IEEE COMMUNICATIONS MAGAZINE

SATELLITE MULTIPLE ACCESS PROTOCOLS

C. RETNADHAS

Protocol for accessing satellites efficiently: a tutorial.

SATELLITES provide a convenient medium for data communication between widespread geographic areas. Compared to a terrestrial data network, the satellite system has wide bandwidth, high accuracy of transmission, and high availability of transmission medium. The main disadvantages of the satellite system are the inherently long transmission delays (270 ms one way, the effect of local weather conditions and interferences, and the high cost of the system. Technological advances can reduce the cost and the effect of weather conditions on the transmitted signal. The effect of long transmission delay can be minimized by using effective transmission protocols. Because of the above advantages, satellite technology has aroused a great deal of interest in recent years [7],[14].

This paper presents a tutorial on the various protocols used in satellite data transmission. The most important characteristic of the satellite system is the ability of the earth stations, located at geographically dispersed areas, to access the satellite to transmit and to receive data. The area covered by a geostationary satellite is a function of the satellite's receiving and transmitting antenna(s). For a large number of users with bursty traffic, a highly efficient way of using the channel capacity is to use multiple access techniques. In a multiple accessed channel, two or more users may nominally share the channel. The satellite system can provide broadcast capability at any given time to all earth stations within its transmission coverage area. The combination of multiple access and broadcast capability makes it possible to configure the earth stations into a fully connected "one-hop" network.

CHANNEL DERIVATION

There are three ways to obtain channels in a satellite system [8]. In the first method, the channels are obtained by using the built-in satellite channelization due to the use of multiple transponders operating in different frequency bands. Each one of the independent transponders in the satellite is designed to accept transmission at a selected frequency band, i.e., the uplink frequency. The satellite carries out a frequency translation to a well-defined frequency band, i.e., the downlink frequency. This scheme thus divides the total bandwidth of the satellite into well-defined channels. The advantages provided by this scheme are reduced interference problems and improved reliability in that the possibility of losing all the channels due to satellite failure is small.

The second method uses the basic multiple access techniques of frequency-division multiple access (FDMA), time-division multiple access (TDMA), and code-division multiple access (CDMA).

A number of multiple access protocols are presented. In the final analysis, it is cost which will dictate which protocol is suitable for a particular application.

A simple form of obtaining an FDMA channel is to divide the bandwidth of a transponder into separate nonoverlapping subchannels, with each user assigned a separate subchannel. In FDMA, each user has access to a dedicated portion of the channel at all times. The main advantages of FDMA are that it is simple to implement in that no real-time coordination among transmission of data is needed and can be used to transmit either analog or digital signals. For bursty traffic, the channel utilization is poor. This scheme is cost effective for applications that involve point-to-point trunking.

In TDMA, each user is scheduled to transmit in short nonoverlapping intervals. Therefore, a TDMA scheme requires some form of frame structure and a global timing mechanism to achieve nonoverlapping transmission. For this reason, a TDMA system is more complex to implement than an FDMA. However, an important advantage is the connectivity. This is obtained because all receivers listen on the same channel, while all sources in a TDMA system transmit on the same common channel at different times.

The third method uses dynamic sharing of a channel using demand assignment techniques. This method may be used for circuit-switched voice traffic or packet-switched data traffic

---

The author is with the Department of Quantitative and Information Science, Western Illinois University, Macomb, IL 61455.

0163-6804/80/0900-0016 $00.75 © 1980 IEEE

[8]. In this paper, we confine ourselves to packet-switched data traffic only. The packet-switched data traffic system can be divided into random access, implicit reservation, and explicit reservation systems. In the following discussion, we will assume that the whole of a transponder bandwidth is devoted to multiaccess operation, the up channel at one frequency operating in multiaccess mode and the down channel at another frequency operating in broadcast mode. The earth stations which are visible to the satellite antenna transmit packets at the full available bandwidth. The satellite after frequency translation retransmits packets at the full available bandwidth. The downward packets are received at all earth stations within the satellite's coverage. The earth stations identify packets destined to them by looking at the packet header address. All packets addressed to other stations are ignored and those addressed to the station are passed on to it.

