Project	**IEEE 802.16 Broadband Wireless Access Working Group** <http://ieee802.org/16>
Title	**Feedback and scheduling strategies in 802.16m Relays**
Date Submitted	**2008-09-12**
Source(s)	Eldad Zeira eldad.zeira@interdigital.com Ron Murias InterDigital Communications, LLC
Re:	SDD Session 56 Cleanup, in response to the call for PHY details: "Any parts of Section 11 (PHY) that are incomplete, inconsistent, empty, TBD, or FFS."
Abstract	The contribution considers the pros and cons of centralized and distributed scheduling and associated feedback from a HARQ latency point of view. It argues that both offers advantages under different operating conditions and proposes that 802.16m support a dynamic in-band switchover between the two modes.
Purpose	To be discussed and agreed by TGm
Notice	
Release	
Patent Policy	

Feedback and scheduling strategies in 802.16m Relays

Eldad Zeira

Ron Murias

InterDigital Communications LLC

Introduction

Introduction of relaying into a cellular system opens up new operational possibilities for optimization of system operation as well as new challenges associated with efficient operation of such systems. A major feature which needs addressing is the operation of the Hybrid ARQ (HARQ) protocol – in particular the management of scheduling and feedback across the transmitter of origin (BS or MS) and the relays. The appropriate design choices in these cases depend greatly on which relaying method is adopted. We illustrate the challenges and trade-offs by considering a simple example: downlink transmission with only a single relay: BS → RS → MS.

A traditional relaying method (which we shall call a Non-cooperative 2-hop) in this case is for the BS to send information to the RS. Once the RS has received this information it sends it to MS without any further assistance from the BS. This method is clearly appropriate for coverage extension either beyond the cell edge, or into "black holes" – areas within the cell (e.g. inside buildings) where the BS signal does not penetrate.

At a high level, we have 2 choices as to HARQ scheduling and feedback for Non-cooperative 2-hop. We can allow the BS to retain control over scheduling ("centralized scheduling"). However, this has certain downsides: the BS is required to use signaling overhead to schedule the RS transmission after the RS has the data; the ACK feedback (which may be available only at the RS) must be relayed to the BS, which results both in overhead and delays in data delivery. An alternative approach is to allow the RS to schedule data transmission to the MS autonomously once it acknowledges receipt of data to the BS ("distributed scheduling"). This avoids the 2 issues mentioned above, and in the case of Non-cooperative 2-hop it may be argued that this is the more efficient method of operation.

On the other hand, it is well known that relays can provide significant QoS benefit to MSs that are within "range" of the BS but are experiencing degraded performance – e.g. MSs at cell edge in a cell where interference is the primary limitation on performance and the BS-RS link is comparable to the BS-MS link. (that fact has been demonstrated in our recent contributions [1]). To do so, the RS must cooperate with the BS in transmitting to the MS. For example, in our simple example, once the RS has the data, BS and RS transmit cooperatively, forming a distributed transmit antenna array (open or closed loop). We shall refer to such a scheme as Cooperative 2-hop (or Coop 2-hop). Clearly, distributed scheduling is not appropriate for Coop 2-hop – the coordination required can only be achieved with centralized scheduling.

Currently, existing system design, such as, e.g. the 802.16j specification, support both a distributed and a centralized mode of operation. However, the switch between the two is done using out-of-band control signaling. In many scenarios the need for such switching may occur rather frequently as MSs move, therefore the process is cumbersome, slow and carries significant overhead. .

A second consideration is feedback. Each hop of feedback introduces additional delay into the HARQ operation and this should be minimized. In a non-cooperative 2-hop scheme this is done by reverting to distributed scheduling. Under the Coop 2-hop this is not possible (the BS must get the feedback). However, in this case there is an underlying assumption of a BS-MS link that is sufficiently good to carry some data – and therefore good enough for the small amount of feedback required. This could be utilized to allow the MS to broadcast feedback simultaneously to RS and BS thus reducing the delay penalty associated with relaying. As with scheduling, a switch between the two modes currently requires out of band signaling, and is slow, cumbersome and expensive.

In this contribution we propose a mechanism for a fast, low-overhead in-band signaling for "seamlessly" optimizing the downlink connection to the MS through one or more RSs. Moreover, our scheme is transparent to the mobile. In our simple (BS-RS-MS) example, this provides a seamless transition between 3 modes: distributed scheduling with 2-hop feedback; centralized scheduling with 2-hop feedback and centralized scheduling with direct feedback. In a more general case, we support creation of seamless (from the perspective of the MS) hybrid modes which allow flexible optimization of the connection. This seamless scheme could be enabled by the BS when the conditions for it are met.

Note also that in [1] we have demonstrated additional benefits of BS transmitting to both RS and MS at the same time, termed multicast stage. It should be noted that this mode of operation could also be achieved with the proposed scheme without additional standards changes and thus become a scheduling decision.

Terminology

We believe the following definitions to be consistent with agreed-upon 802.16m terminology, however for clarity will state it here.

An MS is said to be *connected* to a BS if there exists a logical link between the two – the physical link may be direct, through relays or both.

An MS is said to be *associated* to a BS and/or RS if there exists a sufficient good link between the two for exchange of lower layer control information – i.e. HARQ scheduling and feedback can be exchanged.

