An Evolved Cellular System Architecture Incorporating Relay Stations

Jerry Sydir, Intel Corporation
Rakesh Taori, Samsung Electronics

ABSTRACT
Shrinking cell sizes, primarily a result of keeping up with the increasing demand for higher data rates, are stretching thin the notion of our traditional cellular system architecture. More and more base stations are serving smaller and smaller areas (cells) which in effect is driving up deployment costs. The option of deploying relay stations is increasingly beginning to look like a solution to the problem of providing a cost-effective way to extend the coverage and capacity in a cellular network. A relay station can be used to extend the point-to-multipoint link between the base station and mobile stations. Relay stations connect to the base stations without wires and are expected to be deployed to cover smaller geographic areas. Primarily the deployment costs but also the equipment costs are expected to be substantially lower than those of base stations. In this article we first introduce the relay station and discuss the place it is likely to carve out for itself in the traditional cellular architecture. We highlight the important economic and performance benefits relay stations could potentially offer, and outline a few scenarios where relays are likely to be deployed in the beginning. The multihop relay standard developed by the IEEE 802.16 working group is then used as a basis to provide an overview of the relay-enhanced cellular architecture and the key choices that can be made in developing relay support within a cellular system — a precursor to what can be expected in later releases of the Mobile WiMAX system. Finally, we discuss some future directions in the development of relay systems.

RELAY STATION IN THE CELLULAR ARCHITECTURE
Mobile Internet, a term that has been used to describe the capability of delivering broadband data service to an end user’s terminal while the user is on the move, is driving up the demand for higher data rates. The goal of mobile Internet is to deliver to mobile users the same experience currently enjoyed by users at home or in the office over a wired broadband connection. The need for higher data rates has traditionally been met by increasing the bandwidth of the radio frequency (RF) carriers, better channelization (ability to divide the radio spectrum into narrower slices), better modulation techniques, but above all by spectrum reuse, which has been achieved by dividing the coverage area into smaller cells. The shrinking of cell sizes has resulted in larger numbers of base stations (BSs) per unit area and has left operators wondering about the scalability of the traditional cellular architecture.
In the traditional cellular architecture a given coverage area is divided into smaller areas called cells. A BS, often located at the center of the cell, provides coverage to the mobile stations (MSs) within the cell. The BS is connected to the core network via a backhaul connection, typically provided by a wired or point-to-point microwave link. The area within a cell can further be subdivided into sectors. Within each sector, the BS communicates with MSs that are associated with it, using what is referred to as a point-to-multipoint (PMP) link. The PMP link here refers to a specific type of multipoint link whereby a central device (BS) is connected to multiple peripheral devices (MSs). Any transmission of data that originates from the central device is received by one or more of the peripheral devices, while any transmission of data that originates from any of the peripheral devices is only received by the central device. Each BS manages the allocation of resources to support communications between itself and the MSs it serves. The BSs are informed about the resource allocation by the BS. Coverage within a sector is commonly enhanced by the deployment of analog repeaters. Repeaters are simple devices that receive the signal transmitted by the BS (as well as the signals from neighboring cells, if any) and with very little delay indiscriminately amplify and forward this signal.
The traditional cellular architecture mentioned above can be enhanced using devices called relay stations (RSs), which intelligently relay data between the BS and MSs wirelessly. The BS communicates with a multitude of RSs within its coverage using the PMP mode mentioned above, and the RSs in turn also communicate with MSs associated with them using the PMP link. The BS maintains overall control over the RSs and MSs associated with it, although the implementation of individual control functions...
**RS Benefits**

The deployment of RSs offers performance and cost benefits over the use of BSs in a traditional cellular network. Performance benefits include improvements in coverage and/or increases in capacity. Cost benefits are realized through reduction in the cost of providing service, which in turn is obtained due to a reduction of the cost of equipment, site development, and backhaul.

In discussing RS benefits, we first examine the performance related aspects. The performance improvements that can be achieved due to relay are based on two factors. The first factor is the increase in frequency reuse that results when the BS and RSs within a sector each communicate at the same time to different MSs using the same frequency resources. The second factor is the increase in effective capacity of a multihop wireless link as the number of hops is optimized to match the distance between the communicating nodes. Oyman showed that for any distance that separates two nodes communicating using wireless protocols, there is an optimal number of hops that maximizes the effective capacity of the link between the two nodes [1]. The effective capacity of the link takes into account the capacity of the individual hops and the resources used to transmit the same data multiple times, once for each hop.

