Load Forecasting of a Space-based Telecommunications Network: NASA's TDRSS

Allen J. Levine
Networks Division
NASA GSFC, Code 532.1
Greenbelt, MD 20771
E-mail: Allen.J.Levine@gsfc.nasa.gov

David H. Joesting
AlliedSignal Aerospace
1 Bendix Road
Columbia, MD 21045
E-mail: David.H.Joesting@atsc.allied.com

Abstract

Planning to assure adequate resources exist for low-earth orbit satellite telecommunications through NASA's space-based Tracking and Data Relay Satellite System (TDRSS) is an on-going activity at Goddard Space Flight Center (GSFC). The software tools used in this process were originally developed to analyze, and phase down, an older NASA tracking ground network. The tools have been enhanced over many years with retention of the original ground-based network capabilities. The analysis techniques involve modeling possible network configurations and simulating expected satellite tracking, telemetry, and command (TT&C) telecommunications requirements on a priority basis. The key to generating representative deterministic schedules is simulation procedures that efficiently support complex generic requirements while at the same time generate results that are consistent with what could be expected by the actual network under similar conditions. Over the years this approach has proven itself although there is a bias to overestimate expected loads due to unforeseeable satellite launch slips. This type of architecture can be easily and quickly applied to hybrid space-based and ground-based networks of any practical complexity.

At the Goddard Space Flight Center (GSFC), network capacity planning analyses and capability assessments of low-earth orbit (LEO) satellite telecommunications support have been performed for at least 25 years. These activities originated in support of the Spaceflight Tracking and Data Network (STDN), a NASA network of approximately 20 ground stations. This network was composed of antennas of diverse sizes and encompassed a wide range of radio frequency (RF) bands ranging from VHF to S-band. Since then NASA has expanded to the X- and K-bands with greatly increased data transfer speeds. In the late 1970's in an effort to consolidate operations and reduce costs, while taking advantage of emerging technologies, NASA turned to the Tracking and Data Relay Satellite System (TDRSS). Using an existing application designed for ground station planning through simulation and analysis of spacecraft support schedules, workarounds were developed for these new "ground stations" whose elevations were at geosynchronous altitude. For the most part this worked; however, we found that there were new attributes at play and inherent inefficiencies using the current processes; so new capabilities needed to be added.

Rather than start over and produce new tools from scratch, the Network Planning and Analysis System (NPAS), as the existing system was called, was re-engineered. This integrated set of C and FORTRAN programs initially operated on IBM mainframes and has since been migrated to Hewlett-Packard Series 9000 workstations using X-windows' interfaces. It was originally developed and maintained for NASA by Bendix Corporation, now part of AlliedSignal Aerospace. In the latest version, the NPAS software has been modularized to ease future upgrades. It has retained all previous ground station capabilities, as well as incorporating new relay satellite service operational capabilities and constraints including advanced scheduling algorithms. The compelling reason for keeping the ground network (GN) capabilities in NPAS during the re-engineering was the need to analyze the transition/phase-down from a GN to a space network (SN) which was occurring at that time. Ground station reductions and closures, coupled with the service transition to the relay satellites, needed to be planned carefully in order to minimize impacts to the network's customers. Even after the transition was complete, GN capabilities were retained, because of anticipation of combined GN/SN supports. In recent years, the concept of "service provider" flexibility has become increasingly valuable with the evolution of cross network service provision concepts. It is possible to have a constellation of SN relay satellites in mid-altitudes, and also use cross-links to ground stations or high altitude SN relay satellites for completion of telecommunications links.

