Assessing the Impact of Real-time Communication Services on the Space Network Ground Segment

Marc Sanchez Net, Iñigo del Portillo, Bruce Cameron, Edward Crawley
Massachusetts Institute of Technology
77 Massachusetts Ave 33-409
Cambridge, MA 02139
617-682-6521
{msnet,portillo,bcameron,crawley}@mit.edu

Abstract—**Communication networks to support space missions were originally architected around non-real time data services. In fact, missions have always required real-time services (e.g. telemetry and command), but the bulk of scientific data being returned to Earth has typically been highly delay tolerant. Nevertheless, future robotic and human exploration activities are rapidly pushing towards low latency, high data rate services. Examples can be found both in the near Earth domain (e.g. near real-imagery through NASAs LANCE program) and the deep space domain (e.g. HD video from Mars). Therefore, the goal of this paper is to quantify the effect of new real-time high data rate communication requirements on the ground segment of current communication networks.**

To that end, we start by analyzing operational schedules for NASAs Space Network (SN) in order to characterize the utilization of the overall network in terms of total data volume and contact time, as well as identify current mission drivers. These results are compared against proposed network requirements for future robotic and human near Earth exploration activities in order to quantitatively assess gaps in the SN capabilities. Using these results, we implement a rule-based expert system that translates SN-specific operational contacts into high-level data requirements for the ground segment of the network. We then exercise the expert system in order to derive the requirements that future exploration activities will impose on the SN. Finally quantify the impact of real-time data delivery services across NASAs ground segment by computing the wide-area network cost for different levels of data timeliness.

TABLE OF CONTENTS

1. INTRODUCTION

Motivation

NASA's Space Network [1] provides the necessary infrastructure to deliver communication and navigation services to robotic and human exploration missions operating in the near Earth domain. Over the last years, several studies have formulated and analyzed the space of possible network architectures that can efficiently improve the amount of data successfully returned from the spacecraft to the ground (see, for instance, references [2], [3] or [4]). Results indicated that data rate improvements on the space segment would have to be accommodated in order to provide service to high demanding missions such as DESDynI or HyspIRI [5], while it was assumed that the ground segment could be easily upgraded to match these increasing communication rates.

In this paper we revise this assumption and assess the impact of increased data return profiles on the ground infrastructure of a space communication network, with specific emphasis on NASA's Space Network (SN). In particular, we focus our attention on missions that will require services characterized by both high data rate and latency-constrained communications. Examples of such services are High Definition (HD) video for astronaut support, as well as other emerging applications such as telerobotic servicing or deliverance of real-time Earth imagery.

Literature Review

Ground Network Architecture—The ground segment of space communication networks has been studied in a wide variety of contexts. For instance, references [6] and [7] focus on architecting ground networks in order to maximize the optical space-to-ground link availability. In other words, they study how to optimally allocate the network's downlinking functionality across multiple ground stations in order to ensure that communications with the network's spacecraft will not be disrupted due to cloud coverage.

In contrast, reference [8] compares the architecture of ground network with respect to the level of service provided to its users, where service satisfaction is measured in terms of daily data volume, latency and network cost. It is argued that capacity increase through ground antenna deployment is two orders of magnitude lower less costly than that of a space based system (EUR 1M vs. EUR 380M), and one order of magnitude less expensive if up to six ground stations are build to improve the coverage of the ground system (EUR 25M vs. EUR 380M). However, the cost of deploying and maintaining the wide-area network (WAN) that interconnects all ground stations is not considered, thus biasing the validity of these estimates towards favoring the purely ground infrastructure.

On the other hand, the architecture of a network's ground segment has also been studied from the perspective of allocating signal processing functionality across different nodes in the network [9]. Results in that case indicate that ground architectures that distribute the demodulation and decoding of

978-1-4799-5380-6/15/$31.00

incoming signals across all ground stations results in cheaper alternatives due to reduced WAN bandwidth requirements. Furthermore, it also demonstrates that significant cost savings can be achieved by regulating the store and forward functionality for delivering large scientific data products collected by the spacecraft.

Latency-constrained Applications—Several studies have reported that future spacecraft and their missions would largely benefit from (near) real-time communication services. For instance, reference [10] summarizes the findings of the Latency Study initiated by NASA's Earth Science Division in order to assess the latency requirements for the latest Decadal Survey missions [5]. It was conducted using an interview-based expert elicitation with both scientists and engineers. Results indicated that over 40% of the respondents considered the optimal latency for an Earth Observation data product to be less than 2 hours as opposed to the traditional 6 to 12 hour requirement. Note that this latency requirement includes all data processing delays and thus translates to a 20 to 30 minute latency requirement for NASA's ground communication network [11].

Similarly, human-operated robotic space exploration missions would also greatly benefit or can only be undertaken with the support of real-time services. The rationale for needing them is largely grounded on human physiology, particularly what reference [12] refers to as *cognitive timescales for space exploration*. In a nutshell, researchers have found that in order for a human to successfully undertake any interactive activity remotely, the maximum allowable round-trip latency has to be limited to less than 500ms. This is particularly relevant in the space environment due to the large propagation delays that the signal has to overcome.

Applications with remote interactivity in the space domain are generally divided in three categories: Teleconferencing, robotic servicing and telepresence. Teleconferencing refers to real-time interaction between an astronaut and a mission control center, or between two astronauts. It can be based on a voice-only service, thus requiring limited data rates and up to 200ms latency [13]. Alternatively, enhanced teleconferencing experience can be obtained through the transmission of real-time high-definition HD video, which in turn results in large bandwidth requirements - see for instance [14] or [15].

On the other hand, robotic servicing or in-space servicing refers to the ability of repairing, refueling, assembling or cleaning space assets through a remotely controlled mission [16]. As indicated in references [17] and [18], the success in-space servicing is based on the ability to control robotic spacecraft from a remote location, be it a ground control center or a space-based station. Finally, telepresence includes all processes and activities that enable humans to conduct research activities in remote location through a controlled robotic agent [19]. Several articles (e.g. [12], [20]) propose mission concepts based on telepresence and demonstrate the added value of having humans in the loop as opposed to the current purely robotic-based exploration.

Research Goals

Based on the surveyed literature, this paper addresses the following research goals: First, it defines the architecting problem for the ground segment as a constrained combinatorial optimization problem. Second, it proposes a simulation engine that is flexible enough to quantify the effect of increased data rate and latency requirements on the space network's ground segment. Next, it validates the tool against the current implementation and operations of the SN. Finally, it uses the combinatorial formulation together with the simulation engine to quantify the required network capacity in order to successfully satisfy the requirements imposed by future near Earth missions, with specific emphasis in the 2020 to 2030 decade.

