APPLICATION MANAGEMENT AND RESOURCE SCHEDULING IN SOFTWARE DEFINED BANDWIDTH ON DEMAND NETWORKS

Dragomir D. Dimitrijević
Consultant, Belgrade, Yugoslavia

Dale Berisford
Telogy Networks, Inc., Germantown, USA

I INTRODUCTION

Satellite communications [1] is an evolving technology with a variety of telecommunications applications [7]. A comprehensive overview of satellite communications systems may be found in [6]. Description of satellite communications sub-systems is described in [2] and [4]. Authors in [3] emphasize advantage in broadcasting that satellite communications systems have over terrestrial communications systems.

Advances in Very Small Aperture Terminal (VSAT) technology provided fertile ground for effective use of satellites in global communications. There are several projects [5] competing for their position in global communications. They differ in communications technology such as Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), or Code Division Multiple Access (CDMA). They also differ in satellite deployment strategy: small number of large satellites in high geostationary orbits (35,860 kilometers) or large number of small Low Earth Orbit (LEO) satellites (between 500 and 1,400 kilometers). The proposed projects differ in their satellite network topology: individually operating satellites or satellites mutually connected in a mesh network topology.

No matter which project wins, the remaining issue will be user's prospective of effective use and management of such a satellite network. A satellite communications network consists of a number of satellite resources such as bandwidth and power as well as terrestrial resources such as modems. The effective use of satellite power and bandwidth favors Bandwidth On Demand (BOD) assignment over permanently assigned circuits. This means that satellite resources are assigned only when they are actually needed and used. In such a way, a higher utilization of resources may be achieved and they may be available to a larger number of potential users.

This paper describes a software defined BOD network architecture. The "software defined" means that the network configuration, as perceived by end-users (e,g,, connectivity and network services), may be dynamically modified from a centralized network management center. Furthermore, end-users may remotely request modifications of network configuration tailored to their evolving needs. Changes may be executed immediately, or may be scheduled for a later time. The network management system keeps track on current and future availability of satellite and terrestrial resources, and accordingly may or may not grant user requests.

The paper is organized as follows. The BOD networks, their architecture, and services are described in Section II. Resources in BOD networks and their management are described in Section III. Algorithms for resource scheduling are described in Section IV. Conclusions are given in Section V.

II BOD NETWORKS

In this section we describe BOD networks, their architecture, services, users, and desirable features of their network management systems.

II.1 NETWORK ARCHITECTURE

Figure 1 depicts the network architecture. Connectivity of a full mesh network is achieved via one or more satellites. There are multiple sites which may communicate via the satellite(s). Each site has a modem which is used as a device controller. It controls other, third party modems, at the site. Overall control of the network is achieved from the centralized network management center. It controls remote

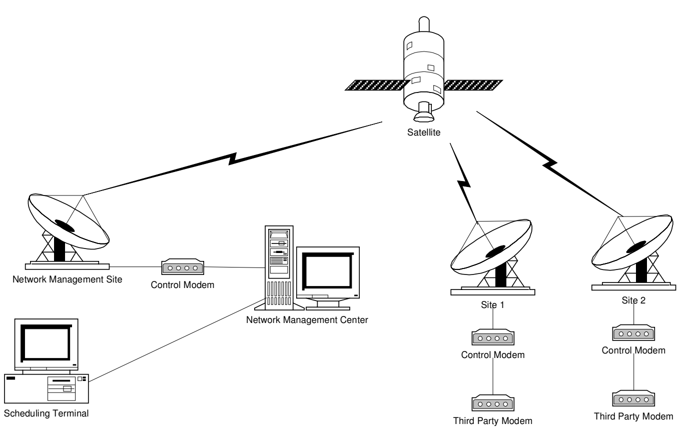

Figure 1: BOD Network Architecture

devices in the network by communicating with other control modems via its own control modem.

Communication in the network is achieved using the Single Channel Per Carrier (SCPC) technology. When a request for a circuit arrives, it is established using the control modems. Multiple messages are combined in a single channel using the FDMA.