RS deployment enhances the coverage and capacity in areas where the capacity of the direct link between the BS and MSs is low. Such areas can exist at the cell edge (e.g., MS1 in Fig. 2) or in the shadows of large objects such as tall buildings (e.g., MS8), within the buildings themselves, or underground. RS deployment enhances coverage in areas where the capacity of the direct link between the BS and MS is zero (e.g., MS2 in a coverage hole or MS7 beyond the edge of the cell). RS deployment enhances capacity throughout the cell due to increased frequency reuse.

It is important to understand that the enhancements outlined above are measured rela-

---

**Figure 1. A sector in base station coverage enhanced using relay stations.**

(e.g., scheduling) may be centralized at the BS or distributed between the BS and RSs. The RSs operate using a store and forward paradigm. The RSs receive the data selectively in specific time/frequency allocations indicated by the BS, decode and process the data, and subsequently transmit (relay) this data in different allocations that occur later in time. RSs can range in intelligence and capability. In the simplest form the RSs might simply demodulate and subsequently remodulate the data prior to relaying. In the most advanced form RSs can decode the data, process the decoded data (which may involve operations such as error handling, fragmentation, packing, rescheduling according to quality of service (QoS)), and subsequently re-encode the data in accordance with the conditions on the next link prior to relaying. Furthermore, RSs may implement varying levels of control functionality.

Figure 1 illustrates an example of a sector enhanced using RSs. The area immediately surrounding the BS is the BS’s coverage area. The MSs in this area associate with the BS and are served directly by the BS. Each RS has a coverage area within which it serves MSs. The coverage areas of the BS and RSs can overlap to varying degrees.

As Fig. 1 shows, the path between the BS and an MS can consist of one or more hops. The RSs can be thought of as being arranged in a multi-level hierarchical structure with the BS at the top of the hierarchy. RSs at the first level of the hierarchy are attached directly to the BS. RSs at the second level are attached to the first-level RSs and so on. In its simplest form the topology of the sector enhanced using RSs is a tree, with a single path between the BS and each MS. In a more complex topology, such as a mesh topology, more than one path can exist between the BS and one or more MSs, RSs on the same level may commu-

...
The primary advantage of deploying RSs in terms of the cost is expected to come from the differences in the cost of the backhaul. When an RS is deployed, instead of a BS with a wired backhaul connection, there are no direct backhaul costs.

Figure 2. Example RS deployment scenarios.

In order to discuss the deployment of RSs, we outline a few example deployment scenarios that are expected to be used during different phases of the evolution of the network. Figure 2 illustrates some of these scenarios. Additional relevant information on RS deployment scenarios can be found in the IEEE 802.16j Usage Models document [3].

The first example deployment scenario is relevant during the initial buildout of the network. As a network is first deployed, RSs can be used to reduce costs when the subscriber base is being established and system capacity is not an issue. In this scenario the operator selects some cell sites for the deployment of BSs, while deploying RSs within other cell cites. As the network adds subscribers and requires more capacity, selected
RSs are replaced by BSs to increase the capacity of the network.

Another deployment scenario involves the usage of RSs to enhance coverage within the network. RSs can be deployed in a regular pattern to improve coverage across the sector, or in specific locations to provide coverage in the shadow of large buildings or other physical obstructions.

Yet another deployment scenario involves the deployment of RSs to provide cost-effective wireless backhaul in a dense cell deployment. RSs can be used to reduce the cost of such a deployment because they allow the operator to trade off the cost of backhaul with the usage of some of their air link resources for supporting relay links.

Another potential deployment scenario is for providing in-building coverage within a home or small office (e.g., MS6 in Fig. 2), often referred to as a femto cell. Coverage within a femto cell can be provided by a small BS or an RS. RSs can be used where a wired backhaul is not available.

There are a number of more novel deployment scenarios in which RSs can be used. For example, in-vehicle coverage can be provided by a mobile RS deployed within a vehicle such as a train, bus, or ferry to MSs in the moving vehicle. RSs can also be used to provide temporary coverage during a concert or in the aftermath of a disaster where the physical infrastructure may have been disabled.

IEEE 802.16J: The First Relay Standard

The use of relay has been specified for IEEE 802.16 systems by the IEEE 802.16 WG. At the time of writing this article, the IEEE 802.16j amendment had been completed by the IEEE 802.16 WG [4]; approval for publication by IEEE SA Standards Board was expected by the end of March 2009. In this article we refer to [4] as the 16j draft.