Paper Structure

The remainder of this paper is structured as follows: First, section 2 formulates the problem of architecting the ground segment of a space communication as a constrained combinatorial problem over a set of predefined functions. It also describes the different tools utilized to generate meaningful traffic patterns for space networks, as well as the results of the validation case against the current SN system. In turn, section 5 focuses in a forward-looking case study where the increase in bandwidth requirements for the SN is quantified as a function of the data rate and latency requirements imposed by high demanding Earth observation mission under formulation such as DESDynI or HyspIRI. Finally, section 6 concludes the document by summarizing the obtained findings and identifying areas of future work.

2. Problem Formulation

System Definition

Figure 1 provides a notional view of a space communication network layered in three planes: The space segment, the ground segment and the terrestrial infrastructure. The space segment, if it exists, is composed by a set of relay spacecraft that communicate with the network customers through space-to-space RF or optical links. In contrast, the ground segment is composed by a set of assets (antennas, operation centers and end-users) that cooperate in order to send and receive the necessary information to successfully control and operate the remote spacecraft. Finally, the ground network is supported by a terrestrial infrastructure, typically outsourced to commercial networking companies (e.g. AT&T), that provides low level functionality to ensure data delivery across different points-of-presence (POP) located at each ground segment asset.

Given this description of the system, this paper is particularly interested in the ground segment of the space communication network, while the terrestrial and space segment are assumed to be outside the scope of interest. Note, however, that these two parts of the system either provide services or impose requirements on the ground segment, thus being fundamental to successfully support services in the overall space communication network.

Problem Formulation

Based on the definition of system architecture, i.e. "the allocation of physical/ informational function to elements of form, and the definitions of interfaces among them and with the surrounding context" [21], we formulate the problem of architecting the ground segment of a space communication network using the following assumptions:

- The system's surrounding context is composed by both the network's space segment and the WAN network, data link and physical layer. From a practical standpoint this assumption translates to the use of standard space-link protocols [22] for the space-to-ground link , as well as a standard IP infrastructure to move data across the different elements of the network (see figure 1). It is also assumed that the terrestrial infrastruc-

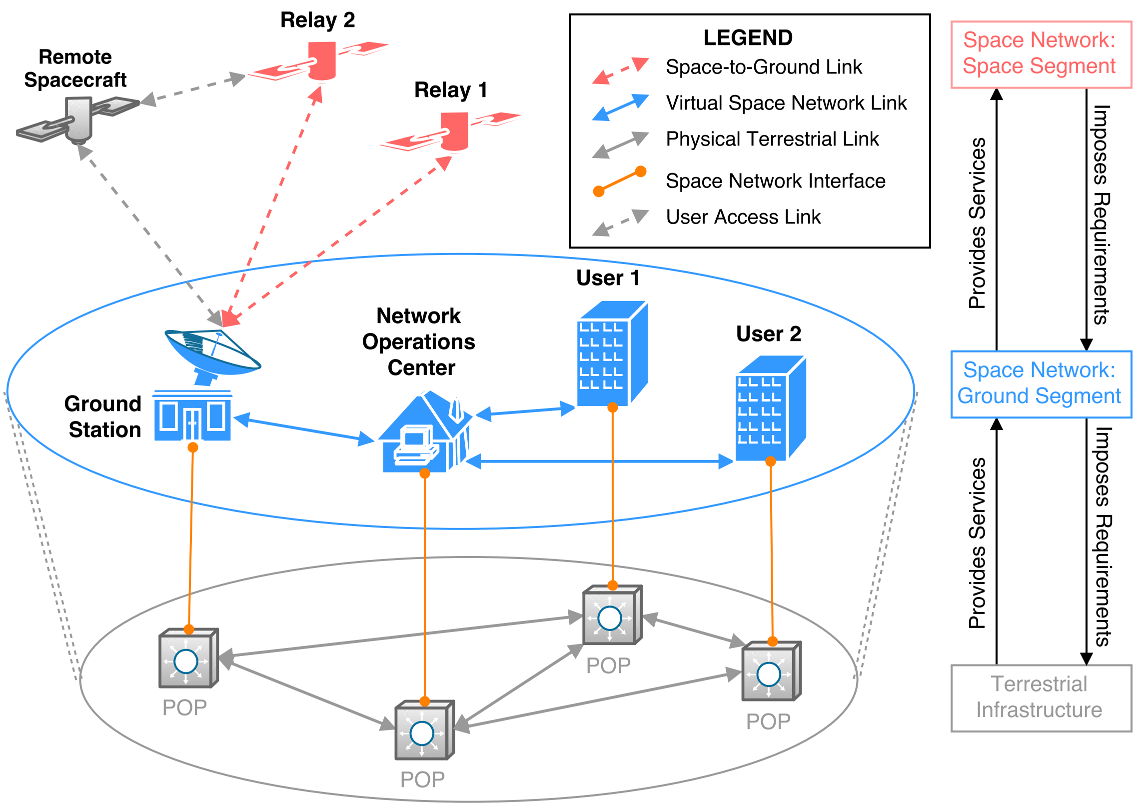


Figure 1: **Space Communication Network**

ture has been optimized (e.g traditional TCP protocols have been replaced to avoid problems with high bandwidth-delay links [23] [24]) to deliver a certain quality of service (QoS) defined by three parameters, bandwidth, maximum delay and availability (see for instance [25]).

- Once the information reaches the ground, seven canonical functions need to be accomplished in order to successfully deliver the data to the end-user: sampling, beamforming, (de)modulating, (de)coding, framing, routing and store and forwarding [9]. Figure 2 provides a summarized overview of these functions and the order in which they need to be executed to successfully deliver data from a remote spacecraft to a ground user.

- The form of the ground segment is defined by a set of geographically dispersed nodes connected through dedicated lines. Five canonical nodes are available: a sampler, a receiver, a router and a leveler, as well as a transparent node that does not recover digital information from the received signal. Table 1 provides the mapping between each node type and the previously introduced functionality.

Based on this assumptions and the insights from references [26] and [27], we formulate the system architecting problem for the ground segment of space communication networks as an assigning problem in which each node in the network is assimilated to one of the canonical nodes from table 1. As an example, consider the original unspecified architecture on the left-hand side of figure 3. Multiple network architectures can be obtained by assimilating nodes N_1 to N_4 to a sampler, receiver, router or leveler. It is immediate to see that for a network of N nodes and M canonical node types a total of N^M architectures can be generated if no constraints are assumed.

On the other hand, note that in this paper we are not interested in sizing the topology and links that define the terrestrial infrastructure. Rather, we are interested in determining the appropriate bandwidth for the virtual space network links (see figure 1) such that latency-sensitive services can provisioned across the network. In other words, we are not interested in sizing capacity between two POPs of the terrestrial infrastructure. Instead, we want to quantify the bandwidth requirements at the ground segment level, i.e. between ground stations, network operation centers and final users.

