



the capabilities of large systems to small stations, more applications will move toward VSATs and mobile earth stations, while the need for even smaller stations will accelerate (2). The decrease of station size and cost, combined with a global deregulation of the telecommunication industry, should make VSAT systems even more attractive service providers.

The main factors that account for the popularity of VSAT networks are:

- VSAT networks offer a cost-effective means of implementing high-quality and reliable communications to locations that are not well served by terrestrial facilities.
- VSATs can be used as end-to-end digital networks, with very good bit error rate characteristics [in the range  $10^{-7}$  to  $10^{-10}$  bit error rates (BER)]. Not all terrestrial networks are end-to-end digital, and their BER performance is often worse.
- There is no need for complex routing, and despite the long satellite propagation delay, response times are usually less than 3 s, and there is little variance in delays.
- VSATs are the technology of choice for multicasting services (i.e., point-to-multipoint simultaneous broadcasting of information).
- VSAT networks are virtually insensitive to geographical separation, unlike terrestrial networks, where the network cost is proportional to the distance between the network's nodes. It is easy to plan and implement a network expansion, and expansion costs are usually predictable. Also, VSATs can be installed rapidly and moved to new locations as needs change.
- VSATs provide a high degree of security and network management and control to the customer, and the ability to bypass any local network
- The expensive satellite capacity and hub facilities can be shared among multiple users and applications.

To sum up, VSAT networks are continuing to grow lucratively beyond all industry forecasts. Able to transmit information quickly, efficiently, and cost-effectively, VSAT technology is now critical technology in countries lacking the telecommunication infrastructure, while developed nations are using VSAT technology for thousands of new applications in education, government, agriculture, electronic mail, and financial markets. The recent increased interest in providing high speed connection to the internet over satellite combined with the planned deployment of several broadband satellite systems targeting business and consumer markets will also contribute to an even more rapid growth of VSAT-based services.

## HISTORICAL EVOLUTION OF VSAT NETWORKS

In 1972 the US Federal Communications Commission (FCC) opened the way for domestic satellite systems, and since then there has been a great evolution in the market for private satellite networks. There are four major types of private satellite networks: trunking, business TV, interactive data VSAT, and data broadcast. There is an increasing amount of overlap, and networks are often built for combined applications. The recent rise of VSAT networks in particular has been dramatic. It is estimated that there were around 120,000 VSAT

terminals installed or on order in 1994, and these are expected to grow to 600,000 by 2000. It is expected that terminal sales revenues will more than double, from \$400 million in 1996 to \$910 in 2003. Roughly 82% are located in North America, 6% in Latin America, 6% in Europe, 5% in Asia-Pacific, and 1% in Africa. Europe, Latin America, and Asia-Pacific are the most prolific regions for VSAT expansion, for reasons that differ from region to region.

The relatively slow evolution of VSAT systems in Europe was caused by the strict regulations and monopolies that, until recently, had governed telecommunications in this region. The gradual relaxation of these regulations in Western Europe combined with the need to connect Eastern European countries, where the telecommunications infrastructure is not well developed, has created numerous opportunities for providing fast and inexpensive services via VSAT systems (3). Apart from the fact that VSATs provide the ultimate monopoly by-pass weapon, another influencing factor in Europe is the unique need of service providers to ensure standard levels of service across the entire European continent: VSATs have the capacity to offer a homogeneous international private network.

Rapid deregulation and privatization of government-controlled telecommunication carriers is also driving the VSAT expansion in Latin America. The need to connect locations where no infrastructure exists is also an important factor in Latin America, Africa, and some parts of Asia. In parts of the Asia-Pacific region, which have experienced a very fast economic growth but currently have no reliable telecommunication infrastructure, VSATs are offering a viable solution, mainly for data transmission, for companies with a number of geographically separated branches. The specific geography in Asia plays a major part in the VSAT expansion. In countries such as Indonesia and the Philippines, which consist of a large number of islands, satellite communications are the most viable way to provide telecommunications. The Asian VSAT market is currently domestic with no plans for pan-Asian VSAT networks because there is little commercial or other co-operation between the countries of the region and no regulatory standardization. A unique aspect of Asian VSAT networks is the rapidly increasing demand for telephony on top of the existing data networks. Despite problems such as time delay and loss of quality caused by compression, voice overlays to VSAT networks seem to be the only immediate option for voice communications in Asia, in locations where there is little or no reliable terrestrial infrastructure. There is clearly a need to make this service efficient and reliable and to develop techniques that improve the quality of service these systems can provide.

## OVERVIEW OF VSAT NETWORKS

### Space Segment

The majority of today's telecommunication satellites are located in a geostationary orbit, on an arc 36,000 km above the earth's equator. The scene is set to change significantly in the next few years with the introduction of Low and Medium Earth Orbit (LEO, MEO) satellites. A major limitation of satellite networks is the 0.27 s signal propagation delay for every satellite hop in geostationary systems. Delays are shorter but still significant in MEO (typically 0.080 s to 0.170 s) and LEO

(maximum 0.1 s). These delays play a major part in the protocol design and network optimization.

Space segment was initially available from international consortia, (e.g., INTELSAT, INMARSAT), but as demand for service increased, a number of regional (e.g., EUTELSAT in Europe) and domestic systems (e.g., TELECOM for France) started offering such services. Also, in recent years, a number of companies, mainly in the United States, own or lease satellites that carry their own or their customers' traffic. The military have dedicated satellites, operating in different frequency bands and with performance requirements considerably different to civilian systems (4).

Satellite transponders operating in the Fixed Satellite Service (FSS) typically have bandwidth in the range of 36 MHz to 72 MHz with Equivalent Isotropically Radiated Power ( $r$ ) levels of 30 dBW to 52 dBW. In VSAT systems, both power and bandwidth are limiting resources, and the network designer must always take this into account (5).

### Frequency Bands for Satellite Communications

The International Telecommunications Union (ITU), an agency of the United Nations, is responsible for allocating frequency bands for all forms of radio spectrum. Since 1959, when the first allocation was made for satellite communications, World Radio Conferences (WRCs) have revised and extended these according to changes in demand. WRCs are periodic meetings of delegates from the world's countries to discuss the international table of frequency allocations. These meetings are held every two years.

The first band used extensively for FSS satellite communication was the C-band (6 GHz uplink/4 GHz downlink frequency). The Ku-band (uplink 14 GHz, downlink 11 or 12 GHz) came into use during the mid 1980s, as satellites with higher power became available. At this band, the terminals can be considerably smaller, but there is more sensitivity to rain fade. Most VSAT systems in operation today and DBS TV satellites use portions of the Ku-band (6).

Congestion in the Ku-band and the need for delivery of broadband services via satellite in the 1990s made it necessary to move to a higher band; a number of systems in the planning stage will use the Ka-band (uplink 27 to 30 GHz, downlink 18 to 20 GHz). The main problem at this frequency is the fact that rain attenuation becomes significant (7,8) because the molecular water vapor absorption resonance frequency is located at 22.3 GHz. The European Space Agency's Cooperative Data Experiment (CODE) using the Ka-band transponder of the OLYMPUS satellite was an early example of such a VSAT network (8).

Continuing demand for additional bandwidth has forced commercial satellite system designers to start considering even higher frequency bands, namely the so-called Q-band (33 GHz to 50 GHz) and V-band (50 GHz to 75 GHz). Some military satellite systems already operate in this frequency range.

### VSAT System Components

In this section we briefly outline the various components of a typical VSAT network. A more detailed description of the different components of a VSAT system can be found in Refs. 1, 9, and 10.