RANDOM ACCESS SYSTEM

One of the protocols used for transmitting packets in a random access satellite system is the ALOHA protocol. In this protocol, each one of the earth stations transmit packets as soon as each one of them has a packet to transmit without regard for other stations. Due to the lack of coordination among the distributed ground stations, packets from different stations may reach the satellite at the same time and collide, thereby destroying the information content. Therefore, a subsequent retransmission of the packet is required. Because of the broadcast capability posed by the down channel, the transmitting station will be able to detect any collision. No acknowledgment is necessary in the satellite system in the event of collision. The collided packets are retransmitted after a further random delay in order to avoid the risk of repeated collisions. The maximum channel capacity that is usable is about 18 percent in the ALOHA protocol.

A substantial increase in usable channel capacity can be obtained by using the S-ALOHA (slotted-ALOHA) protocol. In the S-ALOHA protocol, the satellite channel is slotted into segments whose duration is exactly equal to the transmission time of a single packet (assuming fixed size packets). If the earth stations are synchronized to start the transmission of packets at the beginning of a slot, the channel utilization efficiency increases. In the ALOHA protocol, when a collision takes place, the packets may overlap fully or partially. By using the S-ALOHA protocol, the partial overlap is eliminated. Under certain assumptions about the message traffic generated by the earth stations, the channel utilization efficiency is about 36 percent [1],[9]. This increase in channel utilization efficiency is obtained at the cost of increased complexity in control compared to the ALOHA system.

One of the drawbacks of the random access system is the problem of instability. When large numbers of stations are active, excessive traffic leads to more collision. After collision, the channel traffic consists of both the newly generated packets and the retransmitted packets. As the number of newly generated packets increase, the chance of collision increases. This, in turn, increases the number of retransmissions which, in turn, increases the chance of a collision, and a runaway effect occurs; thus, the channel becomes unstable. In the absence of a control mechanism [5],[10],[13], the collision retransmission may produce a congested condition with the system throughput becoming zero. The purpose of the control is to prevent the channel from reaching the unstable condition, while optimizing channel efficiency and performance during normal operating conditions [13].

The low bandwidth utilization of the ALOHA and the S-ALOHA systems have led to many proposals for increasing utilization by means of slot reservation schemes. The object of slot reservation schemes is to reserve a particular time slot for a given station. This ensures that no collision will take place. In general, it may be possible to achieve potentially high channel efficiency using some form of a reservation technique. This increase in channel utilization efficiency is obtained at some overhead cost, either in terms of allocation of part of the bandwidth for reservation purposes and/or increased complexity of the control mechanisms in transmitting stations. All reservations methods use some form of framing approach, and the reservation scheme can be either implicit or explicit.

IMPLICIT RESERVATION

The implicit reservation protocol uses a frame concept to the S-ALOHA channel to permit implicit reservation. A frame may consist of more than one slot. The total number of slots can be grouped into a set of reserved slots and a set of slots which can be accessed using the S-ALOHA contention protocol. Efficient channel utilization is obtained by allowing stations with high traffic rate access to one or more slots from the reserved set of slots in each frame.

The reservation-ALOHA utilizes this principle with implicit reservation-by-use allocation. Reservation-ALOHA uses distributed control, and each earth station executes an identical allocation algorithm based on the global information available from the channel. Whenever a station successfully transmits in a slot, all the stations internally assign that slot in subsequent frames for exclusive use by the successful station. Thus, each station maintains a history of usage of each channel slot for one frame duration. This slot is reserved to this station until the station is finished using it. The stations use the S-ALOHA contention method to access the unassigned slots in each frame. In this scheme, there is no way to prevent a station from successfully capturing most or all of the slots in a frame for an indefinite time period.

