N92-14207

Satellite Communications for the Next Generation Telecommunication Services and Networks

D. M. Chitre
COMSAT Laboratories
Clarksburg, MD 2087

Abstract

Satellite communications can play an important role in provisioning the next-generation telecommunication services and networks, provided the protocols specifying these services and networks are satellite-compatible and the satellite subnetworks, consisting of earth stations interconnected by the processor and the switch on board the satellite, interwork effectively with the terrestrial networks. This paper discusses the specific parameters and procedures of frame relay and broadband integrated services digital network (B-ISDN) protocols which are impacted by a satellite delay. Congestion and resource management functions for frame relay and B-ISDN are discussed in detail, describing the division of these functions between earth stations and on board the satellite. Specific on-board and ground functions are identified as potential candidates for their implementation via neural network technology.

1. Introduction

Next-generation telecommunication networks are currently being shaped as a result of the rapid development of some key concepts and the associated detailed standards for integrated services digital networks (ISDN), both nationally (American National Standards Institute [ANSI] Committee T1) and internationally (International Telegraphy and Telephony Consultative Committee [CCITT] Study Groups XVIII, XI, and I). The ISDN defines a worldwide communications environment encompassing all digital information, not only voice and data, but also facsimile, videophone, videoconferencing, interactive cmputer-aided design/computer-aided manufacturing (CAD/CAM) and any other type of information that can be digitized. There is a symbiotic relationship between the emerging ISDN environment and satellite communications. Digital communications channels provided by various domestic and international satellite networks can form a digital transmission backbone for the establishment of worldwide ISDN. The well-specified ISDN standards on interfaces, signaling procedures, service definitions, and other operational and architecture features will provide a common framework for developing satellite network configurations offering a range of services. The key features of a satellite communications system, namely, agility of communication bandwidth, multipoint/broadcast nature of satellite channels, global coverage and mobility of very small aperture terminal/mobile satellite (VSAT/MSAT) terminals play a very crucial role in the future development of ISDN, intelligent networks, Future Public Land-Mobile Telecommunications System (FPLMTS), and finally, the Universal Personal Telecommunications (UPT) service. UPT is a telecommunications service which will enable the user to establish and receive calls and services on the basis of a unique personal telecommunication number (PTN) across multiple networks at any user-network access whether fixed, movable, or mobile irrespective of the geographical area. The worldwide satellite networks along with VSAT and the mobile satellite networks will go long ways to make UPT service feasible.

Satellite communications can play a significant role in the future national and international telecommunication networks provided the satellite subnetworks are closely integrated with the terrestrial networks. The specific integration issues are discussed in this paper within the context of the networking (and underlying protocol) standards which are going to be the fundamental underpinnings of future terrestrial networks.

This paper will focus on emerging standards in two areas, frame relay and broadband ISDN (B-ISDN); and discusses, in detail, their impact on the satellite networks with on-board processing and switching.

2. Frame Relay

2.1 Service and Protocol Description

Frame relay is a new ISDN packet mode bearer service for data communications at access speeds of up to 2.048 Mbit/s [I.122, I.233].* This bearer service provides the order-preserving bidirectional transfer of layer 2 frames from the source user-to-network interface (S or T ISDN reference point) to the destination network-to-user interface (another S or T ISDN reference point). The data units (called frames) are routed through the network on the basis of an attached label termed, data link connection identifier (DLCI). The DLCI identifies a virtual connection on a bearer channel (i.e., D, B, or H) at a user-to-network or network-to-network interface (UNI or NNI). The major characteristics of this service are the logical out-of-band call control using protocol procedures that are integrated across all telecommunications services and the statistical multiplexing of different user data streams (via DLCI) at the link layer in the user plane.

Flow-control and error-recovery functions are performed on an end-to-end basis by a user-selectable, end-to-end protocol. LAPF (Q.922) is currently being developed for use as one of these end-to-end protocols. The link layer parameters chosen for LAPF are important in determining the effectiveness of the satellite networks providing frame relay service. The parameter for the retransmission timer, T200, with a default value of 1.5 s, can accommodate a one-hop satellite delay. The other parameter, k, (the maximum number of outstanding frames) can have values ranging from 1 to 127. The default value, as currently specified for a 64-kbit/s channel, is seven which is too low. It will result in the inefficient operation of the frame relay service over satellite channels. The value of k should be negotiated to a higher value (such as 40 for a 128-octect frame size) at the call set-up time via the XID procedure described in Appendix III of Q.922.

