Project	**IEEE 802.16 Broadband Wireless Access Working Group <http://ieee802.org/16>**
Title	**ARQ Proposal for TG3/TG4 MAC Working Document**
Date Submitted	**2001-07-02**
Source(s)	Robert R Nelson Raze Technologies 2540 E Plano Pkwy Suite 188 Plano, TX 75074 Voice: (972) 516-1283 Fax: (972) 578-9081 bnelson@razetechnologies.com
Re:	**Task Group Review of IEEE 802.16ab-01/01** **IEEE 802.16 Task Groups 3 and 4** **PHY and MAC Layers for IEEE P802.16a/P802.16b**
Abstract	An ARQ algorithm is presented that builds on the ARQ contributions made to date but provides alternatives to certain undesirable issues regarding the current proposal. Although some detail is omitted, an essentially complete ARQ solution is presented.
Purpose	For consideration for inclusion in the TG3/4 MAC working document.
Notice	
Release	
Patent Policy and Procedures	

ARQ Proposal for 802.16 TG3/4

Bob Nelson
Raze Technologies Inc

Introduction

The ARQ material included in the TG3/4 MAC working document to date prescribes a technique for data blocking and sequence number assignment, and also provides a variety of data structures with the intent of supporting multiple ARQ algorithms. The material also increases the number of situations where a subscriber may "bandwidth steal" from bandwidth grants to outbound service connections, sending ARQ control information to the base station instead of connection data as intended by the base station scheduler.

The purpose of this document is to comment on these issues, and finally, present an ARQ operating scheme that builds on material appearing in the working document and is believed to be appropriate for the 802.16 operating environment.

Multiple Algorithms

An attribute assumed by anyone purchasing "standard-compliant" equipment is that the equipment being purchased will be interoperable with similar pieces of equipment from other manufacturers that conforms to the same standard. The fact that 802.16 intends to provide an umbrella for multiple incompatible algorithms, while requiring that only one of the choices be implemented, diminishes its stature as a "standard".

One of the goals of 802.16 has been to craft a single MAC that can span the collection of physical layer definitions created for the targeted operating environments. The obvious benefits of doing so are minimizing the complexity (and sheer size of the standard), realizing the synergy of as many people as possible concentrating on the subject at hand, and eliminating the duplication of effort that would be involved in defining and validating completely disjoint but similar MAC definitions. The inclusion of multiple algorithms contradicts this goal.

One mark of a good system design is finding that complexity has been reduced whenever and wherever possible. The question to be answered by the committee is whether or not the added complexity associated with support for multiple algorithms in the standard documentation and required in all implementations (negotiation at connection establishment, etc.) is technically justified.

Bandwidth Stealing

The current draft 802.16 standard allows a subscriber to "steal" a portion of a bandwidth grant intended for transmission of service connection data and use it instead to send one or more bandwidth requests (as separate messages) to the base station in the granted bandwidth interval. The TG3/4 draft proposal expands on this capability by allowing ARQ information for inbound connections to be encoded as an ARQ sub-header or packed message partial payload and be included for transmission as part of any outbound message that is available.

This enhancement violates what seems to be a prime design principle of the current MAC draft, that all data within a message is related to the CID appearing in the message header, and as a result, routing of received messages can be performed based on the CID value. With the enhancement in effect, message header CID routing is no longer possible. Instead, the entity responsible for distributing inbound data to the proper handler is required to parse the contents of each received message and take the appropriate action based on the message content found.

Embracing this strategy forces implementations down one of two undesirable paths. Either received data handling must be centralized with each message parsed by a single omnipotent module, or knowledge of all the intricacies of message payload construction must be distributed to the appropriate system components (which must then be updated if/when new options are added).

Operating Constraints

There has been much discussion of how TG3/4-based equipment should function and the environment in which it must operate. For purposes of ARQ, the discussion can be reduced to the following salient points.