The approach that we at GSFC have taken to analyzing a network's forecasted load is a deterministic rather than a statistical or probabilistic one. Using the Goddard Trajectory Determination System (GTDS) for planned orbit computations, a database containing a given network configuration and mission support requirements and constraints, and scheduling algorithms that closely mimic real-world operations, the tools provide valid schedules of communication services for periods of future time. We have found that the true strength of the deterministic method is its ability to quickly model adverse or exotic requirements, for example, simulating the real-time support of earth land boundary viewing. This is a highly appropriate procedure for investigating the sensitivity of mission requirements satisfaction as network resources are varied. Scenarios defined at any level of support or constraint complexity may be quickly modeled, analyzed and reported with confidence.

With regard to network resources, the physical characteristics of simulation modeling of a single Tracking Data Relay Satellite (TDRS) is somewhat different than the modeling of a ground station. Ground stations typically contain one or more single frequency antennas with both a forward (uplink/command) and return (downlink/receive) capability. Only one user satellite is supported at a time on each antenna. The current version of the space-based TDRS consists of two single access (SA) antennas that each support two different RF telecommunication bands, S-band and Ku-band, both forward and return services. These SAs operate independently of each other and, except in rare situations, a single SA is only used by one user satellite at a time. In addition, a TDRS contains a multiple-access return antenna (MAR) which supports up to 20 return services at a time. (Current TDRSS ground terminal implementation limits each TDRS to 5 MAR services.) The maximum datarate on any MAR is nominally 50 kilobits per second (100 kilobits if the I & Q channels are used at the same time) which is considerably less than on the SA. Lastly, each TDRS has a multiple-access forward antenna (MAF) that is separate from the MAR. The MAF can be used independently of the MAR, however only 1 such service per TDRS is available.

One factor that is somewhat unique to the TDRSS arena is the prototype event with its possible complex assignment of support services. A prototype event is a set of planned telecommunications services from the 3 different TDRS antenna types defined above: SA, MAR and MAF. The services are known as KSAF, KSAR, SSAF, SSAR, SMAF and SMAR. (The first letter refers to the RF band, the second and third defines Single Access or Multiple Access, and the fourth means forward or return.) A prototype event can include any of the three antennas, using any of the services as well as 1- or 2-way tracking services, concurrently and consecutively. Satellites, such as Landsat-5 and the Hubble Space Telescope have reasonably complex prototype events. An example showing the time relationships between services within one possible prototype event is given in Figure 1.

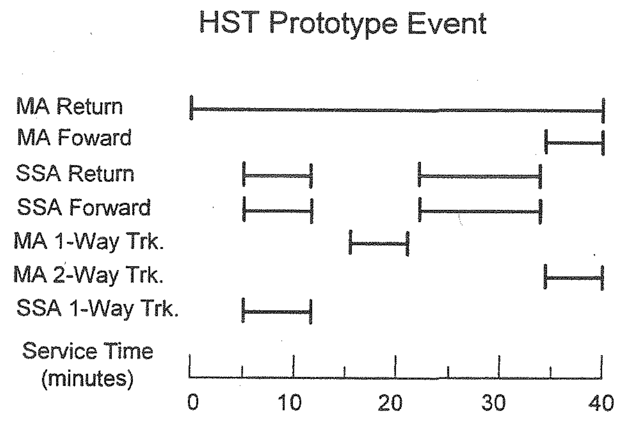

Figure 1

Typically, a TDRS view period of a LEO satellite is relatively long, in most cases approximately 50 to 55 minutes. As a consequence, when one satellite's support requirements conflict with another, the possibility exists of shifting one or more of the services in time within their visibility to resolve any conflicts. Therefore, allowing user requirement definitions of temporal tolerances or windows (within practical limits, of course) is important. Overlapping visibilities by multiple TDRSs provide even more flexibility in resource allocation for any scheduling simulation or even in real operations, if the system takes advantage of it.

Also fairly unique in the SN world are the powered flight launches/expendable launch vehicles, aircraft (powered and balloons), and land or fixed systems that look for long uninterrupted service, even at high data rates (at megabits per second). Support for these types of projects may be beyond the scope of ground stations, not only on a coverage basis, but also because the user ties up a single resource for longer periods of time. This type of support seems to make optimization of a representative schedule easier in the fact that there are fewer combinations of resources. At the same time it can have the opposite effect, this single extended use of resources may make the overall load on the network such that a true optimal solution does not exist (not all users can get all the support they desire.)

