Project	IEEE P802.16 Broadband Wireless Access Working Group
Title	**System Requirements Assuring That Point-to-Multipoint Broadband Wireless Access Networks Are Agnostic to User and Network Protocols**
Date Submitted	24 June 1999
Source	Ray W Sanders CircuitPath Network Systems P O Box 24950 Los Angeles, CA 90024 Voice: (310) 476-5063 Fax: (310) 471-7854 E-mail: rws@Circuitpath.com
Re:	This is a response to the Call for Contributions for System Requirements by Brian Petry and Gene Robinson dated 20 May 1999
Abstract	This document is a tutorial presentation describing a MAC-layer architecture whose objective is to assure that the 802.16 standard is agnostic to network and user protocols. This includes defining a new MAC Bandwidth Allocation Sublayer that can support any existing MAC layer (including but not limited to those normally associated with IP, ATM, STM and MPEG). It has the prospect of supporting new, as yet undefined, protocols within shared media networks in general, and Broadband Wireless Access networks in particular. The approach described provides Bandwidth-On-Demand and robust Quality of Service operation based on minislot scheduling principles being developed by a number of vendors.
Purpose	The purpose of this document is to propose that the approaches and definitions described herein be adopted as a part of the 802.16 System Requirements.
Notice	
Release	

System Requirements Assuring That Point-to-Multipoint Broadband Wireless Access Networks Are Agnostic to User and Network Protocols

Ray W. Sanders
CircuitPath Network Systems

Introduction

The original purpose of IEEE 802 protocols was to provide standards that would be capable of supporting packet-based data traffic in a world where a dichotomy existed between asynchronous data and isochronous voice communications. The solution to supporting both types of service was to build separate networks for each. The dichotomy still exists, but, now, it is mandatory that all communications services be combined into a single network fabric.

There have been many proposals for providing convergence of communication network services. These can be roughly categorized as 1) IP-based, 2) ATM-based and 3) MPEG-based. So far, there is no approach that seems to answer all needs.

Robust Quality of Service (QoS) operation is a compelling need that must be included in future networks. There have been a number of proposals made over the last ten years to solve QoS problems, the latest of which is the DiffServ protocol efforts of the IETF, an approach that shows somewhat more promise than many of its predecessors, but still, like others, is complex. It is now clear that QoS needs can best be satisfied by focusing on Admission Control and Flow Control strategies instead of relying only on internal network buffers and complex network software to reduce network congestion and to provide multi-level priority differentiation among services. This submission focuses on this important conclusion and proposes a MAC layer protocol approach that can provide a solution.

A simple extension of existing 802 protocols can result in superior Admission-Control/Flow-Control capability, and, concomitantly, can be implemented to interface with existing (or future) protocol standards. Properly defined, it will become a *lingua franca* that enables networks to support any current or future user or network protocol.

Point-to-Multipoint Broadband Wireless Access Reference Model

The reference model for a point-to-multipoint Broadband Wireless Access (BWA) cell (possibly as a part of a multi-cell network) is assumed to include the following:

- A BWA cell consists of a single base station with one or more remote stations.
- BWA operation can include Frequency Division Duplex (FDD) or Time Division Duplex (TDD) operation (or a combination thereof).
- Either symmetrical or asymmetrical bandwidth allocation between a BWA base station and its remote stations is allowed for duplex operation.
- All bandwidth allocations within a BWA cell are controlled by its base station.

What makes an 802 standard truly successful?

All 802 standards have been conceived with the idea that their presence enables multiple vendors to participate in markets whose growth is enhanced by the standards. How well this result obtains is a measure of the success of a standard committee's efforts. To focus our efforts, it seems instructive to first try to answer two questions. One is, "What 802 standard has most influenced the development of the communications industry?". The other question is, "Why??".

Arguably, “802.3” is the answer to the first question. Ethernet, in its many variations, has become the *de facto* standard for local area network connections and is now the common means of interconnection for a wide variety of terminal and network equipment.

There may be a number of answers to the “Why” question, but many would say that it is something along the following lines: “The 802.3 standard places minimal restrictions on supporting any type of digital signal whether its origin be data, voice or video.” In other words, the 802.3 approach is quite agnostic to both user and network protocols.

