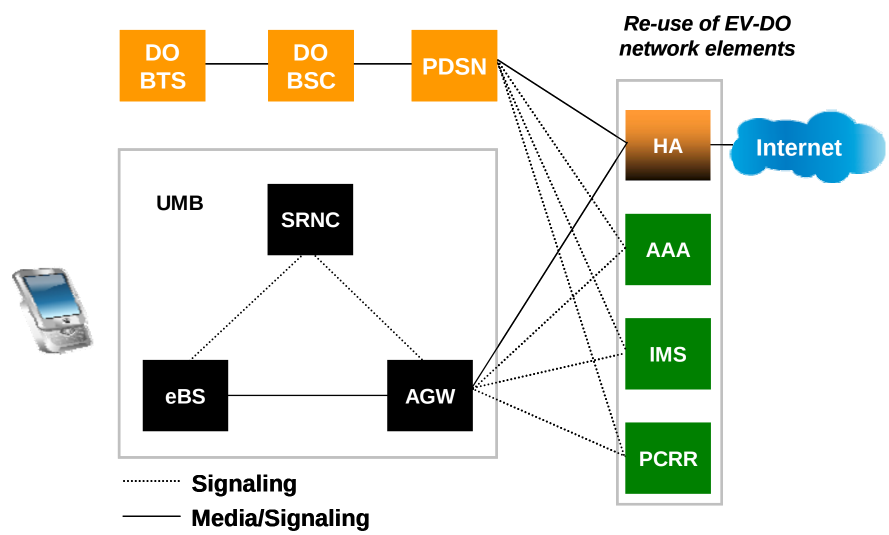


Figure 2. A UMB converged access network, depicting the connecting interfaces between the UMB network and an EV-DO network.

Central to the UMB radio access network is the eBS, which in many respects is comparable to a combination of today's 3G BSs and BSCs. Fitting the flat architectural model, the eBS allows IP to extend to the BS, thereby lowering latency and enabling faster switching of the AT between base stations.

The eBSs of the UMB radio access network connect to a common access gateway (AGW), which provides a point of IP attachment for the AT to the packet data network. The mechanics of the interconnectivity process between different elements for mobility support are addressed in section four of this paper.

3.1 UMB Network Elements

The following are the definitions of the key elements in a UMB network:

3.1.1 AT: Access Terminal

The AT is the subscriber device that provides IP data connectivity to the user, typically a mobile phone, PDA or mobile-enabled laptop.

3.1.2 eBS: Evolved Base Station

The eBS provides the over-the-air (OTA) signaling and user-data transport that is used by the AT for connectivity to the radio access network. Additional functions of the eBS include:

- Providing a layer 2 attachment point for the AT
- Acting as a layer 1 attachment point for both forward and reverse links
- Encryption/decryption of packets at the radio-link protocol (RLP) level for OTA transmission/reception
- Scheduling for OTA transmission
- Header compression

The eBS also provides these important functions:

- Forward-link serving eBS (FLSE)—Serving eBS for the forward-link physical layer
- Reverse-link serving eBS (RLSE)—Serving eBS for the reverse-link physical layer
- Signaling radio network controller (SRNC)—Session anchor eBS that stores the AT's session information and serves as a permanent route to the AT
- Data attachment point (DAP)—Receiving eBS for all the data packets from the common AGW

Additionally, an eBS has visibility into the user's IP packets and can optimize OTA scheduling or perform other value-added functions.

3.1.3 AGW: Access Gateway

The AGW provides the user's point of IP connectivity to the network. That is, the AGW is the first-hop router for the mobile terminal. The AGW performs layer 3 services and above, including hot-lining, policy enforcement, and more.

3.1.4 SRNC: Signaling Radio Network Controller

The SRNC maintains radio-access-specific information for the AT in the CAN. The SRNC is responsible for maintaining the session reference (session storage point for negotiated air-interface context), supporting idle-state management, and providing paging-control functions when the AT is idle. The SRNC is also responsible for access authentication of the AT. The SRNC function may be hosted by an eBS or may be located in a separate (radio-less) entity.

