1982

Satellite channel sharing and its implications for computer networks

David Storch

Satellite Channel Sharing and Its
Implications for Computer Networks

David Storch
College of Applied Science and Technology
Graduate School of Computer Science
Rochester Institute of Technology
August 1982

Master of Science in Computer Science

Thesis Approval Form

This is to certify that **David Storch** has submitted a thesis entitled:

Satellite Channel Sharing and Its Implications for Computer Networks

to the faculty of the School of Computer Science and Technology in partial fulfillment of the requirements for the degree of Master of Science.

Approval:

	(John L. Ellis)	(date)
	(Roy S. Czernikowski)	(date)
	(Peter H. Lutz)	(date)

6-9576

TABLE OF CONTENTS

PART I: DATA COMMUNICATIONS AND SATELLITE NETWORKS

Introduction

The computer revolution, if you will, which has been evolving over the past three to four decades can be viewed as having occurred in roughly two stages. The first stage was the development of the hardware, the software and applications centered around the single computer installation dedicated to local use. Time-sharing would be included in this stage. The sharing of resources within this environment is provided for by the operating system.

The second stage, which is in its infancy, centers around the sharing of resources (hardware, software and information) between different computers or between computers and computer connected devices which are geographically separated. The computers and/or computer controlled devices will subsequently be referred to as termini. The sharing takes place by transmitting data over lines of communication known as the transmission media. Transmission media include telephone lines, micro-wave links, fiber optics, and broadcast media such as radio and satellites.

A computer network is comprised of the termini, the transmission media and any other hardware as well as the software necessary to implement communications. A simple network can consist of two termini communicating over telephone lines using modems. In such a network, communication can be implemented by simple "hand-shaking" protocols with no change in the operating system and very little additional software. However, most networks consist of numerous termini and require the addition of considerable hardware and extensive software to implement communications. Questions of priorities, scheduling, error checking and control and efficient utilization of the transmission media must be addressed in the design of

the hardware and the software. In many instances, separate processors are necessary to handle these functions.

Data Communications

For the purpose of this paper, data communications will refer to the information transmitted over the communication lines between termini in a computer network. Fitzgerald and Eason (2, pp. 6 - 7) give a sense of the scope of data communications with the following classifications: source data entry and collection (e.g. airline reservations, banking); remote job entry (e.g. manufacturing, inventories, sales); information retrieval from data banks; conversational time-sharing (e.g. general problem solving); message switching (e.g. mail delivery , memo distribution and image and facsimile transmission); real time data acquisition and process control (e.g. manufacturing control and traffic monitoring and control); and interprocessor data exchange.

Data communications have some unique characteristics which make the efficient sharing of transmission media difficult to implement. Data communications are generally sporadic or bursty. The length of the transmissions will vary widely. However, in many instances they are apt to be very short. For example, a voice communication (e.g. a telephone call) is usually several minutes in length. On the other hand a data transmission may be a few seconds to much less than a second in length. The time between data communication transmissions, called interarrival time, also varies widely. In many instances interarrival time may be much longer than message duration. In addition data communication transmissions will vary widely in density, bits per second. In an interactive session between a user and a computer, actual communications between the two would be occurring only a small percentage of the time. Interactive transmissions will usually be light density transmissions (several hundred bits per second). Communications transferring information on magnetic tape or disks will almost always be high density transmissions

(thousands of bits per second). Voice and TV transmissions would not be considered data communications[1]. Although one would not want to preclude analog transmissions, the majority of data communications are digital .

Data Communications Growth

A number of factors are contributing to a burgeoning demand for data communications capability and capacity. More and more activities (particularly in business, governmental and research endeavors) are becoming computerized. Many organizations have branches which are geographically dispersed. Many or all of the branches within an organization may have their own computers. It is usually advantageous, if not essential, for information related to the organization to be quickly available. It may be particularly desirable to make information centrally available for management and planning. There are usually economic incentives for establishing data communications. It may be possible to delegate different computing functions to different computers within the organization. For example, one computer may handle payroll and accounting functions and another may handle data base operations. With this task specialization it is often possible to effect savings by avoiding redundancy and by using simpler hardware configurations and less complex software. Integrity of information and continuity of processing are other factors. Data communications may permit back-up in the event of equipment or software failure. And often data communications can provide for much easier and less costly accommodation of the demands of future growth.

In many instances individual employees of companies are generating data communications traffic. For example, salespeople are using computer networks to more efficiently keep their customer accounts current. In addition they are able to much more quickly check inventories, place orders, and generate invoices. Service

[1]The purpose here is to characterize data communication transmissions rather than to classify a particular type of transmission. Many data communication transmissions are carried over telephone lines. Therefore it is inappropriate to equate voice communications and telephone communications.

representatives are also using computer networks. For example, Xerox is developing a data base which contains information on each machine it has sold and the service calls which have been made. Prior to a service call, the service representative interfaces with the data base and inputs information about the machine and the particular problem. The data base then returns information which may expedite the service call. Upon completing the service call, the service representative adds information to the data base. Manufacturing and design can use the data base to check for recurrent problems which may need correction. In the future much work will be done at a home office utilizing communications with the regular office[2].

Considerable data communications traffic is and will be generated as a result of the refinement and development of information retrieval systems. An information retrieval system provides information relevant to a large computerized store of information (data base). Generally such a system is not used by a single user or organization. The emphasis of most information retrieval systems is the provision of information relevant to a request or query rather than the specification of a particular extract from a data base (although this may be provided). Many systems provide the user with a sophisticated query language as well as other user services.

Information retrieval systems have developed in a number of areas. Early systems provided information related to reports, books, and journal articles. Typical queries would include requests for information about books or articles on a particular subject or by a particular author. Several systems have developed which provide specialized information for particular disciplines. METADEX, for example, provides information about literature on metals. ERIC provides information about reports and journal articles related to educational research. A large internationally used system is MEDLARS (Medical Literature Analysis and Retrieval System). MEDLARS uses remote terminals and telephone lines and provides an interactive

[2]Teleconferencing, which is perhaps not data communications in a rigorous sense, is also presently receiving considerable attention.

search capability as well as a special query language. A number of on-line systems have been recently developed which provide legal research information. For example, WESLAW is a nationally used system which provides a compilation of American case law.

Information retrieval systems are very costly in terms of time, money, and man-hours to develop and implement. While some limited ones are magnetic tape services, the trend is toward large, on-line interactive systems. For information retrieval systems to be economically feasible, they will have to be available to a large number of users many of whom will be located great distances from the data base.

Moreover, the trend toward computerizing and communicating information is quickly spilling over from institutional endeavors to the general public. Heretofore the digitization and computer processing of information has been a relatively expensive process. With the rapid technological developments and the rapid decrease in hardware costs, the computerization and communication of personal information becomes desirable and attractive. Already the electronic transfer of funds is commonplace. The computerization of catalogs and the placement of orders from a home terminal seems imminent. Libraries of the future will probably be data banks containing books on magnetic tape or disks which individuals will access via a home terminal and a communications set-up.

Satellite Transmissions

The use of satellite transmissions as a transmission media holds great potential for meeting existing and future demands for data communications. As a transmission media, satellite links offer relatively high capacity. Present day satellite transmissions support 12 to 15[3] channels capable of transmitting 50 to 60 million bits per second each. In the presently utilized satellite transmission

[3]As will be subsequently discussed, a satellite channel is generally divided into several discrete, mutually exclusive transponder channels.

frequency bands, satellites are allocated a 500 MHz band (500,000 KHz). To realize the full potential of such a bandwidth, it should be noted that a voice communication is typically allocated approximately 4 KHz. One study (17, p.52.1.3) concludes that in the future a single satellite could provide a capacity of 41,000 voice circuits.

Even more significant is the flexibility which satellite networks offer when compared with point-to-point terrestrial networks. A satellite in a geostationary orbit 36,000 km. above the earth can transmit to stations on about one-third of the earth's surface. Thus satellite networks allow communications between widely dispersed users. Satellite networks (by virtue of being broadcast in nature) can offer point-to-point, point-to-multipoint (multiple receiver) and multipoint-to-point communications. Satellite networks permit structures which can be easily enlarged or altered. It is even possible to have mobile user locations.

Furthermore, satellite networks can offer high reliability and integrity. Satellite transmissions exhibit very low transmission error rates. Typical error rates in the 1 in 10,000 to 1 in 100,000,000 range are attainable. Satellite networks are highly connected. In many instances complete networks, wherein each terminus can directly communicate with any other terminus, are possible.

Satellite Communications Networks

A satellite communications network will be considered to be any network which utilizes satellite transmissions at some point in the network. Such a network will consist of a terrestrial component (subnet) and a satellite component. A number of ground stations have antennae which transmit signals to and receive signals from the satellite (Figure I). Users will communicate with ground stations through the terrestrial satellite (T-S) interface. The ground stations may be isolated users. However, as is more often the case, they will probably be the concentration point for a local or long distance network such as those used in interna-

tional telephone or television transmissions.

