Chapter 3

Isochronets and Route Division Multiple Access (RDMA)

3.1 Introduction

3.1.1 The Problem

HSNs set forth novel challenges in designing a network architecture, the most important of which have been detailed in Chapter 2 and summarized in Figure 2.3. From the figure, it can be concluded that many proposed architectures address some of the challenges, but fail to tackle core aspects of HSN operations. The goal of this chapter is to develop a new HSNSA from the requirements set forth in Chapter 2.

3.1.2 Main Results

Isochronets. Isochronets is a novel HSNSA that satisfies the requirements in Chapter 2. The main guideline behind the Isochronets switching architecture design is to eliminate any processing in intermediate network nodes that is dependent on frame headers. This goal is achieved by making network operations independent of frame structures.

Route Division Multiple Access (RDMA). RDMA is the novel switching architecture in Isochronets. Switching in Isochronets happens based on time, similar to how traffic

lights control road intersections. Traffic lights enable routes solely based on time, according to pre-set schedules. Accordingly, switching is Isochronets is mainly accomplished by mapping node input port to output port configurations based on time. The net effect is to divide bandwidth among routes. RDMA does not impose any frame structure and, consequently, multiple frames pertaining to diverse protocols can be switched without any adaptation inside the network. Additionally, by controlling the manner in which switching takes place, Isochronets may offer a spectrum of guaranteed QoS.

3.1.3 Chapter Organization

The reminder of this chapter is organized as follows. Section 3.2 presents Isochronets and their switching technique: RDMA. Section 3.3 evaluates the suitability of Isochronets for HSNs using the qualitative measures defined in Chapter 2. Section 3.4 positions RDMA in context with respect to PS and CS. Section 3.5 summarizes the main characteristics of Isochronets.

3.2 Isochronets and Route Division Multiple Access (RDMA)

Isochronets is a novel switching architecture tailored to addresses the issues discussed in the previous section. Isochronets *coordinate* traffic motions in the network over time, in similarity to how green waves coordinate traffic motions in roads. In the latter scenario, coordination is accomplished through synchronized traffic lights to create routes for uninterrupted and collision-free motions from source to destinations. Over time, the traffic lights are switched to cover all necessary routes.

In Isochronets, coordination is similarly accomplished by enabling a set of routes

through periodic configuring of network switches. Switches set up routing trees, that is, routes from all nodes to a given destination. A frame is switched by following its path through an enabled tree. Coordination is among routing trees to solve routing ambiguities. That is, two trees that share a link in the same direction cannot be enabled concurrently. The coordination mechanism is called *Route Division Multiple Access (RDMA)* because it divides network bandwidth among routing trees. Routing trees become available periodically for a given interval of time or *band*.

3.2.1 Operation

In RDMA, the basic construct used to schedule traffic motion is a time-band (*green-band*) assigned to a routing tree, as depicted in see Figure 3.1. The green-band is an interval of time during which network switches are configured to route according to a particular routing tree (that is, the routing tree is *enabled*). During the green-band (shaded), a frame transmitted by a source will propagate down the routing tree to the destination root. If no other traffic contends for the tree, it will move uninterrupted, as depicted by the straight line. Additionally, intermediate nodes perform no processing that is dependent on frame header contents.

The green-band is maintained at switching nodes through timers synchronized to reflect latency along tree links. Synchronization is per band size, which is large compared to frame transmission time. It can thus be accomplished through relatively simple mechanisms. Routing along a green-band is accomplished by configuring switches to schedule frames on incoming tree links to the respective outgoing tree link. A source sends frames by scheduling transmissions to the green bands of its destination.

