JOURNAL OF AEROSPACE COMPUTING, INFORMATION, AND COMMUNICATION
Vol. 1, October 2004

Satellite Data Networks

Eytan Modiano*
Massachusetts Institute of Technology, Cambridge Massachusetts 02139

I. Introduction

NOWADAYS we take the availability of worldwide communications for granted; communication satellites bring us live coverage of events from around the world and undersea fiber-optic cables are used to provide global telephony service. However, as recent as the 1950s transatlantic communications were limited to a few dozen voice circuits, and real-time communication with much of the world was not even possible. The potential of man-made satellites to provide global communication coverage was first envisioned by Arthur C. Clarke in his seminal article "Extra-Terrestrial Relays".[10] He described a satellite in synchronous orbit being used for providing communication relay services within any two points on the hemisphere and a constellation of three such satellites for providing global coverage. Clark argued that using a satellite in synchronous orbit would be more cost effective and provide much better coverage than alternative terrestrial communication techniques.

Thirteen years later, the first communication satellite, Score, was launched and used by President Dwight D. Eisenhower to broadcast a prerecorded Christmas message around the world. That early experiment was followed by a number of experiments with communication satellites in the early 1960s that provided early glimpses of the Global Village, including televising parts of the 1964 Tokyo Olympic Games. Later, in 1964, agreements were signed that created the International Telecommunications Satellite Organization (INTELSAT). Shortly afterwards, the first INTELSAT satellite, Early Bird, was launched to provide commercial satellite service. Early Bird had the capacity for 150 telephone circuits and an additional 80 hours of television broadcast. Early Bird was followed by numerous other satellites that provided worldwide telephone and television service. Today, hundreds of communication satellites are used to provide a variety of services, from military communications to voice telephony, television broadcasts, and Internet access.

Satellite networks play an important role for the purpose of data delivery. They are very effective at broadcasting data over large geographic locations, and are an effective means for reaching remote locations lacking in communication infrastructure. Satellite networks are critical to our national interests both for military and civilian applications. The military depends on satellite communications for robust and reliable communications in hostile environments. On the civilian side, many rural locations, out of the reach of fiber-based networks, depend on satellites for access to high data rate communication services. As our world continues to progress toward globalization, satellite networks will continue to play a critical role in providing a rapidly deployable, reliable, and affordable communications infrastructure.

However, present day satellites are limited in their ability to provide high data rate communication services due to the limited availability, and high cost of, satellite resources such as power, energy, and frequency bands. Moreover, present communication satellites were designed almost exclusively for supporting stream traffic such as voice, video or bulk data transfers, and are not efficient for the transmission of "bursty" data traffic such as Internet traffic. With data traffic constituting an increasing fraction of the demand for communication services, future satellite systems must be designed to effectively support emerging data applications. Doing so requires a paradigm shift from traditional circuit switched technology, used for voice communication, to packet switched technology, used in data networks. With support from NASA and the Department of Defense my research group in the Aeronautics and Astronautics Department and the Laboratory of Information and Decision Systems is working to address these important issues. Our research efforts are aimed at significantly increasing the data delivery capacity

*Associate Professor, Department of Aeronautics and Astronautics Laboratory for Information and Decision Systems, 77 Massachusetts Avenue; modiano@mit.edu; http://web.mit.edu/modiano/www/index.html. JACIC Associate Editor, AIAA Member.

of satellite networks through the use of efficient resource allocation and protocol designs coupled with the appropriate hardware designs.

Traditionally, the functions of a data network are divided into layers where each layer operates independently from other layers. For example, the physical layer is responsible for transmitting bits over a communication channel; the link layer is responsible for transmitting packets over a link; the network layer is responsible for routing packets across the network, and higher layers are responsible for end-to-end data delivery. The layered approach simplifies network operations, but often results in degraded performance. In order to make satellite-based networking viable, technologically and economically, the architecture of future satellite networks must be optimized across the different layers of the protocol stack. Protocols for satellite networks must be designed to take into account the unique characteristics of satellite systems: long propagation delays, limited energy and power, relatively high channel error rates, and time-varying channel conditions. In what follows, we discuss some of our recent accomplishments in the area of resource allocation and protocol designs for satellite networks.

II. Transmission Scheduling for Multi-beam Satellites

An important problem that we are addressing involves the transmission of information from space to Earth ground stations. This data may be gathered in space (e.g., weather, surveillance, or space exploration images) or simply data that is being relayed from the ground through a satellite network. The information gathered in space must be transferred to one of multiple ground stations located at different geographical locations that can be reached by routing the data to a satellite that has the desired ground station in its view. The choice of which ground station to transmit the information to and along which route is governed by available resources along the route, the utilization of the route, and weather conditions that may affect the link quality. This problem is complicated by the fact that the data can be delivered to one of multiple ground stations and that modern day satellite systems employ narrow antenna beams for communication where each beam covers a different geographical location on the ground. The channel quality for each of the beams may be vastly different due to local weather conditions.