The Satellite Subnet

The subnet[4] will consist of a number of ground stations communicating with each other via the satellite. Each ground station consists of a satellite interface message processor (SIMP), baseband equipment and Radio Frequency (RF) equipment. One or more terrestrial networks is assumed to be providing an error free terrestrial signal to the SIMP. The SIMP performs several functions. It will implement message access to the satellite subnet and error handling protocols within the subnet. Some channel sharing protocols require global timing of all ground stations

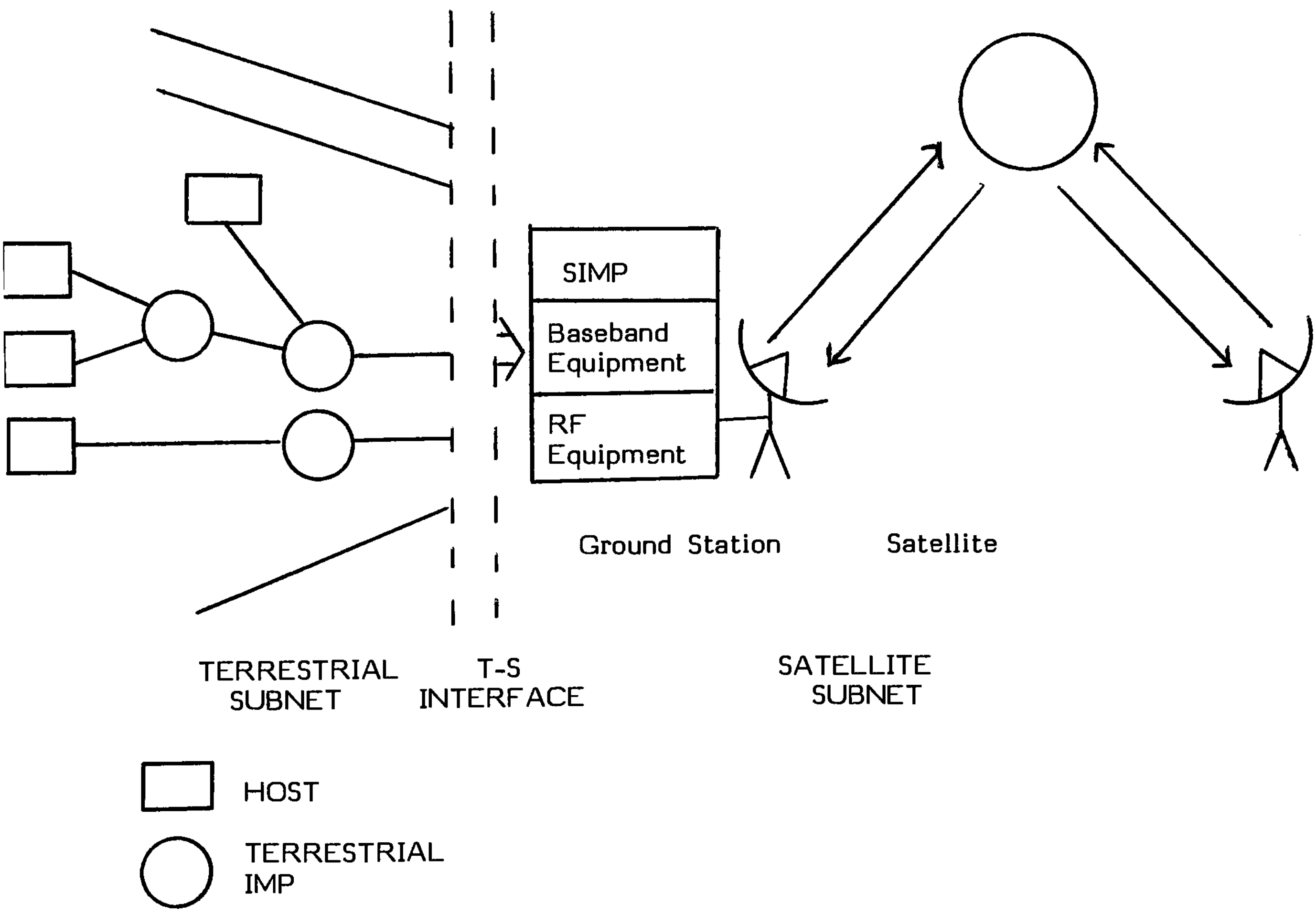

Figure I: The Satellite Subnet

[4]This is a simplified version of a structure proposed for INTELSAT (19).

which the SIMP's would implement. The SIMP buffers each message which has been accepted by the subnet. The SIMP will also hear transmitted packets, extract control information, and send messages intended for that ground station to the appropriate terrestrial network. Protocols may call for acknowledgements which the SIMP would send.

The baseband equipment will include hardware for modulation[5] of terrestrial signals (baseband signals) to the RF carrier to form the RF signal for transmission and for demodulation of the RF signal for reception. It may also include multiplexing/demultiplexing[6] equipment if channel sharing is implemented. The RF equipment essentially transmits and receives the signals to and from the satellite. A link will consist of a sending earth station, the receiving earth station, and the transmission path connecting the two.

Telecommunications satellites are primarily rebroadcasting stations. A satellite consists of a set of amplifiers (transponders) which receive signals (the uplinks), amplify them, and rebroadcast them (the downlinks). The transponder channel (hereafter called channel) is that portion of the RF satellite channel which is allocated to a particular transponder.

---

[5]Modulation is used here in a global sense to denote the superimposing of a given signal, either analog or digital, to a carrier signal forming the signal which is transmitted.

[6]Multiplexing is used globally here to mean combining several signals.

PART II: SHARING A TRANSPONDER CHANNEL

Implementing Satellite Transmissions in Computer Networks

Heretofore, satellite communications have been used extensively for voice and video communications. Most of the developments have come from the physicist and the engineer in improved satellite and transmission technology (e.g. improved transponders and antennae). Most of the innovations have been directed toward developing a few high capacity links. And there are many technological improvements looming on the horizon. Examples include improved transponder design to utilize higher frequency bands and a number of satellite advances which can be implemented as a result of the success of the space shuttle. Other areas receiving attention include more efficient methods of encoding, error checking and control.

However as a result of the burgeoning demand for data communications and the potential advantages which satellite transmissions offer in many applications, the computer scientist is becoming involved. Channel sharing to promote the efficient utilization of a satellite channel is practically a must in any application. It is especially necessary for data communication transmissions. The task of the computer scientist is to develop the protocols to effectively implement channel sharing so that satellite transmissions are a viable alternative as a transmission media. Protocols are needed which result in high channel utilization without incurring excessive message delays. Furthermore, the designers of a particular network will ultimately have to determine if satellite transmissions are a viable alternative.

Increased utilization of satellite channels is presently receiving a great deal of attention. There are two foci for this attention. One approach strives to increase utilization through increasing the number of links for a given satellite network.

This approach involves the sharing of a satellite channel by a number of users. The second approach strives to increase utilization through a more efficient assignment of users to transmission facilities[7]. These two approaches are not mutually exclusive. For example, more efficient assignment will often permit an increase in the number of links, particularly in data communications applications.

Multiplexing and Multiple Access

One of the primary factors affecting the channel utilization in a satellite communications network is the multiplexing technique, if any, used. Multiplexing refers to the process of assembling and combining multiple users' signals into a single signal. Efficient use of the RF channel capacity in a satellite generally mandates the use of multiplexing at some point in the network.

Several signals can be combined to form a single baseband signal. This has traditionally been called multiplexing (and hereafter will be referred to as Baseband Multiplexing). Baseband Multiplexing generally falls into one of two categories: Frequency Division Multiplexing (FDM) and Time Division Multiplexing (TDM). Baseband Multiplexed transmissions are termed Multiple Channel Per Carrier (MCPC) transmissions. Those which are not Baseband Multiplexed are called Single Channel Per Carrier (SCPC) transmissions.

Multiplexing can also occur in the satellite subnet at the RF level[8]. This is known as Multiple Access. An RF signal or link between a sending earth station and the receiving earth station consists of an RF carrier[9] onto which the baseband signal has been modulated. Each RF carrier will contain "address" information so that the message it carries will be detected by the appropriate receiver. Any technique enabling the transmission of more than one RF signal over a given transponder channel would be considered Multiple Access.

---

[7]Many of the considerations related to assignment are analogous to those in operating systems.

[8]Baseband Multiplexing can also occur in the satellite subnet.

[9]These carriers are not necessarily continuous.

Two Multiple Access Qualifications

Mention should be made first that all satellites consist of multiple transponders or amplifiers, each with a frequency band distinct from other transponders in the satellite. As previously mentioned, the bandwidth allocated to a satellite for transmissions is wide. Size, weight, power and application considerations make it inefficient to utilize a single transponder for the full bandwidth[10]. Furthermore, multiple transponder design provides greater satellite reliability because of the small probability of multiple transponder failure. Typical satellite configurations divide the allocated bandwidth into twelve to fifteen transponder channels with bandwidths of from 30 to 36 MHz. Each transponder is a separate entity. The existence of different users using different channels is not considered Multiple Access[11].

It is also important to distinguish Multiple Access from what is termed Host Access. Host Access refers to user access to the satellite subnet across the T-S interface. While Host Access is a very significant parameter in over-all sharing and efficiency, Multiple Access can for the most part be considered independently of Host Access.

The Primary Function of Multiple Access

The primary function of a Multiple Access technique is to provide for increased utilization of the satellite channel by increasing the number of links which can use the channel[12]. In accomplishing this, Multiple Access must

[10]Articles (22) and (23) discuss in some detail the parameters of transponder operation.

[11]This may change. As a consequence of more sophisticated equipment and techniques, on-board processing and switching in the time and/or frequency spectra may become possible. This would necessitate an expansion of the definition of Multiple Access to include the sharing of different transponder channels.

[12]There are many methods for increasing channel capacity. For example in digital transmissions, a change in the coding/modulation technique from two phase to four or eight phase PSK (phase shift keying) can increase the capacity of a channel many fold times. However without an appropriate Multiple Access mechanism to enable channel use for a number of transmissions, no advantage is gained. Multiple Access can be viewed as the fundamental mechanism, in and of itself or

provide for detectable and intelligible transmissions between earth stations which are acceptably error free.

This task is complicated by several factors. The first is the characteristics (i.e. short duration, variable interarrival times, variable density) of data communications transmissions. Furthermore, the majority of satellite transmissions have heretofore been voice and video communications. The demand for these has been sufficient to use most of the existing capacity. Most Multiple Access protocols implemented in the future will have to accommodate voice, video and data communications. An advantage of multiple transponder design is that different Multiple Access protocols can be used on different transponder channels. Some methods of sharing involving long link set-up times (e.g. circuit switching), while acceptable for voice and video transmissions, create undesirable or unacceptable delays for data communication transmissions. Much higher error rates are generally acceptable in voice and video communications.

The hardware configurations of different earth stations may also add complications. For example to make the use of satellite transmissions practical in telephone communications to sparsely settled areas, it is necessary to use small, low power transmitters and antennae. Interfacing these transmissions with high density transmissions which necessitate powerful transmitters and large antennae adds to the problem.

Multiple Access and Multiple Assignment

Hereafter a distinction will be made between Multiple Access and Multiple Assignment[13]. Multiple Access, which is considered in Part III, refers to the techniques whereby a number of users can share a satellite channel. Multiple Assignment, which is considered in Part IV, refers to the manner in which Multiple

---

in concert with other techniques, which generates increased utilization and capacity for a given channel.

[13]This is not done in much of the literature. In most instances Multiple Assignment is considered a variation of Multiple Access.

Access is implemented. In other words, Multiple Assignment will refer to the techniques whereby several users utilize a Multiple Access channel.