A subset of LAPF, corresponding to the data link core sublayer is used to support the frame relaying bearer service. The network does not support any procedures above these core functions of Q.922; such as acknowledging frames (within the network), or keeping the sequence numbers. The core functions are:

- frame delimiting, alignment, and transparency
- frame multiplexing, and demultiplexing using the address field
- inspection of the frame to ensure that it consists of the integer number of octets prior to zero bit insertion or following zero bit extraction
- inspection of the frame to ensure that it is neither too long nor too short
- detection of transmission errors
- congestion control functions.

---

* ISDN Recommendations referred to in this paper can be found either in the CCITT Blue Books, ITU, Geneva 1988 or as draft recommendations output from the CCITT Study Group XVIII Meeting in Matsuyama, Japan, 11/26/90–12/7/90 and XI Meeting in Geneva, Switzerland, April 1991.

The address field of the frame consists of at least two octects, containing a DLCI identifying a virtual connection on a bearer channel. The field variables in the address field for the congestion management are as follows.

a. Forward explicit congestion notification (FECN):

This bit can be used by a frame relaying network node to notify the user that the congestion-avoidance procedures should be initiated where applicable for traffic in the direction of the frame carrying the FECN indication. This bit is set to 1 to indicate to the receiving-end system that the frames it receives have encountered congested resources. This indication can be used by the destination-controlled transmitter rate adjustment.

b. Backward explicit congestion notification (BECN):

This bit can be set by a congested network to notify the user that the congestion-avoidance procedure should be initiated, where applicable, for traffic in the opposite direction of the frame carrying the BECN indicator. This bit is set to 1 to indicates to the receiving-end system that the frames it transmits may encounter congested resources. This indication can be used by the source-controlled transmitter rate adjustment.

While setting of the above bits by the network or user is optional, the network is not allowed to clear (set to 0) these bits. Networks that do not provide FECN or BECN will pass this bit unchanged.

c. Discard eligibility (DE) indicator:

This bit, if used, is set to 1 to indicate a request that a frame should be discarded in preference to other frames in a congestion situation. The setting of this bit by the network or user is optional. A network will never clear (set to 0) this bit. Networks that do not provide DE will pass this bit unchanged. Networks are not constrained to discard frames with DE =1 in the presence of congestion.

2.2 Congestion Management

Congestion in the user plane occurs when the traffic arriving at a resource exceeds the capacity of the network. In the frame relay networks, congestion can arise due to users offering in excess of the committed traffic, or due to coincidental peak traffic demands, or, possibly, due to equipment failure and the consequent degraded network capabilities. In these networks, congestion control is achieved via both the congestion-avoidance mechanisms and congestion-recovery mechanisms. The former is used at the onset of congestion via the explicit congestion notifications. For destination-controlled transmitters, the FECN bit is set in the appropriate frames. For source-controlled transmitters, the BECN bit is set in frames transported in the reverse direction (i.e., towards the transmitter). Alternatively, a consolidated link layer management (CLLM) message can be generated providing reverse notification for one or more DLCIs within a single frame. The CLLM is sent on the layer management DLCI in the user plane.

Congestion-recovery mechanisms are used to prevent network collapse in the event of severe congestion. Implicit congestion detection and end-user responses are defined in Q.922 to recover from congestion.

One of the important issues for satellite compatibility in frame relay networks is the implementation of a congestion-recovery mechanism that works efficiently with large propagation delays. Congestion recovery in frame relaying will be done by adjusting the sizes of the LAPF windows that control the number of outstanding, unacknowledged frames. In general, these windows must be relatively large when satellite links are employed in order to account for the propagation time. If the window sizes are too small, a transmitter will constantly be stopping transmission to wait for acknowledgment of previous frames. This results in inefficient utilization of the satellite channel, along with increased delay and decreased throughput for the user.

Appendix I of Draft Recommendation Q.922 discusses the use of dynamic window size to respond to network congestion and describes the congestion-control algorithm.

