Optimizing Data Transmission Capacity on Satellite Links with Non-Zero Bit Error Rates

Hans Kruse

McClure School of Communication Systems Management
Ohio University
9 S. College Street
Athens, OH 45701

1. Introduction

The intent of this note is to report on work in progress on the problem of designing optimal satellite data communication channels for today's data networks. Satellites have been used for data communications for many years. However, both changes in speeds and protocols in use now make it much harder to provide effective data communications on satellite channels. We will begin by reviewing the role satellites are likely to play in data communications. We then look at the properties of current data communication protocols and their interaction with typical satellite channel properties. Finally we summarize experimental data which provide some guidance towards the design of future satellite systems, and suggest approaches to solve the problem.

2. What role are satellites going to play in data communications?

Traditional VSAT networks have typically been deployed as "stand-alone" networks. Their primary purpose was to tie many dispersed company locations to a central site in the most economical way. Speed requirements were modest (often in the 64kbps range), and since the network is isolated, protocols and equipment can be tuned to the characteristics of the satellite channel.

Today, satellites are being positioned for two very different roles from the traditional VSAT: (1) High-speed mobile communications in areas where cellular coverage is unavailable, and (2) hybrid broadband networks in which the satellite link either plays a backup (redundancy) role, or extends the terrestrial network to remote areas without the need for immediate and costly upgrades to the terrestrial infrastructure. Both cases will bring with them higher speeds [2] , as well as the need for the satellite link to function efficiently in a network that is not tuned for satellite channels.

3. What issues arise in satellite data communication?

The problem faced in a hybrid network is easily demonstrated by looking at the OSI model (even non-OSI compliant protocols tend to have a layered architecture in which the same issues arise).

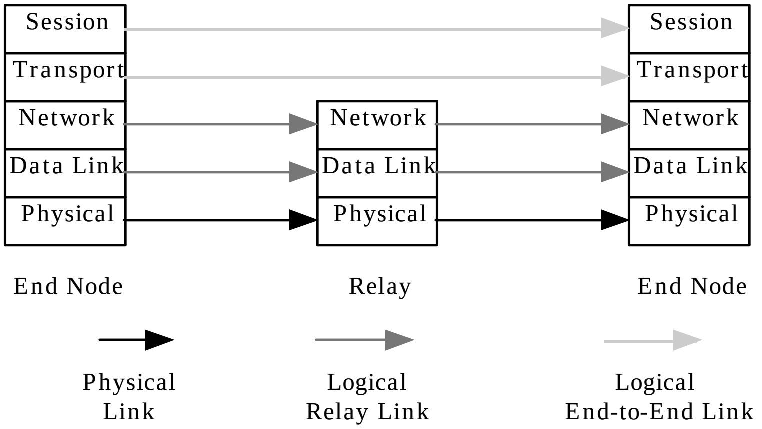

Figure 1.

The OSI model aims to separate functionality to allow more efficient and robust implementations of protocol stacks. Each protocol layer is assumed to be independent from the details of the layers below it, as long as a suitable service is offered by the layer directly below. Each protocol layer in fact is designed as if a link existed directly to the communicating entity. Only layer 1 controls actual physical links. Layer and layer 3 links usually terminate on the next relay node (for example a router), and can therefore be influenced and changed for different parts of the network. Many current VSAT terminals contain built-in routers to optimize transmissions over the satellite link; they are taking advantage of the fact that many protocols permit the insertion of such relay points. Layers 4 and above operate under the assumption that they have access to an end-to-end logical link which is unaffected by intervening relay nodes, and which terminates on the end node with which the source is communicating.

The trend in broadband networks has been to avoid duplication of effort. Specifically, OSI layers 1-3 are typically used to provide best-effort, unreliable services, with error correction accomplished in layer 4, at the endpoints of the logical connection between the two hosts. In addition, layer 5 entities often implement further flow control mechanisms to facilitate, for example, data base synchronization.

Unfortunately, the characteristics of the underlying transport services have a direct effect on the performance of the protocols at the higher layers, particularly if some of the transport links have higher bit error rates or higher latency (round-trip delay) than the protocol designer (at the upper layers) may have envisioned. As an example, one only need to consider a protocol stack which includes a geosynchronous satellite link. Assuming proper adjustments have been made in the transport layer, such links have been demonstrated at layer 4 transfer rates of 155Mbps. If in such an implementation the layer 5 entity implements a stop-and-wait synchronization mechanism (which would have little negative impact on short, error free links), the layer 5 transfer rate over the 155Mbps channel would drop below the 100kbps level, depending on the details of the implementation! These types of problems tend to surface now due to the new roles of satellite channels. The satellite earth station is no longer part of the communication end points, but rather appears as a relay point, where no control can be exerted over the functions in layers 4 and up.