PART III: MULTIPLE ACCESS PROTOCOLS

Introduction

A number of Multiple Access protocols have been developed (18, pp. 8 - 10; 20; 21). They are generally classified on the basis of how the time and frequency spectra for a transponder are utilized. In Frequency Division Multiple Access (FDMA), channel frequency is divided between earth stations (Figure IIa). In Time Division Multiple Access (TDMA), time is divided between earth stations (Figure IIb). In a third class of Multiple Access techniques, hereafter referred to as Full Spectrum Multiple Access (FSMA), a number of earth stations can operate simultaneously using the entire channel frequency band. A user is not delineated in the time or frequency spectra (Figure IIc).

Capacity, Utilization, Throughput, Loading, and Delay

Throughout subsequent discussions an attempt will be made to make a quantitative comparison of the various channel sharing protocols. One of the major concerns and criteria related to choosing a particular Multiple Access protocol is the effective use made of the channel. Measures of effective use generally indicate the proportion of a channel which is used to transmit information. There are a number of factors, or what are termed overheads, which decrease the effective use made of a channel.

In all Multiple Access protocols, some part of a channel (time or frequency) must be used to establish discrete and mutually exclusive user channels. Channel capacity will be used to denote the proportion (usually expressed as a percentage) of a channel which is available for use by the earth stations. For example if 10 per cent of the bandwidth of a channel were used to establish non-interfering user

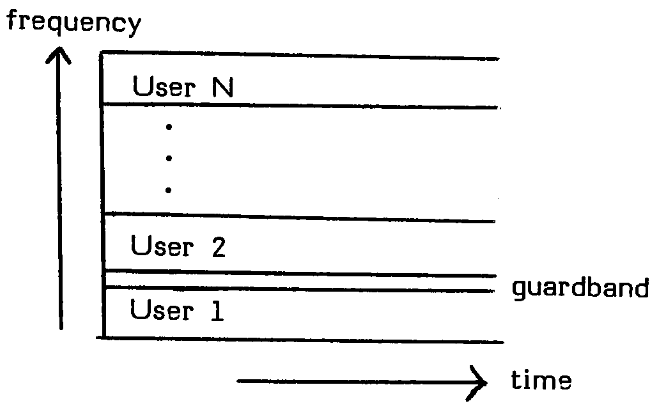

Figure IIa: Frequency Division Multiple Access

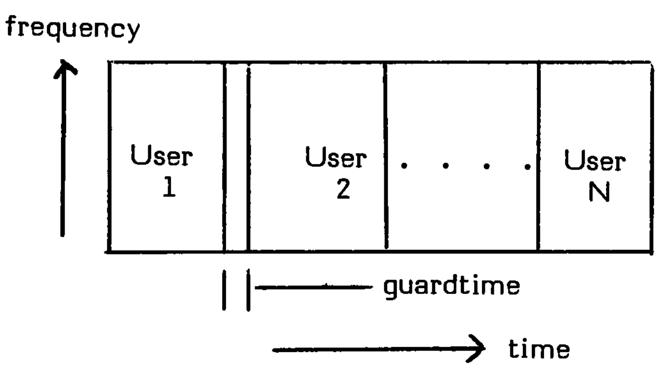

Figure IIb: Time Division Multiple Access

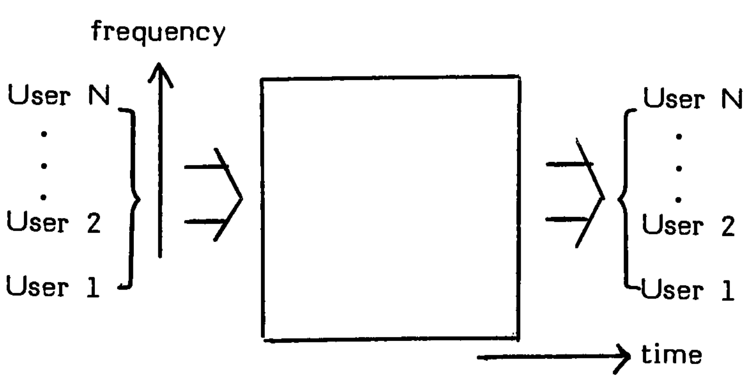

Figure IIc: Full Spectrum Multiple Access

sub-frequencies, then channel capacity would be 90 per cent. Maximum--unattainable when sharing--channel capacity is 100 per cent.

A term related to capacity is utilization. Utilization, in general, will denote the effective use made of a channel for transmitting information or data. There are four factors which determine utilization. The first is the channel capacity resulting from a particular protocol. In addition there may be a number of other overheads besides the overhead necessary to establish user channels[14] For example, some protocols use reservations. Of some form of error checking and correction may be implemented. These overheads decrease the proportion of the channel which can be used for information. A third factor is the percentage of time that accesses are made to the channel. And the last factor is the proportion of accesses which result in "successful" transmissions. Maximum channel utilization is channel capacity. However in most instances, utilization is a much smaller percentage.

A term sometimes used instead of utilization is throughput. Throughput generally ignores some of the overheads in a transmission. Thus throughput would be a slightly higher percentage figure than utilization.

Channel loading will be used to describe the proportion of channel capacity which would be utilized if all accesses to the system were successfully transmitted. For example if all the messages which stations were trying to send would utilize half the channel capacity, then the loading would be .5. Channel loading can be greater than 1. In very general terms, light loading will signify channel loading from 0 to .2; moderate loading will signify channel loading from .2 to .6; and heavy loading will signify channel loading greater than .6.

---

[14]In some instances these overheads are small and quite often tend to be ignored in comparing protocols. However, they can be quite variable and can significantly effect utilization.

Delay will refer to the time that a message is in the satellite subnet. In other words, it is the time from the instant a message enters the subnet until it is received at the other end of the subnet after it has been transmitted over the satellite channel. Delay will consist of propagation delay, queuing delay and blocking delay. Propagation delay refers to the uplink and downlink transmission times. Propagation delay is approximately .27 seconds total. Queuing delay refers to delay in a lightly loaded network due to implementation of the Multiple Access/Multiple Assignment protocols. For example if a particular Multiple Assignment protocol uses a distributed scheduling algorithm, some delay will be necessary for the scheduling of a message. As a network becomes moderately or heavily loaded, a message may experience additional delays due to the unavailability of a free channel for its transmission. This delay will be referred to as blocking delay[15].

Frequency Division Multiple Access

In FDMA (22; 23; 24) the transponder bandwidth is divided into a number of non-overlapping frequency bands (user channels). The bandwidth of each user channel can be based on traffic demands in particular links. Different bandwidths can be used in a single transponder channel. FDMA is implemented by assigning a carrier frequency to each sender. Each sender has exclusive use of its respective frequency band[16]. The receiver(s) obtain intended messages by "filtering" the transponder channel through an appropriately tuned oscillator.

FDMA can be implemented in two ways. The first method (MCPC) frequency modulates a Baseband Multiplexed signal onto a carrier. The second method (SCPC)

---

[15]There is a subtle and sometimes ambiguous distinction between queuing and blocking delays. Often the distinction is not made. Furthermore depending on the Host Access protocols, blocking delay may be experienced in the terrestrial segment of the network before the message is accepted into the satellite subnet. However, it does seem appropriate to distinguish delays inherent in the Multiple Access/Multiple Assignment protocols from those resulting from their implementation in a heavily loaded network.

[16]Some assignment protocols assign users frequency bands when needed from a pool of bands. When a band is no longer needed, it is returned to the pool for use by another user.

modulates each individual user's signal onto a single carrier. In both cases multiple carriers would be transmitted on a single transponder channel.

FDMA was the first Multiple Access technique used in satellite transmissions and is still the predominant technique used. FDMA is the simplest and least costly technique to implement. It makes extensive use of existing hardware and technology. For example if the user's signal is an analog FM signal, only a transmitter and receiver are necessary. Typically no complex ground equipment is needed for encoding, detecting, decoding, timing or storing. Because of its simplicity, FDMA is well suited for light load applications such as rural telephone communications. FDMA satellite transmissions are easily interfaced with terrestrial networks, particularly diverse international ones. Probably one of the most important advantages of FDMA is that it offers great flexibility in its application by virtue of non-uniform band widths. For example, a video transmission (relatively wide bandwidth) and a number of SCPC transmissions (narrow bandwidths) can use the same transponder channel[17].

FDMA Shortcomings

FDMA does suffer some limitations which have lead to consideration and implementation of other Multiple Access techniques. Since band assignment to a sender is a hardware function and not easily changed, a FDMA channel is not easily adjusted for differing traffic demands. Transmissions of different densities or with long interarrival times (typical data communications transmissions) can result in highly underutilized channels with FDMA.

There are several other factors which decrease channel utilization with FDMA. When a number of signals (carriers) are simultaneously present in a transponder,

[17]As will be subsequently discussed, TDMA gives each sender use of the entire channel bandwidth. If applications used narrow bandwidth transmissions (such as thin route telephone communications) it would be necessary to Baseband Multiplex in order to efficiently use TDMA. In some instances this may not be practical.

crosstalk and intermodulation noise occur. Crosstalk occurs when one signal is affected or picks up part of another signal. In intermodulation noise two or more signals form extraneous signals (like harmonics). Both of these can deter from the signals present in the transponder channel and must be compensated for. One compensation is to use guard bands between the frequency bands allocated to the various users. A significant amount of the channel bandwidth must be used for this purpose. Guardbands of approximately 10 per cent of the allocated bandwidth are necessary (22; p. 329). To further decrease the effects of intermodulation noise, it is necessary to decrease ("back-off") amplification. This back-off in transponder amplification results in having to use a transponder in a power range in which it does not make maximum use of the power it consumes.

Another factor limiting channel utilization relates to unequal amplification of signals of differing power levels in a transponder. In FDMA, satellite power is shared by all the signals using the transponders. More powerful uplink signals will receive a disproportionate part of the amplification. Unless this is compensated for, powerful signals will use too much of the available power and weaker signals will become unintelligible. Generally, attempts are made to coordinate or somewhat equalize the power level of signals in the uplink. However, this is difficult to do when sharing a channel between small and large earth stations. Typically a built in margin of power has to be allocated to each band. Power for satellite transponders must be generated by solar cells. Payload limitations of satellite launch systems have limited the power generating capacity of satellites. Ultimately, the number of users of a channel is limited by the bandwidth of a channel. This is a "frequency limited" channel. If the number of users is further limited by the power that a satellite can generate, then the channel is "power limited".

With respect to efficient utilization of a transponder channel for data communications, FDMA suffers two shortcomings. It cannot efficiently accommodate

sporadic transmissions and it is not well suited for digital transmissions. The bandwidth allocated to a sending earth station must be sufficient to accommodate peak traffic demands. This bandwidth will go unused during idle periods. FDMA primarily utilizes frequency modulation to form an analog RF transmission. While FDMA can be adapted to digital signals, it is not efficient to do so. Data communications are primarily digital transmissions. With digital transmissions, noise is not cumulative at repeaters (transponders). And digital transmissions lend themselves ideally to solid state (particularly IC) technology.