Traffic Generation for Space Communication Networks

Traffic patterns for communication networks are typically driven by the type of system under consideration. For packet networks, traffic is typically expressed as two random variables, one that characterizes the probability of having an arrival and another one that characterizes the time it takes for the network to process the packet. This, in turn, allows network engineers to size the network according to queuing [28] and network calculus [29].

In contrast, in mobile networks traffic typically depends on the medium access control (MAC) protocol used [30]. In some cases, this protocol is deterministic in nature and assigns a particular set of resources to the users statically (e.g. TDMA, CDMA). Therefore, in those cases traffic is expressed as the number of unitary resources (time slots, frequency carriers) that the network has to provide as a function of the customer base. On the other hand, random access mechanism approximate the traffic resulting from the (MAC) protocol as a random variable (e.g. new arrivals plus retransmissions in a slotted Aloha system are approximately poisson distributed [28]) that can once more be studied

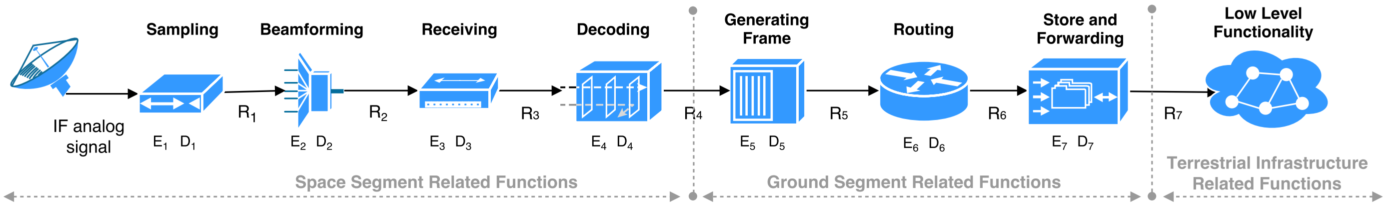


Figure 2: **Network Functional Decomposition**

Table 1: **Architecture Canonical Nodes**

Node Type	Node Functionality	Symbol
Transparent	RF/IF functions (e.g. filtering, mixing)	○ (white)
Sampler	Sampling, Beamforming	● (yellow)
Receiver	Sampler + (De)modulating, (de)coding, framing	● (cyan)
Router	Receiver + Routing	● (orange)
Leveler	Router + Store and forwarding	● (green)

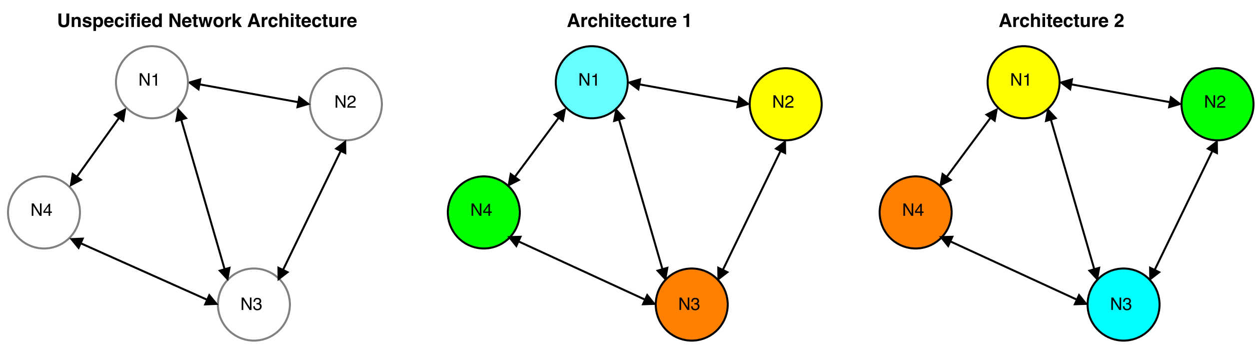


Figure 3: **Notional Network Architectures**

through queuing theory and similar tools.

Nonetheless, space communication network for space exploration activities are particularly challenging for traffic generation purposes due to the scheduling system they implement [1], [31], [32]. In other words, traffic is completely deterministic and serviced according to master schedule that is negotiated with all missions requesting services from the network several weeks in advance.

In order to solve this problem, we combine the service characterization module presented in reference [3] with the heuristic rule-based scheduling system from reference [4] to obtain meaningful traffic patterns for near Earth space communication networks. Note that the first tool transforms high-level communication requirements from any mission into a meaningful concept of operations that can then be scheduled through the heuristic system. Note also that both tools have been previously validated against real operations environments, thus ensuring that the obtained traffic patterns are representative of those supported by networks such as the TDRSS system.

Modeling Real-time Communication Services

Once realistic traffic patterns between the network customers and assets have been obtained, the next step is to obtain a mission-independent model for the types of data that spacecraft return and their associated latency requirements. Based on references [9] and [33], we define four canonical services:

- Audio/Video data, with a latency requirement of 2 seconds.
- Engineering telemetry, also known as spacecraft housekeeping data, with a latency requirement of 5 seconds.
- Quick-look science data, with a latency requirement of 1 hour and 15 minutes. This is representative of missions currently supported by the NASA's LANCE system [34].
- Bulk science data, with a latency requirement of up to 12 hours. This is representative for the vast majority of current Earth Observation missions [10].
- Tracking data, with a latency requirement of 5 seconds [1].

Note that the provided latency requirements are representative of the current system implementation. Therefore, in section 5 we investigate the effect of lowering these latency requirements in the architecture of the TDRSS system and assess the required bandwidth to successfully deliver real-time content to the users.

Modeling Space Communication Networks

Figure 2 provides a summarized view of the different functions that have to be executed in order to deliver data from a spacecraft to its mission operations center and end users. Each function is modeled through two parameters $E_i = \frac{R_{i+1}}{R_i}$ and D_i, $i \in [1, 6]$, which characterize the bandwidth efficiency and processing delay that is required when performing them (see table 2 for the specific values used). Data streams coming out of the store and forwarding functionality (R_7 in figure 2) are delivered to the transport mechanisms of the ground network, which is assumed to have perfect BW efficiency and a maximum latency of 120msec. While the latter is realistic, the former is assumed so that the reported required bandwidth can be interpreted as the *requirement* that the ground network imposes in the terrestrial infrastructure.

Next, we provide a brief description of the different functionality summarized in figure 2 along with the set of assumptions used in order to obtain simplified values for E_i and D_i, $i \in [1, 6]$ in table 2. Furthermore, values for the specific implementation of TDRSS system are provided when necessary.