The algorithm modifies the transmitting-data-link layer entity's transmit window when the congestion is first detected and again as congestion decreases. The congestion-control algorithm is triggered by the loss of I frames. When the data-link layer detects this loss, either by the reception of an REJ frame, or the confirmed T200 expiration event, it invokes the dynamic window algorithm and reduces its transmit window size to a fraction of its original size. The transmit window size is gradually increased until it returns to its original value, k, its value in the absence of congestion.

For satisfactory operation of LAPF (Q.922) over a 64-kbit/s satellite link, the value of k during the absence of congestion should be as large as 40 for a 128-byte frame size (and even larger for higher-rate channels, H_0, H_{11}, and H_{12}). Thus, a reduction in the window size, for example, from 40 to 10 due to the loss of one frame and the gradual increase over the next several round-trip delays will cause considerable reduction in throughput. Notice also that the frame loss is not necessarily due to congestion, but could be due to a bit error.

A somewhat different congestion control strategy, where the transmit window is reduced gradually, for example, by a factor of two at each step, and a more rapid increase will considerably alleviate the problems for the satellite operation without impacting the terrestrial operation in most cases.

Appendix I of Q.922 also discusses the user response on the receipt of the FECN, the BECN and the CLLM.

2.3 Frame Relay Functions For Satellite Subnetworks

Specific functions within the satellite subnetwork are identified here for the support of the frame relay bearer services. These functions are largely independent of the specific details of the satellite network architecture. A generic model of the satellite subnetwork within a frame relaying network is considered. The nodes of the satellite subnetwork consist of distributed earth stations and an on-board packet processor and a switch interconnecting up-link and down-link beams. The interfaces to the terrestrial subnetworks are at the earth stations which perform the frame relay core functions. The on-board processor also implements all the frame relay core functions described in Subsection 2.1. However, the specific congestion-control functions (and algorithms) undertaken at the earth stations and on board differ based on the amount of information available and the complexity of the algorithm.

2.3.1 Functions at the Earth Stations

An earth station of the satellite subnetwork can be an ingress node, an egress node, or just an intermediary network node within a frame relaying network. The following functions are necessary for imminent congestion determination and subsequent proper response.

- traffic monitoring
- congestion determination via thresholding
- FECN, BECN/CLLM implementation
- discard eligibility flag insertion.

Each earth station acting as a frame node monitors the frame traffic on each DLCI connection entering into and going out of the satellite subnetwork. The offered traffic is compared with the committed level at the call setup. The specific traffic parameters are the committed burst size (Bc), excess burst size (Be), and committed information rate (CIR). These parameters are measured over a certain computed time interval (Tc).

The congestion determination can be done by computing the average queue length for each outgoing link, including the up-link to the satellite. Appropriate thresholds T1 and T2 are set such that when the average queue size is greater than T2, the link is considered to be in the state of congestion. Beginning at that time, and continuing until average queue size falls below the threshold, T1, the link is declared to be congested.

Whenever, an outgoing link from the earth station is considered congested, the FECN bit is set to 1 on all outgoing frames on that link. Appropriate CLLM messages are generated for transmission on all of the incoming links contributing the traffic to the outgoing congested link. CLLM can be replaced by frames with BECN bit set to 1, if there is current reverse traffic on those links.

Frames offered in excess of the committed traffic negotiated at the call setup are marked as discard eligible at the earth station before transmitting them forward. However, if the outgoing link is in the congested state, the frames marked discard eligible and buffered for transmission on that link are dropped.

2.3.2 Functions On Board the Satellite

The on-board processor performs the core functions described in Subsection 2.1 and routes the frames based on the mapping tables relating the DLCI number and the down-link channel.

However, the congestion-management functions on board the satellite should be simpler than at the earth stations. Specifically, the detailed monitoring of all the separate DLCI offered traffic and its comparison with the committed level may not be performed on board. The frame traffic buffered for each outgoing down-link is monitored. However, the threshold method of determining a mild congested state of a link may not be cost-effective on board the satellite since it will take some time before the users respond, over the satellite link, to the FECN, BECN, or CLLM messages to relieve the congestion. In the interim, there could be considerable amount of traffic in the pipeline worsening the congestion. On the other hand, setting the threshold value too low will underutilize satellite channels.