EXPLICIT RESERVATION

These reservation schemes try to make better use of the channel bandwidth by explicitly reserving future channel time for transmitting one or more messages for a specific station. To obtain good performance, the ground stations should cooperate with one another to maintain synchronism. Only by conforming to the reservation discipline can the earth stations ensure that packet collisions will either be eliminated or reduced drastically. In the explicit reservation scheme, the earth stations use part of the channel bandwidth for sending reservations for future time slots. This, to some extent, reduces the total bandwidth available for data transmission. By keeping the bandwidth required for reservation proportionately small compared to that available for data transmission, high channel utilization efficiency can be obtained. Compared to nonreservation schemes, more complex control mechanisms are needed in the earth stations. The reservations may be sent in separate time slots which are distinct from the time slots used for data transmission or they may be combined with data transmission (piggybacked) or both. The control technique used to allocate the reserved time slot may be central control, distributed control, or a combination of both.

RESERVATION ALOHA

This scheme makes use of separate time slots for reservation, with the control function distributed in all the stations. The satellite channel is divided into time slots of fixed size [11]. Every $M + 1$th slot is subdivided into V small slots as shown in Fig. 1.

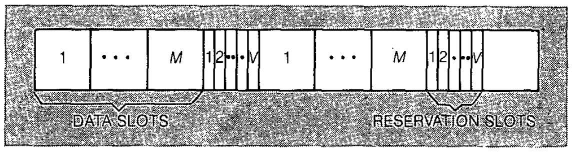


Fig. 1. Satellite channel for reservation ALOHA.

The V small slots are used by all the active stations to send reservations for future time slots and acknowledgments. These V small slots are accessed using the ALOHA contention technique. The M large slots carry reserved data packets.

Whenever a station receives data packets to transmit, it randomly selects one of the V slots and transmits its reservation. This reservation is heard by all the stations. The distributed control in each of the earth stations adds the broadcasted reservation to the existing reservation count. Effectively, all the waiting packets for which a reservation has been made join one "queue in the sky," the length of which is known at all times to all ground stations. The number of reserved data slots that can be reserved in one request range from one to eight. The requesting station has now successfully reserved a sequence of future time slots for data transmission. Once a reservation is made, each one of the stations knows which future slots belong to them, and no other station need concern itself with the details of reservations made by other stations.

Fig. 2 shows an example taken from [11] which illustrates how this reservation scheme functions. Let us assume that the total roundtrip delay for signal travel is 10 slot time and there are five data slots (M) and six small slots (V). If a station transmits a reservation for three future data slots so as to fall in a small slot (ALOHA) at $t = 5$, then all stations will receive this reservation request at $t = 10$ (the roundtrip delay). If no collision has taken place, then the future data slots that can be used for data transmission are easily calculated, provided the current queue length is known. Assuming the current queue length to be 13, then the station which requested the reservation has to wait until 13 data slots have passed by before it can transmit data. In our example, the slots are at time $t = 21, 22$, and 24, 23 being the ALOHA slot. Because it takes 5 slot time for the data packets to reach the satellite, the ground station starts transmission at $t = 16, 17$, and 19.

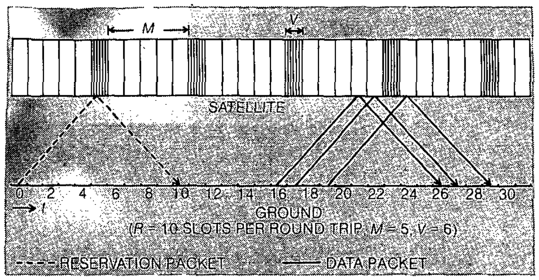


Fig. 2. Satellite channel for reservation.