3.1.5 AAA: Authentication, Authorization and Accounting Function

This functional entity provides authentication, authorization, and accounting functions with respect to the AT's use of the network resources.

3.1.6 HA: Home Agent

The HA is used to provide a mobility solution to the AT in a 3GPP2 packet-data network. However, in an evolved network, the HA may also be used for intertechnology mobility.

3.1.7 PDSN: Packet Data Serving Node

The PDSN is the node that provides the user's point of IP connectivity in the existing EV-DO or CDMA2000 1X packet-data networks.

3.1.8 PCRF: Policy and Charging Rules Function

The PCRF provides rules to the AGW. The purpose of the PCRF rules are to:

- Detect a packet belonging to a service data flow
- Provide policy control for a service data flow
- Provide applicable charging parameters for a service data flow

[4] Mobility Management

UMB benefits from a highly innovative network design, particularly with respect to mobility management in enabling faster handovers, freedom in network scalability and a truly distributed access design. With UMB network architecture, operators can offer full-mobility applications while providing highest QoS. To get an insight into UMB's mobility management, it is useful to review a set of key concepts that are used to handle various mobility scenarios. They are described as follows:

4.1.1 Multi-route

Multi-route is at the core of UMB network architecture. The UMB AT maintains an independent air-interface protocol stack associated with each BS. Each of these protocol stacks is called a route. The AT further maintains a route set consisting of all eBSs that have a route with the AT, and the multi-route is used to signify the multiple routes that the AT maintains with different eBSs. An important feature is that every eBS is set up to be the serving eBS when it is added to the route set. A route set can support a minimum of six routes at any time. When the AT is idle, it maintains one route with the SRNC.

Further, each eBS maintains a connection state associated with each route. The connection state includes parameter values and the state of algorithms that help to maintain the connection between an eBS and an AT.

> Examples: Transmit/receive buffers, sequence numbers in an RLP that provide reliable delivery of upper-layer packets, granted QoS for various flows, granted MAC resources, etc.

Since the AT maintains a different route with each eBS and the connection state is local to an eBS, as the AT hands off from one eBS to another, no connection state is transferred between the eBSs. This significantly reduces the complexity of inter-eBS signaling.

4.1.2 Common Session

Although each eBS has an independent route, all eBSs share a common session with an AT. A common session defines sets of protocol types and protocol attributes negotiated and stored by the AT and eBS.

4.1.3 Personality

The session comprises one or more personalities. Personality defines the protocol types and attributes that are used between the AT and the eBS during communication. While the session is common between the AT and all eBSs in a route set, each eBS independently negotiates a personality to use for its route.

A personality negotiated by one eBS can be used by another eBS with no need for any new renegotiation. This significantly reduces the amount of time required to add a new eBS to the route set. The key outcomes are:

- It is faster to add a new eBS to the route set compared with adding a BS to the active set in existing networks
- Inter-eBS interface is made simple, with only a minimal need for coordination of the eBS configurations

4.1.4 FLSE: Forward-Link Serving Entity

This is the eBS that provides layer 1 connectivity on the forward link.

4.1.5 RLSE: Reverse-Link Serving Entity

This is the eBS that provides layer 1 connectivity on the reverse link.

Figure 3 illustrates the relationship between the eBS and the AT with respect to sessions and personalities.