First, the sampling functionality is specified by the RF/IF front-ends used to digitalize the analog IF signal obtained after the low noise amplifier and RF/IF frequency conversion. In general, the bandwidth efficiency of the sampling functionality depends on the IF signal bandwidth and central frequency, as well as the number of quantization levels used during the A/D conversion process. A general expression can be derived from reference [38]

$$R_1 = f_s \left[\frac{samples}{sec}\right] \cdot Q \left[\frac{bits}{sec}\right] \quad (1)$$

$$f_s = \begin{cases} \dfrac{2f_c - B}{m} & \text{if } \exists m = 2n : f_s \geq \alpha B,\, n > 0 \\ \alpha\,(f_c + B) & \text{otherwise} \end{cases} \quad (2)$$

where f_c is the IF carrier frequency, B is the two-sided signal bandwidth, Q is the number of bits per sample and m is a positive even index such that the Nyquist sampling theorem is satisfied ($\alpha \geq 2$). Table 3 summarizes the sampling rate obtained with equation 2 and signal processing information from references [1] and [9]: 16 quantization levels for low data rate services (S-band) and only four quantization levels for high data rate services (Ku and Ka-band). Furthermore, realistic values for the hardware equipment used in the Space Network were provided in reference [9], thus allowing us to successfully validate the proposed analytic formulation.

On the other hand, the TDRSS multiple access service requires beamforming functionality in order to increase the signal to noise ratio. This functionality is implemented at the hardware level using FPGAs that ingest a stream of samples ($\sim 6Gbps$) and output $\sim$ 200Mbps for each of the 30 available independent beams [35]. Therefore, delays introduced by this functionality will be in the order of microseconds, thus allowing us to idealize the performance of this part of the system.

The receiving functionality is primarily equivalent to demodulation and frame synchronization, but can also include despreading in the case of a multiple access service. The bandwidth efficiency is characterized by (1) the quantization levels used during the sampling process, (2) the modulation levels ($b \frac{bits}{symbol}$) and (3) the number of samples available per symbol which, in turn, is primarily dependent on the symbol duration T_s. Since we do not know the exact modulations that future users will utilize, we idealize the receiving functionality by assuming that the receiver can always successfully demodulate the space-to-ground signal and therefore its output is equal to the coded data rate sent by the spacecraft. Furthermore, the only significant delay comes the original frame synchronization process and will therefore by negligible for high data rate streams but can be significant for low data rate telemetry and commanding streams.

Reference [9] is used to estimate the effect of the decoding functionality. First, three complete frames are assumed to be needed in order for the decoder to be properly synchronized. Furthermore, it is assumed that half of future mission will utilize the 7/8 low density parity check (LDPC) codes, while the rest will still use current concatenated codes. As a result, the average coding efficiency for a space frame will be 0.58, thus indicating that almost 50% of the received bits are redundancy included in order to provide forward error correction capability. Additionally, the routing functionality is idealized and only accounts for appending the destination address to the packets generated from the decoded frames. As a first approximation, we assume this packet header to be negligible (e.g. a typical Internet packet contains 20B of header information for 1.5kB of payload data on average), thus yielding a bandwidth efficiency of approximately 1. Moreover, we assume that computing the appropriate route to the end user is part of the terrestrial network functionality and therefore its effects are included in the 120 msec delay from reference [25].

Finally, the store and forwarding functionality is modeled based on the 2-state Markov leveling scheme provided in reference [37]. The goal of the leveling scheme is to provide a coarse-grain bandwidth estimation mechanism to assess the impact of data product with different latency requirements over the same network. It allows for data to be trickled through the system so that all mission requirements are satisfied, while allowing some information to be delayed in the network in favor of high priority data. Note that reference [37] validates the bandwidth estimates of the leveling scheme against direct simulation of the entire protocol stack and finds minimal errors. Similar results are reported in reference [39] for a similar mechanism to model the TCP protocol.

3. SIMULATION ENVIRONMENT

This section describes the simulation environment developed to quantitatively assess the impact of real-time space communication services in the networks that support them. It is divided in two main parts: First, the simulation engine is described. Then, a simplified case study that exemplify how it works is presented.

ArchNet Description

ArchNet [40] is an simulation tool based on Python [41], SimPy [42] and PyQt [43]. It provides an integrated environment to define a network scenario, estimate the network bandwidth requirements and analyze the simulation results. The simulation engine is primarily based on SimPy, an open-source process-oriented discrete-event simulation library for Python. SimPy automatically handles the prioritization and synchronization of events in during the simulation and provides a well-documented API to extend its functionality. Additionally, ArchNet provides an intuitive graphical interface developed with the PyQt framework that can be used to

Table 2: **Ground Segment Functionality Characterization**

Functionality	BW Efficiency E_i	Induced Delay D_i	Observations
RF/IF Functionality	-	0 sec	-
Sampling	$f(f_c, B)$	0 sec	See equation 2
Beamforming	1	0 sec	From reference [35]
Receiving	$\approx \frac{1}{Q} \cdot f(T_s) \cdot b$	$\approx 1 \cdot \frac{4500}{R_3}$ sec	From references [1] and [9]
Decoding	≈ 0.58	$\approx 3 \cdot \frac{4500}{R_3}$ sec	From reference [9]
Generating Frame	≈ 1.15	$= \frac{10240}{R_4}$ sec	From references [9], [36]
Routing	1	0 sec	Appending destination address
Store & Forwarding	f(leveling scheme)	f(leveling scheme)	From reference [37]
Terrestrial Network	1	$\leq$ 120 msec	From reference [25]

Table 3: **Space Network Sampling**

SN Service	$\mathbf{f_c}$[MHz]	B[MHz]	m	$\mathbf{f_s}[\frac{\text{Msamples}}{\text{sec}}]$	$\mathbf{Q}[\frac{\text{bit}}{\text{sample}}]$	$\mathbf{R_1}$[Mbps] (estimated)	$\mathbf{R_1}$[Mbps] (real)
S-band Single Access	370	20	16	45	16	720	804.66
Ku-band Single Access	370	225	-	1190	4	4760	4710
Ka-band Single Access	1200	650	-	3700	4	14800	14140
S-band Multiple Access	370	6	60	12.23	16	195.73	201.17

visualize both the network architecture and the simulation results. Additionally, a Python console synchronized with the simulation engine is available for high-level complex data processing and plot generation.

ArchNet defines an architecture based on the problem formulation from section 2. The five canonical nodes from table 1 are available to compose a network architecture, defined through an external XML [44] document that contains a list of nodes, their type, and the connections between them. This XML file is ingested by ArchNet, along with another file where the network traffic is defined. This second input file is also external to ArchNet and comes from a scheduling system that assigns contacts between network users and network nodes. Once both inputs are available, ArchNet simulates the network performance and records the data rate and data storage requirements at all network connections and nodes, and then provides a set of menus to visualize the results and analyze their statistics.