The performance of the system is a function of the value M, the number of data slots available between each reservation slot. Assuming that there are N ground stations, and if each one of them is allowed to reserve up to eight slots, the maximum allowed, then some reservations may carry over beyond the next reservation slot if $8N$ is greater than M. The system becomes overloaded if each station is allowed to reserve eight further slots. This increases the queue length of future reservations, thereby increasing packet delay. This situation can be avoided if each ground station is constrained to a limit of eight future reserved slots at any time [4]. Another factor which may degrade the performance of the system is excessive contention for reservation slots. The number of V slots must be related to the number of earth stations and to the likely traffic activity to be expected.

In this scheme, the channel may be in any one of the two states called the ALOHA and RESERVED states. On start up and when it so happens that no reservations are outstanding, the channel is in ALOHA state. In ALOHA state, the channel consists of only slots of type V. It is possible to send acknowledgments, reservations, and even data which will fit into the small slots. In this state, a reservation request may be transmitted in any of the small slots, with no requirement to wait for up to M data slots to pass by. Once a successful reservation has been established, the channel enters the reservation state and any further reservation can be made in the small slot. Once again the channel enters the ALOHA mode if the number of reservations goes to zero.

R-TDMA

This explicit reservation protocol is a modified version of the contention and fixed assignment reservation method used in [2]. This scheme uses a fixed-assignment technique for making reservations and allows the total available channel capacity to be shared among all stations that are busy [14].

Fig. 3 shows the R-TDMA channel. One routing frame on the channel is divided into a number of reservation frames. The reservation frame consists of a set of reservation slots and a number of fixed length data slots. These data slots are grouped together to form a data frame. A reservation frame may have one or more data frames. Each station is assigned a fixed slot in each reservation frame. Each of the stations is assigned a fixed slot in each one of the data frames. Therefore, each data frame has as many slots as there are stations.

Each earth station keeps a reservation table to track the data slot allocation. To make reservation for data slots, the earth station transmits its "new reservation" count in its reservation slot. The stations which do not have data to send place a value of zero in their fixed reservation slots. The new reservation count represents the number of data packets that arrived after the last reservation took place. All the earth stations receive the reservation packet and adjust their reservation table values by adding the new reservation counts at a globally agreed upon time.

The allocation of data slots now becomes straightforward. Those stations whose reservation table entries are not zero transmit their data packet in their fixed slots. The data slots which belong to station with no data packets are assigned in a round-robin manner among those stations with outstanding data packets. The sender for each slot is determined just prior to the slot transmission time. In this scheme, synchronization is acquired and maintained by having each station send its own reservation table entry in its reservation slot.

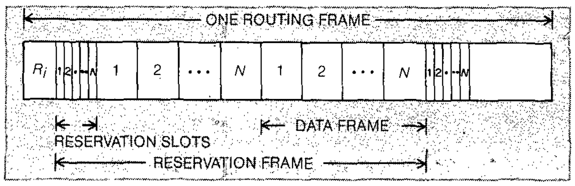


Fig. 3. R-TDMA channel.

CONFLICT-FREE MULTIACCESS (CFMA)

This scheme [6] eliminates all conflict on the satellite multiaccess channel. The channel is divided into frames. Each frame is subdivided into an R-vector, an A-vector, and an I-vector. Fig. 4 shows the frame structure and three vectors. The R-vector is used to request future reservations and is divided into a number of reservation slots. The number of reservation slots in the R-vector is equal to the number of earth stations. Each one of the earth stations is assigned a reservation slot in the R-frame. This avoids contention for the reservation slot. The A-vector is divided into a number of mini-slots which are used to send acknowledgment for previously received packets. An I-vector in a frame is divided into data slots. In this scheme, the maximum number of slots a station may request is equal to the number of slots in the I-frame. Assuming that there are m data slots in an I-frame, the allocation of data slots is based on assigning a priority order for each of the m slots. For example, if the number of stations equals the number of data slots ($N = m$) in an I-vector, then the priority order for each data slot is different. Every station has one data slot for which it has first priority, another for which it has second priority, and so on down to the least priority. If a station does not use its first priority data slot, then a station with second priority to that slot gets a chance to use that data slot. If all stations are busy, then each of the stations will be allocated its first priority data slot and no station will be allocated more than one slot in the above example. The overhead involved in this system does not seem to be high in terms of channel bandwidth.