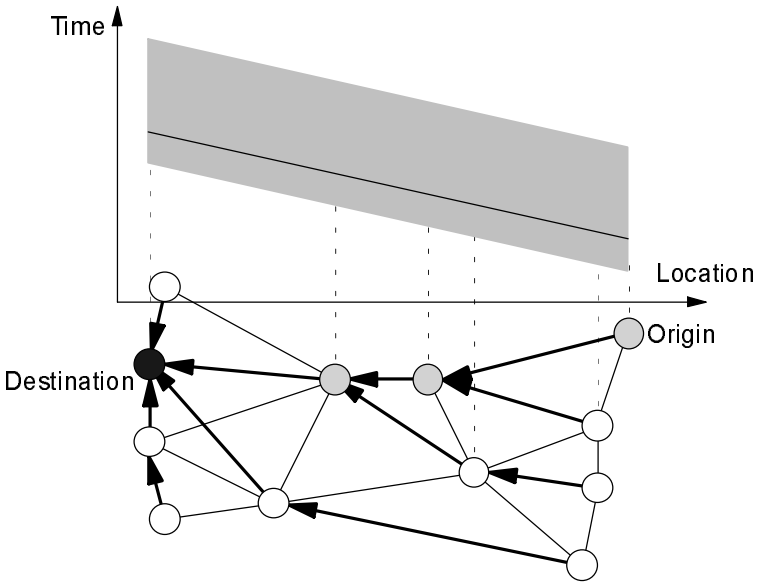

Figure 3.1: Green-band

In similarity to CS slots, green-bands allocate reserved network resources. However, the units to which resources are allocated are neither point-to-point connections, nor traffic bursts, but routes. Routes represent long-lived entities and, thus, processing and scheduling complexities can be resolved over time scales much longer than latency.

3.2.2 An Example

To illustrate the concepts in Isochronet operations, assume that the network topology is as shown in Figure 3.2. A possible routing tree allocation in the network is shown in Figure 3.3. Notice that the trees to A and C do not interfere and can coexist in the same band. The same is true for the trees to B and D. Nonetheless, for example, the trees to A and B interfere because they share links in the same direction.

A possible band allocation is depicted in Figure 3.4. The clock contains 12 units and is divided among the two bands necessary for the allocation in Figure 3.3. Trees A and C are active simultaneously (share the same band) between 12 and 8:15. Trees B and D are ac-

tive between 8:15 and 12. In addition, the tree to B gives priority to source D between 10 and 11.

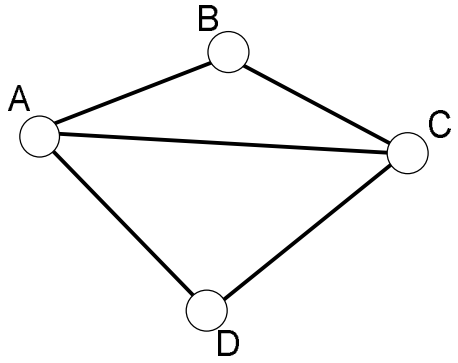

Figure 3.2: Example network topology

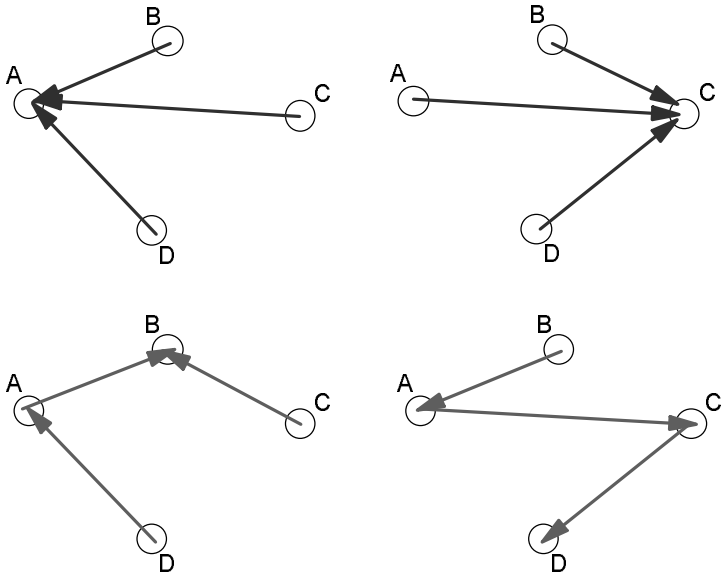

Figure 3.3: Example routing tree allocation

3.2.3 Band Allocation and Contention Resolution in RDMA

The allocation of synchronized time bands to routing trees and resolution of frame collisions are the primitive constructs used by RDMA to control traffic motions and QoS.