Satellite channels can be designed to meet a wide range of bit error rate requirements. One method for improving bit error performance is to increase signal strength. For both mobile and VSAT applications, power consumption and terminal cost issues limit the designer's ability to boost signal strength. The digital encoding scheme of data onto the analog satellite channel is the other area in which bit error rates can be controlled. In this case, more robust encoding schemes with lower bit error rates will require a larger amount of analog spectrum to achieve the same channel data rate. The satellite communications provider is therefore faced with the tradeoff between providing large amounts of channel capacity with relatively high bit error rates (compared to terrestrial channels), versus providing a lower volume of very high quality (low bit error rate) channels. Note also that the bit error rate which the channel is designed for must be achievable over a very large percentage of operating time (equivalent to the availability of terrestrial channels), which requires coding schemes to be designed for adverse weather conditions, especially in the Ku and Ka bands. It seems likely that the coding scheme should be designed to be time-dependent, i.e. in good weather we may want to increase the channel rate per MHz of spectrum, and decrease the ratio in bad weather. As we will see later, it is also desirable for the coding scheme to vary with the type of data being transferred, i.e. we want a service-dependent coding scheme.

Finally, we should note the different latency issues which exist for satellite channels. The latency for geosynchronous orbit satellites is well known to be around 500ms. Latency for a direct round-trip to a Low-Earth-Orbit satellite is about two orders of magnitude smaller. However, any communications link (terrestrial or LEOS) between two different parts of the world will have latencies similar to that of geosynchronous satellites (the circumference of the earth is of the same order of magnitude as the height of the geosynchronous orbit). The difference between the intercontinental terrestrial link versus the same link over a LEOS system lies in the bit error rate performance. The performance of terrestrial links is determined by the media installed and the maintenance of those media. For a satellite link, the performance is only partially fixed by the physical design choices. For digital links, the choice of encoding schemes, which can be time and data-stream dependent, also permits control over the bit error rate. For satellite links, the interaction between bit error rate performance and end-to-end protocol efficiency is therefore a much more detailed design issue than it would be for a terrestrial link.

3. Interaction between satellite links and data communication protocols

In the absence of errors, it is generally well understood that the flow control mechanisms at layer 4 and above must be adjusted for satellite links to allow efficient use of high-speed channels. The amount of adjustments needed depends on the round trip delay time; the protocol parameter is generally a flow control window size with its associated buffers. Figure 2 below summarizes the buffer sizes (which in some sense represent a "distance" between sender and receiver) needed for different channel latencies and transport services. This distance is sometimes referred to as the "bandwidth-delay product".

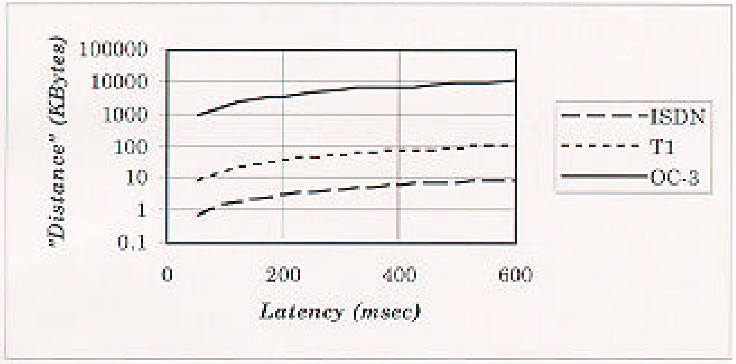

Figure 2

We can use figure 3 below to envision the effect of random and relatively rare bit errors on a protocol which has been adjusted at least in accordance with the buffer requirements shown in figure 2.