Going beyond its initial field of application, Local Area Networks, the 802.3 protocol has been used within a number of fixed wireless implementations. Most of these have been in point-to-point symmetrical FDD duplex configurations. But, in spite of its great success in local area environments and its limited application in fixed wireless systems, no 802.3 variant fits well the needs of point-to-multipoint BWA networks. Without substantial modification, it wastes available bandwidth and has no built-in bandwidth allocation mechanism.

A major objective of the 802.16 Working Group must be to devise a standard that is at least as successful as 802.3. This submission strongly suggests that this can occur if the standard includes within its MAC layer a bandwidth allocation mechanism that results in protocol-agnostic operation.

Is there a common thread among existing Admission-Control/Flow-Control Strategies?

The premise of this submission is that there **is** a common thread among known Admission Control/Flow Control strategies. All approaches depend on end-to-end Bandwidth-On-Demand allocations between a multiplicity of data sources and their respective data sinks.

Scheduling bandwidth between edge nodes of a BWA network is the common thread. How to do it in the BWA environment agnostic to user and network protocols is the challenge.

How can we achieve a Bandwidth-On-Demand MAC layer objective?

The best outcome of 802.16 deliberations will be a simple standard that works with all 802 and other Layer 2/3 network protocols. Based on the work of a number of vendors who have been willing to move forward without a standard in place, there appears to be emerging a common approach that can be standardized to produce the desired result.

Vendors are incorporating “minislots” within a Time Division Multiplex (TDM) format as a critical part of their systems. These minislots have two purposes: One is to allow for multiple bursts of contiguous minislots to support variable length packets.

A second purpose results from the assumption that a base station is responsible for scheduling transmissions in both the *broadcast* (base-station to remote-station) direction and for the *return path* from each remote station to the base station. A relatively small amount return path Bandwidth (some number of minislots per TDM frame) is set aside, enabling remote stations to *request* assignment by the Base Station bandwidth (minislots) for them to send return path traffic. Since it is usually very bandwidth inefficient to assign fixed bandwidth to return path transmissions for long periods of time, a key role for a base station is to *grant* minislots to a remote station in response to a remote station’s requests.

A growing consensus is to implement a *request/grant* protocol that causes a base station to grant bandwidth chunks to any remote station in response to the remote station’s requests.

Simplified statement of BWA network scheduler requirements

One way of looking at Bandwidth-On-Demand operation that satisfies BWA network needs is to realize that we are attempting (metaphorically or actually) to interconnect backplanes at a distance. An analogous situation is that of interconnecting backplanes in stacked switches, multiplexers, routers and similar units. The fundamental problem of interconnecting backplanes either locally or at a distance is illustrated simplistically in Figure 1. Data from an input port at one end of the link must be scheduled to be delivered over the link to an output port at the other end of the link. For input port data, a local scheduler sends information to the remote scheduler as to the time(s) of occurrence of the input port data on the link so that the remote scheduler can deliver the data to the proper output port at the distant end.

Scheduler synchronization is a mandatory requirement of networks involving interconnected backplanes. It is a mandatory requirement to implement Admission Control/Flow Control networking.

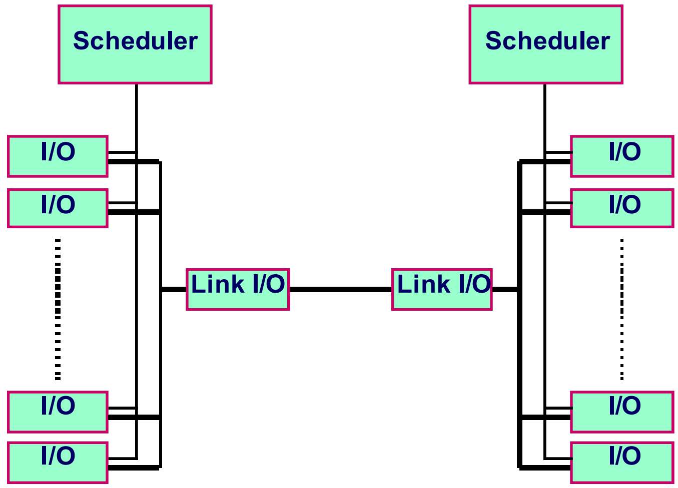

FIGURE 1 SIMPLIFIED BLOCK DIAGRAM OF SYNCHRONIZED SCHEDULERS

A fundamental issue in designing competitive BWA systems is to devise methods that assure the receive side of each local remote station scheduler is synchronized with the send side of its base station scheduler and *vice versa*. [1]

More specific requirements for BWA network schedulers

One of the assumptions of this submission is that a base station must be responsible for scheduling bandwidth for not only its own transmissions to remote stations, but also for return path transmissions. Each scheduler has two parts, a send part and a receive part, so that there are always two synchronized scheduler send/receive parts for every duplex link. Each pair of send/receive parts supports one direction of information flow. This means that if there are ***N*** remote stations, the base station must manage bandwidth scheduling for **2*N*** scheduler pairs [2].