HARQ Timing Under Various Transmission and Scheduling Schemes

In this section we examine the detailed timeline of relaying operation with Non-cooperative and Cooperative relaying under with frame formats proposed in [3]. The goal is to clearly illustrate the following fact. *When possible (i.e. with Non-cooperative Multi-Hop) distributed scheduling can offer significant benefits in reducing both delay and over-the-air signaling overhead.*

Given the number of different options (FDD vs. TDD, sub-frame allocation between UL and DL, processing delay, etc.) we make the following assumptions:

- 1/2 sub-frame is sufficient time for any required processing. This is likely optimistic, but is sufficient for our purposes. In particular, as processing delays increase, the differences between the schemes that we present will be accentuated. We do assume, however, that 1 relay-switching GAP is not sufficient for any processing.
- UL and DL sub-frames alternate in TDD. Again, this tends to minimize delays associated with feedback and a different allocation will only accentuate the differences shown here. On the other hand, this will make our analysis for TDD and FDD the same as the delay to produce and send feedback will be identical for both.
- The GAP which is used for the relay to switch between Tx and Rx is located in the middle of a sub-frame. This is for convenience, it does not affect our conclusions in any significant manner.
- We concentrate on the transmission of a single packet and highlight the three types of transmissions which can occur in connection with such getting a single data packet across successfully:
  - Sub-frames in which data and scheduling is sent.
  - Sub-frames in which scheduling only is sent (for our data of interest – other data is, presumably, transmitted)
  - ACK/NACK feedback.

We begin with the 2-hop case. In this case, we have only an odd-hop relay present and there is no operational difference between the 2 frame structures. Notwithstanding a lack of clear definition for forwarding data and feedback through the relay in 802.16m, a reasonable approach to 2-hop operation with centralized scheduling is depicted in Figure 1. The figure presents a configuration where feedback (ACK/NACK) for a particular signal can be sent in the sub-frame following the sub-frame where it is received.

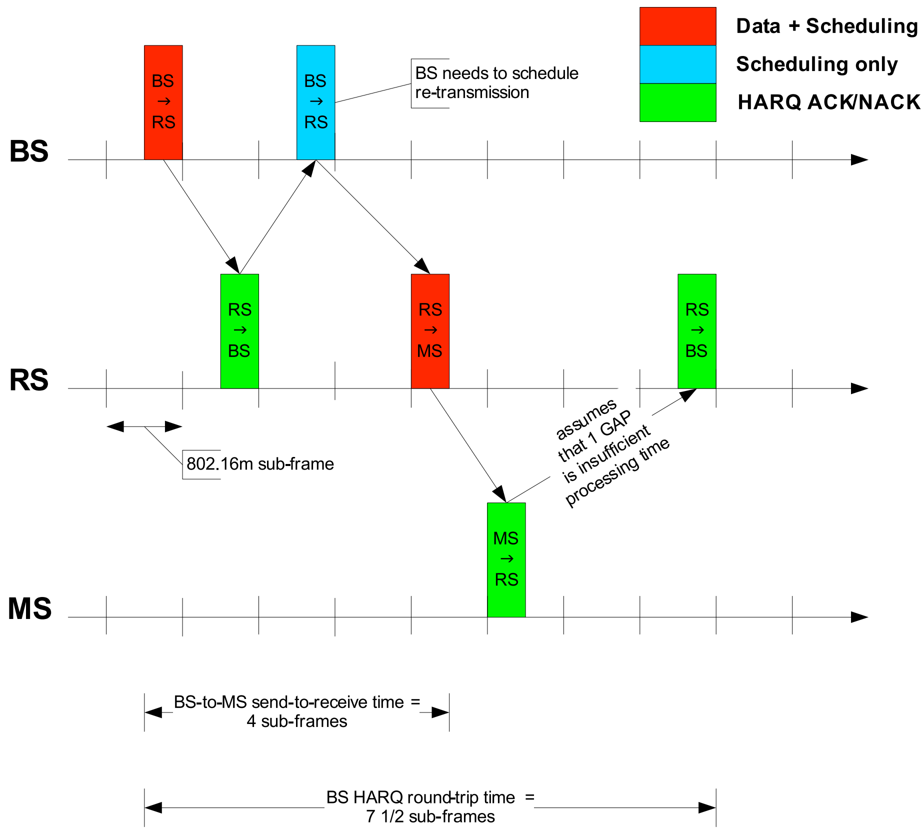

Figure 1. Timing diagram of downlink with Non-cooperative 2-hop and centralized scheduling

Here, we depict the case where no NACK's are sent – i.e. the minimal number of transmissions is required to get the data through. As we observe, it takes 4 full 802.16m sub-frames to get the data through a single relay to the MS. Moreover, it takes 7 1/2 full frames for the BS HARQ scheduler to be informed of successful delivery. This has a potentially significant negative impact both on the required memory and on the complexity of the scheduler.

As we observe from Figure 2 the delays associated with the Non-cooperative 2-hop, are reduced when distributed scheduling is used. The data delivery delay is reduced to 2 sub-frames and the HARQ round-trip delay from the BS point of view is reduced dramatically – to just 1 ½ subrame. One should also observe a dramatic reduction in the signaling overhead over-the-air. Under the centralized scheduling scheme we used 4 signaling exchanges (3 ACK's and one blank scheduling message). With distributed scheduling, we only require 2 ACKs.