Green bands (or *contention bands*) are allocated according to the traffic demands per

tree. The same band can be of different sizes at different nodes in the tree. Indeed, one can view a green band as a resource that is distributed by a node to its up-stream descendants, as long as the bands allocated to descendants are scheduled within the band of the parent.

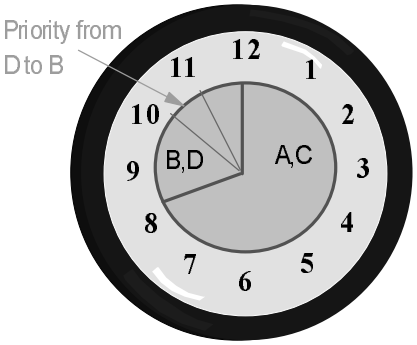

Figure 3.4: Example band allocation

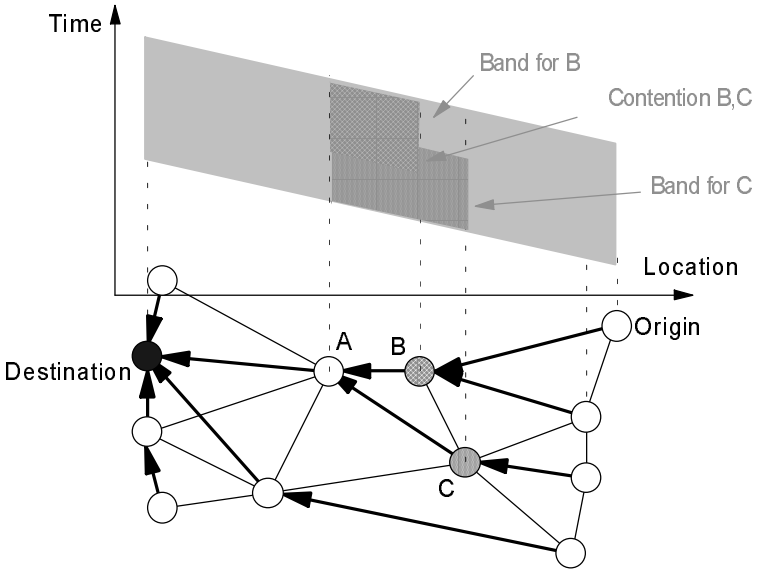

Figure 3.5: Control of contention through degree of band overlap

Figure 3.5 illustrates this concept by showing the band allocation at an intermediate node A to the destination. A partitions its band between its descendants B and C. Notice that the sizes allocated to B and C are different and only overlap partially. In the portion where the bands overlap, traffic from B and C contend in A toward the destination. In

particular, if the bands allocated to two descendants do not overlap, their traffic does not contend. By controlling band overlaps, switches can fine-tune the level of contention and statistical QoS seen by traffic.

Similar considerations can take place in the allocation of priority bands. That is, one may view the bands distributed by node A in Figure 3.5 as priority bands. In such scenario, in the portion when both priority bands overlap, both B and C contend at A. In the other portions, either B or C has priority toward the destination. Notice that, for example, B having priority is equivalent to the whole subtree with root in B having priority to the destination. By finely tuning priority bands, a whole set of priority criteria can be built to solve contention for outgoing links in the trees.

It will be assumed in the reminder of this work that two priority bands in the same node do not overlap. This assures uninterrupted motions during priority bands that are allocated all the way from the source to the destination.