- The limiting factor for TG3/4-based communication is the carrying capacity of the RF link.
- Due to the effects of multi-path and non-line-of-sight operation, a high bit error rate is expected. The error rate will be sufficiently high in fact, that frequent packet losses are expected despite the best efforts of signal processing,.
- Due to the computational complexity of the PHY algorithms and the desire to minimize cost of the equipment, the scheduler will be required to schedule multiple frames into the future so that maps can be forwarded to subscribers in time for them to stage the data intended for transmission. As a result, an active connection may have many messages in the “pipeline” at any particular time.

Algorithm Selection

In consideration of these issues, the ARQ algorithm of choice is one that does all it can to optimize bandwidth usage and also can effectively deal with a transmission pipeline spanning multiple frames. Of the two choices available, go-back-n or selective-repeat, only selective-repeat satisfies these requirements.

Consider the following simplified scenario:

- The scheduler executes once per TDD frame at the start of that frame. (As a result any information gained during a particular frame is available to the scheduler in the following frame.)
- The scheduler is operating three frames into the future (i.e. at time N, the contents of the frame transmitted at N+3 is determined).
- A downlink (base-to-subscriber) service connection is moderately active and has priority such that the scheduler can only include one message for the connection in each downlink frame.

For this scenario the sequence of events might be as follows under go-back-n:

Frame N

Base Station: The connection message scheduled for N is transmitted. The contents for frame N+3 are scheduled.

Subscriber: A transmission error occurs and no message is received by the subscriber. It takes no action.

Frame N+1

Base Station: The connection message scheduled for N+1 is transmitted. The contents for frame N+4 are scheduled.

Subscriber: The message sent by the base station is received and is recognized as being out of sequence. The message is discarded and the NACK is queued for transmission to the transmitter.

Frame N+2

Base Station: The connection message scheduled for N+2 is transmitted. The contents of frame N+5 are scheduled. The subscriber’s NACK is received. The frame N message is marked for retransmission.

Subscriber: The message sent by the base station is received and is recognized as being out of sequence. The message is discarded. The NACK is sent to the base station.

Frame N+3

Base Station: The connection message scheduled for N+3 is transmitted. The contents of frame N+6 are scheduled, the lost message from frame N is included.

Subscriber: The message sent by the base station is received and is recognized as being out of sequence. The message is discarded.

Frame N+4

Base Station: The connection message scheduled for N+4 is transmitted. The contents of frame N+7 are scheduled, the discarded message from frame N+1 is included.

Subscriber: The message sent by the base station is received and is recognized as being out of sequence. The message is discarded.

Frame N+5

Base Station: The connection message scheduled for N+5 is transmitted. The contents of frame N+8 are scheduled, the discarded message from frame N+2 is included.

Subscriber: The message sent by the base station is received and is recognized as being out of sequence. The message is discarded.

Frame N+6

Base Station: The connection message scheduled for N+6 (the original lost message) is transmitted. The contents of frame N+8 are scheduled, the discarded message from frame N+3 is included.

Subscriber: The message sent by the base station is received and is recognized as being the next message expected. The message is handled. An ACK is queued for transmission to the base station.

The system has now recovered from the error and normal data transfer resumes. As result of the error, the data successfully transmitted and received in frames N+1 through N+5 are discarded and must be retransmitted in order no earlier than frame N+6. Net result, data transfer to the subscriber is delayed by 6 frames and bandwidth in 5 frames is wasted.

The sequence of events under selective-repeat (as presented in the working document) for the same scenario are as follows:

Frame N

Base Station: The connection message scheduled for N is transmitted. The contents for frame N+3 are scheduled.

Subscriber: A transmission error occurs and no message is received by the subscriber It takes no action.

Frame N+1

Base Station: The connection message scheduled for N+1 is transmitted. The contents for frame N+4 are scheduled.

Subscriber: The message sent by the base station is received and is recognized as being out of sequence. The message is buffered, a NACK message is queued for the missing frame N message, and an ACK is queued for the frame N+1 message which was successfully received..