1. **Antennas.** The main difference between the hub and the VSAT is the size (and cost) of the antenna. The typical C- or Ku-band hub terminal has an antenna with diameter in the range of 5.6 m to 11 m, costing between \$300,000 and \$5,000,000. VSATs are characterized by much smaller antennas, typically less than 2 m in diameter, and cost less than \$12,000. The EIRP and gain of the VSAT is thus considerably lower than that of the hub terminal.
2. **High-Power Amplifiers.** Traveling Wave Tube Amplifiers (TWTAs) are generally used in hub earth stations because they can have output levels of several hundred watts and are capable of being tuned across an individual satellite uplink band. Klystrons are used for single frequency uplinks, e.g., TV.
3. **Solid State Power Amplifiers.** Solid State Power Amplifiers providing output powers of 50 W to 100 W in C-band and 20 W in Ku-band are now available, taking advantage of recent improvements in semiconductor technology. These power levels are adequate for final-stage amplification in VSAT terminals.
4. **Low-Noise Converters.** These amplify and downconvert the received signal while trying to minimize the added noise. It is very important to make sure that no spurious signals are generated and that phase noise in the oscillators is kept to acceptably low levels for data transmission.
5. **Up/Down Converters.** The function of the up converters is to translate the signal intended for transmission from an intermediate frequency (typically 70 MHz) to a microwave signal [6 GHz (C-band), 14 GHz (Ku-band)]. Down converters perform the reverse operation, translating the received microwave signal [4 GHz (C-band), 12 GHz (Ku-band)] to a lower intermediate frequency.
6. **Modems and Codecs.** Because size and cost of equipment is of paramount importance, modulators and demodulators are incorporated in a single unit (*modem*). Similarly, encoders and decoders are built into a single *codec* unit. Typically, a modulator converts a digital signal into a Phase Shift Keying (PSK) signal and the demodulator reverses the process.
7. **Network Interface Unit (NIU).** An NIU is used to implement the user protocol interface and access to the satellite.

### Modulation and Coding

Although a wide range of modulation schemes are currently in use, constant-envelope modulation schemes such as phase modulation are preferred in satellite communications. These schemes require a constant power level irrespective of the data transmitted; therefore, there is no need to adjust the transponder load or apply smoothing techniques.

VSAT networks were predominantly intended for packetized data transmission; therefore digital phase modulation schemes were the natural choice, in the form of binary phase shift keying (BPSK) or quadrature phase shift keying (QPSK).

VSAT systems used to be power limited (especially in the downlink) so that any underutilized transponder bandwidth

could be used for digital encoding. Forward Error Correction (FEC) coding is one option, where redundant bits are added to the bit stream so that errors can be detected and corrected at the receiver. It is always desirable to keep the coding schemes as efficient as possible, in order to use the expensive space segment at maximum efficiency and keep the cost and complexity of the terminal equipment to a minimum, keeping in mind that the channel usually has a very low bit error rate ( $10^{-7}$  to  $10^{-9}$ ).

### VSAT NETWORK ARCHITECTURE

Figure 1 shows the main components of a typical VSAT network. VSATs communicate with each other and with the hub via a shared forward link, whereas hub-to-VSAT communication takes place via a separate, usually faster link, configured using conventional Time Division Multiplexing (TDM). Although there are different choices for data rates, modulation techniques, and transmission formats among different VSAT networks, there is, in general, an agreement on the use of TDM from the hub to the VSATs.

The multiple access link from the VSATs to the hub, however, has been subjected to a greater degree of variation in data networks built during the past decade. Even though the choice of the data rate, modulation, and encoding techniques and transmission formats have a major impact on the network performance, it is probably accurate to characterize the choice of access technique employed as the primary feature distinguishing one network from another (2), and this is one of the key decisions to be made by the network designer.

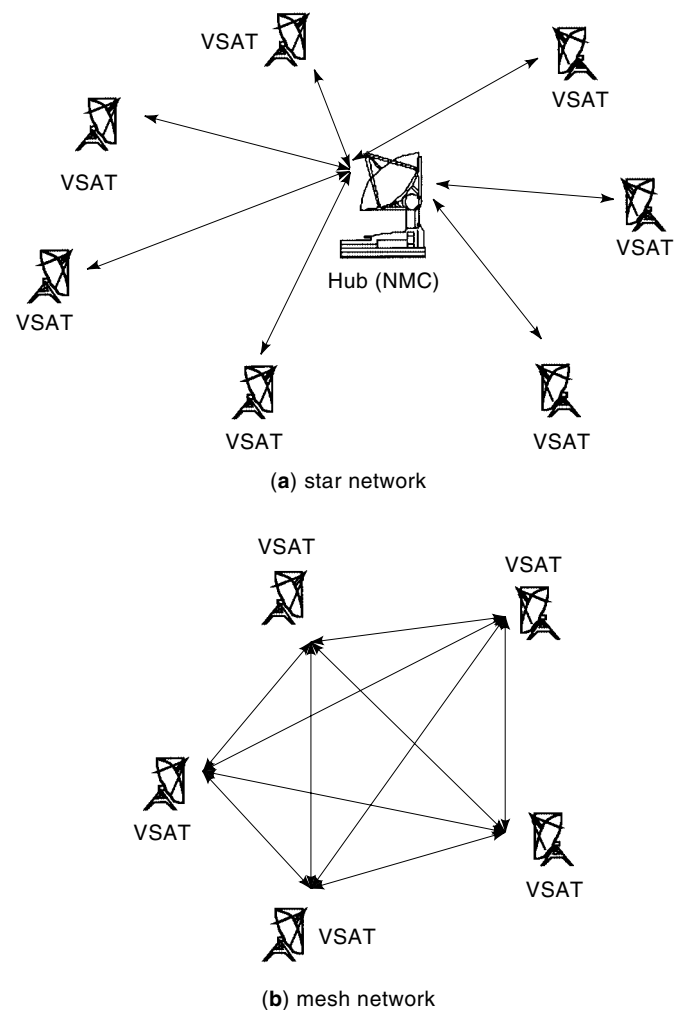
Satellite networks can be classified according to the arrangement of the links between the various VSATs. This is referred to as the architecture, configuration, or topology of the network. The selection of multiple access, multiplexing, and modulation schemes for a particular network depends heavily on the network architecture.

Most VSAT networks use one of two possible network architectures, although some hybrid topologies can occasionally be used for specialized services:

- In a *star* network [Fig. 2(a)], the VSATs always communicate with each other via the hub; this is ideal for an application involving point-to-multipoint transmission (e.g., a hub located at the company headquarters sending information to a number of branches).
- In a *mesh* network [Fig. 2(b)], the hub acts as a Network Management Center (NMC) for channel allocation and policing, but VSATs, having established a connection, talk directly to each other via the satellite. A typical example of a mesh application would be telephone service provision.