RDMA bands can also be *priority* or *multicast*, in addition to contention bands. Priority bands are allocated to sources requiring absolute QoS guarantees, similar to a circuit service. Traffic from a priority source is given the right of way, by switches on its path, during its priority band. Unlike CS networks, however, priority sources do not own their bands. Contention traffic may access a priority band and use it whenever the priority source does not. During a multicast band, the routing tree is reversed and the root can broadcast to any subset of nodes.

One may view these mechanisms to schedule traffic motions via band allocations as a MAC technique. The entire network is viewed as a routing medium consisting of routing

trees. Bandwidth is time- and space-divided among these routes. Sources need access respective trees during their band times, seeing the network as a time-divided medium, much like Time Division Multiple Access (TDMA) [Halsall 92]. This similarity is the reason for the name Route Division Multiple Access.

Within a contention band, frames may contend competing for the outgoing link. The collision resolution mode used is designed in terms of signs "-", "+", and "++". In RDMA- only one frame proceeds, while the other colliding frames are discarded. In RDMA+, only one frame proceeds upon collision, but the others are buffered. Nevertheless, when the band finishes, all buffered frames are discarded. RDMA++ uses the same collision resolution technique as RDMA+, but stores frames beyond band termination and reschedules them during the next band.

Multiple simultaneous routing trees can schedule transmissions in parallel, that is, have simultaneous green bands, depending on the network topology. For an extreme example consider a fully connected network: all trees to all nodes can be simultaneously active without interference. In more realistic examples, significant parallelism can be accomplished. Figure 3.6 shows two non-interfering routing trees.

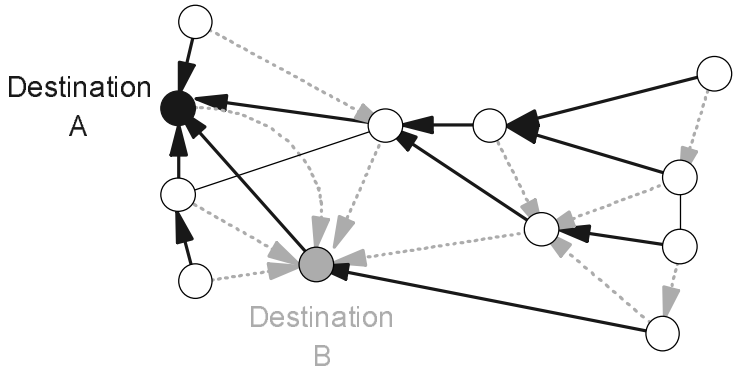

Figure 3.6: Multiple non-interfering trees

An interesting problem to consider is that of selecting clock periods for band repetitions, or the *cycle*. Let U indicate the shortest clock unit used in band allocation. Let P denote the periodicity of the clock measured in U units. For example, let U = 1 μs and P = 125 U; that is, after 125 μs the clock returns to 0. Time may then be indicated in terms of counters similar to seconds, minutes, hours, etc. For example, the time <12, 3>, with the above U and P, means 3 periods 125 μs long plus 12 μs.

Traffic may also vary in terms of typical frame sizes. Consider the choices of U and P above over a 2.4 Gb/s link. During a period of P = 125 μs, some 300 Kb can be transmitted. If the link is equally shared among 3 to 6 trees, this means that each tree can be allocated an average of 50 to 100 Kb. Additionally, since link speeds may vary greatly, Isochronets may wish to use different periodicity over links. For example, a link of 155 Mb/s may use a period of 16 P = 2 ms. Arrivals over this link will be buffered and delivered to higher speed links. Discussion of this general case, however, is beyond the scope of this work.

The band periodicity may vary with the type of traffic served. Low duty traffic such as file transfers may use periods of long duration, whereas interactive voice or video traffic may use much shorter periods.

For the reminder of this work, the physical clock period is fixed at 125 μs. This value is also the sampling rate for voice communications. Bands are allocated as portions of the physical clock.