Frame N+2

Base Station: The connection message scheduled for N+2 is transmitted. The contents of frame N+5 are scheduled. The subscriber's NACK and ACK messages are received. The frame N message is marked for retransmission.

Subscriber: The message sent by the base station is received and is recognized as being out of sequence. The message is buffered. An ACK is queued for the received message.

Frame N+3

Base Station: The connection message scheduled for N+3 is transmitted. The contents of frame N+6 are scheduled, the lost message from frame N is included. The ACK for the message in frame N+2 is received and handled.

Subscriber: The message sent by the base station is received and is recognized as being out of sequence. The message is buffered. An ACK is queued for the received message.

Frame N+4

Base Station: The connection message scheduled for N+4 is transmitted. The contents of frame N+7 are scheduled including the next message in queue for the service connection. The ACK for the message in frame N+3 is received and handled.

Subscriber: The message sent by the base station is received and is recognized as being out of sequence. The message is buffered. An ACK is queued for the received message.

Frame N+5

Base Station: The connection message scheduled for N+5 is transmitted. The contents of frame N+8 are scheduled, including the next message in queue for the service connection. The ACK for the message in frame N+4 is received and handled.

Frame N+6

Base Station: The connection message scheduled for N+6 (the original lost message) is transmitted. The contents of frame N+8 are scheduled, including the next message in queue for the service connection.

Subscriber: The message sent by the base station is received and is recognized as being the next message expected. The message is handled along with the all the messages buffered in expectation of the lost message's eventual arrival.

The system has now recovered from the error and data transfer resumes normally. As result of the error, the only bandwidth wasted was that of the frame in which the original error occurred. Net result, data transfer to the subscriber is delayed by 6 frames, but in terms of calculated data rate (data/frame) only 1 frame delay is visible.

While the example scenarios are idealized, they show a drastic difference in the performance characteristics of the algorithms. Due to a single bit error corrupting one packet, the receiver at the end of the go-back-n connection experiences complete data loss for the interval from the occurrence of the error until recovery is achieved. Also, once link recovery is achieved, data resumes trickling in at one message per frame with arrival of all future packets delayed by the entire length of time it took for link recovery to occur.

In the case of the selective-repeat receiver, because of the single bit error, the receiver application sees no data arrive for the time it takes for the link to recover, but when link recovery is completed, all the data destined for the application arrives one frame late. As a result, subsequent transmissions destined for the receiver arrive only one frame late.

Taking the comparison one step further, consider the effect of an error corrupting every sixth frame of the example data stream. Under go-back-n, the unlucky packet repeatedly affected by the error occurrence is eventually purged due to retry failure. During the entire period of time until this occurs, no data is delivered to the receiver application. The link is effectively dead although most packets are successfully being delivered to the receiver's protocol stack.

In the same situation for selective-repeat, data is buffered until the retry limit is exhausted. Once this has occurred, all the buffered data held by the receiver protocol stack is immediately provided to the receiver application. Again, although the receiver experiences a delay until data is actually delivered, the data and bandwidth lost is limited to only the messages actually experiencing an error.

Proposed ARQ Solution

The proposed ARQ solution is based on the selective-repeat with CACK algorithm and consists of the following components tailored for 802.16 operation:

- Blocking and serialization algorithms presented in the MAC working document.
- ARQ Sub Header short or long (10 bit sequence numbers instead of 11) presented in the MAC working document.
- A MAC management message which is used to update the "received" status of all ARQ managed connections associated with the same data link.
- The addition of a new UIUC value that identifies bandwidth the scheduler has set aside for transmission of uplink ARQ information.
- Retransmission timer managed by the transmitter.
- Retry exhaustion recovery and reset logic.

ARQ ACK/NACK MAC Management Message

The MAC management message for ARQ activities is intended to carry information for all ARQ-related connections associated with a single data link (aka subscriber). A message sent from subscriber to base station provides received/not received information for all downlink ARQ managed connections. A message sent from the base station to subscriber, holds the corresponding information on uplink transmissions from the subscriber.

