CHAPTER 2

Related Work

Personal communications and mobile computing requires a wireless network infrastructure that is fast deployable, possibly multi-hop, and capable of multimedia service support. The wireless network environment can also be connected to wired infrastructure, or just be a stand alone one. The bandwidth problem is a very important issue in the broadband multimedia mobile ad hoc network. The goal of the bandwidth routing algorithm is to find a shortest path that the available bandwidth on this path is above the minimal requirement [5] . The key to the support of QoS reporting is QoS routing, which provides bandwidth information at each source [6] . QoS routing includes the following parts, bandwidth reservation, path bandwidth calculation, and routing.

2.1 Bandwidth Reservation

We consider bandwidth as the QoS since bandwidth guarantee is the most critical requirement for real time applications [7] . "Bandwidth" in time-slotted network systems is measured in terms of the amount of "free" slot. Bandwidth reservation and calculation are the first step for a network to achieve QoS guarantees, it must reserve and control resources. To set up a connection with QoS constraints, a routing path with sufficient resources must be found first. Then the resource allocation can make the reservation along the path. Slot assignment problem can be reduced to the graph coloring problem which is NP complete problem [5] .

2.1.1 CDMA over TDMA

TDMA (Time Division Multiple Access) scheme is generally used in the wireless environment for bandwidth reservation. CDMA (Code Division Multiple Access) can also be overlaid on top of the TDMA infrastructure; that is, multiple sessions can share the same TDMA slot via CDMA, as shown in Figure 2-1. CDMA over TDMA can somewhat avoids the hidden terminal problem because a code assignment scheme is assumed to be running in the lower layer of the system.

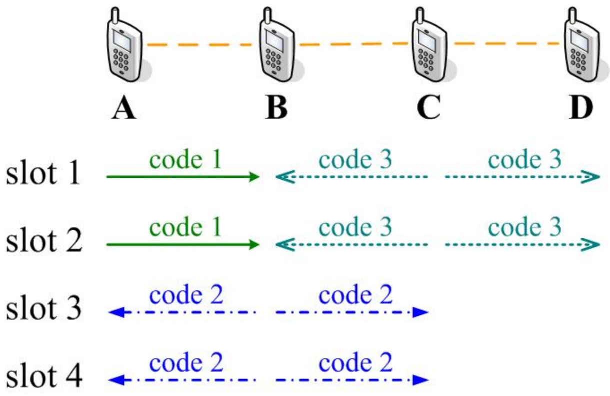

Figure 2-1: CDMA over TDMA.

Each real-time connection is assigned a VC (virtual circuit) since real-time traffic needs guaranteed bandwidth during its active period, each node has to reserve its own slots to the VC at connection setup time. The available bandwidth between two nodes is critical and is used to select a route satisfying the QoS requirement. Furthermore, it's also used to determine whether a new connection is allowed into the network. Routing with a QoS indication is needed in order to manage bandwidth resources efficiently.

Clustering

There is another way to take advantage of frequency reuse property; that is, emulating cellular structure by clustering all the mobile stations into different channels [8] . The clustering algorithm in the control phase of every frame cycle overcomes the change of topology [9] [10] . There is an example of clustered multi-hop wireless network as shown in Figure 2-2 which emulates the cellular structure. After clustering, the nodes in the same cluster use the same channel, different clusters use different channels.

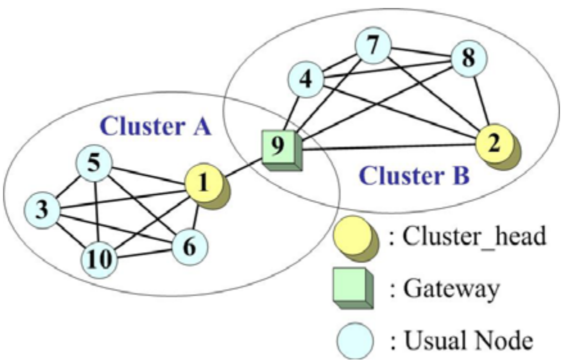

Figure 2-2: Clustered multi-hop wireless network.

2.2 Path Bandwidth Calculation

QoS is meaningful only for a flow of packets between the source and destination, so calculating the current bandwidth to learn about if there is enough resource on this certain path to achieve the QoS guarantee is important. In TDMA, the assumed bandwidth is the number of free time slots.