FDMA Capacity

Transponder back-off and power coordination have generally resulted in power limited channels. However as a result of recent hardware improvements and greater launch system payloads, power is becoming less of a limitation. Present day FDMA channel capacity typically ranges from 60 per cent (many SCPC transmissions) to 90 per cent.

Time Division Multiple Access

A technique devised to better deal with sporadic transmissions which utilizes digital signals is TDMA (1, pp. 284 - 385; 14; 21; 24). TDMA has been extensively implemented in satellite networks in the past few years. In TDMA time is divided into non-overlapping time slots. Each sending earth station is periodically allocated a time slot and is given use of the entire transponder bandwidth for the duration of the time slot. The appropriate receiver(s) obtain their intended messages on the basis of a time gating function. As a consequence of a single signal present in a transponder at any given time, crosstalk and intermodulation noise problems are virtually eliminated. Transponders can be operated at their most efficient power level (near saturation).

Transmissions must not overlap in time at the satellite with TDMA. If two or more transmissions do overlap, they are said to "collide". The information in

overlapping transmissions is lost. To ensure time slots do not overlap, a guard time of from 100 to 200 nanoseconds (1/10 to 1/5 microseconds) is used between time slots.

TDMA Frame

Generally several time slots are grouped together with control information into a "frame". Gabbard and Kaul (24, p. 349) and Dill (25, p. 351) provide a typical frame structure used in TDMA. Each earth station is assigned a time slot. Often the length of the time slot assigned is based on anticipated traffic for a particular station. In this case time slots for different stations will be of different duration. Typically, time slots range from 20 to 30 microseconds. Frames typically range from 125 microseconds to one or two milliseconds (Figure IIIa).

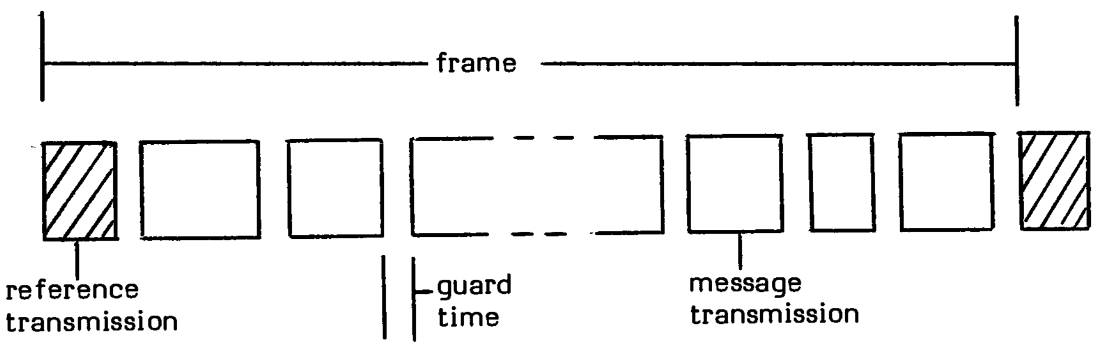

Figure IIIa: TDMA frame

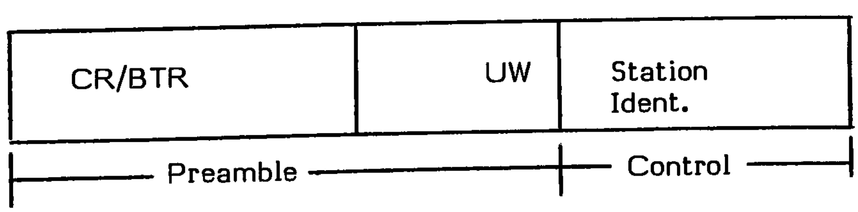

Figure IIIb: Reference Transmission

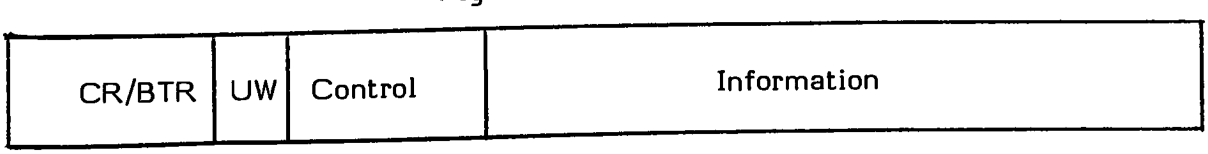

Figure IIIc: Message Transmission

A frame begins with a reference transmission which is used by each of the transmitting stations to time their transmissions. The reference transmission can be sent by either the satellite or one or more of the earth stations. After receiving the reference transmission[18], an earth station will time its transmission with respect to the reference transmission so that it will transmit within the allocated time slot. A reference transmission consists of a preamble and a control section (Figure IIIb). The preamble consists of a carrier and bit timing recovery field (CR/BTR) followed by a unique word (UW) field. The CR/BTR is used to establish demodulation. Detection and verification of the UW establishes timing. The control section contains information related to a transmission such as station identification and other network information. A message transmission (Figure IIIc) will have a preamble, a control section and a data section.

TDMA Shortcomings

TDMA implementation requires some very complex techniques and equipment. For efficient implementation, messages must be broken into "packets". This requires processing equipment as well as buffers at each earth station. Longer time slots and hence longer frames increase channel utilization by decreasing preamble and control section overheads. However, longer frames necessitate larger buffers for storage.

Global timing is required for all stations. Timing in TDMA has two aspects. One, acquisition, deals with locating a transmission time slot. The second, synchronization, involves timing the transmission so that it remains within the appropriate time slot, once located. Most acquisition and synchronization techniques depend on correct detection of the UW in the preamble of each transmission. Developing techniques and hardware to minimize the chances of missing or falsely detecting the UW has been a primary consideration in TDMA design.

[18]TDMA cannot be implemented if earth stations cannot receive a reference transmission.

Techniques such as UW windows and special modulations for the UW have been tested. In the past hardware to perform detection and timing functions has not had the necessary precision or reliability. However, developments in integrated circuitry and digital techniques have resulted in effective timing in the larger earth stations.

TDMA Capacity

The capacity of a TDMA channel depends on the following parameters: frame length, accesses (time slots) per frame, number of overhead bits in the preamble, transmission rate, and guard times. Preamble and control field lengths of 100 to 300 bits are common. Guard times of 15 nanoseconds (.15 microseconds) or less[19] are usual. TDMA has potential capacities as high as 95 per cent or above.

Full Spectrum Multiple Access

A number of Multiple Access techniques (7; 18; 26; 21; 20) have been proposed which allow multiple users to simultaneously share the time and frequency spectra. Included are Spread Spectrum Multiple Access (SSMA), Code Division Multiple Access (CDMA) and Pulse Address Multiple Access (PAMA). In general these techniques are implemented by a coded transformation of the RF carrier signal which can be detected and decoded by the intended receiver(s). For example, in SSMA a pseudo random binary signal is modulated to the RF carrier. The message itself is an additional modulation. The receiver must have an exact copy of the code to demodulate the message. Most FSMA techniques necessitate timing of senders and receivers.

In practice the main advantage of these FSMA techniques turns out to be that they are difficult to intercept or jam. Their implementations have been very limited (mostly in military applications). In theory different users should be able to

[19]Tugal (4, p. 246) suggests that new timing techniques have been developed which eliminate guard times.

simultaneously use the same channel by using different (non-interfering) codes. However, in practice there is some interference including crosstalk and intermodulation noise. This has limited channel capacity and utilization with FSMA techniques. Low data rates (particularly for PAMA) have been experienced in most implementations.

Summary of Multiple Access Techniques

Some advantages of FDMA (especially for its use in low density traffic areas) make it doubtful that FDMA will become totally obsolete. However in most applications, TDMA offers so much greater flexibility and potential for increased channel utilization that it would seem that TDMA will assume pre-eminence as the Multiple Access technique used in satellite data communications networks. TDMA permits full utilization of highly efficient digital encoding and transmission techniques. The integration of data communications with voice and video communications is more effectively done in the time spectrum and TDMA is better suited to this approach. Furthermore, TDMA would seem best suited for use with on-board processing and switching capabilities which are looming on the horizon. While some other Multiple Access techniques (such as SSMA and PAMA) offer some limited and specific advantages, it is not at all clear that they approach the over-all efficiencies of TDMA.

PART IV: MULTIPLE ASSIGNMENT PROTOCOLS

Pre-assignment and Demand Assignment

Multiple Assignment refers to the rules governing the manner in which an earth station gains access to one or more of the channels in a Multiple Access transponder channel. Multiple Assignment protocols are generally classified as either Pre-assignment or Demand Assignment. In Pre-assignment[20] the channels[21] assigned to each earth station are dedicated and more or less permanent. Links which have relatively continuous, constant traffic (e.g. Cable TV) would generally be given Pre-assigned channels. However for links where traffic is variable and/or light, pre-assigned channels result in long idle periods and very inefficient use of the channels. In such instances greater utilization of the channels can be obtained by assigning them on the basis of need. This is known as Demand Assignment[22]. In Demand Assignment a pool of idle channels is maintained. When an earth station has a transmission to send, it may draw from the pool.

In most applications, particularly for data communication transmissions, Demand Assignment can more efficiently utilize satellite capacity than Pre-assignment. Blocking of messages, wherein messages cannot be transmitted due to the lack of an available transmission channel, is less likely to occur with Demand Assignment. This results because each earth station will have access to all unused channels (not just unused channels pre-assigned to that earth station). In some

[20]This is sometimes referred to as static reservation.

[21]Channel is being used in this PART to denote a frequency band or sequence of time slots being used by an individual earth station for transmissions. This should be distinguished from a transponder channel or the satellite channel.

[22]Sometimes this is referred to as Demand Assignment Multiple Access (DAMA).

instances a protocol will use both pre-assigned and demand assigned channels. In this instance utilization can be increased by handling any overflow from pre-assigned segments with demand assigned channels. In most instances for voice and video as well as data communications, a given volume of traffic can be handled by fewer demand assigned channels. Typically demand assigned channels are estimated to have a two to five fold increase in utilization (8, p. 375; 28, p. 388) when compared to pre-assigned channel utilization.