Generic Mac Header	
Management Message Type = ??	Last Received Time (Frame Count)
Last Received Time (Mini-Slot Offset)	

A message consists of the appropriate MAC header followed by a byte containing the management message identifier assigned to this message type and a 32-bit quantity specifying the time of the last message reception included in the message's map data. The MAC header CID field always contains the BASIC CID for the associated data link and direction of message transmission. The message payload is protected by CRC.

The remainder of the message contains information elements that are repeated for each connection for which information is provided.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Connection Id															
Last	Block Serial Number							M0	M1	M2	M3	M4	M5	M6	M7
Last	Block Serial Number							M0	M1	M2	M3	M4	M5	M6	M7
Last	Block Serial Number							M0	M1	M2	M3	M4	M5	M6	M7

Connection Id – Holds the id of the connection associated with the following map records.

Last – The bit value is set to one if the map element is the last one for the connection. For all other map records, the value is set to zero.

Block Serial Number – Provides the base value for the associated bit map byte. The field holds a value between 0 and 127, inclusive.

Mx – Each of the Mx bit values indicates whether or not a block with the serial number corresponding to that bit field { (Block Serial Number * 8) + x } has been received. A value of one indicates the message with the corresponding sequence number has been received. All bit values for sequence numbers outside of the transmission window should be set to zero.

By convention, the first map record always provides the block number of the cumulative ACK value determined by the receiver. The number of map records provided in the message (beyond the CACK record) is left to the discretion of the originator. The most information that will ever be included covers the interval from the CACK to the end of the current transmission window.

For subscribers, the message is only transmitted in response to polls for ARQ information appearing in the uplink map (see next section). In this case, subject to the size of the granted bandwidth interval, the subscriber is expected to provide as much information as possible for each of its inbound ARQ connections. When building the message, breadth of information across all connections is preferred to depth of information for a few connections.

ARQ Grant UIUC Element

Rather than relying on bandwidth stealing or piggy-back mechanisms which tend to disrupt the base station's scheduling activities (as well as other problems already mentioned) bandwidth grants for subscriber-originated ARQ information are explicitly provided by the base station. The grant mechanism is consistent with the mechanisms already in place for station ranging and bandwidth request polling.

Specifically, reserved UIUC code 12, is assigned as the ARQ UIUC. When a subscriber receives a UL_MAP with an entry for this UIUC and its uplink BASIC CID value, the subscriber is expected to send an ARQ management message (as described in the last section) to the base station in the specified bandwidth interval.

Retransmission Timer

Retransmission timer management is maintained by the originating side of each ARQ connection. For each individual message transmitted, a timer is started when the message is sent. If no received notification arrives from the downstream end of the connection when the timer has expired, the message is retransmitted. This process continues until a received notification has been received or the retry counter is exhausted.

Retry Counter

When the retry counter is exhausted, the offending packet is purged along with any other fragments of the same message that are still in queue for transmission or retransmission. A Reset message (format TBD) is sent to the downstream end of the connection instructing it to clean up its end of the connection by dumping all references to the purged packet and its sibling fragments.

The downstream handler acknowledges the Reset message by sending an indication that all packets of the dumped message have been received in the next granted ARQ opportunity. Finally, any messages blocked from delivery by the purged message are released and delivery to the host application is allowed to occur.

Conclusion

Certainly, any of the methods currently proposed in the MAC working document will provide a higher level of data integrity than having no ARQ support at all. However, each algorithm has its own set of strengths and weaknesses in different operating environments, and while admittedly simplistic, the scenario discussed here offers compelling support for the argument that selective-repeat is the best choice for 802.16.

The variation of selective-repeat presented in this paper provides an ARQ solution that works within the confines of the current MAC draft architecture rather than forcing radical changes in philosophy. Furthermore, the methods presented are not simply a mental exercise. The described techniques for received message notification and retransmission have been implemented and work well.