For base-station to remote-stations broadcast links, the base station schedules all embedded-channel traffic from its input ports to meet the system's established QoS specifications. The base station informs the remote stations of base station's forthcoming transmissions. For return paths, the remote stations inform the base station of its needs, whereupon the base station transmits return path scheduling information to the remote stations, taking into account the relative priorities of the requests from all remote stations within a BWA cell.

---

[1] It should be noted that the word "synchronized" relates to the state of each scheduler with respect to the traffic on the link. It does *not* presume that the scheduler and links are slaved to a system or network master clock. One might describe the scheduler synchronization requirement as one that assures *coherence* between the schedulers and the information on the link. This distinction may be important in some circumstances where the link may or may not be slaved to a master clock. For example, supposed that the link is a standard Ethernet connection. The schedulers must be synchronized with the Ethernet frames that may or may not be slaved to a master clock. Simpler systems often obtain from clock-synchronous operation of shared media networks as this approach obviates the necessity for complex software to deal with time stamped packets or ATM cells.

[2] There are cases where scheduling may need occur in only one direction so that only ***N*** schedulers are required. For example, Video On Demand and other "On Demand" broadcast services need share bandwidth only in the broadcast direction. The return path bandwidth can be small and permanently assigned either full time or on a contention basis. The discussion herein focuses on bandwidth scheduling in two directions, but could be modified with obvious changes.

In most instances, it would be wasteful for the base station to assign full-time return path bandwidth. Request/grant protocols are an obvious means by which return path transmissions can be made efficient and satisfy global BWA cell needs.

Proposed Protocol Stack Model

A request/grant protocol need not depend on any specific MAC layer protocol that now exists. This suggests that there is the prospect of defining a generic MAC *Bandwidth Allocation Sublayer* that can be incorporated within a multiplicity of MAC protocols. Based on the suggested protocol stack models of several authors of 802.16 submissions [1], [2], [3], [4], the architectural model shown in Figure 2 appears appropriate.

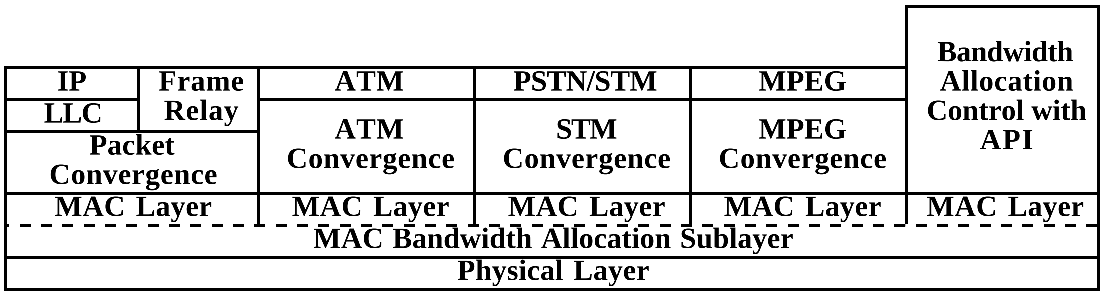

FIGURE 2 PROTOCOL STACK LAYERING INCLUDING A MAC BANDWIDTH ALLOCATION SUBLAYER

Figure 2 shows interface support for IP, Frame Relay, ATM, various possible PSTN and/or other Synchronous Transfer Mode (STM) services, and MPEG. In addition to providing standard network service interfaces, the model includes an interface for Bandwidth Allocation Control. This is not a network protocol interface, but is a control interface by which bandwidth can be managed over signaling channels of a network link. As shown in the figure, it includes an Application Program Interface (API) that becomes a part of the 802.16 standard. The dotted line MAC Bandwidth Allocation Sublayer is the standard interface that supports all services.