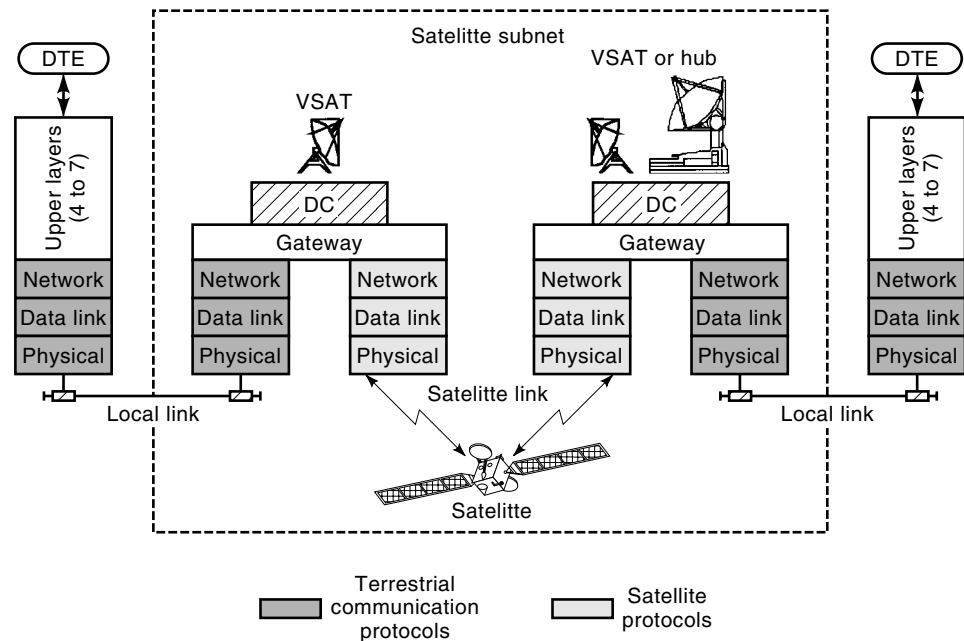
### Communication Protocols in VSAT Networks

In telecommunication networks, the process of communicating is separated into an orderly sequence of software procedures called layers, and the classic example of this process is the attempt to standardize this using the 7 layer Open Systems Interconnection (OSI) Reference Model, defined by the International Organization for Standardization (ISO) with the co-operation of ITU.



**Figure 2.** Pictorial representation of two possible network topologies: (a) Star VSAT network: The network management center (hub) is shown in the center of the network, with a number of VSATs connected with it, forming the points of the “star”. In this case VSATs talk only via the hub, usually sending information to it or receiving data from it. A typical example would be corporate headquarters (hub) and geographically distributed branches (VSATs). (b) Mesh VSAT network: A number of VSATs connected in a mesh. Each can talk to some or all the others, and the network control is assumed by one of the stations, determined in advance. A typical example would be telephone service provision/teleconferencing between a number of remote locations.

A representation of the general layered protocol approach in a typical VSAT network is given in Fig. 3. The VSAT Data Terminal Equipment (DTE), installed at a user’s premises, communicates directly with its local Data Circuit-terminating Equipment (DCE) and indirectly with the remote DTE. DTE to DCE communications use the physical (layer 1), data link (layer 2), and network (layer 3) layers to exchange data. DTE-to-DTE communications rely upon the services provided by the lower layers (1 to 3) and can use the higher layers (4 to 7) for end-to-end communications. The gateways perform the translation process between the internal network procedures (DCE to/from DCE) and the DCE-to-DTE exchange procedures.



**Figure 3.** Protocol layer architecture of a VSAT network connected to a terrestrial network. The satellite subnet is shown in the dotted box. As most VSATs will need to interface to existing terrestrial networks, this shows how the protocol layers would look in such an architecture.

Existing VSAT networks offer performances similar to terrestrial data networks and use the most common user-network interconnections and interface protocols. The fact that VSAT networks operate internally in a packet mode does not limit users to packet-oriented communications because packetizing functions can be performed in the interface units of the VSAT terminals. The signal processing terminal of a VSAT forms the local DCE at one end of the network. At the other end, the DCE is formed by either

- the hub signal processing terminal (VSAT-to-hub communication) or
- the signal processing terminal of another VSAT (VSAT-to-VSAT communication), provided that this is permitted by the network architecture

The packet-switched nature of VSAT systems implies that an error detection code embedded in the packet format causes packets received with errors to be rejected. There are two possible options:

1. In the first option, the acknowledgment process (at the data link layer) actually takes place in the satellite network management software, and each erroneous packet is repeated. This is usually the case in random assignment systems (e.g., using an Aloha assignment process). In this category, the BER accounted for from the link budget should be low enough to avoid frequent repetitions of the messages. Taking a typical numerical example, a  $10^{-7}$  BER corresponds to a packet error rate of  $10^{-4}$  for a 128 byte packet, which is satisfactory. For most data communications, it is essential that all messages are completely error free, but for other services some error rate might be acceptable (e.g.,  $10^{-2}$  frame loss for packetized speech). It is obvious that the acknowledgment process must take into account the long satellite propagation delay.

2. In the second option, there is no acknowledgment process in the satellite network software. This is, for example, the case of the hub-to-VSAT satellite link. Here there is no data link layer, and the error recovery mechanism take place in the application layers (i.e., into the host protocols). For these reasons, any erroneous data will be repeated only after a very long delay. Therefore, the link budget parameters in this case have to be arranged for a much better BER, typically around  $10^{-10}$  giving a packet (of similar size as earlier case) error rate of  $10^{-7}$ .

#### MULTIPLE ACCESS IN VSAT NETWORKS

The satellite bandwidth is a very expensive shared resource, which must be cooperatively used by its users. Thus, in a VSAT environment, there is a need to use appropriate multiple access schemes. Similarly, there may be interference on the downlink from other satellite or terrestrial sources. This interference is usually combatted by spatiotemporal signal processing and better antenna sidelobe performance. However, the downlink stream may contain a collection of information packages in which each package is intended for a proper subset of the set of users. This situation is common in communications via shared media. Thus there is a need to multiplex information so that, when it is sent down from the satellite, it can be properly separated by the users. Although multiple access and multiplexing can be considered as two aspects of the general sharing problem, they can often be designed independently. We will focus here on the multiple aspects with appropriate commentary on multiplexing as needed.

In a typical star network, VSATs transmit data in packets to the hub station using the multiple access capability of the satellite channel. Because there are no direct links from one VSAT to another, any VSAT-to-VSAT traffic must follow a path of two satellite hops from source to destination. Al-

though this represents a serious limitation on the system, it is forced by two main reasons. First, transmit and receive power requirements on the VSATs can be relaxed considerably by the fact that the uplink and downlink benefit from the higher performance capabilities of the hub station. (This implies that the size of the VSAT can be kept as small as possible, with significant economic, regulatory, and planning benefits.) Second, the objective of the majority of VSAT networks is to establish communication from a large number of dispersed users to a central information resource so that a star architecture must be used.

The link from the hub station to the VSATs is usually configured using time division multiplexing. If frequency division multiplexing (FDM) is used in the downlink, several carriers (corresponding to several signals received by the satellite) must be amplified simultaneously for transmission. However, transformer amplifiers are not strictly linear in operation, and the degree of nonlinearity is greatest as maximum power output. The nonlinearity causes input carriers to generate intermodulation products at the amplifier output, which are signals at frequencies other than those input to the amplifier. Such products distort the transmitted signal and cause interference between VSATs. The solution is to reduce the power of input signals to reduce correspondingly the power of the undesired products. Of course, this input backoff reduces the power of the desired output signals as well. Thus, for an FDM downlink, an amplifier must be operated at less than full power. Now if TDM is used with a constant amplitude modulation, there is only one input signal to be amplified, which produces no inband intermodulation products. Furthermore, unlike the case of an earth station transmitting in a TDMA plan, a satellite producing a TDM downlink transmits essentially continuously (whenever there is at least one call/session through the satellite).

The multiple access link from the VSATs to the hub has been subjected to a greater degree of variation. Even though the choice of the data rate, modulation, and encoding techniques and transmission formats have a major impact on the network performance, it is probably accurate to characterize the choice of access technique employed as the primary feature distinguishing one network from another. The main difficulty in satellite access protocols is the long propagation time of the geostationary satellite link which can impose unacceptably long coordination times.