3.2.4 Protocol Support and Interconnection

All stack layers above the MAC layer are delegated to interfaces at the network pe-

riphery. A typical stack organization for Isochronets is depicted in Figure 3.7.

Interconnection of Isochronets can be accomplished via MAC layer bridges using extensions of current well-understood technologies. Conversions need only handle physical layer interfaces and MAC. Above the MAC layer, interconnection becomes transparent. Contrast this with the problem of internetworking two distinct HSNSAs via higher-layer gateways.

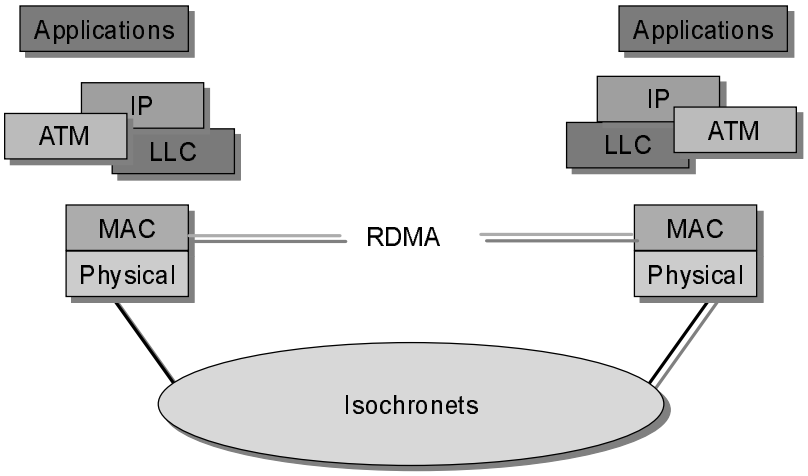

Figure 3.7: Multiple protocol stacks in Isochronets

3.3 Isochronets for HSNs

The main characteristics of Isochronets with respect to the features discussed in Chapter 2 are:

1. *Support end-to-end transport of multi-protocol frames.* Isochronets do not rely on nor need to adapt to any frame structure because frame headers do not need to be parsed inside the network to be switched. Multiple frame structures (e.g., IP, ATM, etc.) can coexist in the network without adaptation.
2. *Support tunable guaranteed QoS of end-to-end transport.* Through band alloca-

tion, sources can be given priority in a tree. When transmitting in this mode, traffic has guaranteed bandwidth and experiences no jitter, loss, nor random delays introduced by contention with other sources. On the other hand, contention bands do not provide any QoS guarantee. By sizing the many kinds of bands at each tree node, an array of services can be tuned. Consequently, QoS is strictly guaranteed through tunable bands.

3. *Support asynchronous, synchronous and isochronous traffic streams.* Synchronization signals are forwarded to network interfaces when bands begin and can be used to synchronize traffic motions. Nevertheless, within a band, motion is asynchronous. Depending on the kind of service a band provides (priority or contention), the end-to-end motion may attain every synchronization need.
4. *Minimize processing and queueing latency in the transport path.* Switching is accomplished by re-configuring nodes over time. Frame headers are not processed to be switched. As a consequence, nodes do not need to process frames at incoming high-speed rates. Rather, nodes need only re-configure themselves periodically. Additionally, queueing only happens to access outgoing links and can be finely controlled, and even avoided completely, by assigning priorities to sources in a band.
5. *Support speed-scalability for link bandwidth ranging from hundreds of megabits to terabits per second.* On the one hand, transmission speeds are independent of control speeds. Control is responsible for node configuration, synchronization, and band allocation. These mechanisms can be implemented completely off-line with respect to transmissions. On the other hand, switching only depends on link speeds for frame detection and contention resolution. These are minimal functions for any switching technique that allows some degree of frame multiplexing. It will be shown in Chapter 6 that these mechanisms can be implemented very efficiently and special techniques can be used to make the dependency on link speeds virtually

non-existent.