A predictive method which does not depend upon implementing a very complex algorithm on board the satellite is, thus, highly desirable. A suitable neural network (NN) implementation scheme could, indeed, be such a method. A feed-forward back-propagation network, with the input to the NN consisting of the traffic pattern on a particular down-link channel, can be trained to give the correct output. The output will be a binary decision whether that link is entering a congested state or not. To train the NN, a number of traffic patterns can be simulated and their impact on the future state of the link observed. With an appropriate definition for the link congestion, different patterns can then be correlated with the predictive behavior of the future state of the link.

The congestion detection based on the NN output can trigger the generation of appropriate FECN, BECN, or CLLM messages and the dropping of frames with discard eligibility set to 1.

It should be noted that the NN technique can also be used at the earth stations for the detection of oncoming congestion.

3. B-ISDN

3.1 B-ISDN Protocols

Although the CCITT work on the formulation of broadband ISDN (B-ISDN) recommendations is still in its early stages, several issues concerning broadband services and network capabilities and requirements have already been agreed upon. Asynchronous transfer mode (ATM) is the proposed transport technique for the B-ISDN. ATM is a packet mode information method which uses fixed-size packets called cells. The cells are statistically multiplexed and are identified as belonging to a particular logical connection by the virtual channel identifier (VCI) that is carried as a label in the header of each cell. A virtual path, identified by a virtual path identifier (VPI), is a grouping of virtual channels. ATM offers the flexibility to support a wide variety of service types (voice, data, or video) and provides efficiency by statistically multiplexing possible bursty traffic. There is a lot of similarity between frame relay and ATM. The major differences are that frames are of variable size while cells are fixed size (48 bytes of user data and 5 bytes of header); and ATM cell header has more network functionality (I.121, I.150, I.361).

ATM adaptation layer (AAL) protocols on top of the ATM layer are being specified for different types of services. Recommendation I.363 defines an assured mode operation for data transfer for a Type 4 (or Type 3) AAL convergence protocol. For the high bit rates envisaged for B-ISDN, the choice of the protocol for the assured mode operation needs to be made after careful analysis.

Two recovery strategies are likely candidates for the Type 4 (or Type 3) ATM adaptation layer convergence protocol for assured operation. They are:

- *go-back-N (GBN)*—retransmit all messages from missing sequence number, and
- *selective retransmission (SR)*—retransmit only messages corresponding to missing sequence number.

In the GBN method, the AAL messages which are encoded for error detection are transmitted sequentially and the acknowledgments from the receiver arrive after a round-trip delay. During this delay, which is the time between the transmission of the AAL message and the receipt of its acknowledgment, N-1 other messages are also transmitted. When the receiver gets an erroneous message or an out-of-sequence message, a negative acknowledgment (i.e., an REJ message) is sent by the receiver. When a negative acknowledgment is received, the transmitter stops sending new messages, backs up to the negatively-acknowledged message and retransmits it and all subsequent messages, thus giving rise to spurious retransmissions of at least N-1 messages.

At high link speeds, such as 45 Mbit/s, this phenomenon causes a severe degradation in both throughput and delay even at low bit error rates (BER) and the performance deteriorates very rapidly as the BER or message-loss rate increases. Furthermore, if the loss of a message is due to mild congestion, the spurious retransmissions will aggravate the congestion giving rise to still more messages being lost causing more spurious retransmissions and so on.

Figure 1 shows the degradation in throughput for the GBN protocol operating over a 45-Mbit/s satellite connection. (Note that the bit error ratio is really an effective cumulative bit error ratio arising out of line errors and packet losses.) Figure 2 is the corresponding curve over a 45-Mbit/s terrestrial connection. The selective retransmission curves are based on simulations of the protocol and the GBN curves are based on analytical results.