To successfully establish a connection, the sufficient bandwidth on an appropriate satellite transponder must be available at the requested time. There must be enough power available, i.e., power may be unavailable due to other activities on the satellite. Also, all participating modems must be available.

When sites no longer need to communicate, the SCPC circuit is automatically released and allocated resources are freed for subsequent use.

Events, e.g., video conferences, may be ordered/scheduled on-line directly by eligible users who have their user accounts created in the system. The users log into the system from remote Scheduling Terminals, and based on the specified requests and availability of the system, order events. For customers without access to Scheduling Terminals, events may be ordered over the phone, and entered in the system by the satellite services provider personnel (Schedulers).

To achieve a complete control of such a network, the system described in this paper keeps track of scheduled changes in the dynamically modified network configuration. Some modems are transportable (located on trucks) and roam from site to site. Also, antennas on satellite transponders may be repositioned to provide different coverage.

The network management center keeps track of ongoing and scheduled upcoming events. It automatically starts/stops/(re)configures remote third party modems. Network operators may monitor state of the network in real-time, view network alarms, events in progress, upcoming scheduled events, and upcoming scheduled changes in the network configuration.

The described network architecture is an excellent framework for management of a flexible, vendor-independent network.

II.2 NETWORK APPLICATIONS

There are two types of BOD applications (requests for connection):

Applications such as telephone calls may be initiated ad-hoc upon request (on demand). Subject to availability of requested resources, the connection is established immediately. Duration of requests is generally not known in advance.

So called, occasional services, are applications like video conferences and TV broadcasts. They are scheduled in advance and their start and stop times are generally known. The term "occasional" is a somewhat confusing industry-wide term which contradicts the fact that there may be hundreds of evens scheduled per day. There may be some variations on scheduling policy which depend on business practices of a satellite services provider. For example, an event may be scheduled with an "approximate end" when some slack to completion time is allowed. The service provider will attempt to keep slack period available and allow the user to go over time. The user makes the decision about the use of slack time during the event (e.g., over time in a sport event broadcast). However, availability of additional time is not guaranteed at the time of event.

Each event in the network may require one or more transmitting modems. Each transmitting modem may be heard by one or more receiving modems. Depending on the number and function of participating modems, there are three typical cases:

Point-To-Point connection comprises of two modems which may transmit and receive simultaneously (full duplex). This configuration may be used for telephone calls.

Broadcast connection requires one transmitting and several receiving modems. This configuration may be used for TV broadcasting.

Multicast connection may be used for video conferencing. Participants in the event send their audio/video signals to a central location which must have one receiving modem for each participant. Using a video equipment, the received video signals are combined into a single audio/video signal (multiple pictures in a picture) and broadcast back to all participants.

II.3 SYSTEM USERS

There are several types of system users with various levels of privileges:

Administrators maintain the system and have access to all information about the system (e.g., system resources and user accounts).

Schedulers handle orders over the phone or e-mail and generally have access to information about all scheduled events.

Satellite operators monitor ongoing and immediately upcoming events on a real-time monitor. They also handle scheduled repositioning of antennas on the satellite transponders.

General users are customers of the satellite services provider. They may remotely log into the system and order/schedule events. Their view to the information in the system is restricted to general information about availability of the system resources as well as their own events.

III RESOURCES IN BOD NETWORKS

In this section we discuss the problem of resource management and scheduling. First, we define resource management operations. Then we describe general types of resources, and finally we address resources specific to BOD networks.

III.1 RESOURCE MANAGEMENT

The system described in this paper allows multiple users to access (read or modify) information (characteristics/availability) about system resources. For that reason, the information is kept in a centralized database which may be accessed in a distributed (client/server) manner. Its consistency is preserved using standard database/record locking mechanisms.

There are several operations that may be performed with a resource:

Creation is the process of defining a new resource and making it available for subsequent scheduling.

Modification is the process of changing characteristics of a resource. Care must be taken that changes do not conflict with requirements of events which are already scheduled to use the resource.