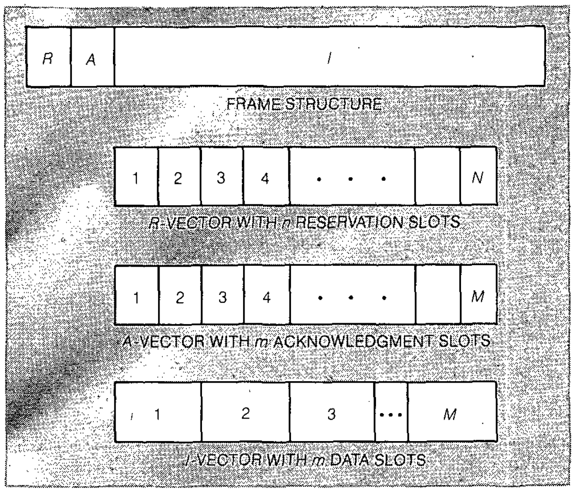


Fig. 4. Frame structure for CFMA channel.

CONTENTION-BASED DEMAND ASSIGNMENT PROTOCOL (CPODA)

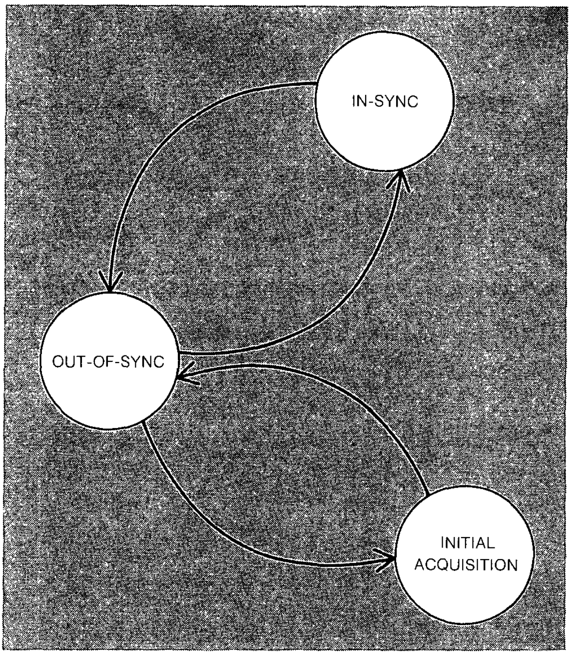
This protocol is designed to handle packetized data and voice traffic [7]. It can handle traffic with multiple priority and delay class distinctions, variable message lengths, and arbitrary load distribution among the stations.

The channel is divided into fixed size frames, and each frame consists of reservation and information subframes. The reservation subframe is divided into fixed size reservation slots. In this scheme, the reservation subframe is allowed to grow or shrink according to the amount of traffic. Therefore, when the number of reservations for the information frame is zero, the reservation subframe expands to occupy the whole frame. On the other hand, when the system is fully loaded, the reservation subframe contracts to the minimum number of slots required to allow reservations by high priority traffic or previously idle stations.

There are two ways in which reservation for information subframes can be done. The first way is to send a reservation in the slots in the reservation subframe. The stations use contention to gain access to the reservation slot. The second way is to send the reservation by piggybacking them in the header field of the reserved message transmission. A maximum of only two new reservations is allowed in each one of the messages. This allows a station transmitting messages to use the piggybacking technique to build their reservation, thereby leaving the reservation subframe free for new entries and/or higher priority traffic.