In general, when performing capability assessments of ground-based networks, RF constraints take a back seat to geometric line-of-sight restrictions. While sufficient signal strength is important, independent analysis could be performed to determine the feasibility of support. Assuming reasonably capable telecommunications equipment on board the spacecraft and on the ground, relatively low- or mid-gain antennas with wide beamwidths can be placed on the spacecraft, minimizing RF blockage and attitude constraint problems. This is not, however, true for services through a geosynchronous relay satellite. Although omni-directional antennas can be used for some telecommunications support, especially at low data rates using the TDRS Single Access (SA) antenna, in many cases a high-gain (directional or pointing) antenna is required by the user spacecraft. Thus antenna blockages and attitude constraints are very important factors in determining the ability of a space-based network to provide service and their definition becomes a significant requirement in performing a meaningful capability assessment.

Telecommunications support requirements can be defined specifically or generically. In NPAS, both specific and generic requests are processed. A specific request is one that defines specific services, times and resources (for example, 20 minutes of SSA forward and return service on a particular TDRS at 1200 GMT). A generic request allows the system some latitude to do the selecting (for example, between 15 and 20 minutes of both SSA forward and return service by any TDRS once per day). A further extension of a generic request is its ability to define recurring requirements, that is, one request statement defining multiple requests (for example, once per orbit, or twice per day separated by at least 12 hours). A robust model would be able to include all of the above defined requirements requests.

One unique situation that manifested itself in the early days of SN analysis that hadn't appeared in the earlier GN studies involves the "per orbit" requirement. An orbit is generally defined as one complete circuit around the earth; early convention typically defined the start and end points as being the successive ascending equatorial crossing. A "per orbit" service requirement definition for a network of ground stations is easy, but with geosynchronous relay satellites, it can become a problem, if long visibilities are required. These relay satellites, like ground stations, move with the earth's rotation; user spacecraft do not. Taking a simple case of a user satellite with a 95 minute orbital period, with no precession, would see the earth shift in one orbit almost 24 degrees. The TDRS visibility therefore intersects and crosses the orbit boundaries. Many times a customer will request "2 TDRS events per orbit", thinking one event per TDRS, that is, one TDRS stationed over the West Pacific (TDW) and one over Brazil (TDE). What they are really saying is "I want two TDRS events every TDRSS viewing cycle." A TDRSS viewing cycle is, however, longer than an orbit since over the course of an orbit, the TDRS has moved approximately 24 degrees. This equates to 6 minutes more; the cycle is 101 minutes. This has important implications for on-board data storage management, and also the estimate on the number of events per day requested (e.g., 14 versus 15 for "I per orbit", 28 versus 30 for "2 per orbit"). The current challenge is to determine a good convention to cover the possible and probable relay satellite constellations. The one typically used in NPAS is the longitudinal midpoint between TDE and TDW in the Zone of Exclusion, which is the region of no coverage by TDE and TDW of satellites under 1200 Kilometers altitude and with low to mid orbital inclination. (For a network of relay satellites at lower altitudes, the orbit cycle convention will probably be better suited.)

A number of solution methods to the telecommunications service scheduling problem exist, most of which involve a priority or service request processing order. The two methods are not necessarily the same. (The priority order indicates a hierarchy while a processing order can allow for adjustment.) Furthermore, one can take the defined requirements and expand them up front (for all users), then "fit" the resultant services all together, like a puzzle, or expand each requirement one at a time and place it in the "schedule" as it is processed. NPAS uses the latter because this provides the maximum flexibility by retaining generic definitions including any level of complexity "up to the last second". The former method appears to have the advantage of simplifying schedule adjustments since all requests are defined prior to "scheduling," however, for all practical applications this method is computationally expensive. In any realistically representative NASA Space Network support scenario, the number of possible puzzle pieces, (time dependent pieces in this puzzle change shape each time related piece/service is "fitted") is enormous. This becomes a substantially larger problem if multiple service prototype events are supported.