To do path bandwidth calculation, every mobile station has to broadcast its own slot condition (reserved or idle). Then a node does path bandwidth calculation first when it receives slot conditions from neighbors. Then according to its own slot condition, it will do some modification on the slot conditions and transfers it with the calculation result. If the routing table has no more space to store the information or the result of path bandwidth calculation is not better than the existing ones, then the received message will stop its traveling in this multi-hop packet radio network. This feature prevents the message from traveling in the multi-hop packet radio network endlessly and from wasting valuable bandwidth.

2.2.1 Bandwidth Calculation using Link/Path Bandwidth Information

The transmission time scale is organized in frames, each containing a fixed number of time slots. The entire network is synchronized on a frame and slot basis. That is, time is divided into slots, which are grouped into frames, as Figure 2-3 shown below.

Figure 2-3: Transmission Time Scale, Frames and Time Slots.

Here the bandwidth information is embedded in the routing table. By exchanging the routing table, the end-to-end bandwidth of the shortest path for a given source-destination pair can be calculated.

Because only adjacent nodes can hear the reservation information, and the network is multi-hop, the free slots recorded at every node may be different [11] . Here the *link bandwidth* is defined to be the set of the common free slots between two adjacent nodes. And the *path bandwidth* between two nodes, which are not necessarily adjacent, is the set of available slots between them. Figure 2-4 depicts the bandwidth information calculation overview.

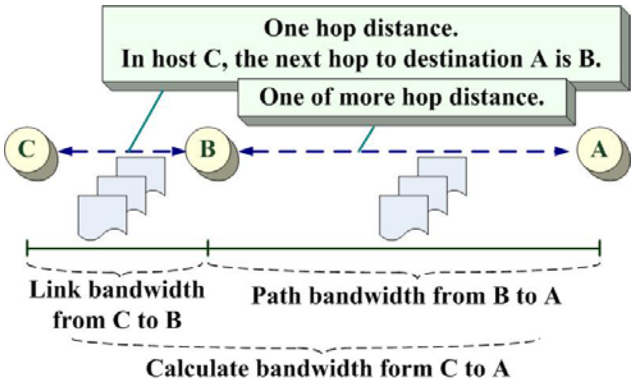

Figure 2-4: Bandwidth information calculation overview.

free_slot(*X*) is defined to be the slots, which are not used by any adjacent host of *X* to receive or to send packets, from the point of view at node *X*. Next, link bandwidth can be further employed to compute the end-to-end bandwidth. This information can provide an indication of whether there is enough bandwidth on a given route between a source-destination pair. Figure 2-5 illustrates an example for calculating the path bandwidth. The source node (Node 0) delivers packets to the destination node (Node 9) through Node 1 to 8. The number of data slots in the data phase is 10, and notation "–" means a reserved slot by other connections and is not available. From this example, *path_BW* (1, 0) = *link_BW* (1, 0) = *free_slot* (0) ∩ *free_slot* (1) = {0, 7, 8}. *path_BW* (2, 0) is calculated from *path_BW* (1, 0) and *link_BW* (2, 1), which is equal to {1, 8}. Then *path_BW* (9, 0) is recursively calculated from *path_BW* (2, 0) and *path_BW* (2, 0), and is equal to {2, 5}. As a result, bandwidth requirement is two data slots per frame in this example. After calculating the end-to-end bandwidth, the data slots need to be reserved from the destination (Node 9) hop-by-hop backward to the source (Node 0). And the reservation wouldn't be released until the end of the session. In the end, Node 0 begins transmitting datagram on completing the reservation.