In a VSAT satellite network, a number of earth stations transmit messages (via the uplink channel) to a satellite in orbit, which then relays the messages down (via the downlink channel) to the earth stations. Because only a single station can successfully transmit a message on a channel at any given time, the stations must somehow coordinate their access to the channel in order to share its use. The distributed algorithm, which defines the rules of this channel sharing is known as the *Channel Access Protocol* (CAP). The problem is further complicated because:

- Stations are distributed, so they must either explicitly or implicitly communicate information to each other to coordinate the channel sharing. However, because the channel is the only communication medium, such coordination requires the use of the channel itself.

- Stations are separated so they can never instantaneously know the present status of other stations in the environment; information of this type will always be at least as old as the amount of time required for the message to propagate along the channel to its destination.

We can classify multiple access schemes according to

1. Channel allocation method:

- *Pre-assigned multiple-access*, in which two stations are permanently assigned the channels required for their connection for their exclusive use. This is usually a connection carrying a heavy traffic load at all times, such as telephone trunk connections.
- *Demand Assignment Multiple Access (DAMA)*, in which the channel allocation changes in accordance with the originating call. The channel is automatically selected and the connection is active only while the call is continued. This system improves substantially the efficiency of the transponder utilization and the overall system performance, in comparison with pre-assigned multiple access, and is especially useful when there is a variation of the traffic load with time at the various stations.

2. Type of transponder sharing:

- *Frequency Division Multiple Access (FDMA)*, where each station has its assigned carrier frequency upon which it can transmit. With FDMA, many users can simultaneously share the satellite. FDMA has been used for decades in satellites for uplinking and downlinking of analog signals.
- *Time Division Multiple Access (TDMA)*, where all stations use the same carrier frequency and bandwidth with time division. Users are assigned positions in a quickly repeating schedule for transmitting on a common frequency to the satellite. The ability to buffer digital data and to maintain tight synchronism have rendered TDMA a practical access technique.
- *Code Division Multiple Access (CDMA)*, where all stations share simultaneously the same bandwidth and recognize the various signals by some type of code identification. CDMA is an application of spread-spectrum technology. It uses pseudonoise patterns, or codes, which are used to change quickly the characteristics of the transmitted signal at a rate usually greater than that of the bit stream to be transmitted.

Because of the bursty nature of typical VSAT traffic, random access schemes are often used in VSAT networks. The Aloha random access scheme (11) and its numerous variations (12) is especially important in VSATs, with its simplicity of implementation and its suitability for bursty traffic loads. In Aloha, new packets arriving at a station are transmitted, and when two or more of these transmissions overlap, they mutually destroy each other. There is feedback from the receiver (collision/no-collision binary feedback) so that all users learn whether or not a collision occurred. When a collision takes place, the “colliding” senders time-out and retransmit their packets after a random waiting time, selected independently at each station. There is also a slotted version of Aloha in which messages are transmitted only at regular time inter-

vals (“slots”), and this has the effect of doubling the effective throughput because packets can be either successful or suffer complete collision, as opposed to the possibility of partial collision in the unslotted case. Figure 4 shows the concept of the Aloha protocol.

Relatively simple schemes such as Aloha have been very successful in networks that carry a particular type of traffic, which is usually fairly bursty (e.g., short interactive messages at random intervals) and where it would be wasteful to have a fixed bandwidth assigned to each user. They can provide low access delays for lightly loaded channels and are simple to implement. The need to be able to offer a wider range of services (and especially incorporate stream-type traffic, like speech) means that more sophisticated, “second generation” protocols must be developed, which will be flexible and adaptive to particular traffic demands.

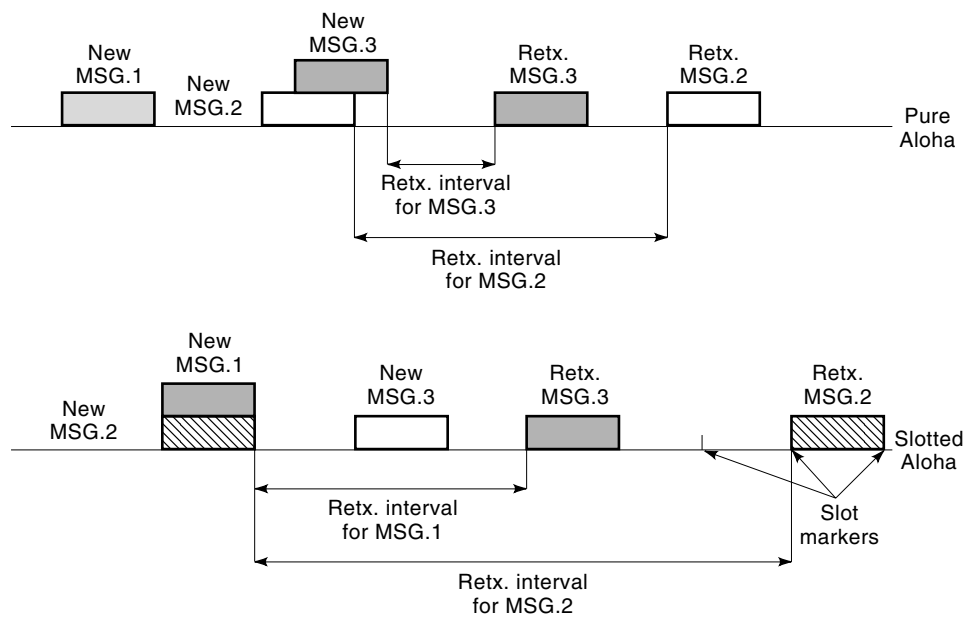
Selective reject (SREJ)—Aloha is an unslotted random access technique that can achieve a maximum throughput comparable to that of Slotted Aloha but without requiring timing synchronization (4,5). It involves a subpacketization of messages in conjunction with a selective reject ARQ retransmission strategy. The operation is similar to unslotted Aloha, but only the collided parts of messages are retransmitted. It is well suited for networks handling frequent short, interactive-type messages and periodic longer file transfers (13).

Another important performance advantage is the apparent insensitivity of SREJ-Aloha to the new message length distribution. This is particularly important when the input traffic consists of a wide variety of applications that could have different lengths (e.g., single-packet messages), very long file transmissions, and a mix of uniform and exponentially distributed medium-size data messages. This performance could be further enhanced by appropriately adjusting the various protocol parameters, such as the retransmission policy and the subpacketization size, to changes in the traffic load.

In a VSAT scenario, the key criteria in the protocol selection are

1. The *efficiency or throughput* of channel sharing. This can be defined as the proportion of the time “useful” traffic (i.e., information) is carried by the channel.
2. The *access delay*. This is defined as the time between the arrival of a message at a VSAT and the start of its successful transmission on the channel.
3. The *protocol stability*. This is the operating region in which the protocol performance remains bounded and relates to avoiding the possibility of long-term congestion or unstable operation.
4. *Robustness* in the presence of channel errors, fading, and possible equipment malfunction.
5. *Operational properties*, which include considerations of system start-up, monitoring, network expansion, addition of different traffic types.
6. *Cost and complexity of implementation*. This relates to the cost and complexity of the various parts of hardware and software required for a particular VSAT system.