6. *Support efficient internetworking with existing stacks and media.* The fact that no frame adaptation is necessary together with the fact that Isochronets may operate on any media (electronic or optical) because it does not process frames make interconnection with existing protocol stacks simple.
7. *Support scalability for large size networks (ranging from ten to hundred million end nodes).* Isochronets offer a degree of scalability between PS and CS. Bandwidth is dedicated (as opposed to available on-demand as in PS), but to routing trees (instead of circuits as in CS), thus enabling a large degree of multiplexing and consequently supporting scalability in number of nodes.
8. *Support efficient interfacing with end nodes.* Band allocation is the single mechanism used to finely tune traffic demands in the network. This simplifies network interface. The interface must receive the synchronization signals from the network indicating the beginning of bands and forward them to upper layer protocol stacks. It also must be able to perform the usual reception and forwarding of frames.
9. *Support efficient bandwidth sharing and flexible adaptation to service demands.* Bandwidth sharing can be controlled through the band allocation mechanism. The band of a tree may be presented in smaller sections to the sources and can thus be used to control the level of contention in the network. On one extreme, all the band is available to all sources, maximizing network resource use. On the other extreme, the band is partitioned in non-overlapping intervals at the sources, thus maximizing QoS. In between, an array of performance behaviors can be carefully tuned.

3.4 RDMA between CS and PS

This section positions RDMA in the context of PS and CS.

If the band associated with a routing tree consists of priority-bands only, that tree is operated in an optimized CS mode. That is, each source is allocated a circuit (priority band) to the tree root. The form of CS supported by Isochronets is superior to traditional CS as circuits only get priority over band usage but do not own it. In a situation where the entire band has been allocated to priority bands serving real-time isochronous traffic, non-real-time data traffic may take advantage of underutilized parts of the band.

Consider now an Isochronet operating in RDMA++ contention resolution mode. If the entire band is allocated to contention traffic, frames moving down the tree will be stored and forwarded as in an ordinary PS network. The form of PS supported by Isochronets is advantageous to traditional one in a few ways. First, Isochronets support virtual cut-through mechanisms as frames arriving to a free switch will continue without store-and-forward delays. Second, no headers are processed in Isochronet switches. Third, buffered frames are aggregated into larger units and transmitted at once, improving the efficiency of buffer retrieval. Fourth, contention happens only among frames to the same destination (and not among non-correlated traffic).

Isochronets, it may be argued, could potentially under-perform PS networks due to the time-division of bandwidth among routes. In situations where significant traffic bursts are randomly generated at multiple routes while other routes are empty, the bandwidth committed to unused routes will be underutilized while the routes serving a burst may have insufficient bandwidth to handle it. A PS network would have permitted the traffic burst to move into the network and utilize its entire band without pre-allocation. Typically, however, admission-control policies will prevent large bursts from entering the net-

work. Such mechanisms as leaky-bucket [Sidi et al. 89] reduce the effective bandwidth available to any given source. A PS network governed by admission policies that limit source bandwidth, presents no advantage over an Isochronet that limits the bandwidth to sources through pre-allocation to routes.

In summary, at the two extremes, Isochronets compare favorably with CS or PS networks. In-between, Isochronets can be configured to span a spectrum of switching techniques of superior performance characteristics to both.

3.5 Summary

Isochronets substantially differ from other HSN architectures. Isochronets virtually eliminate the network layer, reducing it to the MAC layer. As a result: (1) no frame processing (for routing, switching, etc.) is required in the network; (2) there is no need for adaptation layer between different protocol stacks at network interfaces; (3) internetworking is reduced to MAC layer bridging; (4) the network can adapt to the frame sizes and arrival statistics of sources.

In Isochronets, network control functions are entirely separated from transmission activities, which render them capable of: (1) transmission-speed elasticity—transmission speeds can be arbitrarily faster than control speeds; (2) distance elasticity—the network can extend over local, metropolitan, and wide areas; (3) bandwidth-heterogeneity—the network can incorporate links of different transmission speeds; (4) media independence—electronic or all-optical implementations.