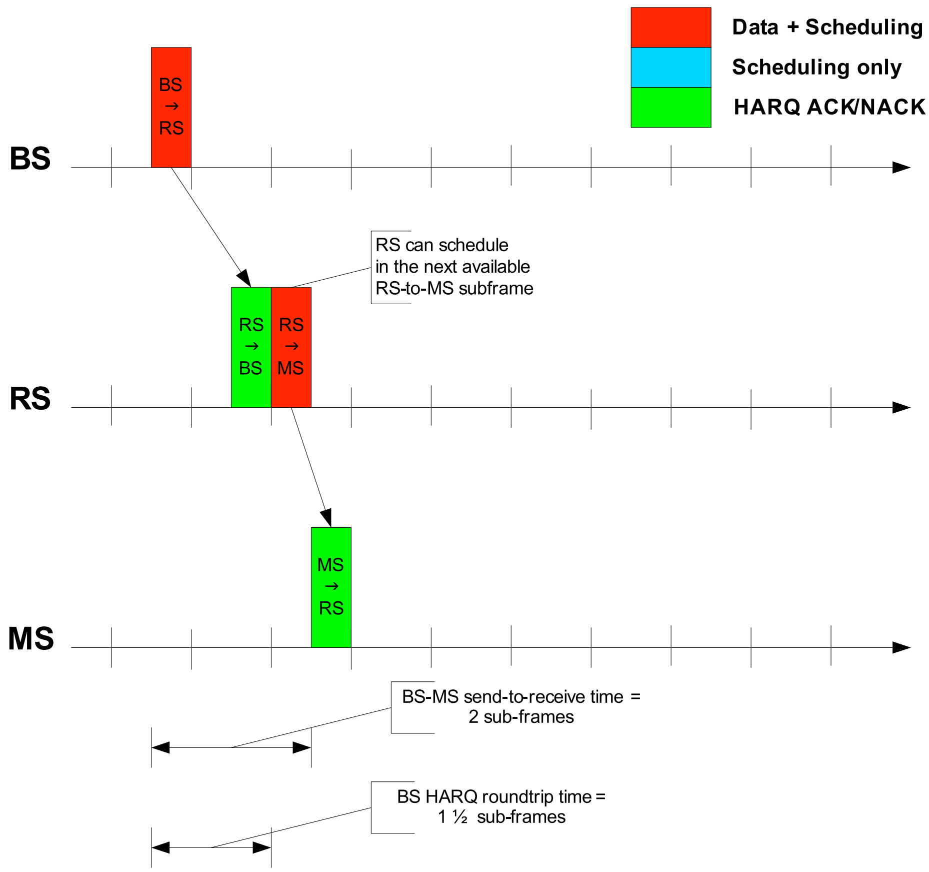

Figure 2. Timing diagram of downlink with Non-cooperative 2-hop and distributed scheduling

To further illustrate the advantages offered by distributed scheduling in the 2-hop case, we now consider the case where a re-transmission to the MS is required (shown in Figures 3 and 4). We observe here that the penalty for centralization of scheduling is now very significant. In particular, we can easily see that each re-transmission to the MS requires a roundtrip delay of 6 sub-frames with centralized scheduling versus only 2 sub-frames with distributed scheduling. The impact of the HARQ buffer in the BS is even more significant.

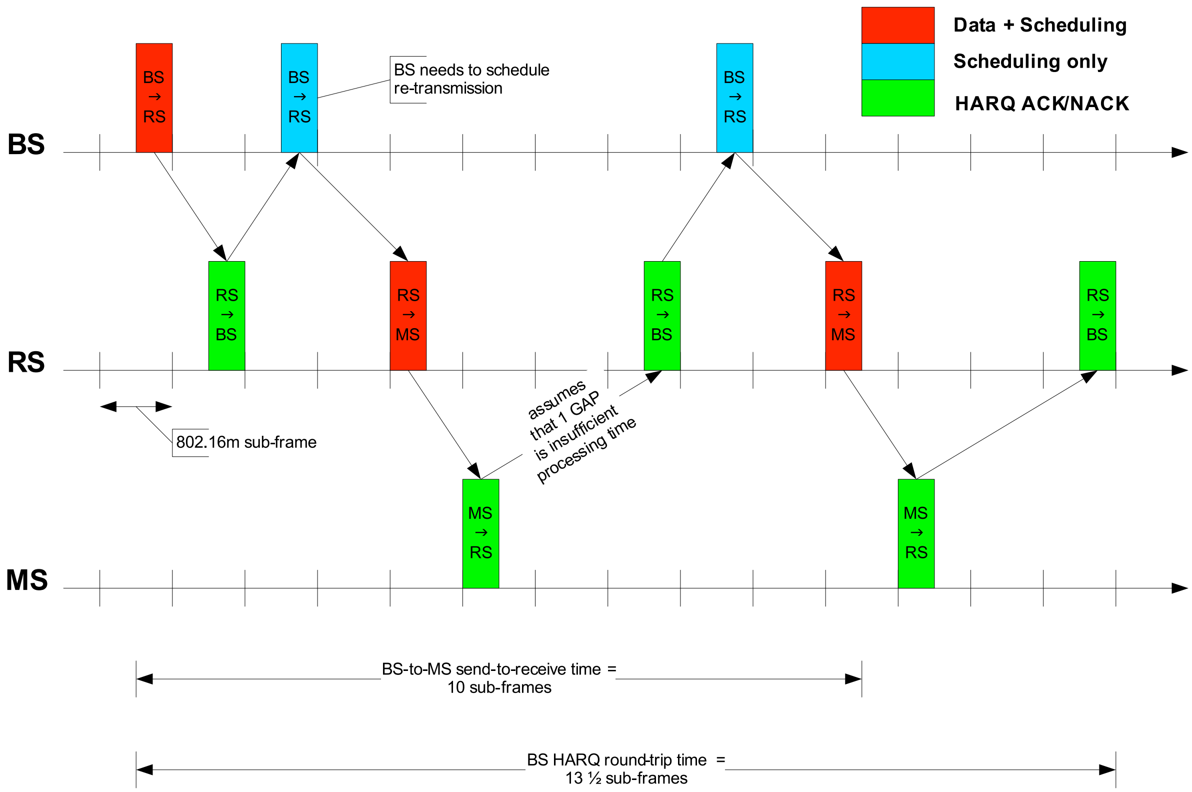

Figure 3. Timing diagram of downlink with Non-cooperative 2-hop, re-transmissions to the MS and centralized scheduling