Applications supported by the model

The model works best when applied to information *flows* as opposed to individual packets. A flow can be one of two things. For Constant Bit Rate (CBR) and Variable Bit Rate (VBR) applications (such as voice or video), it is a continuous stream of (clocked) information that is transmitted from an information source to a corresponding sink. For packetized data traffic, it is a flow of (one or more) packets that belong to a single message or data file.

It is likely that any future protocol will fit within the dichotomy of burst-flow or continuous-flow traffic. Even if packet-based protocols become the dominant network interface (that most observers believe is likely), the flow models based on the ATM definitions (CBR, VBR, ABR and UBR, with appropriate sub-classifications) appear sound.

The proposed approach allows assignment of bandwidth for each flow. Predetermined priorities for each flow class (and subclass) determine the relative amount of assigned bandwidth among the classes.

The model is the equivalent of turning on a physical circuit for each flow. However, it is very **un**like a conventional circuit switch. The model allows the bandwidth of any circuit to be varied dynamically. In other words, the bandwidth of any circuit, once established, is not static. It is varied according to the intensity of the flow. This bandwidth scheduling technique satisfies QoS needs at both the flow level and the class level.

The proposed bandwidth scheduling model is simple but powerful and is uniquely applicable to BWA networks.

Mechanisms for dynamically adjusting bandwidth

A preferred mechanism for dynamically adjusting the bandwidth for each flow is to implement a means by which the number of minislots per flow in a TDM frame can be changed based on each flow's instantaneous traffic needs and their service level guarantees.

This approach is the equivalent of connecting a clocked circuit to each flow where the clocking rate can be fixed, varied or quickly turned on and off. As a result, STM-based, cell-based and packet-based traffic all can be supported by a single simple-to-understand mechanism — assigning flows a variable number of minislots per TDM frame minislots.

One of the most important features of the approach is its ability to support robust Admission Control/Flow Control capabilities as a part of QoS strategies. This feature is illustrated in a simulation example in the next section.

Simulation example for ABR packet traffic

Figure 3 shows an example of the ability of the proposed model to handle packet flows. The simulation assumes that all required buffering takes place at traffic entry nodes using the proposed bandwidth control mechanism. In other words, there is no need for large statistical-multiplexing buffers internal to a BWA network. All buffering occurs only at network edges.

Each of the 20 horizontal lines in the figure represents one or more flows of information from an information source to an information sink. The simulation focuses on two service classes to which some aggregate bandwidth must be shared. The objective of the system is to make the latency QoS for the two classes equal. The first class (Class 1) consists of 500-byte packets and the second class (Class 2) consists of 1500-byte packets. The latency of the packets in the two classes is to be the same even if the total aggregate bandwidth is restricted. Packet flows belonging to Class 1 are shown on the first 9 lines of the figure while the remaining 11 lines belong to Class 2. The small triangles show when an incoming packet is placed in a class buffer. The distance between the triangle and the following packet exiting a buffer shows the delay encountered between the packet arrival into the buffer and its subsequent transmission on the outbound link. The distance between a triangle and the start of is proportional to the amount of time the packet remains in the buffer before starting transmission.