ArchNet Example

Figure 4 provides a simplified example to illustrate how ArchNet works. The network is simply composed of two nodes, a ground station (N_1) and a mission operations center (N_2). Data sent from a satellite to the ground station is sampled, demodulated, decoded and routed to the MOC. When possible, the data is also stored and forwarded to the MOC at a lower data rate to reduce the bandwidth requirements in the connection between both entities.

Assume that the ground station N_1 has multiple antennas and can support multiple spacecraft at the same time, all of which are controlled by the same MOC. In particular, 4 spacecraft communicate with N_1, each one with a pass that starts and ends at different times and has a given data rate. These passes are store and forwarded independently at the ground station

Figure 4: **ArchNet Example**

resulting in the 4 passes depicted in figure 5. In order to obtain a total estimate for the N_1-N_2 connection bandwidth, ArchNet's simulation engine keeps track of starting and ending events for all passes and then aggregates them to produce the green dotted line in figure 5. Note that aggregating even four passes results in a non-trivial bandwidth profile over time that would be hard to estimate without a discrete simulation engine. However, because ArchNet simulates flows of data at the pass level, the execution time for 15 days of operations and 800 passes per day is reduced to less than a minute, thus making it suitable for high-level architectural studies.

Overall Research Approach

Based on the simulation environment presented in section 3, we now describe the overall research approach to assess the impact of real-time communication services on the Space Network ground segment:

1. Based on current operational traffic patterns, we simulate the bandwidth requirements across the different elements of the Space Network and define an initial baseline scenario.

2. We estimate the types and missions that will utilize the Space Network in 2030. For each of them, we characterize the quality of service (QoS) they request as a combination of data volume (or data rate) and latency.

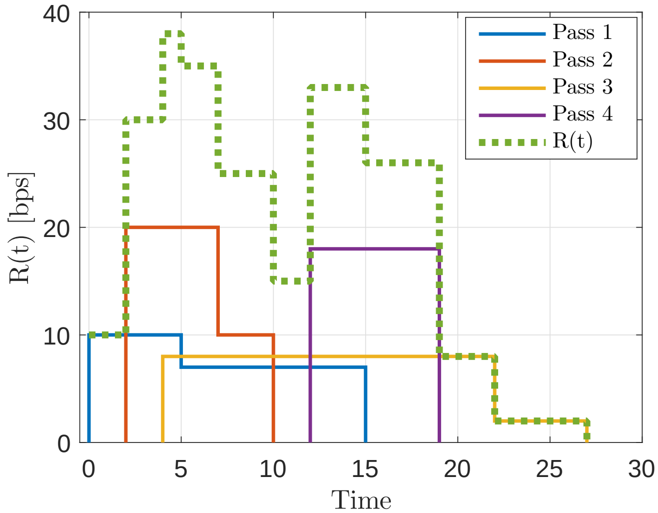


Figure 5: **Bandwidth Estimation Example**

3. We then use the tools presented in references [3] and [4] to transform this set of users into traffic patterns for the Space Network.

4. Finally, we simulate the operation of the system for a given time span (typically between a day and a month) in order to obtain the bandwidth and storage profiles that will be needed in the different elements of the system. Furthermore, we compare these values with the current status of the network to identify potential bottlenecks and suggest infrastructure improvements to alleviate them.

4. Baseline Scenario

Traffic Generation

The validation case study is based upon realistic traffic patterns obtained from the TDRSS scheduling system, from 05/08/2012 to 05/20/2012 (see reference [45] for an in depth analysis). Each day, an average of $\sim$ 800 contacts were granted by the network, resulting in $30-40$ Tbit returned to Earth, as well as approximately 400 hours of scheduled service. These include both forward and return services, where the forward contacts are used exclusively for commanding the spacecraft. Since the amount of data they generate is at least one order of magnitude smaller than that of the return services [45], they will be omitted in all performed simulations.

Furthermore, the schedule obtained through the SN's system is structured according to the services that the network provides (e.g. SSAR, KuSAR, KaSAR [1]) and, as a result, must be adapted to the service model presented in section 2. Next, we summarize the set of heuristics used for this adaption process:

- A single band contact contains 10% engineering data and 90% science data.
- For dual band contacts, the higher frequency band always carries science data while the lower frequency band carries engineering data (see rule 1 as an example).
- For dual contacts using the SSAR and SMAR service, the former carries all science data and the latter carries engineering data.
- Science data is split into quick-look and bulk science data according to 5%/95% relative fractions. Noteworthy exceptions to this rule are NASA's Earth Observation System (Terra, Aqua and Aura). In that case, all science data is assumed to be quick-look since it is delivered through the LANCE system.
- The ISS is assumed to deliver multiplexed engineering telemetry as well as voice/video through its low data rate contacts.
- Satellites are divided into three nominal regions, each one supported by one ground stations. The Atlantic Ocean Region (AOR) and Pacific Ocean Region (POR) have two satellites, while the Indian Ocean Region (IOR) only comprises one TDRS satellite.

Rule 1: **Simultaneous SSAR and KuSAR Contact**

```
(defrule MAIN::SSAR-and-KuSAR
   (exists SNEVENT (service=SSAR) (time=?t1)
                   (duration=?d1) (datarate=?dr1))
   (exists SNEVENT (service=KuSAR) (time=?t1)
                   (duration=?d1) (datarate=?dr2))
   =>
   (assert EVENT (service=eng-data) (time=?t1)
                 (duration=?d1) (datarate=?dr1))
   (assert EVENT (service=science-data) (time=?t1)
                 (duration=?d1) (datarate=?dr2))
)
```

Current Space Network Architecture

Figure 6 depicts the current Space Network architecture as a function of the canonical nodes presented in table 1. Each TDRSS region is connected to a ground station that performs all signal processing functionality required to recover the original information sent by the spacecraft. Three ground stations are available, the White Sands Ground Terminal (WSGT), the Second TDRSS Ground Terminal (STGT) and the Guam Remote Ground Terminal (GRGT). The first two are located in the White Sands complex close to the Network Control Center Data System (NCCDS) and therefore data across them is transmitted through an inexpensive local-area network (LAN). In contrast, data from GRGT is returned to the the NCCDS through a costly WAN connection.

Note that the depicted architecture assumes that the TDRSS satellites only perform RF and IF functionality. In reality the second generation TDRSS placed the beamforming functionality on-board the satellite, but this approach was later replaced by the current ground-based beamforming strategy in order to facilitate the deployment of improved signal processing technology as it becomes available. Since the beamforming functionality has a minor impact in the signal delay, this modeling error will not affect the obtained results and is therefore deemed insignificant.

Finally, data reaching the NCCDS from GRGT is immediately transmitted to the end user as it was originally stored and forwarded at the ground station node. In other words, the NCCDS acts a router for this fraction of the data. In contrast, data downlinked from the WSGT and STGT has only been received and therefore is assumed to be stored and forwarded in the NCCDS to minimize the bandwidth cost from the White Sands complex to the other NASA centers. This duality is notionally depicted in figure 6 by node colored both green and orange.