Legacy Support

An important thing to understand about the 16j draft is the constraint of legacy support. Legacy here refers to the IEEE 802.16e standard (or simply 16e) [5], which defines the air interface on the BS-MS link. The 16j project was approved under the restriction that the MS specifications shall not change. This means that as far as the MS is concerned, the access link through an RS (the RS-MS link) should look identical to the BS-MS link. This restriction had a large influence on the scope and architecture of 16j. Packet headers and access link protocols could not be modified. Only strictly backward compatible changes could be made to the PHY layer and the MAC layer constructs that are visible to the MS, although the relay link protocols could be different. To be backward compatible, the frame structure for 16j had to be designed in such a way that it allowed access and relay link communications to be multiplexed within an RF channel. Another consequence of this constraint was that certain concepts such as client relay (an MS relaying data on behalf of other MSs), for instance, became out of scope by definition.

Protocol Architecture

For the relay link, 16j defines a complementary set of tightly coupled protocols that extend the 16e protocols across multiple hops. A high-level representation of the protocol architecture is illustrated in Fig. 3. This representation is not a part of the 16j draft, but is included here for illustrative purposes. The protocols between the RS and MS are defined in 16e, while the relay link protocols (BS-RS link in the figure) are defined in 16j. In some cases the 16j protocol simply consists of relaying the 16e messages between the BS and MS, while in other cases new messages have been defined in the 16j protocol to allow additional communications between the RSs and the BS. An important assumption in 16j is that the BS has/obtains all the information relevant for operations such as network entry (of MSs and RSs), RS selection, path selection, and handover, and thus maintains overall control in the entire BS cell. Necessary signaling, not the corresponding algorithms, has been defined (PHY allocations and MAC messages) on the relay link to facilitate the aforementioned level of BS control.

Relaying Modes

The 16j draft supports two types of relaying operations: transparent and non-transparent. The draft specifies both centralized and distributed scheduling and security models. It supports topologies that are greater than two hops, but restricts the topology to a tree structure. The hop count here is measured in terms of how many hops the MS is away from the BS. In the remainder of this section we describe the basic operation of transparent and non-transparent relay.

We use the term transparent relay because the presence of the RSs is essentially transparent to the MSs. The BS transmits downlink (DL) synchronization and control channels to the MSs directly, while the RSs may relay select unicast data transmissions to the MSs as determined by the BS. The hop count in the case of a transparent relay operation is limited to 2. The BS and RSs within a sector all operate on the same RF carrier.

Figure 4 illustrates the types of transmissions that occur in transparent relay mode and the transparent relay mode frame structure. To enter the system, an MS first synchronizes with the DL of the BS and receives the DL control information. The MS first performs ranging with the BS. During this process, the transparent RSs in the sector listen to the ranging transmissions of the MS and report the measured signal strength to the BS. The BS uses this feedback from the RSs to determine which, if any, of the RSs can be used to relay data to/from the MS. The MS completes the network entry process with the BS, with the BS mak-
ing the decision whether to allow the MS to enter the network. As part of this process security associations and connections are established between the BS and MS. The RSs are not aware of the security associations and connections, and do not have access to the keys or connection identifiers.

The BS schedules the usage of all air link resources within the sector. When scheduling data transmissions, the BS determines whether the data is to be transmitted directly to the MS or whether one of the RSs should be used to relay the data. The BS makes this determination by comparing the effective capacity of the direct link between itself and the MS with the effective capacity of several two-hop links through each RS. If the BS decides to relay the data, it transmits the data to the chosen RS, instructing it to forward the data in a specified allocation later in the frame. Only the RS-MS transmission is reported to the MS. Selection of the modulation and coding scheme (MCS) is performed by the BS for both transmissions. The scheduling of uplink (UL) data is similar.

As an MS moves around the sector, the BS may determine, based on the channel quality information it receives from the RSs and MSs, that data to/from the MS should be relayed by a different transparent RS. This decision is made without an explicit handover process. The BS simply begins to instruct a different transparent RS to relay data to/from the MS. When the MS moves into the coverage area of a neighboring BS or sector, handover is performed in the same manner as in a traditional cellular network.

When deploying transparent RSs, the coverage areas of the BS must be designed to allow the MSs to receive the control channels at the cell edge. In this configuration the addition of RSs does not improve coverage or extend the range. Instead, the deployment of transparent RSs can only increase capacity.