The price which must be paid for Demand Assignment is greater overhead. Earth station hardware and software must be more complex. Most demand assigned protocols utilize time division. This necessitates global timing of earth stations and, in most instances, necessitates considerable buffering. There is generally also a time cost. Particularly for lightly loaded networks, additional time is required in executing demand assigned protocols (e.g. transmitting reservations). And in some instances a channel using demand assigned protocols may become unstable. In this case so many users are accessing the system that they block each other and utilization goes to zero.

In some applications pre-assigned protocols will be more efficient when compared with demand assigned protocols and their associated overhead. For example, in a high traffic situation where the satellite channel is a trunking connection between two large terrestrial components, FDMA using pre-assigned channels and MCPC signals may be the simplest, most efficient and cost effective approach.

Pre-assignment Protocols

Pre-assigned FDMA is very simply implemented. An earth station is given one or more frequency channels for its exclusive use. In general one would not expect queuing delays[23] since each earth station has an exclusive channel (or channels) for its use. However, one would expect extremely poor utilization of the

[23]However, a message may still experience blocking delays.

transponder channels, particularly for a data communications network with infrequent users.

Pre-assigned TDMA[24] is more complex to implement because of the requisite global network timing. An earth station would be given one or more time slots for its exclusive use. Additional time slots would be received in a round robin fashion in subsequent frames. This can result in some significant queuing delays in a data communications network. One would typically expect a number of infrequent users in a data communications network. As a result each transponder channel will have a large number of users. Since each user must be assigned at least one time slot in each frame, time frames would be relatively long. A queuing delay of up to one frame could be anticipated for a communication which requires one time slot for transmission. For communications which require several time slots for transmission, one could anticipate a queuing delay of several time frames. However, queuing delays should not grow exponentially since each user is given transmission time in each frame.

To develop a feeling for the channel utilization of an actual satellite network implementation using dedicated channels, it is useful to look at an analysis done of traffic load projections for the INTELSAT IV[25] serving North, Central and South America and Africa and Europe (The Atlantic Basin). Excerpts of a chart presented by Puente and Werth (28, p. 385) projecting the traffic between Argentina and various countries served by INTELSAT IV are presented in Figure IVa.

Figure IVb presents the percentage of time that the circuits would be used if they were dedicated to traffic on links between Argentina and the respective countries. Figure IVb provides a rough upper bound of the utilization which could be expected for a data communications channel using pre-assigned channels. The

[24]This is usually referred to as a Fixed-TDMA (F-TDMA) protocol.

[25]INTELSAT is an international consortium of satellite users. It offered the first commercial satellite service in 1965 (INTELSAT I, Early Bird). INTELSAT IV was the first satellite to offer Demand Assignment Multiple Access.

Traffic between Argentina and:	Number of Circuits	Paid Minutes per Day
Canada	2	70
France	7	350
Mexico	4	84
United States	52	8882

Figure IVa: Traffic Projections for INTELSAT IV

Traffic between Argentina and:	Per Cent Channel Utilization
Canada	2.4
France	3.5
Mexico	1.5
United States	11.9

Figure IVb: Channel Utilization for INTELSAT IV

traffic analyzed in Figure IV is primarily voice communications which can efficiently be provided for with uniform bandwidth channels. These communications would certainly be much less variable than typical data communications.

Demand Assignment Protocols

On the basis of the traffic analysis excerpted above in Figure IV, INTELSAT initiated a demand assignment protocol using FDMA on their INTELSAT IV satellite. This SPADE protocol (discussed below) is analogous to circuit switching using different frequency bands. A significant shortcoming of SPADE is that it introduces a minimum queuing delay of 240 milliseconds in establishing a link. For most data communication transmissions this results in inefficient channel utilization. It may also result in an unacceptable delay. In general polling schemes result in unacceptable delays in a satellite network. Furthermore, the high data rates of many data communication transmissions require a wide band channel. In general assignment

protocols useful for satellite transmission of data communications rely on the use of a single wide band channel shared on a time division basis by Demand Assignment. Inasmuch as there are advantages in having a single signal at each transponder (see FDMA discussion), the wide band channel shared is assumed to be a transponder channel.

Most demand assigned TDMA protocols separate earth station transmissions into uniform time slots called packets. The overhead (e.g. headers) in breaking up a long transmission into packets is generally not excessive[26].

SPADE

As mentioned above the SPADE[27] protocol (28) is a Demand Assignment protocol using FDMA. An INTELSAT IV transponder maintains 800 channels for use upon demand. A distributed scheduling algorithm is implemented at each earth station. When a call request is received at an earth station, the SPADE equipment at that earth station selects an unused frequency pair from a pool of unused channels. The receiving station is notified over a Common Signaling Channel. Until the transmission is completed, the channel is unavailable for use for other transmissions. Upon completion of the transmission, the channel is returned to the pool and all earth stations are notified of its availability.

ALOHA

One of the first demand access schemes[28] using time division was called ALOHA (30). It was proposed by Norman Abramson for use in a radio broadcasting network at the University of Hawaii. However, ALOHA can be implemented in a satellite network (assuming a transmitting station can receive its own transmission). It is the one exception to typical TDMA approaches in that it does not partition

[26]Since messages are buffered at an earth station, there is no significant increase in buffering required for this.

[27]Single Channel Per Carrier, Pulse Code Modulation, Multiple Access, Demand Assigned Equipment.

[28]ALOHA is sometimes referred to as a random access scheme.

time into slots. Whenever a particular earth station receives a message, it transmits it. If another station transmits at any time during the former's transmission, the two transmissions will "collide" and the information in the transmissions will be lost. A sending station listens for the downlink transmission of its message. If the transmission isn't heard within a certain time, the sending station retransmits it. Using this process an earth station is said to "contend" for a free channel. Generally these retransmissions take place at random times to avoid repeated collisions.

ALOHA has some potential in that it is a relatively simple protocol to implement. No global timing of earth stations is required, and it is not necessary to implement a complex scheduling algorithm. In lightly loaded networks ALOHA is quite effective because there is no queuing delay. Furthermore, in lightly loaded networks there is a small probability that one user will interfere with another. As a network becomes more heavily loaded, collisions become more frequent and eventually utilization suffers. Blocking delay becomes very large (see instability below). The maximum channel utilization is approximately 18 per cent (see Figure V).

Slotted-ALOHA

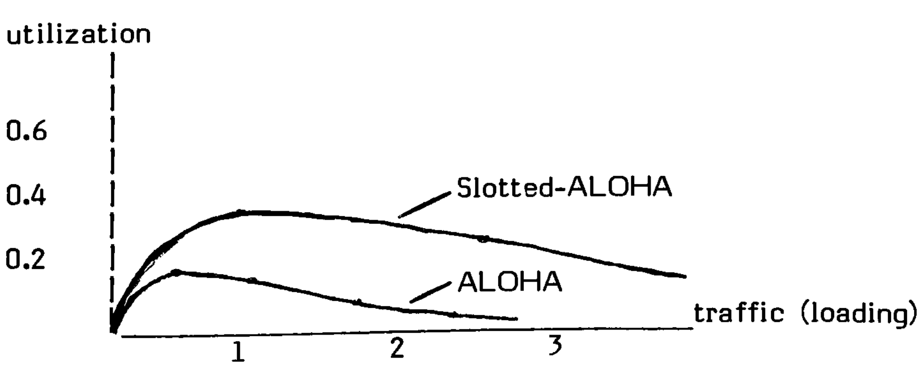

Figure V: Channel Utilization--ALOHA and Slotted-ALOHA

In an ALOHA protocol if all transmissions are of the same time duration, say T, then any particular transmission is vulnerable to any other transmission begun within a time period of 2T. This vulnerability time can be halved and the maximum channel utilization can be doubled (see Figure V) if the time spectrum is partitioned (slotted) into time slots of duration T. Under the Slotted-ALOHA protocol, each station is permitted to start a transmission only at the beginning of a time slot.

With Slotted-ALOHA the maximum channel utilization is approximately 36 per cent. The price for the increased utilization is considerable hardware overhead for global timing of all earth stations. In addition transmissions would suffer a slight queuing delay. Slotted-ALOHA is also unstable without appropriate control mechanisms.

Instability

ALOHA and Slotted-ALOHA are examples of a class of protocols known as contention protocols. Any channel sharing protocol wherein users may be in conflict or interfere with each other's transmissions is a contention protocol. Contention protocols are of particular concern because, as channel loading becomes heavy, interference may become rampant and self-perpetuating. Channel utilization may virtually become zero.

Slotted-ALOHA is typical of a contention protocol. For light channel loading most transmissions will be successful on their first try. However as loading becomes more heavy[29], more and more transmissions will collide. As a result more and more transmissions will have to be retransmitted creating what is termed a "backlog". What evolves is two queues of transmissions: the new transmission queue and the retransmission queue. Both new transmissions and retransmissions

[29]In a data communications network even if loading is light or moderate on the average, there is a large likelihood that loading will become heavy at some point.

contribute to the channel loading or what is termed "offered load". If channel capacity is not sufficient and if the offered load is greater than one (transmission per slot), more and more collisions will occur and throughput becomes very small. With this there is a commensurate increase in backlog. A channel is said to be unstable whenever backlog increases with no increase in the rate of new transmissions.

Channels with only a few users may be inherently stable. In this case channel capacity is sufficient so that any temporary increase in backlog can be handled. However even in this case, transmissions may incur large delays during heavy loading. Most implementations will require controls to avoid instability. These controls must be directed at limiting the offered load (new transmissions and/or retransmissions) when the channel is in a potentially unstable area[30], i.e. when the offered load is greater than 1. Most control algorithms are effected through controlling the frequency or probability with which a transmission is sent from the new transmission queue or the retransmission queue.

One example (29, pp. 536 - 541) of instability control is the controlled Slotted-ALOHA implemented on SATNET[31]. Each earth station maintains a new and a retransmission queue. At the beginning of a time slot, each station will transmit a new or a backloged message with probabilities P(n) or P(r) respectively. A dynamic[32], distributed algorithm measures the offered load over a period of time and adjust P(n) and P(r) accordingly. When the offered load is less than 1, P(n) = 1 for all stations. When the offered load becomes greater than 1, both P(n)

---

[30]However the controls should not limit the offered load if it is less than 1. In this case increasing the offered load will increase the utilization of the channel (see Figure V).

[31]SATNET is a satellite network jointly undertaken by agencies in the U.S., Great Britain and Norway. It is a sophisticated network designed to implement and evaluate various Multiple Access/Multiple Assignment protocols.