The protocol that offers the best combined performance based on these criteria will be considered the “optimum” choice for a particular system/traffic scenario. It is important to note that in VSAT networks, the efficiency of bandwidth use is not the only factor affecting multiaccess technology. Individual stations generate a relatively small amount of traffic, so that transmission cost per VSAT is overshadowed by the cost of the VSAT station and its operation. In addition, power and not bandwidth is the most limiting resource in a VSAT network, permitting the use of relatively inefficient protocols (such as Aloha) without significant impact on system capacity. Therefore, for the interactive data VSAT network environment, low delay, simplicity of implementation, and robust operation are generally of greater importance than the achieved bandwidth efficiency (2).



**Figure 4.** Pictorial representation of channel events in Aloha-type protocols. Messages originating or retransmitted from same station are represented by same color blocks. Collisions, resulting in retransmissions are represented by overlapping blocks. Note that for the slotted case only complete overlaps (collisions) are possible, while partial overlaps can take place in the unslotted case.

### PROVISION OF INTEGRATED SERVICES USING VSATs

Today's technology allows a wide range of services (data, voice, facsimile, etc.) to be transmitted using a low-speed digital circuit. High volume T1/E1 satellite links are no longer the only means for integrating communications, and a number of small- and medium-volume users can take advantage of the savings a fully integrated service offers. The main driving force to integrate communications is not the technology itself, but the cost savings and efficiency that data/voice integration offers. Integration at low data rates allows a company to combine multiple voice, fax, data, and LAN traffic over a single satellite link using a VSAT network.

A key technology to low-speed network integration is speech compression. A 64 kbit/s satellite link can carry a single Pulse Coded Modulation (PCM) voice signal but no other traffic. Using a combination of digital signal processing (DSP) technology and silence suppression however, voice signals can be compressed to as low as 4.8 or 2.4 kbit/s; therefore, more than one call can be multiplexed over a low-speed link. Further efficiency improvement can be achieved by LAN/data connectivity devices with compression technology. A typical network may carry a combination of packet data and LAN traffic of a bandwidth of about 38.4 kbit/s. With 2-to-1 data compression only half of that is required for transmission. In a similar way as with voice, a fax transmission that would normally require a 128 kbit/s link could be digitized to run at 9.6 kbit/s over the network.

### Integrating Various Types of Data Traffic

First we consider a typical interactive, inquiry/response-type data traffic scenario, with relatively low and bursty traffic volume at each station. In general, for a random access scheme to be a feasible choice, the average data rate at each station needs to be orders of magnitude lower than the available channel speed. For simplicity, we assume only a single terminal is connected to each VSAT, although it will not be difficult to extend our analysis to the case of several terminals multiplexed at each VSAT.

### Voice/Data Integration

Different classes of traffic have different performance requirements. Stream-type traffic (voice calls, file transfer) usually require a collision-free allocated channel, whereas random access transmission could work well with small data messages. It is therefore necessary to develop an access scheme that accommodates all traffic classes in an efficient way. One way of treating this problem is to develop a combined random access/channel reservation protocol and to try to make it adaptive to changing traffic mix.

In this case we consider a traffic scenario consisting of three types of service:

1. *Small Size Data Message Transmission*, typically single packet messages. These could be updates of the value of a particular quantity (e.g., share price) or specific requests from a central database (e.g., number of items in stock).
2. *File Transfer Transmission*. Connections for file transfer or other relatively long data transmission (e.g., complete list of prices for items on sale).

3. *Voice Calls*, typically business calls of short duration (e.g., a mean call time of 120 s), using low-bit-rate coded voice for efficient use of channel capacity.

### Dynamic Channel Allocation

Most existing VSAT networks handling speech use dedicated channels that can be predefined or allocated on demand. One way to optimize the overall network performance would be to develop an efficient dynamic allocation scheme. A possible way to achieve this is discussed next.

Assuming we have, for example, a bandwidth of 64 kbit/s for the overall VSAT network, we can separate this into 8 channels of 8 kbit/s. If we have only data traffic in the network, a random access scheme, like the ones described earlier, could operate on the channel. If a request for a voice call arrives however, a portion of the bandwidth is allocated to this, whereas the remaining bandwidth continues to operate for the data transmission at a lower channel speed. The system can have a maximum of seven voice calls at any time because there should always be a channel for reservations and data transmission using random access. If there are no free voice channels available, a busy tone will be sent with the acknowledgment, and the caller must redial to establish a new connection. Because for such a system the blocking probability is given by the Erlang-B formula:

$$P_B = \frac{\frac{(s\rho_v)^s}{s!}}{\sum_{k=0}^s \frac{(k\rho_v)^k}{k!}}$$

where  $s$  = Number of channels,  $\rho_v$  = Voice traffic load/channel.

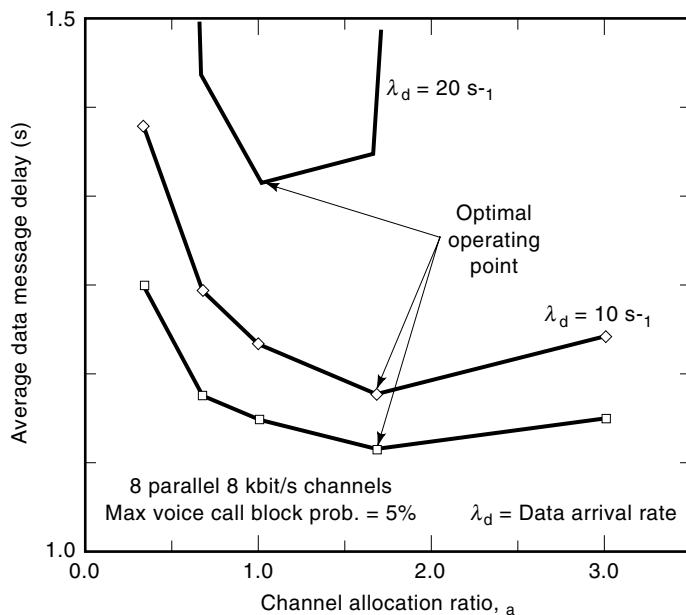
We can thus estimate the maximum call arrival rate the system can handle for a mean call-handling time and a specific number of voice channels. By plotting the end-to-end mean data message delay for various data arrival rates and a particular voice call blocking probability the optimal channel allocation ratio  $a$  = (No. of Reservation Channels/No. of Message Channels) for a particular traffic load can be determined (Fig. 5).

There is clearly a need to optimize the channel allocation (i.e., number of channels allocated to voice calls or random access reservations) to suit the traffic mix. If this mix is known in advance, we can adjust the allocation of reservation and voice channels accordingly. If, however, the traffic mix changes considerably, it would be beneficial if we could adjust this allocation at regular time intervals. By defining a maximum blocking probability for voice calls, we can estimate the data message performance for various loads and determine the channel allocation ratio that provides "optimum" performance (14,15). The network management center can then periodically update the channel allocation based on the traffic load and inform the stations accordingly.

### Compression and Multiplexing Techniques

We next focus on current research on techniques that can reduce the bit rate for telephone signals to 16, 8, 4.8 kbit/s or even lower and thus maximize the efficient use of the satellite channel (even if this results in some degradation in the quality of service). This implies the use of sophisticated coding





**Figure 5.** Plot of channel allocation ratio (a)  $V_s$  average message delay (in seconds) for various message arrival rates and a maximum Blocking Prob. of 5% for voice. Figures like this can be used to determine the optimal operating point (local minimum point for each curve). This information can then be used by the network management center to allocate channels for particular traffic loads so that operation remains close to this point (and delays remain minimum) for different total message arrival rate (represented by the different curves).

schemes that require a considerable processing time (tens of milliseconds) and need a specific internal frame time for packet arrangement management. It is obvious that the Quality of Service required would determine the required minimum bit rate.