Baseline Scenario Results

Figure 7 presents the required capacity for the different links in the current Space Network architecture given the baseline

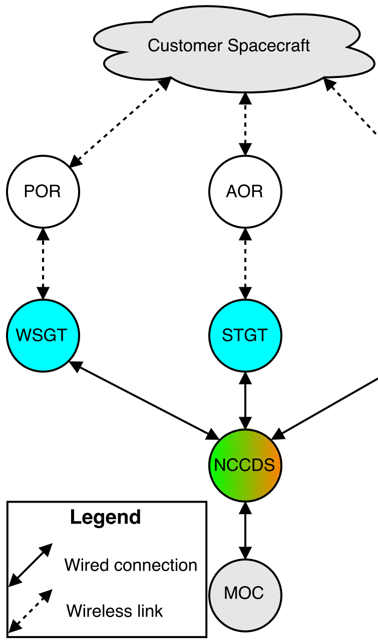


Figure 6: **Current SN Architecture**

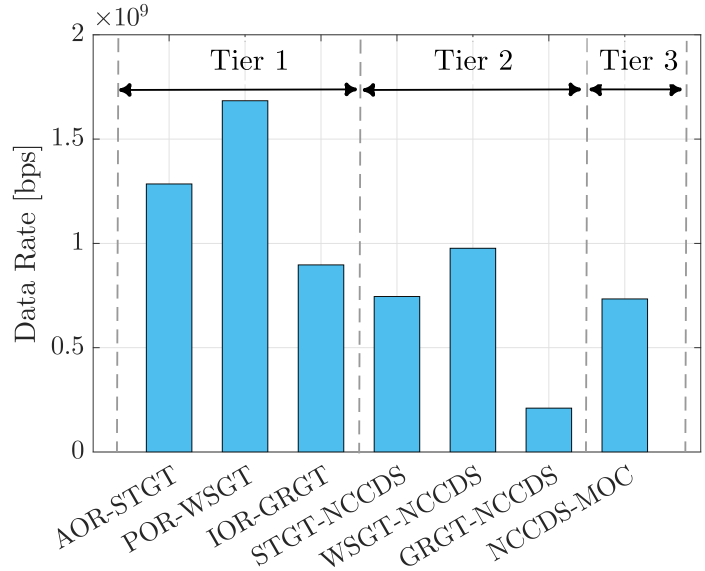


Figure 7: **Baseline SN Capacity**

demand. Links have been grouped in three tiers: Tier 1 corresponds to all space-to-ground links. Note that the required data rate are, in some cases, greater than the current capacity provided by the Ku-band link between a TDRS satellite and the ground. This is due to the fact that each region pools together the traffic of more than one satellite.

Tier 2 corresponds to all links between the ground stations and the network control center. In this case, the bandwidth requirement of the input and output lines for WSGT and STGT differs mostly due to the inefficiency introduced by the forward error correction (FEC) mechanism in the space-to-ground link. In contrast, the GRGT output bandwidth is largely inferior to the amount of data downlinked instantaneously thanks to the smoothing effect of the store and forward functionality (see figure 8).

Finally, tier 3 corresponds to all links between the NCCDS and the rest of the NASA networks (e.g. connections to JPL or GSFC). Once again, it can be observed that the store and forward functionality implemented at the NCCDS is able so successfully reduce the total required bandwidth and therefore minimize the cost of this part of the network.

5. Supporting Real-time Communication Services in the Space Network

Traffic Generation

In order to simulate real-time traffic for the Space Network we make the following assumptions:

- In 2025 approximately the same number of missions will utilize the system.
- Scientific missions are classified in three categories: High, moderate and low data volume (see reference [3]). The vast majority of missions belong to the two latter categories, while only three missions belong to the former.
- High data volume missions increase the amount of data returned to Earth by one order of magnitude. This is representative from transitioning from Terra ($\sim$ 200Gbit/day) to DESDyni ($\sim$ 40Tbit/day).
- All scientific data is considered quick-look science and therefore has to be returned in less than one hour and thirty minutes.
- The ISS still drives the return of voice and video data to the ground, albeit sporadic support to other human activities is also provided (e.g. Orion capsule).

Table 4: **Network Requirements**

Scenario	Daily Data Volume [GB]	Average Latency [h]
Baseline	4714	6.2
Data-Driven	23080	6.2
Real-Time	23080	1.3

Based on this traffic pattern rules, table 4 details the overall network requirements as the total data volume to downlink per day and the average latency of offered services. Three scenarios have been defined: Baseline corresponds to the current state of the system (see section 4). Data-Driven refers to a futuristic scenario where the missions' data volume has significantly increased but the latency requirement has not. Finally, the real-time scenario depicts a situation in which both the data volume and latency requirements have evolved towards more stringent values. Note, for instance, that the total daily data volume increased by a factor of approximately 5 in the two futuristic scenario. Furthermore, the average latency has been reduced by almost 80% in the last one.

Current Architecture

Figure 9 presents the estimated capacity required at each SN link for the three proposed scenarios (see table 4). In this