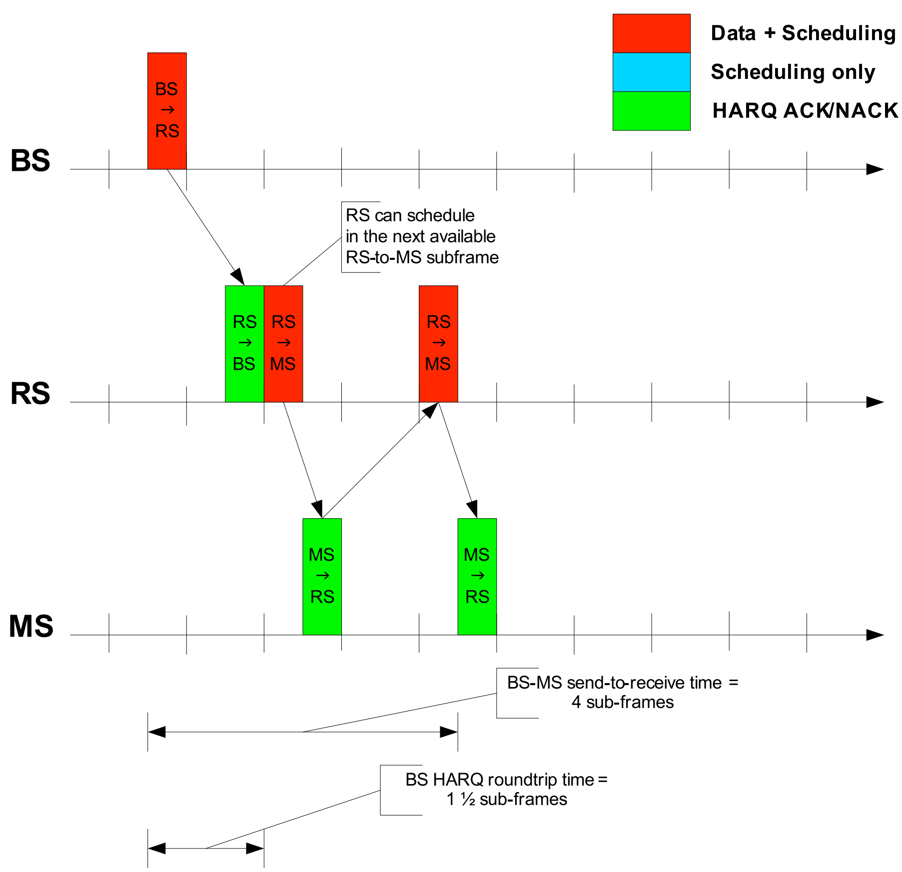

Figure 4. Timing diagram of downlink with Non-cooperative 2-hop, re-transmissions to the MS and distributed scheduling

Let us now consider cooperative transmission, in which case we must resort to centralized scheduling. A timing diagram for this is shown in Figure 5. In this case we cannot expect significant delay reductions, as these are associated with centralized scheduling. However, some overhead reduction is seen because the relay does not need to forward the ACK/NACK. Moreover, transmission of data in the cooperative phase, requires lower power from BS and RS and may have a potentially positive impact on the overall system performance (this is indeed confirmed by simulation, see, e.g. [1,2]).