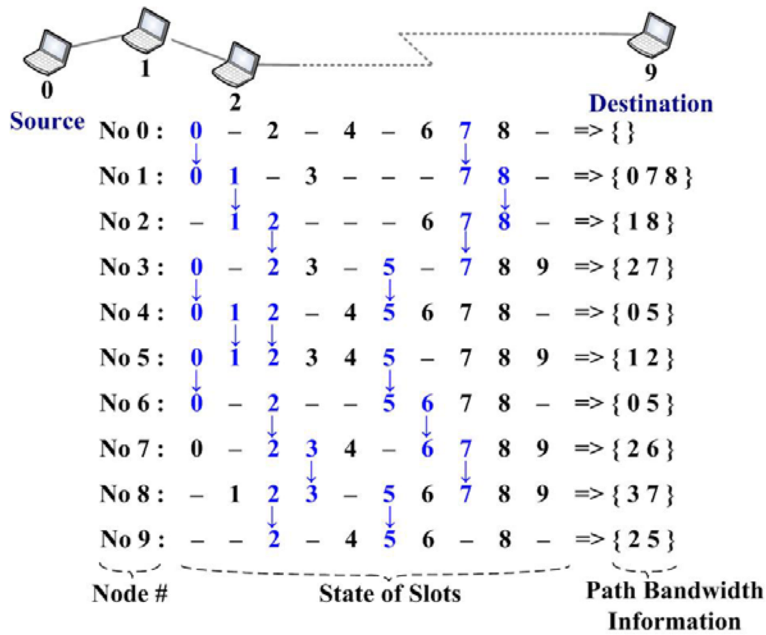

Figure 2-5: Path bandwidth calculation example.

In general, to compute the available bandwidth for a path in a time-slotted network, one not only needs to know the available bandwidth on the link along the path, but also has to determine the scheduling of the free slots. Just like Figure 2-5 illustrated, that path bandwidth from No 0 to No 9 is not intuitively equal to the minimum of link bandwidth on this path. Slot scheduling would influence the utility rate of the network bandwidth; if the free slots weren't scheduled well, path bandwidth might be zero even if the link bandwidth is not zero. So it is important to schedule free slots in path bandwidth calculation step, and there would be more path bandwidth if there is a superior slot scheduling scheme. To resolve slot scheduling at the same time as available bandwidth is searched on the entire path is equivalent to solving the satisfiability problem (SAT), which is known to be a NP-complete problem [12] .

2.3 Routing

In the case of multi-hop networking, most routing protocols for packet radio networks can be categorized as being *before-demand (Table Driven)* or *on-demand* protocols.

2.3.1 Table Driven Routing Protocol: DSDV [13]

The DSDV (Destination-Sequenced Distance-Vector) routing algorithm is based on the idea of the classical Bellmen-Ford Routing Algorithm with certain improvements [14] . It is a hop-by-hop distance vector routing protocol requiring each node to periodically broadcast routing updates [15] . The key advantage of DSDV over traditional distance vector protocol is that it guarantees loop-freedom.

Every mobile station maintains a routing table that lists all available destinations, the number of hops to reach the destination and the sequence number assigned by the destination node. The stations periodically transmit their routing tables to their immediate neighbors. A station also transmits its routing table if a significant change has occurred in its table from the last update sent. So, the update is both time-driven and event-driven.

Because of “before-demand” bandwidth calculation, a host can decide either to accept or reject a new call immediately without any delay for the virtual circuit setup.

Rerouting When the Path Broken

A topological change in the wireless mobile network may destroy the VC; then the connection control must reroute or reestablish the VC over a new path. The routing protocol must be capable of finding new routes quickly when a topological change destroys existing routes. So maintaining secondary paths that can be used immediately when the primary route fails is important. Figure 2-6 shows that the primary route fails, and the secondary route becomes the primary route; then another standby route is constructed. The secondary (standby) route is easily computed using the DSDV algorithm, and it must be decided by computing bandwidth information from all neighbors.

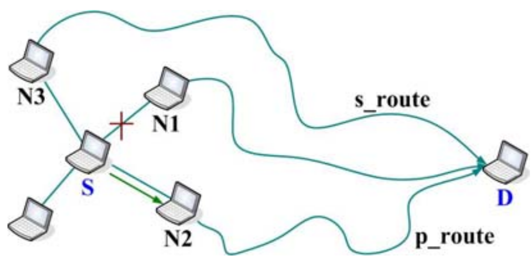

Figure 2-6: Rerouting when the path broken.

2.3.2 The On-Demand Routing Protocol: AODV [26]

Due to the rapidly changing availability of resources and processing delay, the on-demand routing is another choice for bandwidth calculation. On-demand routing can also save the control messages for maintaining inactive routes, but there will be delay for the virtual circuit setup. AODV (Ad Hoc On-Demand Distance Vector) routing algorithm is an improvement on the DSDV algorithm which minimizes the number of broadcasts by creating routes on-demand as opposed to DVDS that maintains the list of all the routes. It is essentially a combination of both DSR [27] and DSDV; it borrows the basic on-demand mechanism of Route Discovery and Route Maintenance from DSR, plus the use of hop-by-hop routing, sequence numbers, and periodic beacons from DSDV.