The second relaying mode is referred to as non-transparent relay. The term non-transparent is used because in this mode the MSs are aware of the existence of the RSs, in principle. In 802.16j, however, the MSs cannot distinguish between BSs and non-transparent RSs (i.e., MSs perceive a non-transparent RS as just another BS). In contrast to transparent relays, addition of non-transparent RSs can improve coverage, extend range, and/or increase capacity. The 16j draft defnes both centralized and distributed scheduling for non-transparent relay. Here, we restrict the description to non-transparent relay with distributed scheduling.

The major types of transmissions that occur in non-transparent relay mode are illustrated in Fig. 5. Both the BS and RS support a PMP link for communicating with the MSs (access link) and a PMP link for communicating with the other RSs (relay link). The access links of the BS and RSs within a sector are distinct from the perspective of the MSs. The MSs within the network associate with either the BS or one of the RSs. MSs, which associate with the BS, receive all DL transmissions directly from the BS and transmit their UL transmissions directly to the BS. MSs, which associate with one of the RSs, receive all of their DL transmissions from this

---

1 In 802.16 ranging refers to a collection of processes by which the MS and BS maintain the quality of the RF communication link between them [5].
In a sector with non-transparent RSs, the BS and the RSs establish the topology of the sector by determining the route between the BS and each RS. The topology is modified in response to the entry or exit of the RSs.

RS and transmit their UL transmissions to this BS. The RS relays MS data to/from the BS through zero or more intermediate RSs. The length of the path between the BS and MS can be two hops or more. Routing decisions are generally less frequent than in transparent relay and are decoupled from the scheduling decisions. The BS and RSs within a sector can operate on the same RF carrier or different RF carriers.

In a sector with non-transparent RSs, the BS and RSs establish the topology of the sector by determining the route between the BS and each RS. The topology is modified in response to the entry or exit of the RSs. To enter a network with non-transparent RSs, an MS synchronizes with the DL of the BS or one of the RSs. If an MS synchronizes with the DL of an RS, it performs ranging with the RS. When ranging has been completed, the RS notifies the BS of a new MS that is trying to enter the network. The remainder of the network entry process is carried out between the BS and MS, with the RS relaying messages between them. As part of this process security associations and connections are established between the BS and MS. The route for these connections is determined according to the topology that has been established within the sector. The RSs along the path of each connection store the connection identifier associated with that connection.

Figure 5 also illustrates the frame structure used in non-transparent mode for the case where the topology is limited to two hops. We see that the BS and RS transmit DL synchronization and control channels at the same time-frequency location in the frame. The transmissions within the DL access zone are performed according to the 16e protocol. From the perspective of the 16e MS, the DL access zone appears identical to a 16e DL zone in a 16e radio frame. In the DL relay zone, the BS transmits data to the RSs attached to it. From the perspective of the 16e MSs, this zone appears to be a 16e DL zone with no allocations (appears to be empty). Similarly, the UL access zone is used for communications on the access link (MS-RS or MS-BS). Transmissions within this zone are performed in accordance to the 16e specification, and to a 16e MS this zone appears like a 16e UL zone in a 16e radio frame. The UL relay zone is used for RS-to-BS transmissions. To the MS this zone appears to be a 16e UL zone with no allocations.

The BS and each RS independently perform MCS selection and schedule the transmissions on the PMP link between themselves and their downstream neighbors (MSs and RSs). MAC protocol data units (MPDUs) are transmitted between the BS and MS along the path that was selected when the connection was established. Each RS along this path receives the MPDUs and determines the next hop. The RS may decrypt and manipulate the MPDUs in order to aggregate individual MPDUs into larger MPDUs or split large MPDUs into multiple smaller MPDUs for the sake of achieving greater efficiency.
As an MS moves around the sector it may move from the coverage area of one of the infrastructure stations (the BS or an RS) to the coverage area of another infrastructure station that is in the same or a different sector. This triggers a handover procedure that from the perspective of the MS is identical to the handover procedure specified in 16e. The handover procedure is always controlled by the BS even if the MS is handing over to/from an RS.