Within the NPAS standard priority scheme, prototype event scheduling is performed based on longest service length first and then station priority (if any). In order to simulate the complex scheduling optimization that may occur within any single user's own scheduling operations, a special "geometric optimization algorithm" may be invoked for a set of generic requirements. This process is designed to maximize the total scheduled time for that generic request. The process starts with an initial greedy schedule with local optimal solutions. If the total support is less than the desired minimum, then a limited depth first search with backtracking process is invoked. During its operation, it considers local sub-optimal solutions in order to generate a better global solution. The limit to which the depth of searching is allowed may be controlled by the modeling analyst based on time considerations.

Periodically, an indirect route is taken in examining the ability of the SN to provide service to a prospective new customer. Analyzing the resource "free" time, or unutilized time, in terms of service gap time durations and frequency, often provides valuable insight to the ability of the network to service that additional customer. If a new project comes along and indicates a desire for services that do not impact existing customers, then a quick analysis can be performed to determine whether or not the new customer's needs can be met. Even with the possibility of impact, an experienced analyst can make preliminary judgements as to the probable success of the additional services.

An important consideration is the resultant accuracy of any simulated operational schedule. It would be possible to develop computer schedule solution procedures using generic satellite support requirements that would in effect be too accurate. The purpose of the network loading analyses, whether SN or GN, is to simulate the performance of the real network with expected resource configurations and mission support requests. For the most part in the SN, the real operational environment consists of individual satellite projects expanding their own requirements into events with fairly specific start times and durations for each services. These sets of events are sent in on a weekly basis for a support week up to 3 weeks in the future. Conflicts between any two users are settled manually by operations controllers, typically on a priority basis. Rejected events are returned to the satellite project for modification and resubmittal. Due to the different responses by the projects based on their own internal priorities, this is an iterative process which cannot be directly simulated.

The NPAS process has shown itself to fairly closely mimic the results of that iterative process without introducing too much optimization that would bias the results towards overestimating either the network's capacity or an individual satellite's expected support. However, there are other times when a near optimal schedule may be required. Instead of introducing the up front expansion of the users' requests with that procedure's known computationally intense requirements, a second approach has been taken in the NPAS. The resultant schedule from a normal conflict free NPAS simulation, which represents a good operational schedule typically well within 95% of the true optimal schedule, is taken as the starting point for a optimization run. All missions' requirements are expanded and missions requirements that are not 100% satisfied in the starting schedule are revisited. Event time movements of all missions on required network resources are attempted within tolerances. Tolerances here are determined by intersections of the original visibilities and the mission's stated generic requirements. Backtracking to predefined levels of complexity is used in this process. Based on the amount of computer time available, a better to near optimal representative schedule is obtained.

In the capacity planning process, it is important to analyze not only individual network configurations and mission support profiles, but to also obtain an understanding of the sensitivity of the resultant loading to the various support parameters. One area of concern is that the results of solution schedules are somewhat dependent on the relative alignments of the user satellites in their orbits. NPAS capability assessments typically begin with a coverage variability analysis portion that attempts to identify worst, best, and average (expected) cases. This is accomplished by the use of Monte Carlo simulation methods, which vary the defined mission orbital parameters in a hundred or more simulations. The satellite orbit profiles from each of the three cases are made available for use based on the purpose of the assessment.