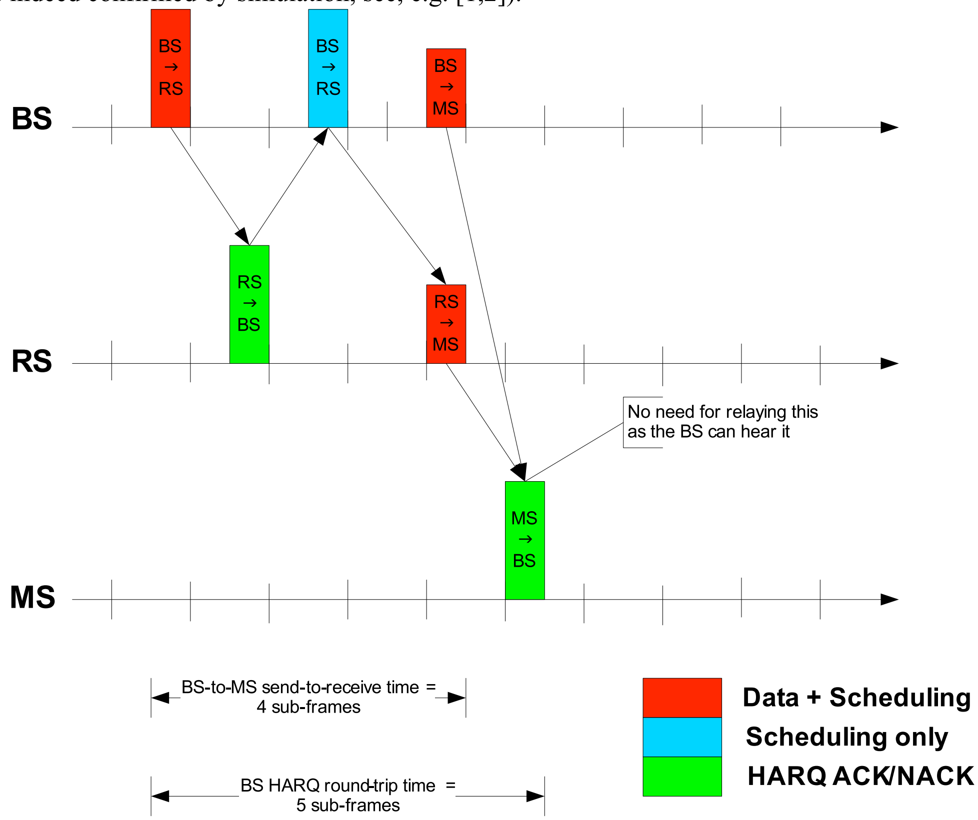

Figure 5. Timing diagram of downlink with Coop 2-hop and centralized scheduling

It would appear from Figure 5 that enabling direct feedback from MS to BS provides only marginal improvement to the latency of the system. However, the advantages are not apparent because we have not considered re-transmissions to the MS. We consider these in Figures 6 and 7. Specifically, both Figures depict a Coop 2-hop transmission scheme with re-transmissions to the MS. However, Figure 6 depicts a case where the relay must relay the NACK to the BS, while Figure 7 depicts a case where this is received directly by the BS. From these we observe that an additional delay of 2 sub-frames per re-transmission is incurred if the feedback is to be forwarded to the BS.

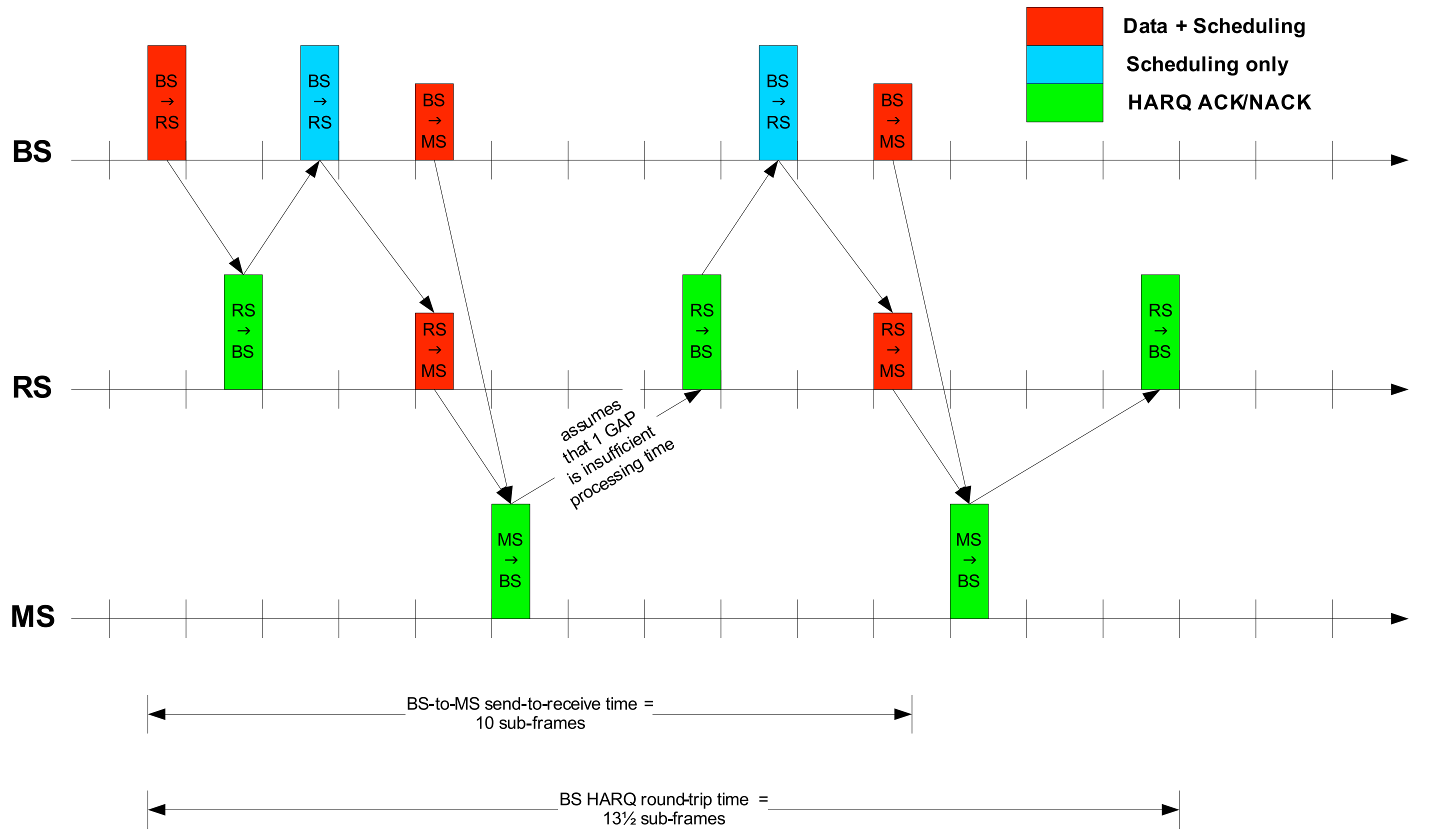

Figure 6. Timing diagram of downlink with Coop 2-hop and centralized scheduling, and 2-hop feedback