Deletion is the process of removing a resource from the system. Care must be taken of events which are scheduled to use the resource.

Inquiry is the process of finding available resources subject to specified characteristics like availability in time or physical characteristics of a resource. The user must be able to partially specify needs and gradually narrow down the selection. For example, a partially specified time requirement would be duration of an event (flexible start time) while a complete time requirement is start time and duration (or equivalently start and stop time). Based on the specified requirements, available resources must be listed in a prioritized way so that less valuable (less flexible) resources are scheduled first.

Scheduling is the process of reserving an available resource, thus making it unavailable for subsequent requests. A special type of scheduling is blocking resources in certain intervals so they cannot be used although they are not assigned to any particular event (e.g., maintenance). An important characteristics of a scheduling algorithm is the amount of storage needed to keep information about scheduled events.

Unscheduling is the process of freeing scheduled resources.

Garbage collection is the process of removing useless information from the system. For example, information about an available time-slot should be removed after the end time of that slot has passed.

III.2 General Resource Types

There are three different general resource types:

Indivisible resources are scheduled on all-or-nothing basis. At any point of time, a resource is either entirely available or entirely assigned to a particular event.

Range divisible resources are scheduled in a two-dimensional scheduling space. One dimension is time while the other dimension is application specific. In our case, the other dimension is frequency. A scheduled event is represented as a rectangle in the time/frequency domain meaning that other events cannot use the resource in the specified time and frequency range.

Percentage divisible resources may be shared among various events at same time. Unlike for range divisible resources, it is irrelevant which particular section of the resource is scheduled. In our case, power on the satellite is the percentage divisible resource. Availability of such a resource may be represented as a bar chart. A request may be granted if the requested amount of power is below the bar chart (available power) in the requested time interval.

There may be application specific variations of the way the three resource types are scheduled.

III.3 BOD NETWORK SPECIFIC RESOURCES

In this section we discuss BOD network specific resources and their characteristics.

III.3.1 SATELLITE/TRANSPONDER BANDWIDTH POOLS

The reservation system described in this paper may handle multiple satellites. Each satellite has multiple transponders. Each transponder has one or more bandwidth pools. The bandwidth pools are resources which are scheduled as range divisible (two-dimensional) resources. The hierarchy of information is shown in Figure 2.

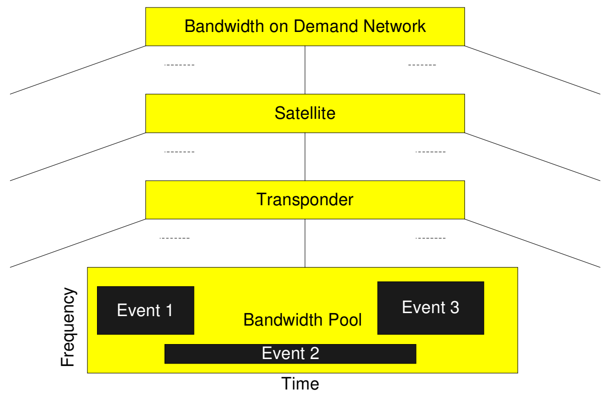

Figure 2: Satellites/Transponders/Bandwidth Pools Hierarchy

Characteristics of transponders, relevant to their proper selection, are as follows.

Uplink access is the general area on the Earth from which the transponder may be reached. An example may be "North America".

Downlink access is the general area on the Earth which may be reached from the transponder. It is also known as "transponder's footprint". These areas may be prioritized. For example, let transponders A and B have downlink access "Europe" and "North Europe". Transponder A has a broad beam and transponder B has a spot beam. In case we need to reach London, transponder B is a better choice since transponder A has broader coverage and should be preserved for later use by requests which cannot be handled by transponder B (e.g., broadcast to London and Athens).

In practice, two transponders may be "networked" and used together to cover very large areas such as multiple continents. For example, transponder A may have uplink access in North America and downlink access in Europe. Transponder B may have the opposite uplink and downlink accesses. If networked together, they may be used to broadcast signal from one continent to both continents. In the system described in this paper, we consider such transponders as a single valuable (high priority) transponder whose downlink access is "North America/Europe".