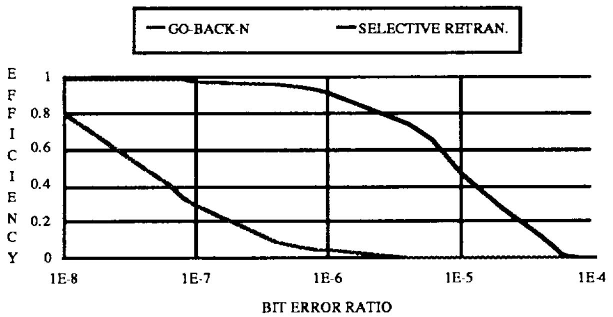

FIGURE 1. *Efficiency of a 45 Mbit/s Satellite Link for a Frame Size of 4,096 Octets*

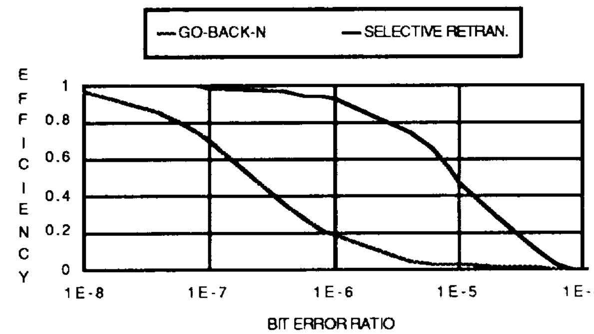

FIGURE 2 *Efficiency of a 45 Mbit/s Terrestrial Link for a Frame Size of 4096 Octets*

The GBN strategy, which has a certain simplicity for implementation, however, performs very poorly at high speeds which are likely to be encountered by the AAL convergence protocol. It will thus be desirable to remove the option of GBN error-recovery procedure and introduce new generation of simple-to-implement selective-retransmission protocols.

3.2 Traffic Control and Resource Management

The advantages of the ATM, however, can be seriously hindered if effective traffic and congestion control techniques are not implemented by the network. At the high-speed cell transport level, the congestion control techniques proposed in frame relaying networks are not adequate. Additional mechanisms like admission control (called acceptance/rejection) need to be used in conjunction with bandwidth enforcement and flow control.

Since the B-ISDN, using the ATM technique, is designed to transport a range of traffic classes with a widely varying traffic and quality of service (QOS) requirements, it is essential to have several levels of traffic control capabilities such as:

- connection admission control (CAC)
- usage parameter control (UPC)
- network parameter control (NPC)
- priority control
- congestion control.

Connection admission control is defined as the set of actions taken by the network at the call setup phase (or during call re-negotiation phase) in order to establish whether a virtual channel connection (VCC) or a virtual path connection (VPC) can be accepted or rejected. On the basis of the connection admission control outcome in an ATM based network, a connection request for a given call is accepted only when sufficient resources are available to establish the call through the whole network at its required QOS and to maintain the agreed QOS of existing calls.

UPC and NPC perform similar functions at different interfaces such as UNI and NNI.

UPC/NPC is defined as the set of actions taken by the network to monitor and control (user) traffic in terms of traffic volume and cell routing validity. Its main purpose is to protect network resources from malicious as well as unintentional misbehavior, which can affect the QOS of other already established connections by detecting violations of negotiated parameters. The possible parameters could be cell-peak rate, average rate, burstiness, or peak duration.

Priority control is based on the fact that the user is allowed different priority traffic flows by using the cell loss priority (CLP) bit in the ATM header (I.150).

Congestion control can work in two modes, preventive and reactive, depending upon the state of the network. In the preventive mode, connection admission control takes into account the current load on the network and rejects the call request. In the reactive mode, techniques based on the discarding of cells with the high CLP value or cells carrying violation tags, coupled with congestion notification can be used.

3.2.1 Functions at the Earth Stations

Figure 3 illustrates the B-ISDN satellite subnetwork (enclosed in the ellipse). The B-ISDN controller at an earth station performs a number of functions including traffic monitoring and UPC (if the earth station is an ingress node) or NPC (if the earth station is an intermediary node), connection admission control, violation tagging for the cells, congestion prediction/ detection, congestion notification, and discarding of cells.