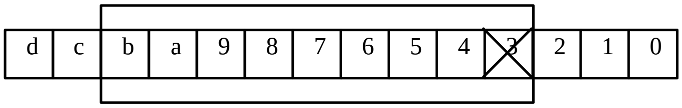

Figure 3

In the figure, the window is shown as it would appear after receipt of the acknowledgment for segment 2. Since segment 3 is damaged, no acknowledgment for it is forthcoming, and the window cannot move. Either through a time-out or through the receipt of a negative acknowledgment, the sender will very shortly after this point resend segment 3. Assuming a steady flow of data (such as a file transfer), at the point shown in figure 3 an amount of data equal to the bandwidth-delay product will have been sent following segment 3. Note also that it will take one full round-trip time for the acknowledgment of the successful retransmission of segment 3 to return to the sender, at which point the window can be moved. (There is of course a small chance that the retransmitted segment will again be damaged, or that the acknowledgment is lost. We ignore these cases for this simplified discussion).

What happens next depends on the protocol implementation, the protocol design, and the actual window/buffer size configured. We can discuss the limiting cases of a continuum from "best-case" to "worst-case".

In the best case, the sender is configured with a buffer size of at least twice the bandwidth-delay product (so that transmission can continue until confirmation of the retransmission is received). In addition, in this best case the receiver returns complete “state” information to the sender, so that a selective retransmission of only segment 3 can take place. The impact of a moderate number of errors in this case if very small, with the exact results depending among other things on the segment size (which determines the amount of data to retransmit).

Experimental data for such a configuration exists in a COMSAT study of the SSCOP protocol [4] . Properly configured, this protocol showed little performance degradation on a 10Mbps channel operating with bit error rates as poor as 10^{-4}.

Worst case performance results from two possible situations. First, if the window/buffer size is set roughly to the delay-bandwidth product, no transmission can take place between the retransmission of segment 3 and the receipt of a valid acknowledgment. If the protocol is configured for Go-Back-N operation, the data transmitted since segment 3 (equal to one delay-bandwidth product) has to also be retransmitted. In both cases effective throughput is strongly effected. The difference between the two causes for this scenario is that the buffer size is somewhat easier to increase, compared to the implementation of a selective retransmission method.

Consider an example of the worst case scenario. If an error where to occur roughly once per bandwidth-delay product, the protocol performance would drop to 50% of the error-free case. In the case of a T1 service, this would be at a bit error rate of 1×10^{-6}. Any satellite link supporting such a protocol would need to perform at a significantly better bit error rate to avoid severe degradation of throughput.

Most protocols will fall between these two extremes. As mentioned before, SSCOP comes close to the best case scenario. TCP/IP is a more complex case. The original implementation[3, 9] , which is still found in many desktop operating systems, will be close to the worst case. Improvements to TCP/IP have been proposed and implemented in a few cases[5, 7, 10] . Implementation will differ and performance will depend critically on the details of the implementation.

Additionally, some protocols (for example TCP/IP) will misinterpret the satellite delay and the segment loss as network congestion. In that case a self-imposed throughput reduction may be observed in addition to the effect of the retransmissions.

We have collected data [1] on TCP/IP performance using the NASA ACTS system[1]. A T1 link over the satellite was deliberately degraded to simulate the effect of a severe weather fade. The channel was operating at a bit error rate of 6×10^{-6}. For this study we have used the “stock” TCP window size of 24KBytes found in most Sun OS implementations. In lieu of increasing this window size, we have chosen to split the data transfer (FTP was used in all

---

[1] The author gratefully acknowledges the assistance provided by the NASA Lewis Research Center staff in these experiments.

cases to transfer a 2MByte file) over multiple TCP connections, creating an "effective" window size equal to 24KBytes times the number of connections. This was done to see if we could mitigate the effects of transmission errors described earlier to some degree.

Figure 4 shows the results of this experiment, together with the results obtained on an error-free link. The 4 connection case represents a buffer size roughly equal to the bandwidth-delay product (ACTS is a geosynchronous satellite). In the error free case performance levels of after that buffer value is reached, the remaining increase is due to the fact that a larger number of connections help overcome the effect of TCP's congestion control algorithm. We can see that the TCP performance in all cases is decreased by a factor of 2-4.