Two natural questions arise: first, routing, namely to which beam should the data be transmitted; and second, how to allocate the satellite's transmitters to the different downlink beams and how to optimally allocate power to the beams. Our main contribution in this area has been the optimal solution to the joint problem of routing and power allocation. In Ref. 1 we developed a joint routing and power allocation algorithm that maximized the satellite network's overall data throughput. The algorithm makes routing and power allocation decisions based on the number of packets in the buffers corresponding to the different beams. The optimal algorithm allocates more power to beams with more packets in their corresponding buffers; and packets are routed to the beam with the fewest packets (i.e., the least congested beam). A nice feature of the algorithm is that the solution to the routing and power allocation problems can be decoupled. That is, the routing decisions can be made independently of power allocations. However, it is interesting to note that the optimal physical layer power allocation must take into account network layer buffer occupancy. This goes in contrast with the traditional layered view of network protocols; where functions at the different layers are decoupled.

Another important performance metric for transmission scheduling is average message delay. To that end, in Ref. 2 we developed a delay-optimal transmitter scheduling policy for a multi-beam communications satellite. We show, using stochastic coupling techniques, that under an on-off channel model that represents a satellite system with intermittent connectivity, a scheduling policy that allocates the transmitters to the longest queues minimizes average packet delays. This result is among only a few known results on delay optimal transmitter scheduling.

III. Optimal Energy Allocation

A related problem that we are addressing is that of energy allocation and admission control. Typically, a communications satellite is equipped with solar panels that gather energy from the sun to be used for satellite operations. Since, at times, the satellite may not be in view of the sun, satellites are also equipped with rechargeable batteries to store energy. However, with a limited battery capacity, efficient use of satellite energy is critical. In a communication satellite, downlink transmissions consume a large portion of the satellite's energy. It is therefore important to make prudent decisions regarding the transmission of data by taking into account the amount of energy available onboard the satellite as well as anticipated future demands for energy. Due to energy limitations, the satellite may not be able to serve all of the requests that it receives. Moreover, some transmission requests may consume more energy than others and different customers may offer a different monetary payment for service. Therefore, our objective is to select the requests to be served that maximize overall "revenue". Towards that end, in Ref. 3 we developed a Dynamic Programming (DP) formulation for deciding which requests to serve based on available energy onboard the satellite, future energy inputs from the solar panels, as well as the expected requests for future service. We are able to show that the DP value function is a concave function of available energy, and obtain solutions for the optimal consumption schedule by solving the DP recursion in a computationally efficient manner. Our results indicate that expected revenue can be increased by more than a factor of two when compared to the commonly used greedy algorithm that chooses to serve requests as long as energy is available. This new approach can have a significant impact on the operation of future satellite data systems, as it will allow operators to make service decisions that maximize their revenue.

Beyond revenue maximization, it is important to develop energy allocation schemes that meet quality of service requirements on delay, throughput and fairness. To that end, in Ref. 4 we developed a transmission scheduling algorithm for delay-sensitive data (i.e., subject to a deadline constraint) that minimizes energy consumption and maximizes data throughput by scheduling transmissions based on channel conditions. The algorithm can significantly reduce energy consumption by avoiding transmission when channel conditions are poor. This approach is of particular importance to NASA for space exploration missions where energy efficiency is critical.

IV. Efficient protocol design

Perhaps the most critical challenge to increasing the transmission capacity of satellite-based networks is the design of protocols for communication over a hybrid network that consists of both space and terrestrial components. Most terrestrial networks today use the Transmission Control Protocol (TCP) as a transport layer protocol. TCP employs a congestion control scheme that is based on end-to-end windows. That scheme assumes that lost packets in the network are a result of congestion and hence decreases its effective transmission rate in response to lost packets. In a satellite network, where packet losses are likely to be due to transmission errors, this response is inappropriate and significantly reduces throughput (by as much as 90%). Much research has gone into developing new protocols that would be more effective than TCP when used over satellite links. Unfortunately, the telecommunication industry is reluctant to make significant changes to Internet protocols; hence progress in this area has been painfully slow.

As a result, satellite systems often employ satellite specific protocols that operate beneath TCP, for improving performance over satellite links. For example, link layer retransmission protocols are used to eliminate packet errors and Media Access Control (MAC) protocols are used for efficient sharing of the satellite channel. In Ref. 5 we developed a link-layer retransmission protocol for making efficient use of the satellite channel. The protocol dynamically adapts the packet size based on estimates of channel conditions in order to minimize the need for retransmissions.