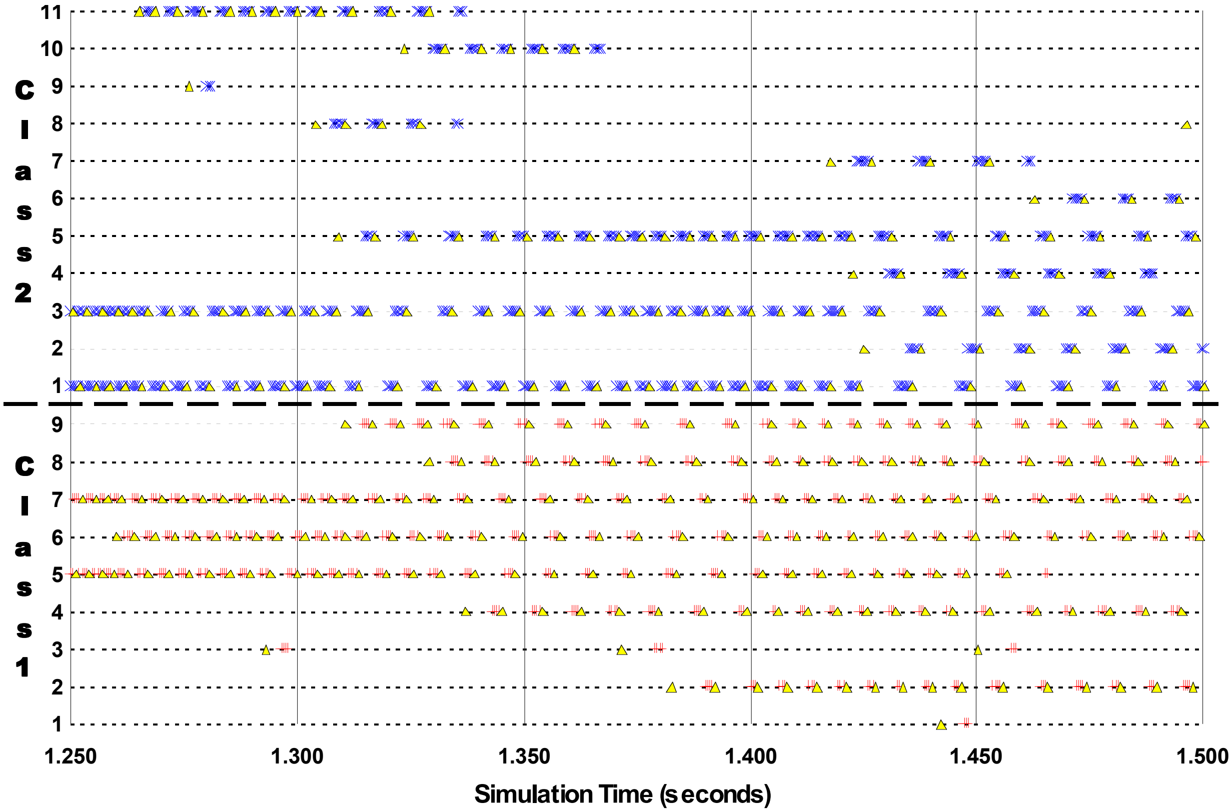

FIGURE 3 SIMULATION EXAMPLE OF ADMISSION CONTROL/FLOW CONTROL OF PACKET TRAFFIC

During the 250 msec period shown, the traffic is quite “bursty”; the number of packets increases dramatically toward the end of the period compared to the beginning. For example, at 1.250 seconds there are just two packets from Class 1 and two from Class 2 present within the buffers. At 1.450 seconds, there are eight Class 1 packets and six Class 2 packets for a total of 14 packets in the buffers. The instantaneous traffic load is 3.25 times greater at 1.450 seconds than it is at 1.250 seconds In other words, the traffic intensity has increased by a factor of 3.25 in a period of only 200 msecs. The time a packet stays in a buffer is clearly greater at the right of the figure than it is on the left. But, it is also clear that the latency for each traffic class is nearly the same — the intended QoS result.

An important aspect of the mechanism illustrated in the figure is the built-in “back-pressure” capability the approach provides. This effect is seen at the right in the figure where the total traffic has increased substantially with no concomitant increase in aggregate bandwidth available. Obviously, the mechanism could be used to adjust the aggregate bandwidth of the combined flows (if the priority of the classes compared to other classes permits) as well as adjusting *pro rata* the flows for the two traffic classes shown.

The figure illustrates that assigning bandwidth at the MAC layer is a simple way to implement Admission Control/Flow Control among service classes. Although, for simplicity, the two service classes are assumed to be at the same (latency) priority level, this restriction is **not** a requirement of the approach.

The traffic presented to the mechanism originates from a diverse set of sources (representing a combination of Web-based, E-Mail-based and File-based traffic) that are connected to the system over connections of vastly differing data rates — all the way from 14.4 Kbps modems to 100 Mbps Ethernet. This diversity can be seen by the differing packet arrival rates into the system. [3]

The approach can be used in many ways to schedule bandwidth. The simulation assumes that the relative bandwidth allocated to each service class is adjusted at a periodic rate of 20 times per second, that is, at fixed 50 msec periods. The vertical lines show 50 msec intervals where bandwidths are readjusted. The total available aggregate bandwidth is assumed to be 10 Mbps which, on the left, results in almost no delay. On the right of the figure, the offered load produces delays of the order of 10 to 15 msecs.

Clearly there is nothing magic about 50 msec intervals. Bandwidth update intervals can be tailored to specific traffic needs and can be as small as one msec or less (a very unusual case) or they may be much larger. A "much larger" example is non-real-time VBR traffic. This case requires periodic bandwidth update intervals that may be many seconds long. Averaging VBR bandwidth over long periods minimizes peak-to-average bandwidth requirements and results in very small network signaling requirements.