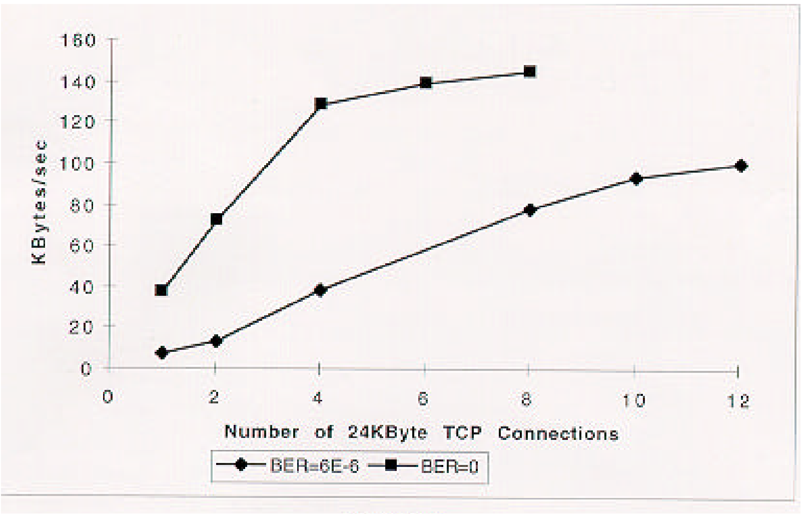

Figure 4

It should be noted that the ACTS system allows two coding schemes[6, 8] . Error rates of the magnitude used in our test could have been corrected by switching to the more robust version (the "coded" mode in ACTS terminology). For the TCP/IP applications shown above, throughput would have increased by a factor 2-4 as stated. At the same time the spectrum demand for the T1 channel would have increased by a factor 4.

Studies using ATM over satellite at DS3 (45Mbps) rates have been reported by COMSAT[4] . The authors of the study point out correctly that satellite links often exhibit burst error characteristics, especially if convolution coding methods[11, 12] are used. Under these conditions, the study suggests that ATM AAL1 traffic may require bit error rates better than $1x10^{-10}$ unless other steps are taken to protect the integrity of the ATM headers. Ironically, transport layer protocols like TCP and SSCOP perform better under burst error conditions than under purely random error distributions.

4. Conclusions

Data available so far indicates a variety of bit error rate requirements for various protocols. In addition, experience with K_a band satellite equipment has shown wide variations of bit error rates in varying weather conditions. The resulting demands on satellite channel design are quite complicated.

In the simple example of the ACTS link described earlier a change in signal encoding would increase spectrum demand by a factor of 4, in return for a 10dB increase in signal margin, sufficient to make the link error free. Based on the experimental data, SSCOP would not have profited much in throughput. TCP/IP would have performed better by a factor of 2-4. Finally, an AAL1 ATM link would probably have required the coding for proper operation (no T1 level data is available).

It appears likely that future digital satellites will need the ability to operate in a number of different signal encoding modes. These modes must certainly be selectable as a function of signal strength, for example during weather-induced fades. In addition it appears desirable to have coding schemes which can be applied to each user's data stream selectively, based on the protocol in use.

5. Bibliography

[1] Allman, M., S. Ostermann, and H. Kruse. *Data Transfer Efficiency Over Satellite Circuits Using a Multi-Socket Extension to the File Transfer Protocol (FTP).* in Proceedings of the *NASA ACTS Results Conference.* 1995. Cleveland: unpublished.

[2] Hughes DirecPC. 1995, http://www.direcpc.com

[3] Stevens, W.R., *TCP/IP Illustrated: The Protocols*, Vol. 1. 1994, Addison-Wesley.

[4] *Demonstration of Asyncronous Transfer Mode (ATM) via Commercial Satellite.* COMSAT Technology Services 1993; Demonstration Technical Report .

[5] Jacobson, V., R. Braden, and D. Borman. *TCP Extensions for High Performance.* LBL 1992; Internet RFC 1323.

[6] Bauer, R. and T. vonDeak. *Advanced Communications Technology Satellite (ACTS) and Experiments Program Descriptive Overview.* NASA Lewis Research Center 1991.

[7] Jacobson, V., R. Braden, and L. Zhang. *TCP Extension for High-Speed Paths.* LBL 1990; Internet RFC 1185.

[8] Naderi, F.M. and S.J. Campanella. *NASA's Advanced Communications Technology Satellite (ACTS): An Overview of the Satellite, the Network, and the Underlying Technologies.* NASA 1988ACTS Publication 100.

[9] Comer, D.E., *Internetworking with TCP/IP Volume I, Principles, Protocols, and Architecture*, 1988, Prentice Hall.

[10] Jacobson, V. and R. Braden. *TCP Extensions for Long-Delay Paths*. LBL 1988; Internet RFC 1072.

[11] Chambers, W.G., *Basics of communications and coding*, 1985, Oxford: Clarendon Press.

[12] Schwartz, M., *Information, Transmission, Modulation, and Noise*, 1980, New York: McGraw Hill.