**ADVANCED CONCEPTS**

In the previous section we described the major operations of transparent and non-transparent relay in 16j. The 16j draft also specifies the operation of all basic 16e protocols for relay, such as hybrid automatic repeat request (HARQ), ARQ, bandwidth request, connection management, sleep, and idle mode. In addition to specifying basic relay operations in transparent and non-transparent relay mode, the 16j draft also specifies a number of advanced relaying techniques. We mention them here briefly to provide a more complete picture of the scope of the draft. Cooperative transmission between the BS and RSs is supported. The BS and/or RSs can use virtual multiple-input multiple-output (MIMO) techniques to transmit data to an MS. Mobile RSs are also supported in the 16j draft. Handover procedures are specified for the RSs and the MSs associated with them. Finally, the concept of RS groups is specified within the draft. Multiple RSs are combined to form a group that appears to the MS as a virtual RS. The RSs within the group collaborate in transmitting data to the MSs associated with the group.

**FUTURE DIRECTIONS**

The 16j draft is the first attempt at specifying relay support in a broadband wireless system (e.g., the Mobile WiMAX system). Relay support in 16j was designed to be transparent to a 16e MS. Because of this restriction, the implementation of the transparent mode as defined in 16j is likely to be complicated. The MS is not aware that it is receiving control messages (including ranging) from the BS and data traffic from the RS. The inability to control interference between the BS and RSs is another limitation of 16j. As MSs cannot distinguish between BSs and RSs when measuring the signal strength of interfering stations. In future relay systems MSs may be made aware of the existence of the RSs, which is likely to help the MSs make more informed decisions related to network entry and handover, for instance.

Another limitation of 16j is the large number of options specified in the draft. At the time of writing this article, the IEEE 802.16m TG (which is developing the next-generation 802.16) is near the completion of the system description document for 16m. Relay is included as a feature, and MSs are aware of the existence of relay. Future enhancements related to relay support are likely to focus on relaxing some of the restrictions placed on 16j and the need for simplicity.

Looking a step further, perhaps beyond the timelines of the 802.16m task group, MSs can be made to operate as RSs, relaying data for other MSs. Other potential directions might be to move from a tree topology toward a more general mesh, or define more distributed control functions to allow the network to operate in a more efficient manner as cell sizes decrease and the number of RSs increases. In this situation, self-organization capabilities will make the network easier to operate, and more distributed handover and network entry decisions will make the network more efficient. Other improvements might include improvements to relay link efficiency and the ability for an RS to receive and transmit at the same time.

**REFERENCES**


**BIographies**

**JERRY SYDIR** (jerry.sydir@intel.com) is a senior research scientist in the Wireless Communications Laboratory at Intel Corporation, where he is involved in research on advanced wireless network architectures for WiMAX systems. He has led the research and standardization effort for relay in IEEE 802.16 and was active in the development of the IEEE 802.16j amendment. He has worked on a variety of hardware and software projects in the telecommunications area, including hardware and software architecture of 802.16e baseband processors, network processors, control plane software, and protocol stacks. He received his M.S. degree in systems engineering and B.S. degree in computer engineering from Case Western Reserve University. His professional interests include wireless communications, wireless network architectures, and network protocols.

**RAKESH TAORI** is a principal engineer in the Digital Media and Communications (DMC) R&D Centre at Samsung Electronics, Suwon, South Korea. He is currently involved in research, development, and standardization of the 4G air interface technologies pertaining to the MAC layer. He was an active contributor in the standardization of multihop relay in IEEE 802.16 systems (802.16j amendment) and is now contributing in the development of the IEEE 802.16m amendment on advanced OFDMA air interface for the IEEE 802.16 system. Prior to joining the Samsung DMC R&D center, he held research positions at Samsung Research (2004-2008, South Korea), Ericsson Research (2000-2004, Sweden and the Netherlands), and Philips Research Labs (1992-2000, The Netherlands). Over the past 17 years, he has performed research and standardization work in the area of media coding and wireless systems. In the area of media coding, his primary focus was low-bit-rate parametric coding of speech and audio signals. In the area of wireless systems he has contributed to research and standardization in wireless PANs (Bluetooth and UWB) and wireless LANs (802.11n, WLAN mesh), and is currently active in the area of wireless MANs. He has contributed to several standardization organizations (MPEG, ITU-T, ETSI, Bluetooth SIG, and the IEEE) and has served in the organizations in various roles. From August 2004 to November 2005 he served as chair of the technical steering committee of the WiMedia Alliance. He obtained his B.Eng. degree in control and computer engineering, and M.Phil. degree in digital signal processing and communications from the University of Westminster in London, United Kingdom.