[32]Stability could also be guaranteed by making P(n) = P(r) = 1/N (where N is the number of stations), but this would result in long delays and inefficient utilization particularly when the offered load is less than 1.

and P(r) are decreased to decrease the offered load. Hence this reverts to the conventional Slotted-ALOHA when the offered load is less than 1. Measurements on SATNET have demonstrated that this algorithm effectively deals with unstable situations. Furthermore even when the channel is in an inherently stable condition, the algorithm much more quickly eliminates excess backlog.

Demand Assignment Using Reservations

For TDMA over 90 per cent of the channel should be available for the transmission of information. In this context ALOHA and Slotted-ALOHA make very inefficient use of channel capacity. To improve utilization many protocols have been devised which use part of the channel bandwidth (more appropriately some of the time slots) and/or a scheduling algorithm to reserve future time slots. If a reservation protocol is to be effective, it must minimize the bandwidth used for reservations. In theory these reservation protocols can effect greater channel utilization than Slotted-ALOHA during heavy network loading. However, they introduce a queuing delay because the reservations are subject to the same satellite propagation delay which other transmissions are subject to. In addition significant queuing and blocking delays may occur in a moderately or heavily loaded network because of delays in getting reservations.

Reservation protocols generally necessitate much more complex earth stations. As a result of the long propagation delay in a satellite network, most reservation protocols use distributed scheduling algorithms which must be implemented at each earth station. Most reservation protocols require buffering at each earth station to keep track of reservations (i.e. reservation tables).

Reservation ALOHA's

Several protocols (3, pp. 271-272) have been proposed which combine Slotted-ALOHA with some form of reservation. In one variation[33] a frame consists of

[33]The first two variations of Reservation ALOHA are sometimes referred to as implicit reservation schemes.

time slots with one time slot for each station (there may be additional "un-owned" time slots). An earth station may transmit in any time slot which is not reserved for its owner. When two or more stations simultaneously transmit, a collision results. In the subsequent time frame all stations except the owner must desist from transmitting during that time slot and must continue to do so until that time slot is unused for one frame.

This protocol approaches the efficiency of TDMA for a heavily loaded system where most users are accessing the network continuously. As a data communications network protocol it does exhibit some inefficiencies. For light loading it is not as efficient as Slotted-ALOHA because of the waiting periods for unused frames. In addition for long transmissions, one would expect the same queuing delay (due to long frames) that Fixed-TDMA exhibits for channels with a large number of users. One would not expect significant blocking delays.

Another variation of Slotted-ALOHA which avoids the problem of long time frames makes a station the owner of a time slot by use. Once a station successfully transmits during a particular time slot, it owns that slot until it has finished transmitting. When a time slot is not used during one frame, it becomes available for any user during the next time frame. Messages would experience some queuing delay with this protocol. And there could be some significant blocking delays because it is possible for one or a few stations to capture most of the time slots. However, one would not expect this to be a major problem in a data communications network. Furthermore, one would not expect this protocol to be unstable.

A third variation uses explicit reservation slots. A frame is divided into data slots and reservation slots. Each reservation slot is divided into subslots. When a station has a message to transmit, it contends for a reservation subslot. If successful (i.e. there are no collisions with other users vying for the reservation slot),

it is placed in a queue[34] by a distributed algorithm and transmits at the appropriate time for its reserved data slot.

This protocol exhibits a queuing delay resulting from contending for reservations. Furthermore, for a heavily loaded network the reservation contention may become unstable and the system may become congested. Thus significant blocking delays may be incurred.

Reservation-TDMA

Reservation-TDMA is conventional Fixed-TDMA with provisions for reallocating unused time slots to stations which have transmissions to send. One implementation (29) used in SATNET divides a frame into a number of "reservation" subframes (see Figure VI). A reservation subframe is divided into a reservation section and a data section. The reservation section contains a reservation slot (a short time slot) for each earth station. The data section contains data slots which are permanently assigned in a round robin fashion to each station. All stations received the same number (one or more) of time slots. Each station transmits the number of packets it is waiting to send in its assigned reservation slot. And each station can transmit packets in the assigned data slots. A distributed scheduling

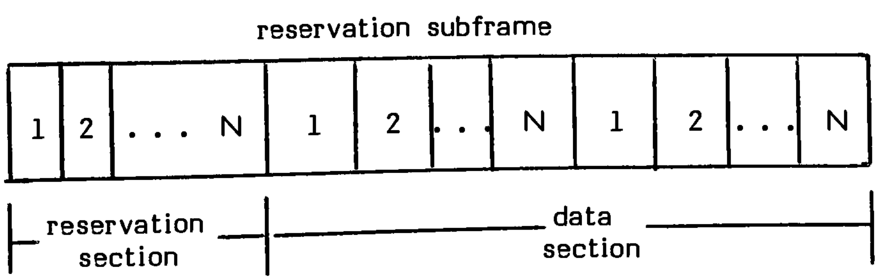

Figure VI: Subframe for Reservation-TDMA

[34]This is often referred to as the First In, First Out (FIFO) protocol.

algorithm assigns unused data slots (data slots whose owners have zero reservations) in a round robin fashion to stations with reservations.

Results from using R-TDMA in SATNET (29, pp. 539, 542) showed that a channel utilization of 90 per cent was attainable. Time delay was two to five fold that for Slotted-ALOHA for light loading. R-TDMA exhibited considerably less delay than Slotted-ALOHA for moderate loading. And R-TDMA is not unstable for heavy loading.

CPODA

A number of variations, some very sophisticated, of these protocols have been proposed. One implemented on SATNET (15, pp. 19, 20; 31) which has received considerable attention is CPODA[35]. CPODA is essentially a Reservation-ALOHA protocol. A frame is divided into an information subframe and a control subframe (used for reservations and acknowledgements). A station may make reservations by contention in the control subframe or by "piggybacking" then in the header of a reserved message transmission. A sophisticated, distributed scheduling algorithm controls how a station may make reservations and allocates information slots on the basis of reservation requests which include priority and delay criteria for the message.

There are three significant characteristics of the CPODA protocol. The first is that the control subframe is allowed to expand when there are few data transmissions. This speeds up reservation acquisition and decreases queuing delay. The second is that reservations can be piggybacked in data transmissions. The efficacy of this is demonstrated by the fact that delay actually decreases as a channel becomes moderately loaded (when compared with a lightly loaded channel). The third characteristic is that the reservation process involves consideration of delay and priority criteria. This would certainly seem to offer greater potential for high

[35]Contention Priority Oriented Demand Assignment.

channel utilization, particularly for a moderate to heavily loaded channel, than a pure FIFO or contention reservation scheme.

PART V: IMPLEMENTATION CONSIDERATIONS

Implementation Parameters

As has been previously discussed, satellite transmissions offer a number of advantages in their use in computer networks. Principal among these are: wide geographic coverage, high capacity channels, architectural flexibility, high connectivity and low error rates.

But there are disadvantages. The cost of using satellite transmissions in a network may be significant. All projections of anticipated demand indicate availability of satellite channels will fall far short of need in the decades to come. A high capacity satellite data channel will be very costly, especially if implementation doesn't provide for efficient sharing of the channel. Furthermore, the hardware and software for earth stations are costly, especially if they must implement global timing or a scheduling algorithm[36]. Limited availability is apt to be another problem in the future.

However, the principal disadvantage which must be addressed when considering the use of satellite transmissions in a network is message delay. Ultimately, the nature of the particular data communications will determine the appropriateness of a satellite network.

Delay Comparisons

Delay in a satellite channel depends upon a number of parameters. Included among these are the channel sharing protocol, the number of users of the channel,

---

[36]In many instances, however, a network implementation will be part of a large terrestrial network which uses a satellite channel as a trunking connection. The pro-rated cost of the channel and the earth station may not be a significant consideration. But the delay considerations discussed below will still be.

the number of users accessing the channel (loading) and (for time division protocols) the length of a message. Figures VII and VIII compare delay versus throughput for various values of these parameters. Figure VII (32, p.410) compares S-ALOHA, R-ALOHA and F-TDMA for 10 and 50 user. Figure VIIa shows delays for 10 users for messages which are one packet in length. Figure VIIb shows delays for 50 users and for messages which are one packet in length. Figure VIIc shows delays for 10 users and for messages which are an average of 4.5 packets in length. Figure VIII (33, p. 45.4.2) compares S-ALOHA, F-TDMA, R-TDMA and CPODA for three users and for messages which are one packet in length.

Delay Considerations

From Figures VII and VIII it can be seen that most of the protocols exhibit delays of 1.5 seconds or less in a lightly loaded[37] network. However since data communications are generally bursty in nature with long interarrival times, one would expect a channel to have a number of users in order to produce acceptable utilization. It is very doubtful that any satellite network transmitting data communications could ensure there would not be times when loading became moderate to heavy with the commensurate increase in delay.

What has to be looked at then in considering the appropriateness of satellite transmissions for a particular computer network implementation is the maximum delays which would be acceptable for the efficient transmission of information. Both maximum average delay and maximum worst case delay must be considered.

Average delay denotes the delay which can be expected for a vast majority of transmissions. In most interactive sessions response times of 2 to 4 seconds (assuming local echoing) are considered acceptable. Response times greater than this decrease efficiency and increase user error rates. Since response time involves round trip transmission time, a maximum message delay of from 1 to 2 seconds

---

[37]The only way to ensure delays of less than 1 second is to provide for a dedicated user channel (e.g. SPADE).