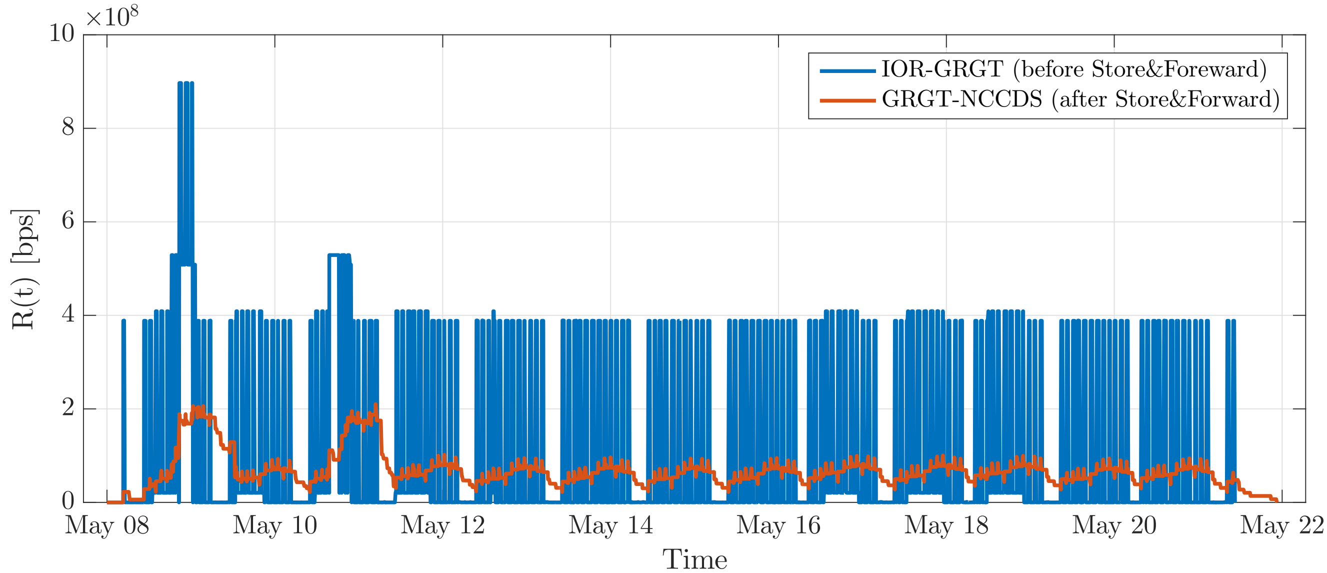


Figure 8: **Smoothing Effect of Store and Forward at GRGT node**

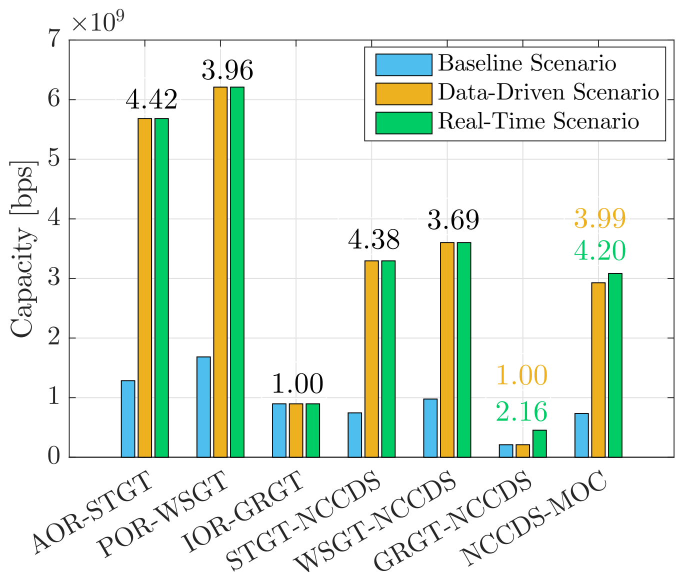


Figure 9: **SN Capacity under Different Scenarios**

case, capacity is computed as the maximum data rate in a given link over time:

$$C = \max R(t) \quad (3)$$

Furthermore, for each link we also provide the multiplicative factor

$$\gamma = \frac{C_{scenario}}{C_{baseline}} \quad (4)$$

that summarizes the increase in capacity required in order to successfully satisfy the network data volume and average latency requirements. When γ is color-coded in black the same value is applicable to both the data-driven and real-time scenario. Otherwise, γ has been colored according to each specific scenario.

We first note that the system capacity is largely driven by the network daily data volume requirement and scales approximately linearly with that parameter. In other words, since the data-driven and real-time scenarios have to return 4 more data than the baseline scenario, the capacity of most links has to also be increased by approximately this factor. Notorious exceptions are the IOR-GRGT and GRGT-NCCDS links. For the former, it has been assumed that high demanding missions without continuous coverage requirements (such as DESDyni) schedule all their contacts through ground stations located in continental US so as to minimize the cost the of WAN connection between Guam and the US. On the other hand, the increase in data volume for the data driven scenario has no effect on the cost of the intercontinental line GRGT-NCCDS, while the capacity requirement more than doubles in the real-time scenario.

While this might seem a minor problem from a capacity perspective, it actually translates into major affordability problems for the network. In particular, conversations with Space Network managers have indicated that the cost of the GRGT-NCCDS is approximately 3 times as high as that of leasing bandwidth over US continental lines. Therefore, the cost of doubling the capacity in the GRGT-NCCDS line results in at least a 600% cost increase on that part of the ground segment. In contrast, the cost of the other continental lines will increase at most 400%, specially if reductions in the cost per bit ($\$/bps$) are obtained when leasing high data rate lines.

Space-based Store and Forward Architecture

Up to this point we have analyzed the ability of the SN architecture in order to satisfy future user demand. Our findings indicate that the required network capacity with the current architecture scales approximately linearly with the returned daily data volume. Furthermore, significant cost increases should be expected if users transition to a real-time demand model where the store and forward functionality becomes less effective.

In order to mitigate these problems, in this section we analyze the suitability of a network architecture where the TDRS

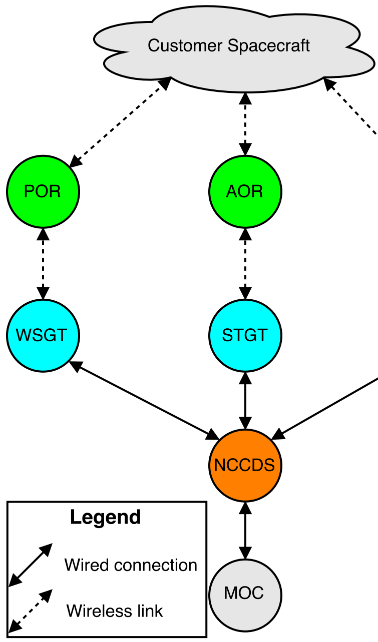


Figure 10: **Future SN Architecture**

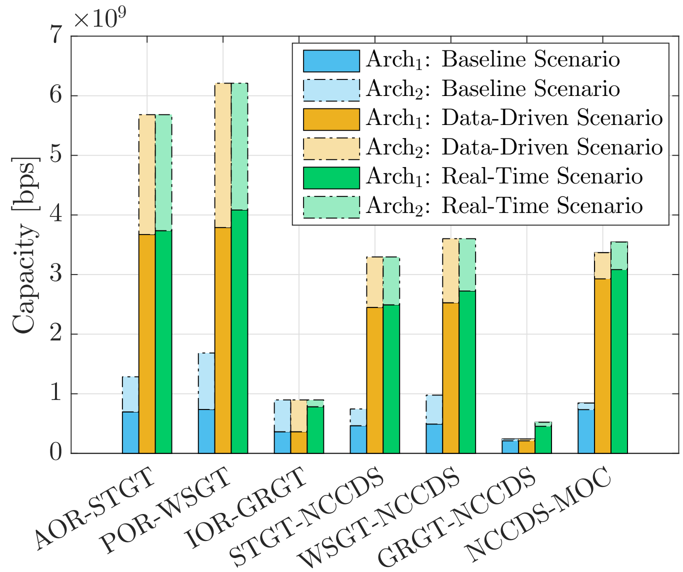


Figure 11: **Store and Forward Architecture Capacity**

satellites have enough processing capability to successfully modulate the space-to-ground link according to the user's latency requirements. Therefore, we have transitioned towards a space-based store and forward architecture (see figure 10) where the ground segment acts as mere receive and forward node. Routing for WGST and STGT is performed at the NCCDS since they are all under the same LAN, while GRGT also provides the routing functionality so that data destined for users in Asia and Europe could potentially be routed directly without clogging the Guam to White Sands line.