Our present research explores the interaction between TCP and these lower-layer protocols so that protocols can be designed for efficient joint operation. For example, TCP relies on a timeout mechanism to determine when to retransmit packets and when to activate flow control. The presence of a link layer retransmission protocol can adversely impact TCPs behavior by inadvertently triggering a TCP timeout due to a link layer retransmission. In order to better understand such interactions, we developed analytical models for the performance of TCP in the presence of link layer retransmission protocols[6] and packet multiple access protocols.[7] Our models demonstrate the benefits of satellite link layer protocols and allow for the optimization of the protocols to achieve dramatic improvement in overall system performance.

V. Routing in Low Earth Orbit Satellite Constellations

To efficiently support bursty packet data traffic, future satellite systems may be designed using packet switching techniques. In Ref. 8 we consider a Low Earth Orbit satellite network with packet transmission capabilities. Each satellite is equipped with a limited number of transmitters for communicating with its neighboring satellites and some limited buffer space for storing packets while they await their transmission. Due to the random nature of packet traffic, contention for transmission inevitably occurs as multiple packets arrive at a satellite. In Ref. 8 we developed and analyzed routing algorithms and transmission scheduling schemes that attempt to maximize the throughput of a Low Earth Orbit satellite system. In particular, we show that a scheduling scheme that gives priority to packets that are closest to their destination, when combined with minimum distance routing, results in maximum system throughput.

The total capacity required by a satellite network to satisfy the traffic demand and protect it from failures contributes significantly to its cost. In Ref. 9 we considered the problem of minimizing the total capacity required on the satellite cross-links to satisfy the traffic demand and provide sufficient capacity to recover from failures (either satellite or link failures). Assuming a uniform traffic model, we use a method of “cuts on a graph” to obtain lower bounds on capacity requirements and develop algorithms for routing and failure recovery that meet these bounds.

VI. Future Outlook

So far, our research has been supported by NASA and the Department of Defense. NASA's interests in satellite communications are clear, as satellites will continue to play a critical role in NASA's space exploration missions. Similarly the DoD continues to rely on satellite communications for both battlefield communications as well as providing a communication infrastructure during deployments in remote or hostile locations.

Perhaps the greatest opportunities for satellite-based networks are in the commercial marketplace. Today, most locations in the world are out of reach of a broadband communication infrastructure. For example, a company that wishes to open a new facility in a remote location in Mexico or China may have to wait years (if ever) before it can obtain high data rate connectivity. Such delays can impede and prevent economic development. Even in the United States most homes cannot receive high data rate Internet access because they are not within close proximity of high-speed infrastructure. Satellites are in a unique position to provide these services for very much the same reasons outlined in Arthur C. Clark's 1945 article. Satellites can provide rapid and global connectivity and can be much more cost effective and practical than a terrestrial-based infrastructure. Key to the success of future satellite systems is the ability to develop architectures and protocols that can dramatically improve network capacity at reduced costs.

References

[1]Neely, Michael, Modiano, Eytan, and Rohrs, Charles, "Power Allocation and Routing in Multi-Beam Satellites with Time Varying Channels," *IEEE/ACM Transactions on Networking*, Feb. 2003.

[2]Ganti, Anand, Modiano, Eytan, and Tsitsiklis, John, "Transmission Scheduling for Multi-Channel Satellite and Wireless Networks," *Allerton Conference*, Oct. 2002. Extended version submitted *to IEEE Transactions on Information Theory*, July 2004.

[3]Fu, Alvin, Modiano, Eytan, and Tsitsiklis, John, "Optimal Energy Allocation and Admission Control for Communications Satellites," *IEEE/ACM Transactions on Networking*, June 2003.

[4]Fu, Alvin, Modiano, Eytan, and Tsitsiklis, John, "Optimal Energy Allocation for Delay-Constrained Data Transmission over a Time-Varying Channel," *IEEE Infocom 2003*, April 2003.

[5]Modiano, Eytan, "An Adaptive Algorithm for Optimizing the Packet Size Used in Wireless ARQ Protocols," *ACM Wireless Networks*, Aug. 1999.

[6]Liu, Chunmei, and Modiano, Eytan, "On the performance of TCP Congestion Control over Satellite Links with ARQ," *Computer Networks*, Dec. 2004.

[7]Liu, Chunmei, and Modiano, Eytan, "An Analysis of TCP over Random Access Satellite Links," *IEEE Wireless Communications and Networking Conference (WCNC)*, Feb. 2004.

[8]Sun, Jun, and Modiano, Eytan, "Routing Strategies for Maximizing Throughput in LEO Satellite Networks," *IEEE Journal on Selected Areas in Communications*, Feb. 2004.

[9]Sun, Jun, and Modiano, Eytan, "Capacity Provisioning and Failure Recovery for Low Earth Orbit Satellite Networks," *International Journal on Satellite Communications*, June 2003.

[10]Clarke, Arthur C., "Extra-Terrestrial Relays," *Wireless World*, Oct. 1945, pp. 305-308.