In ground station analyses, downtimes for system maintenance can be modeled as well as failure modes. In general, however, failures can be ignored or de-emphasized since repairs can be made. With a relay satellite, it is more difficult. When a hardware failure occurs onboard, there is no way currently to effect a repair, redundancy becomes critical. As a result, failure modeling and replenishment needs analysis are needed to complete the work. The TDRS reliability models are generally analyzed at the subsystem level, although the actual reliabilities are defined down to the component level. The four subsystems defined are the Payload, Bus, SAC East, and SAC West. The first 2 subsystems are critical items for any telecommunications services to be provided, so they are single points of failure, although there are redundant components and noncritical elements within each. The 2 SACs are the 2 Single Access antenna systems, and therefore are single points of failure only for the services provided by only that antenna. Therefore, it is possible for a TDRS to have only one SA antenna in the modeling. Currently, the analysis does not go down to the service level, although in real life the TDRS SA systems are used even when some services are not available. TDRS-3, for example, has only one fully functional SA antenna system, the other SA has S-band services and Ku-band return services, the Ku forward being nonfunctional. This capability is currently being implemented in the modeling; the only question remaining is how to present the results and how to judge the viability of TDRS service. Part of the decision is a management one - is the service cost effective?

Historically, load forecasting of the SN has been reasonably accurate for near term planning; long term forecasting was somewhat uncertain. However, with the changing environment, with NASA becoming more focused in its overall planning and improvements in reliability in its systems, forecasting is likewise improving. (Refer to Figure 2.) In the earlier days, with a more robust and optimistic budget, the representative NASA enterprises projected a number of study missions, each of which was felt had a reasonable chance of acceptance. At that time, it was more difficult to predict the number and types of missions to receive commitments due also to the optimism and also due to relative inexperience in forecasting. For the Space Network, a drastic difference between prognostication and reality occurred when the Challenger accident happened. Launches by the Space Shuttle stopped for almost 3 years, commercial satellites (with possible SN customers) were taken off the Shuttle manifest, and a TDRS was lost, being a passenger on that ill fated launch. As it turned out, most of the projection was just delayed approximately 3 years, so a slide could actually be applied to the projection curve.