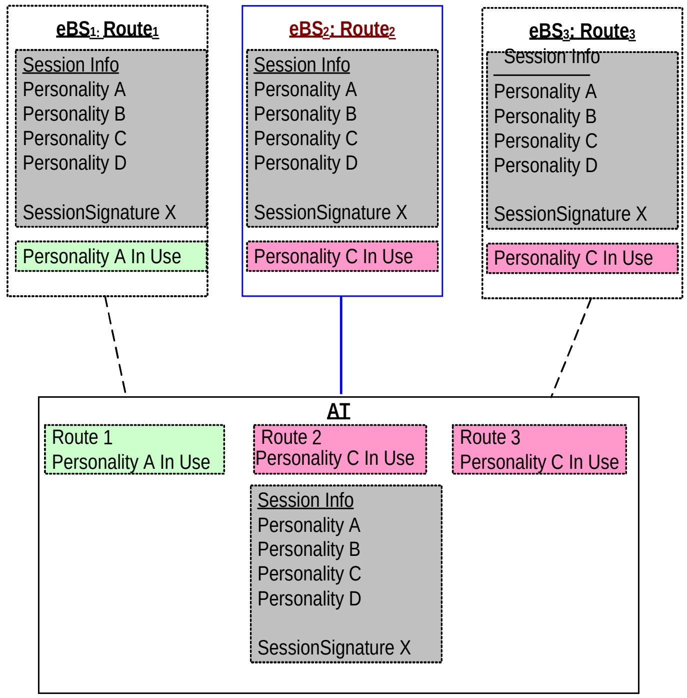


Figure 3. Example depicting a route set of an AT and the eBS selection of different personalities. The AT and eBSs share a common session, whereas each route can choose its own personality.

4.2 Layer 1 Handoff Mechanism

Given the definition of key functions and concepts, the following explains the mechanics of a layer 1 handoff:

- Based on measurements of channel conditions, the AT determines an appropriate eBS in the route set to serve as the layer 1 point of attachment (one for FLSE and one for the RLSE)
- The FLSE and RLSE functions may be served by the same eBS or different eBSs
- The chosen FLSE and RLSE are communicated by the AT to the network using physical-layer channels that support fast switching of FLSE and RLSE as channel conditions change
- When the AT switches from one eBS to another, it uses the route associated with the new serving eBS to transfer packets

The distinguishing features in UMB network design are as follows:

- Because the target eBS is set up to be the serving eBS when it is added to route set, and because no connection state info is exchanged between eBSs, the inter-eBS interface is vastly simplified, enabling fast switching of layer 1 FLSE and RLSE entities during a handoff.
- The system provides QoS to latency-sensitive applications without resorting to compromises using a lower frequency reuse.
- The AT maintains an active set within which it is allowed to switch the FLSE and RLSE with fast physical-layer signaling. The eBSs are added to the active set ahead of switching time, during which connection setup is completed with the new eBS. This allows very fast switching between eBSs, as the target eBS is ready to accept the AT's traffic when the AT switches to the eBS.

After the AT switches from the source eBS to the target eBS in a layer 1 handoff, it is likely that there are remnants or fragments of data packets in the route to the source eBS that still need to be delivered to the AT. A tunneling mechanism ensures zero loss of packets during handoff.

4.2.1 Tunneling

The tunneling mechanism provides zero-loss seamless handoffs in layer 1 eBS switching, without the need for a BSC or centralized controller.

- A layer 2 tunnel between the eBSs is used to deliver the fragmented data packets or fully buffered packets (left at the previous source eBS) to the target eBS
- A layer 3 tunnel between the eBSs is used to deliver unfragmented IP packets buffered at the previous eBS to the target eBS, for OTA delivery to the AT
- The target route then carries the packets OTA on behalf of the source route and delivers the packets to the AT

4.3 Essentials of Layer 2 Communication

Once an eBS is in the route set, the eBS will need to exchange messages with the AT to maintain the connected state, even though the current route is not serving. Layer 2 communication facilitates this process.

In traditional networks, if an AT has to communicate with a non-serving BS, the interface is such that it is necessary to translate the protocols between the BSs. For instance, the OTA protocol for the serving BS is converted into a network protocol. In the UMB network design, the protocols between the AT and an eBS in the route set are independent of the delivery service provided by the serving eBS. The inter-eBS interfaces are kept simple:

- A layer 2 communication is established between an eBS and the AT as soon as the eBS is added to the route set. A relationship is established between the new eBS and the AT using the personality's defining protocols and attributes.
- All exchanges between the AT and an eBS in the route set are communicated through a layer 2 tunnel between that eBS and the serving eBS. This is also called blind tunneling, where the serving eBS blindly delivers the packets to the AT without interpreting the content.