One such speech-coding algorithm is the code-excited linear predictive (CELP) coding (16), which has been shown to produce good-quality speech at bit rates below 16 kbit/s. There is an increasing number of CELP-based coders developed (17). The basic CELP algorithm applies vector quantization of the excitation signal to achieve efficiency in coding, a technique described as fractional bits per sample coding. However, the high complexity of CELP and its relatively low robustness to transmission errors means that the basic algorithm has to be extensively revised to meet the constraints of a VSAT application. Because of the high propagation delays the speech signals experience, the speech codec implementation should also reserve processing capacity for echo cancellation. Note that this is a rapidly changing field, and new coding schemes offering better performance and higher compression are being continuously introduced.

#### Multirate Coding for VSATs

It is possible to use a speech coder with multirate capability with a VSAT system. In such a system, if only packetized speech needs to be transmitted, a higher-rate codec can be switched on to provide better-quality speech. The arrival of data messages at the VSAT could trigger a switch of the speech to a lower-rate codec that will allow the multiplexing of data packets on the same channel, using the bandwidth

that becomes available from the higher speech compression, at the expense of lower speech quality.

#### Re-Using Speech Silences for Data Transmission

Having discussed the problem of sharing a common satellite channel among a large number of VSATs, we next turn our attention to maximizing the efficient use of this resource. We look at the possibility of integration of voice and data on a second level, over the same channel.

Observations on the nature of speech (18) show that a speech source creates a pattern of active talkspurts and silent gaps. There are principal spurts and gaps related to the talking, pausing and listening patterns of a conversation. There are also “mini-gaps” and “mini-spurts” caused by the short silent intervals that punctuate continuous speech. A commonly used approach is to represent the voice source as a two-state Markov modulated poisson process (MMPP) (19). The mean (exponentially distributed) durations of the states and the values of the Poisson arrival rates in each state are sufficient to define the Markov process. Statistical analysis on a number of conversations shows that the “active” period covers approximately 40% of the time, whereas 60% of the time consists of a mix of long and short silences. Therefore, by using a speech activity detector close to a speech source, one can distinguish between active and silent parts in a conversation and allow re-use of the channel when a silence is detected. With a “slow” detector, one can distinguish between active periods and long silences (two states), whereas with a “fast” detector, three states can be observed: active, long silence, short silence.

If we have a low bit rate vocoder with voice activity detection, assuming this takes four to five 30 ms speech frames for a silence detection to take place, we can take advantage of the silence intervals that are longer than two frames to transmit data packets from the same source. This will help us make a more efficient use of the channel and will reduce the transmission delay of long files. These can be broken into smaller fixed-size packets, given a sequence number, and transmitted during these silences. The waiting time before the data transmission is completed should ideally be as low as possible, although it is not a critical limitation in this type of file transmission.

The basic parameters that define the operation of the voice/activity detector (VAD) are:

- The *threshold level* above which speech is assumed to be present (typically  $-30$  dBm to  $-40$  dBm for a fixed threshold level switch).
- The time taken by the voice switch to operate on detection of speech (typically 6 ms to 10 ms).
- The *hangover time* during which the voice switch remains activated after the cessation of an active speech burst (typically four to five frames).

The threshold level needs to be set to a low value to avoid missing large portions of speech at the beginning of a talkspurt and at low speech levels. This of course makes the system susceptible to high ambient noise levels, which is one of the problems that must be taken into account. In order to eliminate the possibility of cutting out speech midbursts, a

further condition is applied, by adding a hangover stage to the VAD output.

The operation of the VAD is based upon these basic assumptions (20):

- Speech is a nonstationary signal. Its spectral shape usually changes after short periods of time, typically 20 ms to 30 ms.
- Background noise is usually stationary during much longer time periods, and it changes very slowly with time.
- The speech signal level is usually higher than the background noise level, otherwise speech is unintelligible.

Based on these assumptions, a VAD algorithm can be developed that can detect silence gaps and distinguish background noise (with or without speech). Assuming that in most VSAT systems the background noise is relatively low, a simple fixed energy threshold can be used to detect the silence regions (unlike mobile systems, where there is a high and variable noise environment that needs to be compensated by a more adaptive algorithm). Various implementations of VAD systems for CELP coding can be found in the literature (17). We can therefore use a multirate codec to avoid congestion in cases where traffic load is higher, whereas we can revert to better speech quality service when we have less traffic. It is important to fill the silence periods with generated noise to prevent background noise silences causing unnatural sounding telephone links.

#### ATM VIA SATELLITE

Asynchronous Transfer Mode (ATM) has been adopted as the main technology for the implementation of the Integrated Broadband Communications Network (IBCN). However, the deployment of an ubiquitous terrestrial infrastructure to support this technology could probably take many years and the traffic demands on this network are as yet unknown. Satellite networks offering broad geographical coverage and fast deployment appear to be an attractive option for the early deployment of the IBCN and could play a major role in its development, provided a number of difficulties arising from the nature of satellite systems can be overcome (21). A more detailed discussion of ATM over satellite can be found in Refs. 22 and 23.

In ATM, information flows in fixed-size blocks called cells, each consisting of a header and an information field. Cells are transmitted over virtual circuits, and routing is based on the virtual circuit identifier (VCI) contained in the cell header. Slots are allocated to a call on an asynchronous (demand-based) manner and the bandwidth is efficiently used because no bandwidth is consumed unless information is actually being transmitted. ATM can accommodate variable bit rate (VBR) services and can be used to improve bandwidth efficiency by statistically multiplexing traffic from bursty sources. ATM can also accommodate circuit-oriented and continuous bit rate (CBR) services by allocating bandwidth based on a fixed rate for a connection, given that sufficient resources are available (24).

ATM has two major aspects: the multiplexing aspect (achieved by the segmentation into standard-size cells) and

the switch management aspect that ensures that the quality-of-service (QOS) guarantees are met for each of the multiplexed traffic commodities. The multiplexing aspect is easily handled over the satellite channel, provided the appropriate modifications to the cell structure are made (some of these were just outlined). The QOS aspect, however, is more challenging. If no switch management is to take place on-board the satellite, then that aspect can also be handled (on the ground) by viewing the satellite link just as the traditional, long-propagation repeater ("bent-pipe") link. However, this is limiting considerably the potential role of the satellite. If on-board processing and switching is introduced in future satellites, this would give new degrees of freedom that provide flexibility and potential performance improvement. However, managing the on-board switch to ensure QOS guarantees is a daunting task. The packets (or cells) of each multiplexed traffic commodity encounters errors and delay on the uplink that must be taken into account by the switch. Thus an intelligent, dynamic switch management process must be developed in order to provide the appropriate priority handling to each cell (25).