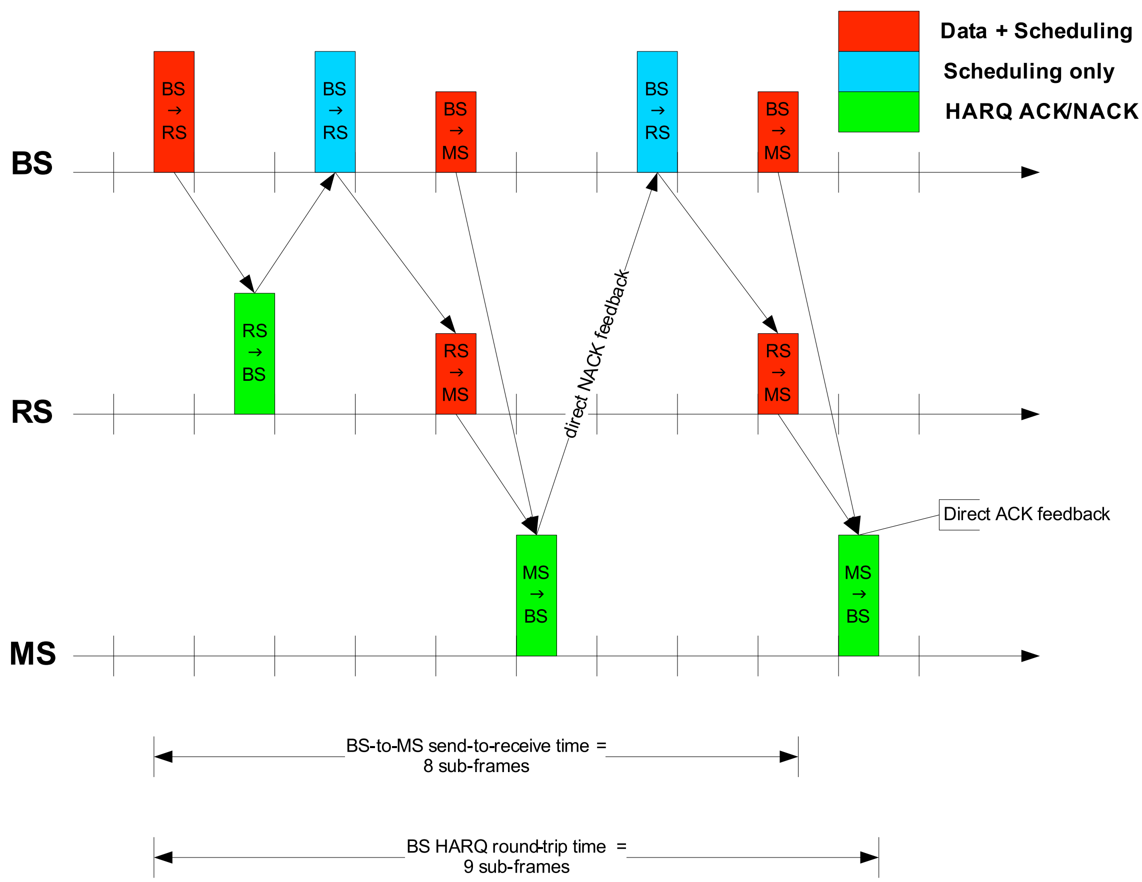

Figure 7. Timing diagram of downlink with Coop 2-hop and centralized scheduling, and direct feedback

To illustrate further differences we consider a 3-hop case, in which case we need to address the two frame structures proposed in [3] separately. As it is clear that the major timing advantage results from the difference between centralized and distributed scheduling we concentrate on the simple 3-hop and illustrate the differences in scheduling only. In Figures below we concentrate on Frame Structure Option 1 [3] and illustrate what happens with 3-relays in a Simple 3-hop configuration but with centralized relaying. Our conclusions translate directly to Frame Option 2. Figure 8 illustrates this situation in the case where the first transmission in *each hop* is successful – i.e. there are no NACKs, only ACKs in the uplink. This figure clearly illustrates the limitations of centralized scheduling as applied to multi-hop and simple cooperation. This is especially so when contrasted with the timing diagram of the same protocol but with distributed scheduling at each stage (Figure 9).