  Examples: Transactions between the AT and the eBS related to reporting radio measurements, QoS requests and grants, etc.

- Thus, the protocol running between the eBS and the AT is independent of the delivery service provided by the serving eBS, without the need for translation of protocols between eBSs.

4.4 Essentials of Layer 3 Handoffs

As mentioned in section three, an eBS can serve any of the four functions: FLSE, RLSE, SRNC and DAP. The network is designed such that these functions need not be designated to a single eBS, but different eBSs in the route set can serve any or all the four functions for a given connection between the AT and the network.

The AGW acts as the layer 3 attachment point for the AT. The AGW sends packets destined to the AT to the eBS serving the DAP function. If an eBS is assigned to be the DAP function, the DAP eBS and the AGW establish a proxy mobile-IP (PMIP) binding between them.

In the event of layer 1 eBS switching and handoff, and if the DAP eBS is different from the new FLSE eBS, when the AGW transmits packets destined to the AT, the network is designed to enable the DAP eBS to forward the packets to the FLSE eBS using a layer 3 tunnel.

A layer 1 switch can trigger the AT to send a DAP move request to the eBS currently serving as FLSE. Although it is preferable to locate DAP and FLSE with the same eBS, it is not desirable to move the DAP functionality too often, as the move causes binding update loading between the DAP and the AGW and requires moving the session reference eBS. To avoid this, the eBS guides the AT by specifying attributes in regards to the frequency of DAP move requests.

The network is designed to decouple the DAP move or Layer 3 handoffs from a Layer 1 handoff allowing Layer 1 handoffs to happen more frequently than DAP moves.

4.5 Inter-AGW Handoffs

In general, the network is constructed to have a meshed topology, with several eBSs connected to a single AGW. More or less, eBSs in the same route set should be connected with the same AGW. An inter-AGW handoff occurs when an eBS added to route set is linked to a different AGW.

Inter-AGW handoffs are implemented as elegantly as the inter-eBS handoffs, with no data interruption. There are no interfaces needed between AGWs. Seamless inter-AGW handoffs are achieved in a make-before-break fashion, by making use of the layer 2 tunneling mechanisms.

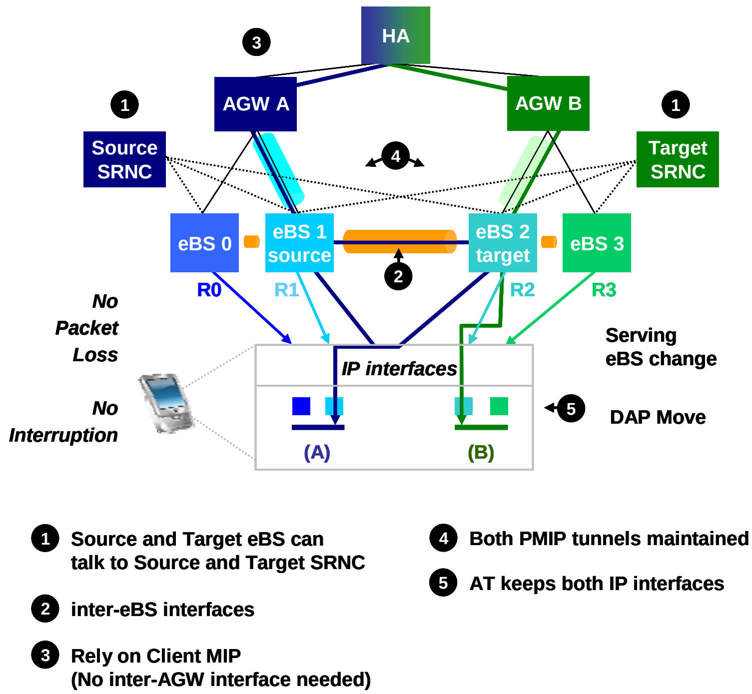


Figure 4. A UMB network in seamless inter-AGW handoffs.