VSATs can be used for:

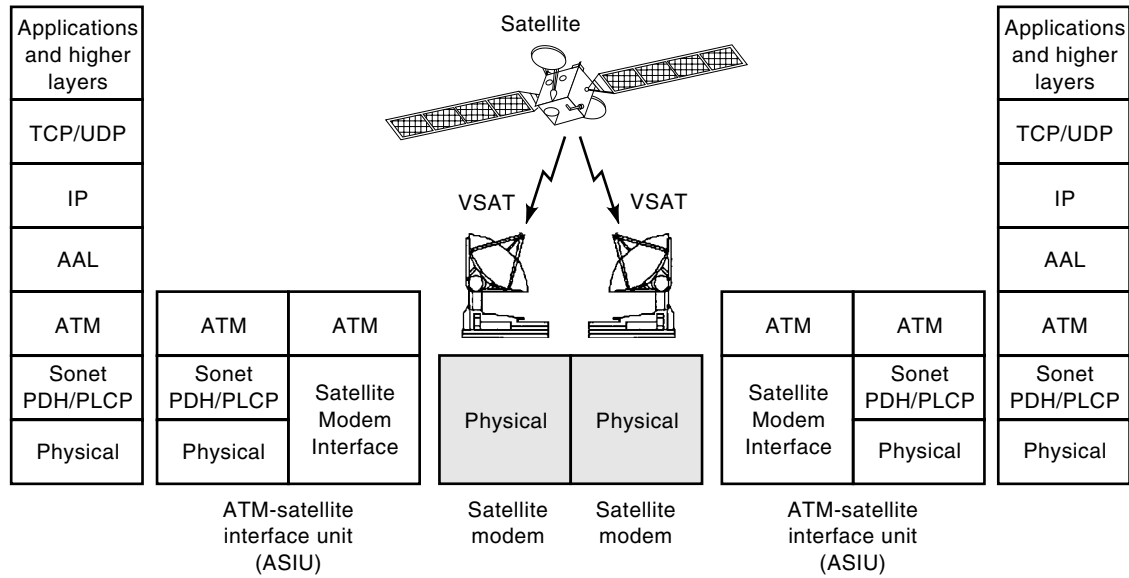
- the interconnection of a few geographically distributed broadband networks, usually called broadband islands.
- the provision of a network interface to a large number of thin-route users that want access to the IBCN.

In order to use VSATs for the provision of seamless interconnection of local area networks (LANs) and metropolitan area networks (MANs) using ATM, a number of problems need to be addressed. Suitable conversion protocols and satellite-ATM interface units (SAIU) between various LAN/MAN architectures need to be developed for efficient and seamless interconnection. Efficient flow control mechanisms (26) are needed to minimize the cell losses, taking into account that the satellite channel is a limiting bottleneck in terms of bandwidth and delay. Finally traffic control mechanisms (22) that take into account the characteristics of the satellite environment need to be developed in order to ensure that QOS guarantees are met. Figure 6 shows the protocol layer architecture for ATM over satellite transmission.

The effect of higher error rates over the satellite channel represents a problem in the integration of satellites with terrestrial B-ISDN (27,28). Another complication is the possibility of bursty errors in a satellite system, especially FEC-based links for high power efficiency. Because ATM header error check (HEC) is able to correct only single-bit errors, the burst errors in the ATM header cannot be corrected. Therefore, there might be a significant increase in ATM cell discard probability, which is defined as the ratio of the number of ATM cells that are discarded because of uncorrectable errors to the total number of cells received. The burst error characteristics can also affect the performance of ATM adaptation layer (AAL) protocols causing severe cell discard at the physical layer, and there is a need to compensate for this by using interleaving mechanisms, error recovery algorithms, or efficient concatenated coding schemes for error correction.

#### EXAMPLES OF NOVEL VSAT SERVICES

A vast and diverse number of applications and traffic types could also be served better by broadband satellites. Distance



**Figure 6.** Protocol layer architecture of a VSAT network carrying ATM traffic. A general VSAT gateway protocol layer structure for a network connection supporting ATM traffic via a repeater satellite is shown. Peer communication between the layers is demonstrated. Internet applications require a TCP/IP connection on top of ATM, while some type of framing (SONET, PDH, PLCP) is required for efficient transmission over the satellite. An ATM-satellite interface unit (ASIU) sitting between the higher layers and the satellite modem is also shown.

education and telemedicine are two important and, for the developing regions of the world, critical services. Transmission of financial transactions, videoconferencing and connection of private business intranets will also be among the main services supported by the next generation systems.

**ASYMMETRIC TCP/IP TO SUPPORT INTERNET APPLICATIONS VIA SATELLITE**

Access to the Internet is either too slow [e.g., dial-up Serial Line Interface Protocol (SLIP)] or too expensive (e.g., switched 56 kbps, frame relay) for the home user or for small enterprises. It is however possible to take advantage of much broader bandwidth available through broadcast satellites and to use a low-cost hybrid (dial-up and satellite) network terminal which can deliver data from the Internet to the user at rates up to 400 kbps (29). An asymmetric Transmission Control Protocol/Internet Protocol (TCP/IP) connection is used to break the network link into two physical channels: a terrestrial dial-up link for carrying data from the terminal into the Internet and a receiver-only satellite link carrying IP packets from the Internet to the user. With a goal of supporting bandwidth intensive Internet applications such as HyperText Transfer Protocol (HTTP), and File Transfer Protocol (FTP), this system has been designed to support any personal computer, any commercial TCP/IP package, any unmodified host on the Internet, and any of the routers within the Internet. The design exploits the following three observations: (1) satellites are able to offer high bandwidth connections to a large geographical area, (2) a receive-only VSAT is cheap to manufacture and easier to install than one that can also transmit, and (3) most computer users, especially those in a home environment, will want to receive much more data than they gen-

erate. IP encapsulation, or tunneling, is used to manipulate the TCP/IP.

The Hughes DirecPC system provides an asymmetric access to the Internet. In this system, a user's inbound information arrives via a satellite link, whereas outbound traffic travels over a conventional modem connection. To achieve the required routing of a user's inbound information from remote Internet hosts to a satellite gateway station, tunneling is used. With tunneling, a user's outbound IP datagram is encapsulated at his machine within another IP datagram. This IP datagram is routed to the DirecPC system, where the encapsulation is removed. The original IP datagram is then changed to that of the gateway so that information from the remote host is returned to the gateway instead of to the user via his Internet service provider. When the desired information from the remote host arrives at the gateway, it is sent over the satellite and received by the user with a small satellite antenna dish. With this system, a downlink rate of 400 kbps can be provided to the user (29).

**Multicasting**

One of the major applications of broadband satellite systems will be the multicasting of information to a large number of dispersed users. Given the maturity and widespread adoption of IP multicast protocols, we can expect a demand for applications in wireless environments, including those of VSATs. Satellite-based videoconferencing could be accomplished by tunneling the IP-multicast messages through the satellite gateways, but this would require the setting up of multiple tunneled virtual circuits between geographically separate users, making group management difficult and using more satellite capacity than would be necessary, if the satellites' on-board switches were to support IP multicast directly (30).

However, efficient use of satellite constellations for group applications requires that the satellites' on-board switches include support for multicast. This appears unlikely in the commercially proposed schemes. Leaving implementation of multicast solely to the IP-routing ground networks, rather than forcing it on both ground and mesh satellite networks, would appear to make the problem of implementing efficient Internet work multicast with a satellite component more tractable.

It is possible to provide LAN clients without Internet access multicast streams using a DirecPC machine as the clan's multicast server. The concept is to have a machine within the Internet receive a multicast stream and forward the tunneled IP packets to the DirecPC machine. The DirecPC machine will then be responsible for receiving the tunneled packets and distributing them to the LAN as multicast packets.

### Web Page Caching

There are plans to increase the efficiency of the Web network connections by carrying the most commonly used data to the Internet service providers (ISPs) by satellite. It is possible to use broadband satellites to distribute commonly requested information to local ISPs, where it would be stored locally or "cached" for distribution to customers. The setup would lessen congestion on Internet backbones and cut communications costs for service providers.