By working with flows instead of packets, the amount of signaling overhead required is much reduced while producing desirable Admission Control/Flow Control capabilities. The approach is facilitated by the fact that bandwidth management for all links within a BWA cell is centralized at the base station of the cell.

This simulation example shows that ABR and UBR packet traffic can easily be carried over the equivalent of synchronized variable-bandwidth circuits. CBR and VBR (both real-time and non-real-time) traffic is obviously easy to support as well. Carrying all traffic can be accomplished without large buffers internal to a network. The approach also facilitates implementation of very robust QoS with simple software.

It is much more difficult to carry CBR and VBR traffic over conventional packet- or cell-based networks that depend on large internal network buffers for bandwidth and priority management. Compared to the proposed approach, conventional packet-based and cell-based operation generally results in less efficient use of available link bandwidth. Robust QoS capability must include provision of low latency guarantees, low cell and packet delay variation guarantees, and low cell and packet loss rate guarantees. With packet- and cell-based networks, these guarantees require complex software that must be installed (and kept updated) at all network nodes through which traffic flows. Even if such complex software can be produced, history shows that this usually equates to low network reliability and high maintenance costs. By requiring matched (updated) software only at network edges, the proposed approach simplifies logistical problems, improves reliability, and reduces operational costs.

Possible Modifications to Proposed MAC Layer Architecture

For many protocols, layers 2 and 3 are not as distinct as in the case of 802-based IP networks. In some cases, modifications as shown in Figure 4 can be more appropriate where convergence layers may be tied directly to the proposed MAC Bandwidth Allocation Sublayer. In this case, it is appropriate to replace the dotted line of Figure 1 by a solid line. This is not to suggest that the sublayer become a new 802 MAC layer standard. Existing 802 standards have needed functionality that should not be replicated as a part of the sublayer. The new MAC Bandwidth Allocation Sublayer is intended only to enhance the functionality of existing standards and not replace them.

[3] Aspects of the model other than input bandwidth of the sources affect traffic arrival rates. For example, the model includes a mix of traffic types such as WWW, SMTP, FTP and UDP. The arrival times of flows from these sources are assumed to follow conform to a Poisson distribution. Discussing these and other issues is beyond the scope of this presentation.

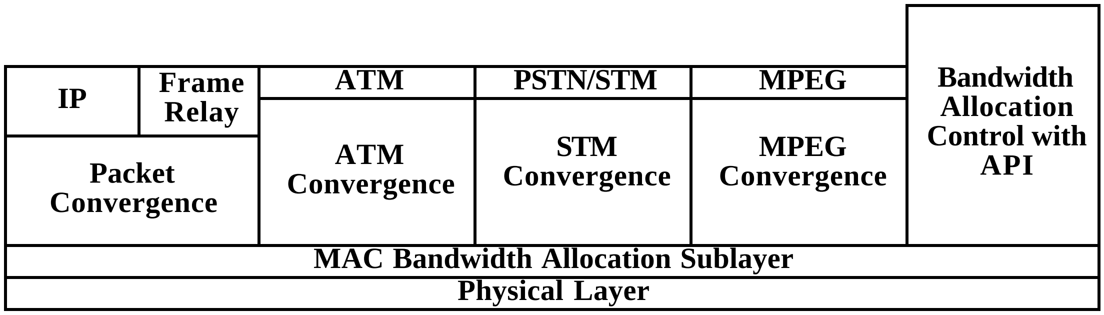

FIGURE 4 POSSIBLE MODIFICATION OF MAC LAYER ARCHITECTURE

For example, suppose that it is desired to support PSTN/STM and Internet traffic on a local area network. The protocol stacks in this case might something look like Figure 5.