For an inter-AGW handoff, layer 2 tunneling is established between neighboring eBSs under different AGWs. During the inter-AGW handoff process, the AT is served by one DAP eBS and two PMIP tunnels are maintained between the eBS–AGW pairs for a short period. The AT, during this short period, keeps alive the two IP interfaces with the two AGWs in handoff. Once the communication is firmly established between the AT and the target AGW, the route set will now have a new DAP eBS with a new PMIP bonding with the target AGW.

4.6 Inter-System Handoffs

The UMB network allows for seamless inter-system handoffs. For example, with a minimal service interruption, handoffs can occur between UMB and EV-DO networks.

For an inter-system handoff, a dual-mode AT has to negotiate an EV-DO session, create a PPP connection and install a TFT through the UMB by establishing link-layer tunneling support between the AT and the EV-DO radio access network. The dual-mode AT, through a tune-away procedure, monitors the signal on radio interfaces with both the EV-DO and the UMB networks, to implement a layer 1 handoff. The AGW can establish a connection with the home agent using a PMIP or client mobile-IP, to implement a layer 3 handoff.

4.7 Paging Design and Functionality

The paging functionality is simpler and lighter in UMB network design. Unlike the traditional centralized-node, BSC-based paging, where a paging area is maintained for the AT, the SRNC takes care of this functionality by tracking the registration of the AT with a single eBS. The following describes the process:

1. **When the AT transitions from connected state to idle state, all routes are deleted except the route to the SRNC**

- The AT then performs distance-based registration with the SRNC
- Distance-based registration involves the AT re-registering if it has moved more than a set distance from the last location it has registered
- For this, the AT sets up a route with a local eBS to tunnel this registration to the session reference (SR) eBS

2. **If a packet arrives for the AT in idle state, the DAP eBS sends a paging request to the SR eBS, and the SR eBS in turn sends the paging request to the last registered eBS**

- The SR eBS's page requests the last registered eBS to propagate the paging request to other eBSs in the neighborhood, within a given radius
- Upon receiving the page, the AT creates a route and sets up connection with the local eBS
- Therefore, the SR eBS also provides key paging support for the AT by just passing the request to the last registered eBS
- Thus, the paging functionality is made lighter by having the SR eBS track only one eBS, which is the last eBS registered by the AT

[5] Conclusions

The UMB solution's flat architecture is a simplified network design that provides superior mobility-management support for bandwidth-intensive traffic. The flat network architecture removes the need for network elements such as centralized BSCs, and significantly reduces the number of network nodes needed for interoperation.

The UMB CAN achieves seamless and fast handoffs, whether they are inter-eBS, inter-AGW or inter-system transitions for the AT, all while minimizing overhead. Thus, the system is well designed to provide QoS for a wide range of applications, including latency-sensitive applications.

The network architecture is designed to keep the interfaces between the radio access network and core network as simple as possible. For example, avoiding the coordination or transfer of connection states between eBSs and avoiding the interpretation of packets bound to another eBS have clearly simplified the inter-eBS interfaces.

This will help operators deploy and scale UMB networks with ease, and will go a long way in facilitating multivendor interoperability.

[6] References

3GPP2 C.S0084-0 v 2.0, Ultra Mobile Broadband™ (UMB™) air interface specification

3GPP2 A.S0020, Interoperability Specification (IOS) for Ultra Mobile Broadband (UMB) Radio Access Network Interfaces

IETF RFC 3095, RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed

Parag Agashe et al., "Ultra Mobile Broadband Radio Network Architecture and Air Interface Upper Layers"