This service is a more efficient way to get Internet data to ISPs because the bulk of Web page requests concern a relatively predictable set of data. Storing that data in caches at the "edge of the Internet" speeds delivery of pages to users and relieves network backbones from the congestion caused by redundant data. To pump the Web's most popular data into a local cache at the ISP, a satellite dish of 1 m diameter would feed data to a receiver at a rate of 4 Mbps.

### SUMMARY AND FUTURE DEVELOPMENTS

VSAT networks have been one of the fastest growing areas in satellite communications. The complex mix of technological innovations, market forces and deregulation of the telecommunication industry will allow the further exploitation of the capabilities offered by these versatile systems. The recent developments of standardized protocols and interfaces for packet-switched networks makes it possible to interface VSATs to existing terrestrial networks and offer the same services available from terrestrial networks at competitive prices. Following the recent drive for integrated service networks, VSAT networks are entering a next generation or phase of development, and their role is starting to change. From specific (mainly data) service providers, they are becoming an efficient way of offering network interconnection and a wide range of two-way multimedia services to remote users.

A number of technical issues need to be resolved so that these new services become financially competitive. The development of adaptive, dynamic protocols for this wide range of services will result in the most efficient allocation of the space segment, one of the most expensive commodities in the service provision, is a critical requirement. Recent innovations, such as the development of cheap, low-bit-rate vocoders with voice activity detection could further improve the efficient use of the channel. Finally, because VSAT networks need to be

made compatible with the developing terrestrial IBCN and to extend to areas such as network interconnection and provision of ATM-based services, we have addressed some of the problems that need to be resolved and provided some suggestions of the important role VSATs could play in the ATM era.

Clearly, improvements in performance are possible if "next generation" satellite technology (on-board processing, low or medium earth orbits) is used that significantly reduces the satellite channel propagation delay. The need to provide full ATM connectivity via satellite will require development of more dynamic and efficient satellite link protocols and coding schemes. Finally, there is scope for more work in developing the interfaces between satellite and terrestrial (fixed and mobile) networks to ensure service transparency and truly global connectivity.

### BIBLIOGRAPHY

1. J. L. Everett (ed.), Introduction to VSATs, in *Very Small Aperture Terminals*, IEEE Telecomm. Series 28, Stevenage, UK: Peregrinus, 1992.
2. N. Abramson, VSAT data networks, *Proc. IEEE*, **78**: 1267–1274, 1990.
3. B. Mignosa, M. DeSilvio, and P. Leone, Present status and perspective of VSAT networks in Europe, *Proc. 14th AIAA ICSC*, March 1992, pp. 1289–1295.
4. D. P. Howarth, Military satellite communications, in B. G. Evans (ed.), *Satellite Communication Systems*, 2nd ed. IEE Telecomm. Series 24, Stevenage, UK: Peregrinus, 1991.
5. D. Chakraborty, Constraints in Ku-band continental satellite network design, *IEEE Commun. Mag.*, **24** (8): 33–43, 1986.
6. R. A. Nelson, V-band: Expansion of the spectrum frontier, *Via Satellite*, 66–76, 145, 1998.
7. E. H. Satorius and L. H. Tong, Analysis of a Rain Compensation Algorithm for K/Ka-band Communications, *Int. J. Satellite Commun.*, **14** (3): 297–311, 1996.
8. M. J. Willis, Fade countermeasures and VSAT Systems for 20/30 GHz, Ph.D. Thesis, Dept. of Electrical & Electronic Engineering, Guildford, Surrey, UK: Univ. of Surrey, 1993.
9. A. Hamid-Rana, J. McCoskey, and W. Check, VSAT technology, trends, and applications, *Proc. IEEE*, **78**: 1087–1095, 1990.
10. G. Maral, *VSAT Networks*, Wiley Series in Communications and Distributed Systems, New York: Wiley, 1995.
11. N. Abramson, The ALOHA system—Another alternative for computer communication, *Proc. AFIPS Fall Joint Comput. Conf.*, 1970, pp. 281–285.
12. D. Raychaudhuri and K. Joseph, Ku-band satellite networks using VSATs—Part I: Multi-access protocols, *Int. J. Sat. Commun.*, **5**: 195–212, 1987.
13. D. Raychaudhuri, Stability, throughput and delay of asynchronous selective reject Aloha, *IEEE Trans. Commun.*, **COM-35**: 767–772, 1987.
14. M. H. Hadjithedosiou and F. P. Coakley, Adaptive multiple access protocols for VSATs providing voice/data services and ATM interconnection, *Proc. 10th Int. Conf. Digital Satel. Comm. 95 (ICDSC-10)*, Brighton, 1995.
15. M. H. Hadjithedosiou, F. P. Coakley, and B. G. Evans, Multi-access protocols for a multiservice VSAT network, *Proc. ICC/SUPERCOM'94*, New Orleans, 1994.
16. M. R. Shroeder and B. S. Atal, Code-excited linear prediction (CELP): High quality speech at very low bit rates, *Proc. ICASP-85*, pp. 1649–1652.

17. A. M. Kondoz and B. G. Evans, A high quality voice coder with integrated echo canceller and voice activity detector for VSAT systems, *Proc. 3rd Eur. Conf. Sat. Comms.*, Manchester, U.K., 1993, pp. 196–200.
18. P. T. Brady, A statistical analysis of on–off patterns in 16 conversations, *Bell Syst. Tech. J.*, 73–91, 1968.
19. M. J. Fischer, Data performance in a system where data packets are transmitted during voice silent periods-single channel case, *IEEE Trans. Commun.*, **COM-27**: 1371–1375, 1979.
20. J. G. Gruber, Delay related issues in integrated voice and data networks, *IEEE Trans. Commun.*, **COM-29**: 786–800, 1981.
21. M. H. Hadjithodosiou et al., Broadband island interconnection via satellite-performance analysis for the CATALYST project, *Int. J. Sat. Commun.*, **12**: 223–238, 1994.
22. I. F. Akyildiz and S. H. Jeong, Satellite ATM networks: A survey, *IEEE Commun. Mag.*, **35**: 30–43, 1997.
23. D. M. Chitre, Asynchronous transfer mode (ATM) operation via satellite: Issues, challenges, and resolutions, *Int. J. Sat. Commun.*, **12**: 211–222, 1994.
24. ATM Forum, *User-Network Interface Specification Version 3.0*, Englewood Cliffs, NJ: Prentice-Hall, 1993.
25. R. Mauger and C. Rosenberg, QoS guarantees for multimedia services on a TDMA-based satellite network, *IEEE Commun. Mag.*, **35** (7): 56–65, 1997.
26. R. Jain, Congestion control and traffic management in ATM networks: Recent advances and a survey, *Comp. Networks and ISDN Sys.*, 1995.
27. H. S. Cheng and G. Fairhurst, Enhancement techniques for low rate satellite ATM, *4th Eur. Conf. Satellite Communications*, November 1997, pp. 232–237.
28. J. Lunsford et al., Link enhancement for ATM over satellite links, *Proc. ICDSC-10*, 1: 129–136, 1995, Brighton, UK.
29. A. D. Falk, A system design for a hybrid network data communications terminal using asymmetric TCP/IP to support internet applications, M.S. Thesis (CSHCN M.S. 94-2), Dept. of Electrical & Electronic Engineering, College Park, MD, Univ. of Maryland, 1994.
30. L. Wood, H. Cruickshank, and Z. Sun, Supporting group applications via satellite constellations with multicast, *Proc. 6th IEE Conf. Telecommun.*, Edinburgh, U.K., 1998, pp. 190–194.

MICHAEL HADJITHEODOSIOU  
University of Maryland