The receiving and transmitting antennas on a transponder may be fixed or moveable. This means that, for example, downlink access may be "North Europe" in one period of time and "Central Europe" in another. To handle this case, uplink and downlink accesses must be scheduled in time. The system described in this paper generates so called "switch orders" and warns satellite operators ahead of time that a transponder's antenna should be repositioned.

Available bandwidth of each transponder may be divided in one or more bandwidth pools. They are scheduled and shared among multiple events as range divisible resources.

III.3.2 SATELLITE/TRANSPONDER POWER

Power is yet another resource which must be scheduled. It may happen that there is enough bandwidth on an appropriate transponder, but a request cannot be granted since there is not enough power available. Power may be either a global resource on a satellite or each transponder may have its pre-assigned budget of available power.

As it was mentioned before, power is scheduled as a percentage divisible resource and its availability in time may be presented as a bar chart.

III.3.3 MODEMS

Modems eligible for participation in events are selected based on the following criteria:

Communications characteristics are frequency range/bandwidth, modulation type, bandwidth overhead etc. The system described in this paper maintains user definable link profiles which are combinations of characteristics bundled under symbolic names such as "16k voice". This feature facilitates selection of appropriate modems.

Modem's location defines network topology. The modem's location influences eligibility of transponders and their bandwidth pools. For example, the user may first request New York as a transmitting location. This will narrow-down the list of eligible transmitting modems to those that are located in New York, and eligible transponders to those which have North America for uplink access. Then, the user may add London as a receiving location. This will further narrow-down eligible transponders to those which have Europe or North Europe as downlink access. If the user adds Athens as the second receiving location, only transponders whose downlink access is Europe will be eligible.

Scheduling of modems is further complicated by the fact that large portion of modems used by satellite service providers are transportable, i.e., their location is not fixed. To handle this case, locations of modems must be scheduled in time. For example, a modem may be in Berlin today, nowhere tomorrow (i.e., on the road), and in Paris in two days.

The system maintains the list of known sites. Generally, granularity of sites is at the level cities, but may also be more or less coarse. The system also maintains a relationship between sites and uplink/downlink areas. For example, London belongs both to Europe (broad beam transponder) and North Europe (spot beam transponder). The system facilitates selection which provides the best coverage or quality of signal on the ground. Sites which are closer to the edge of transponder's footprint have poorer quality of signal. The system maintains this information for each site and each footprint the site belongs to.

IV SCHEDULING ALGORITHMS

In this section we describe scheduling algorithms for uplink/downlink access, modems, bandwidth pools, and power. Due to their complexity, the algorithms will just be outlined. Uplink/downlink access and modems are handled as indivisible resources with slight variations. Bandwidth pools are scheduled as range divisible (two-dimensional) resources. Power is scheduled as a percentage divisible resource.

Information on resource availability is stored in a centralized database. The users (partially) specify their requests and gradually narrow down their choice. The requests, specified via a user-friendly graphical interface, are translated in Structured Query Language (SQL) statements. They are used to inquiry the database on availability of resources. Simplicity and efficiency of the SQL statements was the major factor in the design of data structures used to store information on resource availability.

IV.1 UPLINK/DOWNLINK ACCESS

Figure 3 depicts a sample schedule of a transponder's uplink/downlink access. Three blocks are scheduled during the life-time of a transponder (satellite). Let us assume that we have to reposition transponder's antenna in the middle section so it points to South Europe. We may have to split the middle section to up to three blocks depending on the start and stop times of the South Europe block. The change cannot be performed in case there are events scheduled on the transponder during the new position interval.

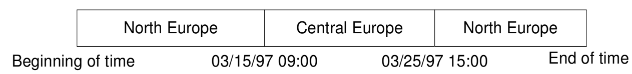

Figure 3: Uplink/Downlink Access Schedule - An Example

The number of records per transponder in the database is equal to the number of changes in transponder's antenna position.