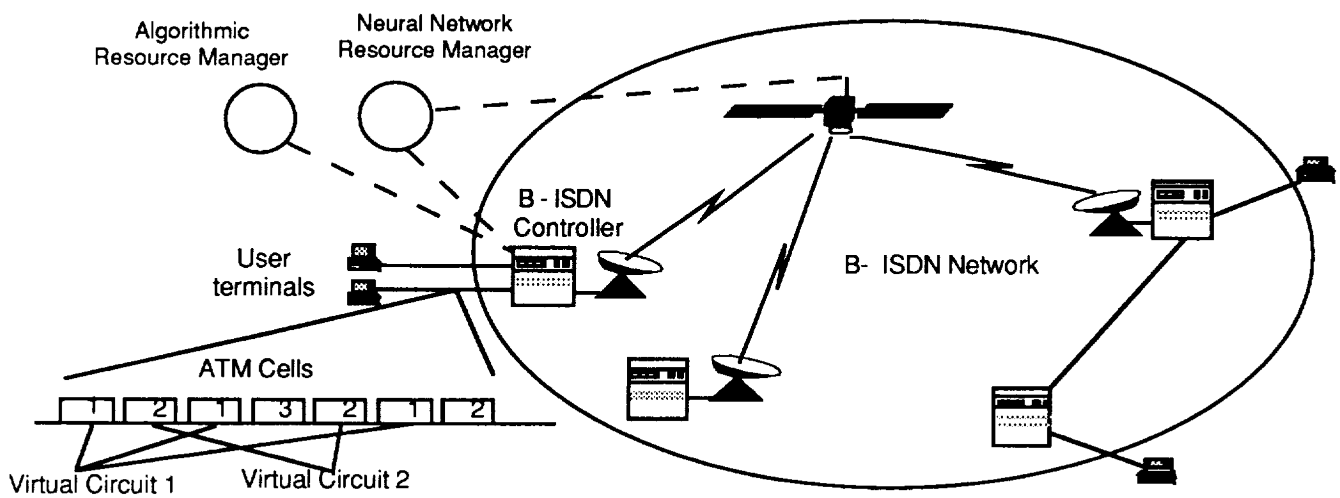

Figure 3: Satellite B-ISDN Network With Onboard Processing and Switching

The ATM cell traffic is monitored for every VPC/VCC to find if it exceeds the negotiated value at the call setup for that connection. The cells in violation can be tagged either using the one bit reserved field or perhaps setting the CLP bit to 1.

Congestion for any outgoing link (including the up-link to the satellite) can be detected either by a classical algorithm (through setting buffer thresholds) or more predictive neural network models can be used. Two different feed-forward back-propagation networks can be considered. The first one will implement the connection admission control. It will decide whether a particular call request, with appropriate traffic descriptors, should be accepted or rejected during a particular state of the network. Notice that the call may have to be rejected even if the current state of the network is not congested. (The acceptance of the call could lead to congestion characterized by exceeding certain levels of cell loss ratio.) The second neural network will predict the future state of the network for the existing calls. In that case, the congestion could arise due to coincidental peak traffics from different sources or the degradation of network resources. The congestion prediction decision of the second neural network can be used to discard cells with violation tags or cells with the high CLP bit and for sending explicit congestion messages as operations, administrations, and management (OAM) cells to the appropriate VPI/VCI cell sources.

3.2.2 Functions On Board the Satellite

The on-board processor routes the ATM cells either individually or as satellite virtual packets (SVP) which consist of a number of ATM cells (with the same CLP) destined for the same down-link channel. (Such concatenation could have been done at an earth station.) The detail monitoring of traffic on individual VPC/VCC may not be performed on board the satellite. However, the ATM cell or SVP traffic buffered for each down-link channel is monitored. The two NNs described in Subsection 3.2.1 can be implemented on board for CAC and detecting the oncoming congestion. Based on the output of these NNs, the NN resource manager, shown in Figure 3, can make a decision regarding the acceptance or rejection of the new call and discarding the cells with violation tags or cells with the CLP bit set to 1. Upon detection of oncoming congestion, explicit congestion notification messages can be transmitted to earth stations contributing to the congestion. The earth stations having monitored the individual VPC/VCC traffic can then generate congestion indication messages for appropriate sources of corresponding VPC/VCC. The sources can then implement suitable flow control.

Notice that for B-ISDN high-speed traffic, the congestion and flow control based on the predictive, but simple to implement, NN on board the satellite is very important to alleviate the delay experienced by the notification messages over the satellite links.

4. Conclusions

Frame relay and B-ISDN services can indeed be provided very efficiently via a satellite network with on-board processing and switching. However, the future development of protocols for these services need to be properly shaped. Finally, the application of NN technology on the ground and on board can be brought to bear upon the congestion and resource management techniques to make optimum use of the satellite resources.