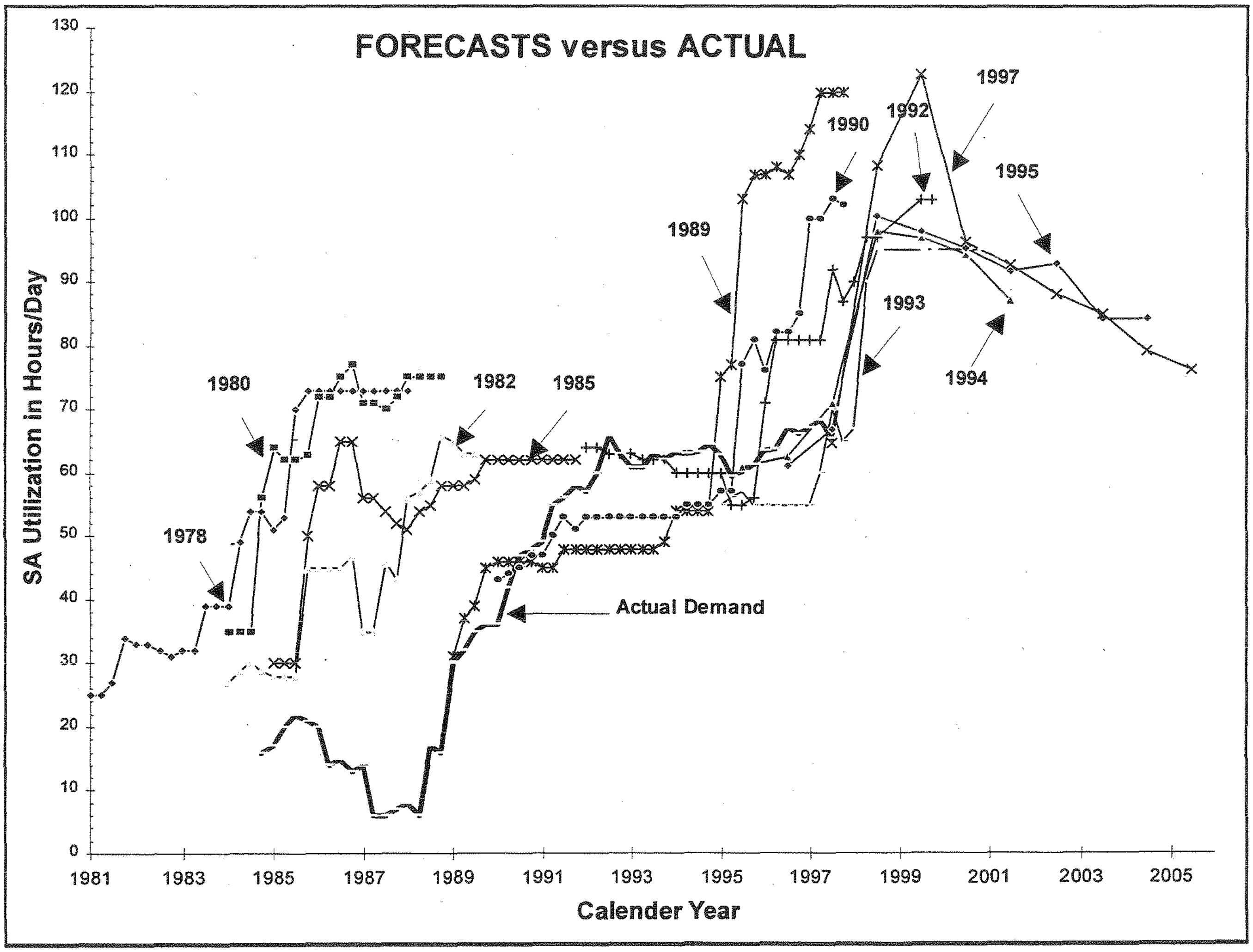

Figure 2

One of the interesting trends in satellite communications capacity planning covers the ground and space networks teaming scenarios. The newer customers are starting to look at direct downlink science/telemetry data, basically getting the principal investigators' data directly, or almost directly, to them through ground links, while using the almost continuous visibility capabilities of space-based stations for spacecraft monitoring and control. Lower data rates (for spacecraft telemetry) and periodic (possibly on-demand) commanding and/or tracking through the TDRSS have been shown to be viable due to the currently forecasted light loading on the MA system.

In addition to this type of analysis, there is the further step of performing GN verses SN trade-off studies. In today's competitive environment, this type of analysis is critical. Obviously, a consistent set of metrics is required; a tool such as NPAS is needed to perform the appropriate analyses. The only other requirement is obtaining from the prospective customers a consistent sets of communications requirements for each support network. It would seem to be a simple matter, however, some problem has been experienced with one or two customers (who may have had a predisposition toward one type of network). Sometimes, more "digging" is needed to provide an objective answer.

Given the changing environment, proven and adaptable tools are required to provide proper capacity planning analyses and capability assessments for telecommunications support. Recognition of the difference between space-based and ground-based tracking systems is necessary as well as the ability to team their support when appropriate. Melding the capabilities into a cohesive model with appropriate analysis is the key to performing effective assessments and recommendations for successful support of all customers.

References

Stern, D. C., Levine, A. J., and Pitt, K. J. 1994. Accuracy Analysis of TDRSS Demand Forecasts. In proceedings of the Third Symposium on Space Mission Operations and Ground Data Systems, 331—318, Greenbelt, MD: NASA Conference Publication 3281.

Simons, M., and Larsson, G. 1994. Analysis of Space Network Loading. In proceedings of the Third Symposium on Space Mission Operations and Ground Data Systems, 1071—1077, Greenbelt, MD: NASA Conference Publication 3281.

Brase, J., and Burns, M. 1997. TDRS Space Network Reliability and Availability Modeling Status Report. Stanford Telecom.