A distributed control is used to schedule channel time for each earth station to transmit messages. The scheduling is done by forming a queue of the desired transmissions from the explicit reservation requested by the stations. The channel scheduling in this scheme is some function of message priority and delay. Thus, a low priority message with a short delay constraint may typically be serviced before a high priority message with a long delay. The ordering, to some extent, is a weighted function of priority and delay.

Each station carries out a consistency check to assure scheduling synchronization. A station is in synchronization when its scheduling decision agrees with the actual transmission in the channel. A station can be in one of three states as shown below.



A station in the in-sync state is in synchronism with the actual transmission taking place in the channel. Hence, it can continue sending messages at the scheduled time. Whenever a station detects a number of inconsistent scheduling within a specified time period, it moves to the out-of-sync state. In this state, the station is not allowed to send any message; instead it carries out channel scheduling and closely monitors the channel. If the station, in the monitoring channel, finds itself in synchronism again within a fixed period of time, it moves back to the in-sync state and participates in message transmission. Otherwise, it moves to the initial acquisition state. In this state, the station listens to the new reservations on the channel and builds up its channel scheduling information. The station does not transmit any message. Once this station has constructed a reservation list compatible with other stations, it can move to the out-of-sync state.

CONCLUSIONS

A number of multiple access protocols have been presented, some of which are undergoing testing for satellite communication. These reservation methods provide a means to increase channel utilization compared to nonreservation schemes. In all the schemes, one must trade off complexity of implementation with suitable performance. Therefore, in the final analysis, it is cost which will dictate which of the protocol schemes is suitable for a particular application.

REFERENCES

[1] N. Abramson, "Packet switching with satellites," in *Proc. AFIPS Conf.*, vol. 42, June 1973.

[2] R. Binder, "A dynamic packet-switching system for satellite broadcast channel," in *Proc. ICC '75*, San Francisco, CA, June 1975.

[3] W. R. Crowther *et al.*, "A system for broadcast communication: Reservation-ALOHA," in *Proc. 6th Hawaii Int. Conf. Syst. Sci.*, Jan. 1973.

[4] D. W. Davis *et al.*, *Computer Networks and Their Protocols.* New York: Wiley, 1979.

[5] M. Gerla and L. Kleinrock, "Closed loop stability control for S-ALOHA satellite communication," presented at the 5th Data Commun. Symp., Sept. 1977.

[6] H. R. Hwa, "A framed ALOHA system," in *Proc. PACNET Symp.*, Sendai, Japan, Aug. 1975.

[7] I. Jacobs *et al.*, "CPODA—A demand asssignment protocol for SATNET," in *Proc. 5th Data Commun. Symp.*, 1977.

[8] I. M. Jacobs *et al.*, "General purpose packet satellite network," *Proc. IEEE*, Nov. 1978.

[9] L. Kleinrock and S. S. Lam, "Packet-switching in a slotted satellite channel," in *Proc. AFIPS Conf.*, vol. 42, June 1973.

[10] S. Lam and L. Kleinrock, "Packet switching in a multi-access broadcast channel: Dynamic control procedures," *IEEE Trans. Commun.*, vol. COM-23, Sept. 1975.

[11] L. G. Roberts, "Dynamic allocation of satellite capacity through packet reservation," in *Proc. AFIPS Conf.*, vol. 42, June 1973.

[12] "Satellite carrier posed for increasing demands," *Commun. News*, Mar. 1979.

[13] F. A. Tobagi *et al.*, "Modeling and measurement techniques in packet communication networks," *Proc. IEEE*, vol. 66, Nov. 1978.

[14] R. Weissler *et al.*, "Synchronization and multiple access protocols in the initial satellite IMP," in *Proc. COMPCON*, Fall 1978.

C. Retnadhas received the Ph.D. degree from Iowa State University. He is a faculty member in computer science at Western Illinois University. His research interests include computer architecture, computer communication networks, and distributed processing.

Dr. Retnadhas is a member of the Association for Computing Machinery and the IEEE.