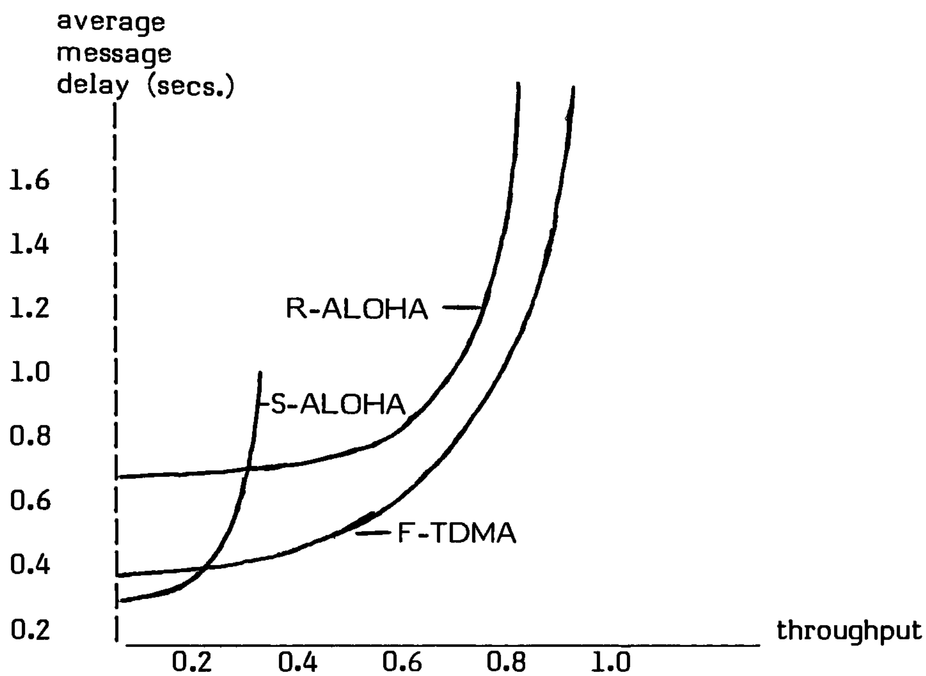

Figure VIIa: 10 users, message length of 1 packet

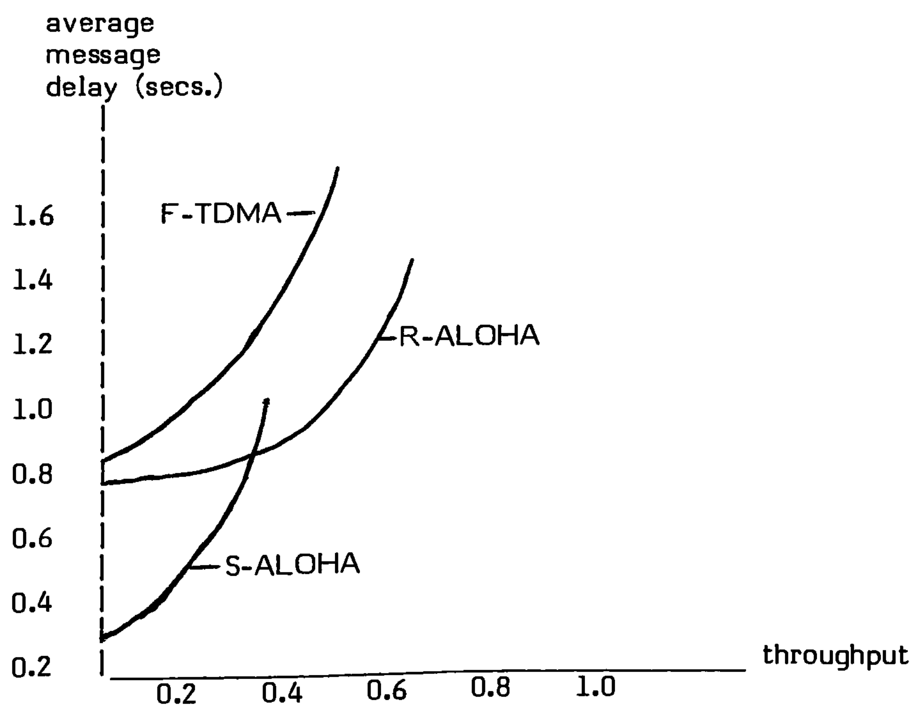

Figure VIIb: 50 users, message length of 1 packet

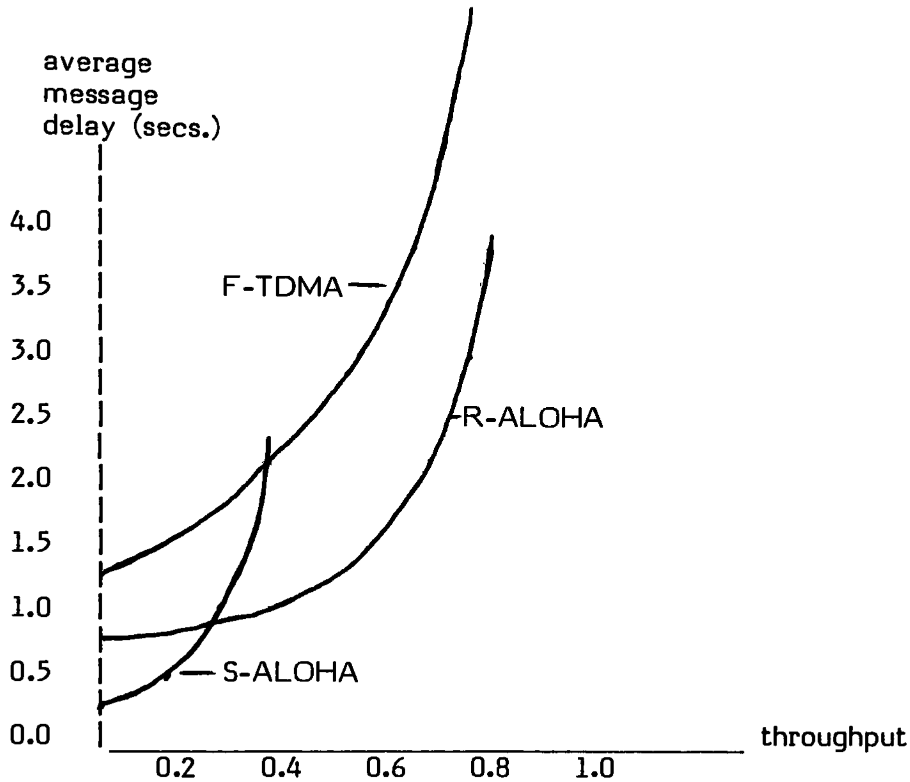

Figure VIIc: 10 users, avg. message length of 4.5 packets

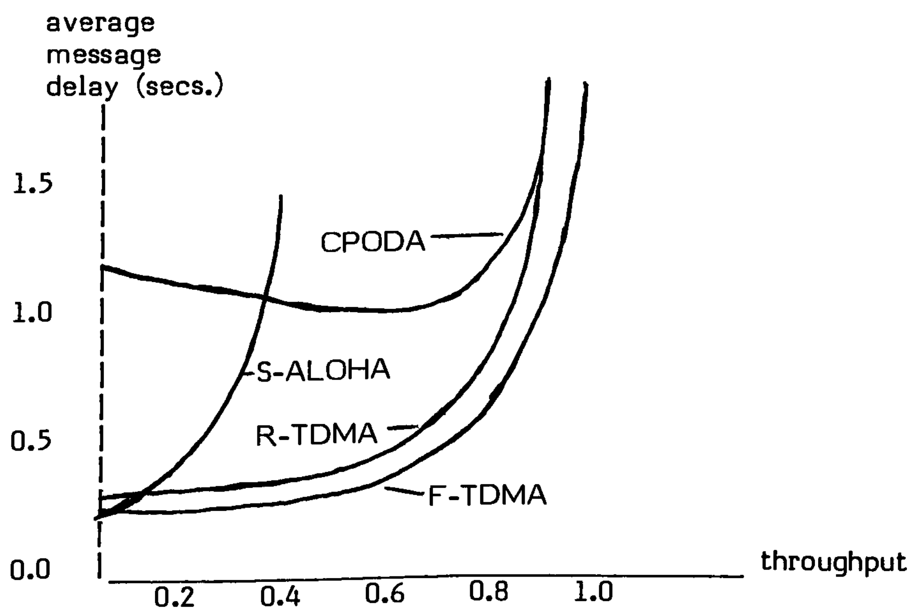

Figure VIII: 3 users, message length of 1 packet

would seem acceptable for interactive communications. For message switching applications average delays from minutes to hours would seem acceptable.

In considering delays for real time data acquisition and process control and interprocessor data exchange, it is useful to divide each category. The following categories will be considered: real time monitoring and data acquisition, real time process control (including monitoring), interprocessor data exchange (concurrent processing) and file exchange. For all of these categories the acceptable maximum average delay is quite variable. In some applications it is possible that the propagation delay of .27 seconds is unacceptably large. For real time monitoring delays of 1 to 2 seconds, or even longer, should be acceptable. For real time process control[38] and for interprocessor data exchange delays of much less than 1 second may be required. For exchange of information on files, several seconds delay should be acceptable. It is assumed that some form of error control will be used and a large amount of buffering would thus be required if longer delays were incurred.

Worst case delay denotes the maximum delay which could be tolerated for the efficient transmission of information[39]. For interactive sessions response times greater than 10 to 15 seconds would generally be unacceptable unless they occurred very infrequently. Hence worst case delays of 5 to 7 seconds should be acceptable. For message switching worst case delays of hours to days should be acceptable. For real time monitoring, delays from seconds to minutes, or even longer, should be acceptable. For real time process control and for interprocessor data exchange,

[38]It would seem that real time applications involving process control would generally be handled in house by a dedicated microcomputer or minicomputer.

[39]Any specific values used in this section are subject to qualification and more appropriately determined when considering a specific implementation. This is certainly true when considering worst case parameters. In most instances a much larger delay could be tolerated s small percentage of the time.

again, delays of much less than 1 second may be required. For file exchanges there may be buffering problems with delays longer than several seconds. Figure IX summarizes the foregoing information.

Conclusions for Implementation

From the standpoint of acceptable delays there seem to be three classes of data communications[40]. The first class of communications is those which should not use satellite transmissions. Included in this class would be real time process control and interprocessor data exchange. Even if such communications were given

data communications	maximum average delay	maximum worst case delay
source data entry	1 - 2 secs.	5 - 7 secs.
remote job entry	1 - 2 secs.	5 - 7 secs.
information retrieval from data bases	1 - 2 secs.	5 - 7 secs.
conversational time sharing	1 - 2 secs.	5 - 7 secs.
message switching	hours	hours - days
real time monitoring and data acquisition	1 - several seconds	seconds - hours
real time process control	less than 1 second	less than 1 second
interprocessor data exchange	less than 1 second	less than 1 second
file exchange	several seconds	several seconds

Figure IX: Maximum Average and Worst Case Delays

[40]It is interesting to note that one particular protocol, CPODA, attempts to deal with this data communications hierarchy.

dedicated channels, the delays due to propagation alone make the use of satellite transmissions questionable. And most likely some form of error handling would be implemented which would increase the delays even more. Furthermore, if any other channel sharing protocol were to be used, one has to make the assumption that heavy loading of a channel will occur at some times. The resulting increase in delay times make the use of satellite transmissions even more questionable.

The second class of communications is those for which long delays have no deleterious effect. Included in this class would be message switching and real time monitoring and data acquisition. For this class of communications over long distances, satellite transmissions hold attractive advantages.