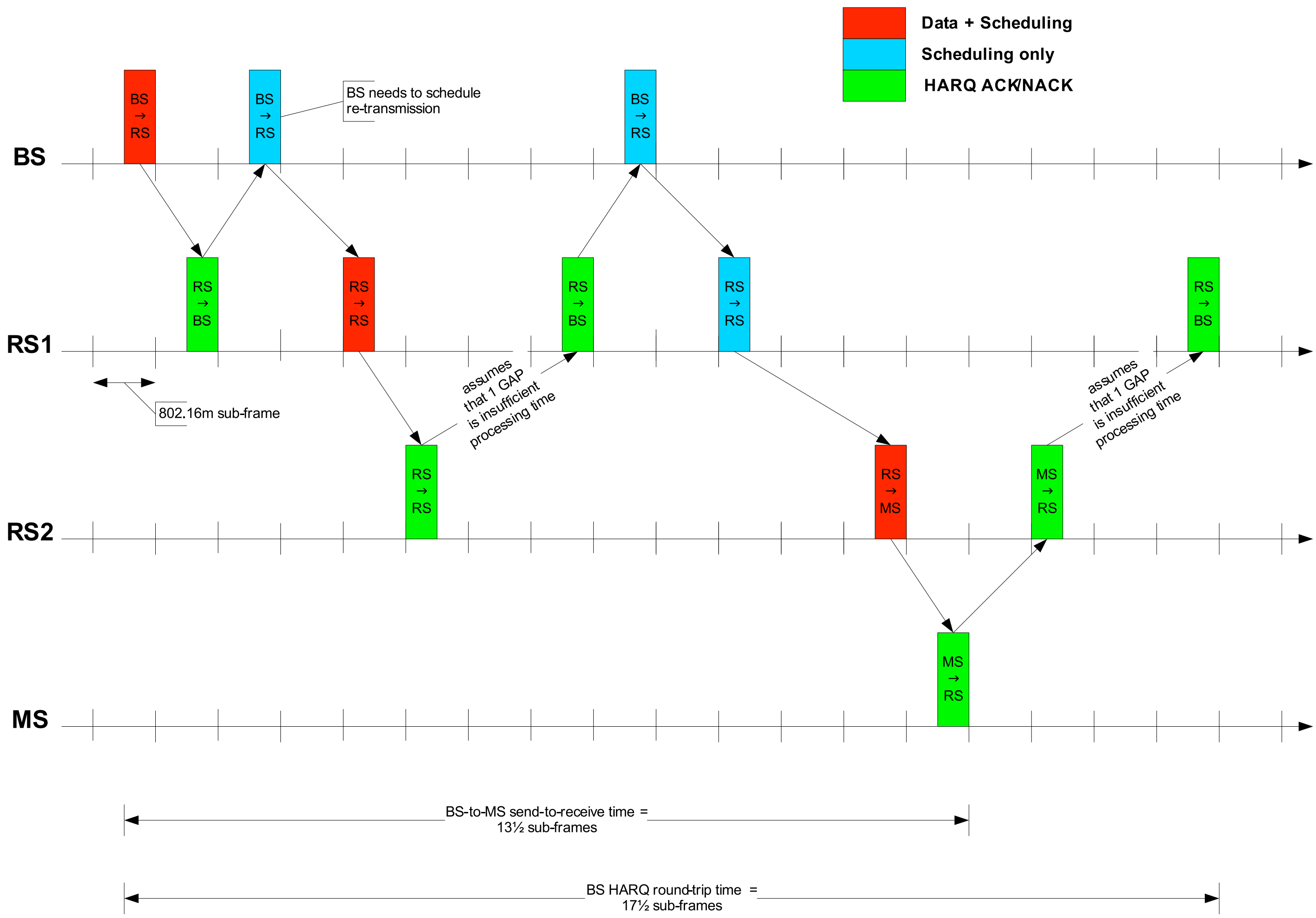

Figure 8. Timing diagram of downlink with Simple 3-hop and centralized scheduling

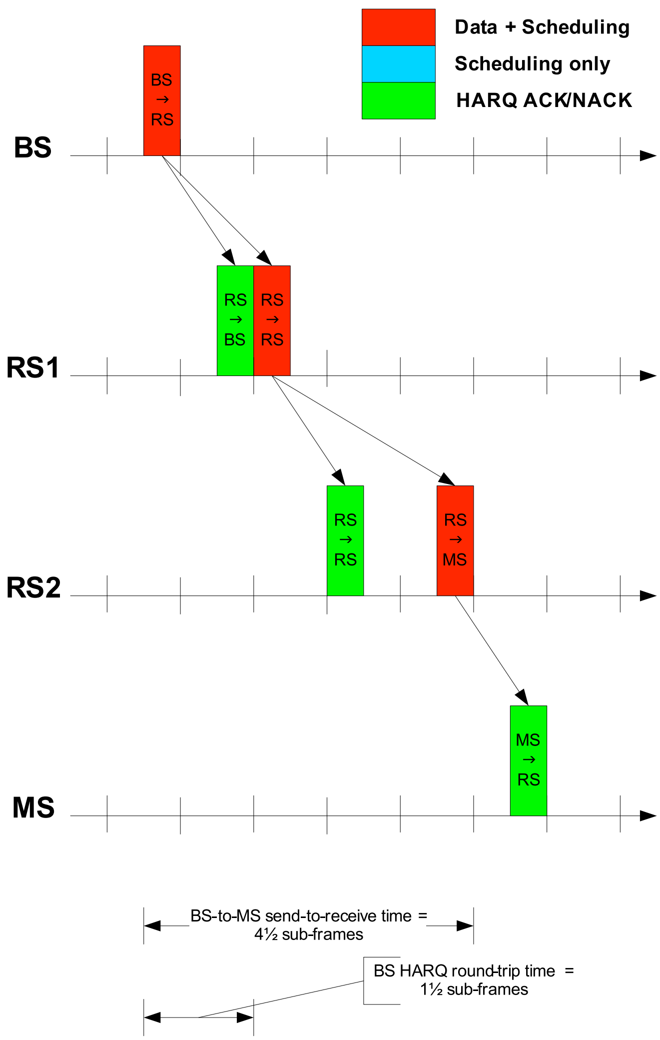

Figure 9. Timing diagram of downlink with Simple 3-hop and distributed scheduling

As expected, cooperative transmission requires centralized scheduling and therefore the timing diagram winds up similar to that in Figure 8. In fact, given the current relaying frame definitions it is not possible to talk of full cooperation between the BS, RS1 and RS2 with either frame Option 1 or frame Option 2. Different sub-layers of cooperation are possible with different frame options. However, it should be clear by now that all of these will require some amount of centralized scheduling.

Observations and Proposal Outline

Based on the analysis presented above we conclude that the optimal mode of operation depends on the relaying/cooperation modes used to transmit to the MS, which in turns depends on "where the MS is located." A Simple multi-hop approach is appropriate for addressing coverage issues, range extension and other key scenarios. In this case, distributed scheduling is advantageous, because it results in significant reduction in delay and over-the-air overhead. On the other hand, as demonstrated in [1,2], cooperative modes can offer significant advantages over non-cooperative 2-hop in those cases where QoS to an MS is severely degraded due to interference. Clearly one would like to support both modes, but these appear to require different architectural approaches to scheduling and feedback:

- In the case of simple multi-hop, the distributed scheduling should be multi-hop and feedback should effectively be multi-hop as well (or, more precisely, "last hop").
- In the case of cooperative relaying, we need centralized. Furthermore, taking advantage of an existing direct signaling connection between the BS and MS offers meaningful reductions in HARQ latency as well as over-the-air overhead.