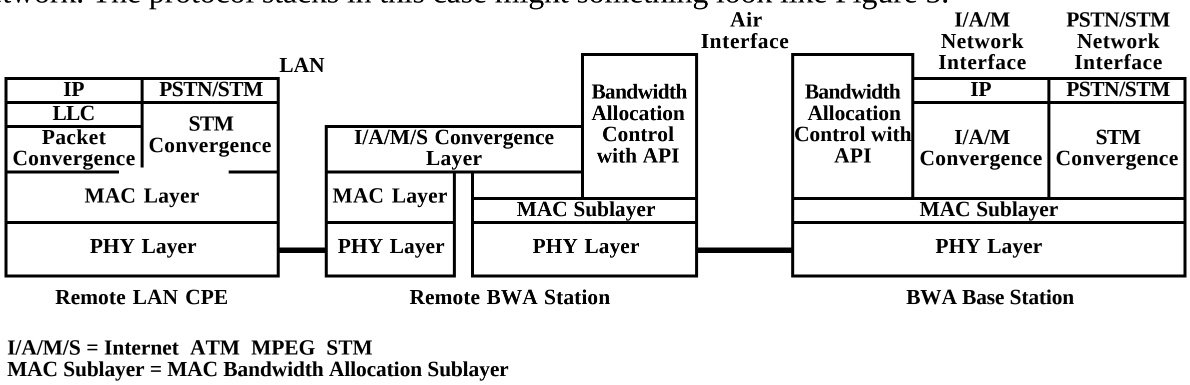

FIGURE 5 EXAMPLE OF PROPOSED ARCHITECTURE WITH LAN EXTENSION

In the figure, a Remote LAN CPE uses an existing MAC layer standard such as 802.3. It is connected by a LAN (wired or wireless) connection to a Remote BWA Station. Within the BWA station, there exists a convergence layer that translates the encapsulation of user traffic to/from the standard MAC layer protocol into two other protocols: either 1) IP, ATM or MPEG for the Remote LAN CPE traffic that is IP based, and 2) STM for the PSTN/STM traffic. The translated traffic interfaces directly with the proposed MAC Bandwidth Allocation Sublayer and is connected by the BWA Air Interface to the BWA Base Station.

How The Proposed Approach Can Stimulate BWA Market Growth

There are two ways that the proposed approach can help stimulate market growth. The first is enhanced efficacy of BWA systems as described above.

The second is market expansion. Specifying a single MAC Bandwidth Allocation Sublayer that supports multiple top-layer MAC and other Layer 2/3 protocols expands the number of vendors who can participate in the BWA market. It opens opportunities for innovation in devising edge-node buffering techniques (such as illustrated in Figure 5) that result in superior Admission Control, Flow Control and QoS capabilities.

The approach creates an important opportunity for small as well as large vendors to introduce new value-added options for both users and network operators. This result would be difficult to imagine if all options for providing BWA networks were limited to a few large system vendors alone. Without defining a standard MAC Bandwidth Allocation Sublayer, each systems vendor would likely invent his own MAC protocol adaptations that would not form the basis for general interoperability so important for stimulating market growth. Limiting interoperability to only existing network and user protocols would limit the robustness of the BWA market.

Conclusion

Defining a new MAC Bandwidth Allocation Sublayer as a part of 802.16 standards leads to a number of benefits. It can result in more efficient use of bandwidth than can more complex alternatives. It also provides a simple mechanism for implementing Admission Control and Flow Control capabilities, both of which are critical to assuring robust QoS operation.

Defining a MAC sublayer as proposed results in the ability to support any existing MAC and higher layer protocol, assuring that the resulting standard will have an important place in the market.

By defining this new sublayer, the overall complexity of BWA systems is reduced. The bandwidth efficiency of systems is enhanced.

The new sublayer creates the opportunity for many vendors to participate in the BWA market so that they can produce creative solutions for supporting standard and emerging protocols. By proceeding in this manner, 802.16 can assure that a robust BWA market emerges.

Finally, the proposed approach should facilitate early completion of a workable 802.16 standard. Focusing the attention of relevant working groups on the proposed approach will reduce debate on many contentious issues.

It is unlikely that early agreement can be reached at the current stage of development of all aspects of MAC and PHY layer alternatives. This is probably a good thing as it leaves room for vendors to prove out different air interface approaches that can be selectively approved as future parts of the 802.16 standard. Focusing on a MAC Bandwidth Allocation Sublayer that can support any of a number of alternatives promises to stimulate the BWA market with low risk to all market participants.

References

[1] James Mollenauer, “Functional Requirements for Broadband Access Networks”, 802.16sc99-5

[2] James Mollenauer, “Functional Requirements for the 802.16 Standard”, 802.16sc99-7

[3] Willie Lu, Charles Bry, Mohan Maghera, “System Reference Model and Protocol Stack for Broadband Wireless Access”, 802.16SC99-15

[4] Imed Frigui, “Services and Performance Requirements for Broadband Fixed Wireless Access”, 802.16sc99-23