Route Discovery

The pure on-demand rule neither maintains any routing table nor exchange routing information periodically. When a source node wants to communicate with another node for which it has no routing information, it floods a route request (RREQ) packet to its neighbors. If the topology exists, a route from the source to the destination RREQ will find it.

Route Reservation

When the destination node receives one RREQ packet from the source node, it returns a route reply (RREP) packet back. As the RREP traverse back to the source, each node along the path reserves those free slots which were calculated in advance. The end-to-end bandwidth reservation is successful when the source receives an RREP; and the VC is established. Then the source node can begin transmitting datagram.

Unsuccessful Reservation

The reservation operation may not be successful when the RREP travels back to the source. There are two different NACK situations, RESERVE_FAIL and NO_ROUTE. All nodes on the route from the interrupted node to the destination must free the reserved data slots in the former case. The source can either restart the discovery process or reject the new call in the latter case.

Connection Breakage

A topological change may destroy a VC; once a route is broken, the breakpoints send a special NACK (i.e., ROUTE_BROKEN) to the source and the destination. Upon receiving the ROUTE_BROKEN, the source restarts the discovery process to reestablish a VC over a new path, and the destination only drops the ROUTE_BROKEN. This procedure is repeated until either the completion of data delivery or timeout. Figure 2-7 summarizes the operation of the admission control over the on-demand routing algorithm.

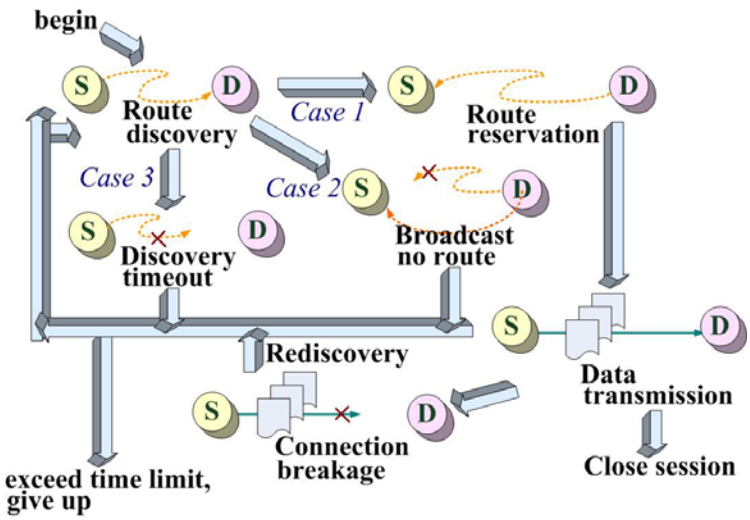

Figure 2-7: Overview of the on-demand routing algorithm.

2.4 SWAN [28]

SWAN is a stateless network model which uses distributed control algorithms to deliver service differentiation in mobile ad hoc networks in a simple, scalable and robust manner. The SWAN model includes a number of mechanisms used to support rate regulation of best effort traffic. It is a best effort MAC and a simple, distributed, and stateless network model that uses feedback-based control mechanisms to support soft real-time services and service differentiation in wireless ad hoc networks. An important benefit of SWAN is that it is independent of the underlying MAC layer, and can be potentially suited to a class of physical/data link wireless standards.

In SWAN, individual nodes along the route from a source to a destination do not maintain any state information regarding the admitting real-time flows [29] . In SWAN, admission control is performed at the source node only. The source decides whether to admit or reject a real-time flow based on the available bandwidth along the path to the destination node. Intermediate nodes regulate their best-effort traffic to meet the QoS needs of real-time flows routed through them. The delay at the MAC layer is used to determine the rate at which to regulate the best-effort traffic.

The SWAN model uses “probing” to obtain the minimal available bandwidth on the path, assuming the routing protocol has found a valid path [30] . The admission control at the source node is then based on the probed bandwidth information. SWAN also proposes to use rate control to manage the best-effort traffic for responding to network congestion. The SWAN model uses “probing” to get the bandwidth information on-demand and thus the call admission delay experienced will be quite large.