Clearly, transition between the two modes can be achieved using out-of-band signaling. However, this is slow expensive and, given the mobile nature of MSs is undesirable. Accordingly, we are interested in defining method to affect transition between the two modes "seamlessly:" meaning

- Fast and in-band, utilizing existing scheduling mechanisms
- With low-overhead

- Transparently to the MS.

We begin with a high-level description of the proposed approach and start by examining the simple example of BS-RS-MS. This is then extended to the general case where multiple-hop relaying is supported. In the next section our proposal are put into concrete form by mapping these onto the 2 relaying frame structures currently under consideration by the 802.16m TG [3]. In all cases we assume that the association and connection remain static (i.e. mobility and inter-relay "handover" are not addressed here).

Consider first then the non-cooperative 2-hop case of a BS-RS-MS connection. In the downlink the process operates as follows:

- Any transmission is scheduled to the relay. Upon successful reception the relay sends a HARQ ACK to the BS.
- The BS behavior depends on whether an MS is associated with it (i.e. whether they can hear each other's L1signaling).
  - If the MS is associated with the BS, the BS continues to schedule transmissions to the MS thus allowing the BS and RS to cooperate. Because the MS is associated with the BS, no ACK relaying by the RS is necessary – the BS should be able to receive it. We note here that cooperation is not required – the BS can schedule a transmission (thus forcing the RS to transmit), while not sending to that MS at the scheduled time.
  - If the MS is not associated with the BS, the BS will signal to the RS to take over scheduling as soon as receives the data. It treats the RS HARQ ACK as an ACK from the mobile and removes the data from its HARQ buffer.

The signaling of scheduling mode to the RS may be done in several ways – e.g. by including a special field with each burst control information field or by sending a special control field following the reception of the burst. It can also be a link property in which case it does not need to be signaled.

As we have seen, extension of the scheme to multiple (>2) hops is straightforward (see also TP below). It should however be noted that not all frame structures support cooperative transmission for this case. Nevertheless we argue that in the vast majority of the cases relaying will be limited to 2 hops and therefore this mode of operation should be supported for any frame structure (but not used for >2 hops)

Text Proposals

TP-1: Synchronization channel for Relays

11.7.2.1.1.9 Synchronization channel for relays

Relays may be instructed by the BS using MAC signaling to transmit a synchronization channel. The structure of the channel shall be similar to the BS synchronization and shall enable MS cell search in the same manner even if BS synchronization channel cannot be received by the MS.

TP-2: Broadcast channel for relays

11.7.2.2 Broadcast Channel (BCH)

…

11.7.2.2.6 Broadcast Channel for Relays

Relays may be instructed by the BS using MAC signaling to transmit a synchronization channel. The structure and information content of the channel shall be similar to the BS synchronization and shall enable MS to associate with the cell search in the same manner even if BS broadcast channel cannot be received by the MS.

TP-3: Scheduling

11.7.2.3.1.2 User-specific control information

User specific control information consists of information intended for one user or more users. Examples of this subclass of information include scheduling assignment, power control information, ACK/NACK information.

Resources can be allocated persistently to MSs. The periodicity of the allocation may be configured.

A group message is used to allocate resources and/or configure resources to one or multiple mobile stations within a user group. Each group is associated with a set of resources. VoIP is an example of the subclass of services that use group messages.

11.7.2.3.1.3 Control information in the case of relays

An MS can be configured to listen to its parent node alone (an RS) or to both its parent and grandparent (an RS or BS, depending on the number of hops) nodes.

In the case that MS is configured to listen to both parent and grandparent nodes: :

- the grandparent node indicates per allocation whether the parent node shall autonomously (distributed mode) schedule the RS-MS transmission. The parent node schedules as indicated.
- If the grandparent node is a relay then its parent mode indicates its scheduling in the same manner.
- The MS attempts to receive control information from grandparent and parent nodes, and
- Attempts to receive data from either or both of grandparent and parent nodes

In the case that MS is configured to listen to parent node alone:

- the BS indicates, using MAC messages, whether the MS parent node shall autonomously (distributed mode) schedule the RS-MS transmission
- The MS attempts to receive control information from the parent node, and
- Attempts to receive data from its parent node

In case of 3 hops or more, the last 2 nodes in the chain may be configured to cooperate in the data transmission to the MS. A group message is used to allocate resources.

TP – 4: HARQ feedback

11.9.1.3 HARQ feedback

HARQ feedback (ACK/NACK) is used to acknowledge DL transmissions. Multiple codewords can be acknowledged in a single ACK/NACK transmission.

11.9.1.3.1 HARQ feedback for use with relays

If an MS is scheduled to listen to the BS to RS transmission then the MS will send HARQ feedback as scheduled by the BS. The HARQ information may be node specific (i.e. a separate ACK/NAK for each node from which data is received) or node non-specific (i.e. a single ACK/NAK to indicate successful reception). The selection of the ACK/NAK mode is determined by the BS using MAC signaling. Node specific ACK/NAK feedback pertaining to a DL transmission could be sent over a single sub-frame or two sub-frames.

The BS may instruct the RS, using MAC signaling, to relay an ACK/NAK feedback from the mobile to the serving BS. Resource allocation for the relayed ACK/NAK will be done similar to a MS ACK/NAK resource allocation. Alternatively, the relay may be instructed to aggregate ACK/NAK that came from different MSs for transmission to the base station.

The BS may also instruct a RS to ACK/NAK its own reception of downlink data.

References

[1] C80216m-Relay-08/024 “Relaying Strategies for 802.16m Multi-Hop Relays”
[2] Deleted
[3, 4] 80216m-08-003r4 Current 802.16m SDD