The third class is those communications for which satellite transmissions are suitable in most instances. Included in this class would be the interactive communications (source data entry, remote job entry, information retrieval from data bases, conversational time sharing) and file exchange. All of the protocols provide acceptable (or better) delays for a lightly loaded channel. Furthermore, most of the protocols provide acceptable delays for a good range of channel throughput.

However, the delays experienced in an actual implementation might make satellite transmissions an unattractive alternative. This would presumably result from too heavy loading of the channels. But in most instances this should be avoidable. Most of the protocols exhibit steep delay-throughput curves for moderate to heavy loading. As a result, a slight decrease in throughput results in a large decrease in delay.

So for the vast majority of data communications over long distances (and in fact the ones which stand to reap the greatest benefits from the utilization of a computer network), satellite transmissions appear to be a viable alternative. And due to the unique characteristics of data communications (particularly their sporadic nature and their long interarrival times in comparison to their duration) it seems

doubtful that satellite transmissions would be economically feasible if it were not for channel sharing.

GLOSSARY

ALOHA: A demand assigned, TDMA, Multiple Assignment protocol using contention in which a ground station transmits a message as soon as it receives it.

capacity: That part of a satellite channel which is available to transmit information after discreet user channels have been established.

channel: A path for data transmissions.
(a) transponder channel: the frequency band assigned to a particular satellite transponder.
(b) user channel: a frequency band or sequence of time slots used by a particular sending ground station.

computer network: Consists of a set of geographically separated computers and computer connected devices which share information; the transmission media over which they communicate; as well as any hardware and software necessary to implement communication.

contention protocol: Protocols for gaining access to a satellite channel which can result in conflicts between users and mutual interference. If two or more ground stations transmit at the same time, their messages will interfere with each other and will have to be retransmitted.

CPODA (Contention Priority Oriented Demand Assignment): A demand assigned, TDMA, Multiple Assignment protocol in which a ground station gains access to a transponder channel for a period of time by contending for reservations.

CR/BTR (Carrier and Bit Timing Recovery): A field of bits used in most TDMA implementations to establish and synchronize modulation and demodulation.

data communications: The information or data communicated over the transmission media in a computer network.

delay: The time from the instant a message enters a satellite subnet until it is received at the other end of the subnet after it has been transmitted over the satellite channel.
(a) propagation delay: A delay of approximately message from a ground station to the satellite and back to the ground station.
(b) queuing delay: The delay incurred in a network due to implementation of a Multiple Access/Multiple Assignment protocol.
(c) blocking delay: The delay incurred in a network due to lack of available channels.

Fixed-TDMA: A pre-assigned, TDMA Multiple Assignment protocol in which ground stations are assigned time slots in a pre-determined, unalterable, round robin fashion. No provision is made for reassigning unused time slots.

FDMA (Frequency Division Multiple Access): A channel sharing protocol which divides channel frequency between users wherein each user has exclusive use of the allocated frequency band.

FSMA (Full Spectrum Multiple Access): A class of channel sharing protocols wherein users simultaneously have use of the entire channel. FSMA is implemented by a coded transformation of the carrier signal which uniquely identifies the transmission. A receiver can detect the transmission if it has a copy of the code.

instability: A channel condition resulting from contention protocols wherein interference becomes rampant and self-perpetuating.

INTELSAT: An international consortium of satellite users.

loading: The ratio of attempted transmissions to channel capacity.

Multiple Access: The techniques and protocols to enable several users to share a (satellite) channel.

Multiple Assignment: The techniques and protocols to enable users to gain access to a Multiple Access (satellite) channel.

MCPC (Multiple Channel Per Carrier): A situation in which two or more terrestrial signals are multiplexed onto a single RF signal.

protocol: The rules and conventions related to the transmission or exchange of data between parts of a network.

Reservation-ALOHA: A demand assigned, TDMA, Multiple Assignment protocol in which time slots are reserved according to some reservation algorithm. The reservation process involves contention.

Reservation-TDMA: A demand assigned, TDMA, Multiple Assignment protocol in which users are assigned time slots in a round robin fashion with the provision that unused time slots are reassigned.

RF transmissions: Transmissions between ground stations and a satellite.

satellite communications network: A computer network utilizing satellite transmissions as a transmission media at some point in the network.

satellite subnet: Part of a satellite communications network consisting of ground stations, the satellite and the satellite channels.

SCPC (Single Channel Per Carrier): A situation in which each RF signal contains a single terrestrial signal.

SATNET: An experimental satellite jointly undertaken by agencies in the U.S., Great Britain and Norway to implement and evaluate various protocols.

Slotted-ALOHA: A demand assigned, TDMA, Multiple Assignment protocol using contention in which a ground station transmits a message in the next time slot after receiving it.

SPADE (Single Channel Per Carrier, Pulse Code Modulation, Multiple Access, Demand Assigned Equipment): A demand assigned, FDMA, Multiple Assignment protocol wherein a ground station is given use of a frequency band upon demand by a distributed scheduling algorithm.

TDMA (Time Division Multiple Access): A channel sharing protocol which divides time between users wherein each user has exclusive use of the entire transponder channel during the allocated time.

time frame: A reference unit of time in many TDMA implementations. Usually a time frame consists of time slots containing message information and time slots containing control information for protocol implementation.

time slot: A reference unit of time in most TDMA implementations. Usually the time spectrum is partitioned into uniform time slots.

transmission media: The lines of communication in a computer network. Transmission media include telephone lines, micro-wave links, fiber optics, and broadcast media such as radio and satellites.

transponder: An amplifier in a satellite which receives a transmission from a ground station, amplifies it and retransmits it.

UW (Unique Word): A field of several bits used in most TDMA implementations to establish timing.

utilization:
(a) in general, the proportion of a channel which is used to successfully transmit information
(b) specifically, the ratio of successfully transmitted message information (excluding all overheads and unused portions of the channel) to the total information the channel can transmit.

BIBLIOGRAPHY

General References:

(1) Freeman, Roger L., Telecommunication System Engineering--Analogue and Digital Network Design, New York: John Wiley & Sons, 1980.

(2) FitzGerald, Jerry, and Eason, Tom S., Fundamentals of Data Communications, New York: John Wiley & Sons, 1978.

(3) Tannenbaum, Andrew S., Computer Networks, Englewood Cliffs: Prentice Hall, 1981.

(4) Tugal, Dogan A., and Tugal, Osman, Data Transmissions: Analysis, Design, Applications, New York: McGraw Hill, 1982.

(5) Heaps, H. S., Information Retrieval: Computational and Theoretical Aspects, New York: Academic Press, 1978.

(6) Dixon, R. C., Spread Spectrum Systems, New York: John Wiley & Sons, 1976.

(7) Cook, C. E., Ellersick, F. W., Milstein, L. B., and Schilling, D. L., eds., "Special Issue on Spread Spectrum Communications", IEEE Trans. on Commun., vol. COM-30, No. 5 (May 1982).

(8) Van Trees, Harry L., ed., Satellite Communications, New York: IEEE Press, 1979.

(9) 1979 Conference Record, (National Telecommunications Conference, Washington, D.C., November 27 - 29, 1979), New York: IEEE Press, 1979.

(10) Inose, H., ed., Evolution in Computer Communications, (Proceedings of the Fourth International Conference on Computer Communications), New York: North Holland Publishing Company, 1979.

(11) Gerald, R. G., and Lum, Y. E., eds., Communication Satellite Systems: An Overview of the Technology, New York: IEEE Press, 1976.

(12) Schnipper, Herbert, "The SBS System and Services", IEEE Communications Magazine, vol. 18, No. 5 (Sept. 1980), pp. 12 - 15.

(13) Carne, E. Bryan, "New Dimensions in Telecommunications", IEEE Communications Magazine, vol. 20, No. 1 (Jan. 1982), pp. 17 - 25.

(14) Chu, Wesley W., Advances in Computer Communications and Networking, Dedham, Mass.: Artech House, 1979.

(15) Retnadhas, C., "Satellite Multiple Access Protocols", IEEE Communications Magazine, vol. 18, No. 5 (Sept. 1980), pp. 16 - 20.

(16) Rosen, Harold A., "Space Telecommunications", IEEE Communications Magazine, vol. 18, No. 5 (Sept. 1980), pp. 5 - 14.

Specific References:

(17) Pelton, Joseph N., "Global Satellite Communications: INTELSAT Faces the Challenges and Opportunities of the 1980's and 1990's", in (9), pp. 52.1.1 - 52.1.5.

(18) Pritchard, Wilbur L., "Satellite Communications--An Overview of the Problems and Programs", in (8), pp. 2 - 15.

(19) Hoversten, E. V., and Van Trees, H. L., "International Broadcast Packet Satellite Services", in (10), pp. 527 - 534.

(20) "Modulation and Multiple Access Techniques", in (11), pp. 82 -85.

(21) Schwartz, J. W., Alin, J. M., and Kaiser, J., "Modulation Techniques for Multiple Access to a Hard-Limiting Satellite Repeater", in (8), pp. 308 - 322.

(22) Dicks, Jack L., and Brown, Martin P., "Frequency Division Multiple Access (FDMA) for Satellite Communication Systems", in (8), pp. 324 - 335.

(23) Ferguson, Myron E., "Design of FM Single-Channel-Per-Carrier Systems", in (8), pp. 336 - 341.

(24) Gabbard, O. Gene, and Kaul, Pradman, "Time-Division Multiple Access", in (8), pp. 344 - 349.

(25) Dill, George D., (COMSAT Laboratories), "TDMA, The State-of-the-Art", in (8), pp. 350 - 358.

(26) Ristenbatt, Martin P., and Daws, James L. Jr., "Performance Criteria for Spread Spectrum Communications", in (8), pp. 360 - 366.

(27) Marsten, R. B., "Service Needs and System Architecture in Satellite Communications", IEEE Comm. Soc. Magazine, May 1977.

(28) Putente, J. G., and Werth, A. M., (COMSAT Laboratories), "Demand-assigned Service for the INTELSAT Global Network", in (8), pp. 382 - 392.

(29) Kleinrock, L., and Gerla, M., "On the Measured Performance of Packet Satellite Access Schemes", in (14), pp. 535 - 541.

(30) Abramson, Norman, "The Throughput of Packet Broadcasting Channels", in (14), pp. 117 - 128.

(31) Chu, W. W., and Naylor, W. E., "Measurement and Simulation Results of CPODA Protocol Performance", in (14), pp. 95 - 101.