IV.2 Modems

Figure 4 depicts an example of a transportable modem schedule. The time-line is divided in time intervals. Black blocks denote busy intervals. Two back-to-back events are scheduled in the figure. White blocks denote available intervals, and gray blocks denote intervals when the modem is incommunicado (e.g., on the road).

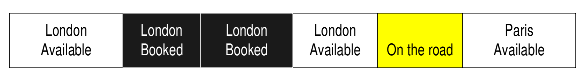

Figure 4: Modem Schedule - An Example

In order to be eligible for scheduling, an available (white) block must completely contain the requested interval, and the modem's location must match the requested location.

Depending on the relationship between requested and available start/stop times, the originally available slot may be completely booked, or split to one available and one unavailable interval, or split to two available and one unavailable interval in the middle of the originally available interval.

If an event is canceled, the freed slot must be merged with any of the adjacent free slots in the same location. This will form a larger available interval.

As a part of garbage collection, any free slot should be deleted after it becomes obsolete, i.e., its end passes.

The number of records per modem in schedule is proportional to two times the number of scheduled events plus the number of times the modem is relocated. This is the worst case (maximum fragmentation of available blocks) is when there are no back-to-back events. The best case is equal to the number of scheduled events plus the number of times the modem is relocated. This happens when all events are scheduled back-to-back.

IV.3 BANDWIDTH POOLS

The two-dimensional scheduling of bandwidth pools is more complex. Figure 5 depicts two-dimensional schedule in the process of scheduling two events. Available (white) and unavailable (gray) blocks are represented as rectangles in the two-dimensional time/ frequency space. In order to be able to schedule a new request (a rectangle in time/frequency domain), prior modification of schedule, it must fit in at least one available block.

Unavailable blocks cannot overlap with any other block (available nor unavailable). Available blocks may mutually overlap. In order to keep the amount of information at minimum, no available block may be a subset of another available block. Note that, in order to better view boundaries of adjacent/overlapping blocks in Figure 5, blocks are slightly shifted with respect to each other.

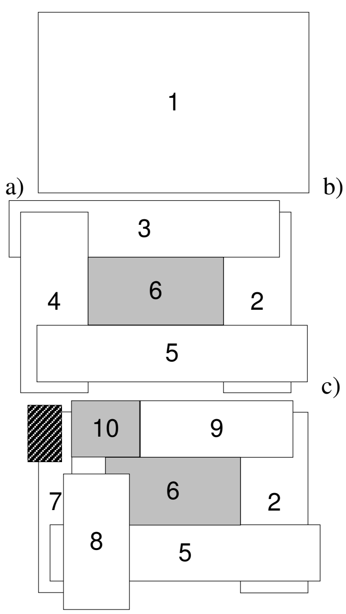

Figure 5: Bandwidth Pool Scheduling - An Example

Figure 5a depicts the initially empty schedule when the entire time/frequency domain is available, and schedule is represented as a single big available block. The first request, block number 6 in Figure 5b, fits in the initial empty block number 1 in Figure 5a. After the first request is scheduled, the initial available block is split into 1 unavailable and 4 available blocks in Figure 6b. As it can be seen, some of the available blocks overlap (e.g., 3 and 4).

The next request, block number 10 in Figure 5c, fits in the available block number 3 in Figure 5b. It also overlaps with available block number 4. After the second request is scheduled, block number 3 is split in three blocks: available block 9, unavailable block 10, and an available small black block. Also, block 4 is split in available blocks 7 and 8 which overlap. Since the small black block in Figure 5c is a subset of block 7, it does not provide any additional information on bandwidth pool availability. Future requests that fit in the black box will also fit in block 7. Therefore, the black box is discarded. Any subsequent request must be a subset of at least one of the white blocks in Figure 5c.

Unscheduling runs in the opposite direction and combines free block in larger free blocks. Garbage collection discards all free blocks prior to (left of) current time.

The number of records per bandwidth pool in schedule (available and unavailable blocks) is proportional to five times the number of events. This is the worst case with maximum fragmentation. The best case is equal to the number of scheduled events.

IV.4 POWER

Figure 6 depicts an example of power scheduling. Prior to scheduling a new request, there were five time intervals (1-5) represented a five bars in the bar chart. The available power is defined by the height of the chart.

A request is issued in the interval defined by the double-headed arrow. The amount of requested power, defined by the height of white rectangles, is smaller than minimum available in the requested interval (bar number 3), and therefore the request may be granted. The same amount of power is deducted from each of the affected intervals. Bars in the middle are just shortened for the same amount. The start of request falls in the middle of interval number 2, so the interval must be split into two smaller intervals. The height of the left bar remains the same while the right bar is shortened for the amount of requested power. The end of request coincide with the end of interval number 4, and therefore this bar does not have to be split to two intervals. However, after bar 4 is shortened, it becomes as high as bar number 5, and therefore they should be merged in a single larger interval.

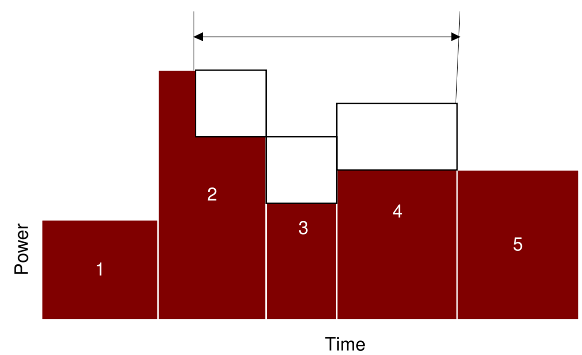

Figure 6: Power Scheduling - An Example

Unscheduling algorithm proceeds in the same way except that the "requested" power is negative. The garbage collection discards all intervals after their ends pass.

The number of records per power pool (one pool per satellite or one pool per transponder) in schedule is proportional to two times the number of events. This is the worst case with maximum fragmentation. The best case is equal to one when, e.g., all events are scheduled back-to-back and all require the same amount of power.

V CONCLUSIONS

In this paper, we presented scheduling algorithms for various resources (satellite power and bandwidth, and terrestrial modems) in a software defined BOD satellite network. They allow ad-hoc bandwidth allocation (e.g., telephone calls) as well as events scheduled ahead of time (e.g., video conferences). The presented network architecture is a unique framework for creating a user-configurable, flexible, vendor-independent network.

REFERENCES

[1] J. Martin, *Communications Satellite Systems*, Prentice Hall, December 1977.

[2] Williamson, *The Communications Satellite*, Adam Hilger, February 1990.

[3] M. J. Miller, B. Vucetic, and L. Berry*, Satellite Communications: Mobile and Fixed Services*, Kluwer Academic Pub., May 1993.

[4] Calcutt, and L. Tetley, *Satellite Communications: Principles and Applications*, Van Nostrand Reinhold, June 1994.

[5] Gilder, “Ethersphere”, Forbes ASAP, October 10, 1994.

[6] Richharia, *Satellite Communications Systems: Design Principles*, McGraw Hill, January, 1995.

[7] D. Roddy, *Satellite Communications*, McGraw Hill, September, 1995.

Abstract: This paper presents algorithms for application management and resource scheduling/ allocation in software defined Bandwidth on Demand (BOD) satellite networks. Resources in such networks are satellite bandwidth and power, and terrestrial modems. Subject to their availability and type of request, resources may be acquired immediately (telephone calls) or reserved for a later time (scheduled video conferences or TV broadcasts).

The architecture is unique in that: (*i*) it is an excellent framework for management of a flexible, vendor-independent network; (*ii*) it allows the end-users to, subject to their system privileges, on-line inquiry resource availability, request network resources, and reconfigure the network; (*iii*) it is built to suit business needs of satellite services providers.

APPLICATION MANAGEMENT AND RESOURCE SCHEDULING IN SOFTWARE DEFINED BANDWIDTH ON DEMAND NETWORKS

Dragomir D. Dimitrijević, Dale Berisford