Figure 11 presents the capacity required to support users from the baseline, data-driven and real-time scenarios for a space-based store and forward TDRSS architecture - also referred to as $Arch_1$. Each link in the SN is depicted using a solid bar plot that indicates the total capacity required. Stacked on top, a transparent bar indicates the extra capacity that is required given the current architecture, referred to as $Arch_2$. In other words, the transparent area of each bar plot indicates the capacity savings incurred by transitioning from the current SN architecture to a more capable space-based store and forward system.

It can be observed that major capacity savings are obtained through the new proposed architecture regardless of the scenario under consideration. These savings are specially notorious in the space to ground links, with a 15% and 33% reduction in the AOR-STGT and POR-WSGT links respectively. On the other hand, it can also be observed that the backbone network (i.e. GRGT-NCCDS and NCCDS-MOC) requirements have increased by a small percentage ($< 1\%$). Indeed, links that before where perfectly optimized because their origin nodes implemented the store and forward functionality are now just routers that cannot store data anymore. However, this effect is considered minimal in comparison to the large capacity savings obtained in the space-to-ground links.

On the other hand, we can also analyze the system's sensitivity to limited latency requirement violation. In particular, let us assume that in the baseline and data-driven scenario up to 10% of the data is allowed to slightly violate the latency requirements. This is plausible because bulk data has already up to 12 hours worth of delay and therefore it is unlikely that users would be heavily dissatisfied if slightly more delay was introduced (e.g. 13 or 14 hours). In contrast, assume that data delivery across the network can never exceed the latency constraint of 1 hour and 30 minutes for the real-time scenario (see for instance reference [46], where the quick-look science latency requirement of 1 hour and 30 minutes is already conservative and should be met with 100% certainty).

Then, the capacity needed in the different links can be computed based on the 90% percentile of $R(t)$ for the baseline and data-driven scenario, while in the real-time scenario it is estimated as $C = \max R(t)$. Figure 12 presents the obtained results. It can be observed that the required network capacity is almost twice as large in the real-time scenario than in the data-driven scenario, thus indicating that both latency and data volume should be considered as fundamental factors when sizing the network. Furthermore, we note that the proposed space-based store and forward architecture can result in major capacity savings for the SN as compared to the current architecture, thus making it more robust towards demand growth and uncertainty in both data volume and date latency requirements.

That being said, several factors might also hinder the viability of the proposed store and forward processing architecture. For instance, placing processing capability at geosynchronous orbit is expensive since it increase the mass and power requirements for the spacecraft bus and introduces new failure risk factors that need to be studied and mitigated. Therefore, it is clear that the trade-off between store and forward functionality in the space and ground segment should be studied. In particular, robustness against network capacity and latency requirement variation should be analyzed in order to identify optimal architectures that can easily adapt to network requirement changes. Factors such as inflexibility and reliability for the space-based store and forward architecture should also be considered, as the impossibility to modify or

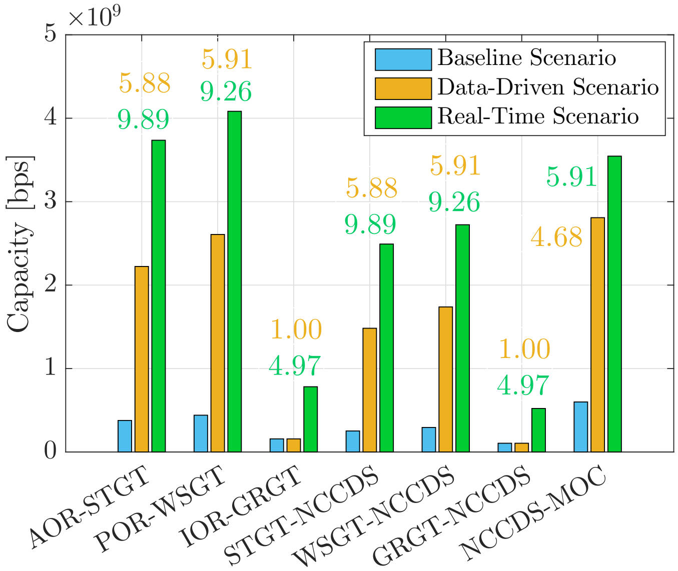


Figure 12: **Latency Sensitivity**

fix satellites once placed on orbit might outweigh the capacity savings obtained through this futuristic architecture.

6. Conclusions

Summary

This paper has focused on assessing the impact of real-time communication services in NASA's Space Network. To that end, we have first formulated the problem as a function of the network's architecture, where architecture is mainly defined by allocating processing functionality (e.g. sampling, demodulating, store and forwarding) across different nodes in the network. Then, we have defined several canonical services that the network should offer in order to successfully support a wide variety of robotic and human exploration activities, and we have characterized their current latency requirements.

Based on this modeling framework, we have initially studied the capacity requirements for the Space Network as a function of the current data volume and latency requirements. Results have indicated that the network can largely benefit from the relaxed latency requirement of most data in order to reduce its WAN bandwidth and cost.

Next, we have studied this same network architecture when subject to both an increase in the amount of data it should return, as well as a significant reduction in the average latency requirement. Results indicate that the combined effect of these two requirements largely penalizes the ability of the network to successfully support its customers. In order to mitigate this problem, we have studied the suitability of a future Space Network architecture where store and foreward processing functionality is placed on-board the TDRSS satellites. Results have indicated that significant capacity saving can be obtained by transitioning towards this architecture, especially in the latency-insensitive scenarios. Finally, we have proven that the proliferation of real-time services in the network largely reduces the savings currently obtained through the store and forward functionality, thus indicating that the combined effect of increase in network daily data volume and decrease in average network latency result in severe capacity limitations for the current system.

Future work

Different areas of research can be envisioned for the future. On one hand, the trade-off between store and forward functionality in the space and ground segment should be further explored taking into consideration both the system cost, robustness and riskiness. Robustness should be used as a proxy to measure the ability of the network to cope with demand changes given that space assets deployed cannot be easily and inexpensively replaced. Similarly, riskiness should be quantify the capacity loss due to failure of a system node and the cost of recovering to nominal performance.

On the other hand, similar analyzes should be studied in the context of the other two NASA networks, namely the Near Earth Network and the Deep Space Network. For the latter, the necessary inputs for the different defined scenarios are currently being compiled and the resulting studies are expected to provide insight into the required architecture and capacity for a network that can successfully support human exploration activities at Mars.

Acknowledgments

The authors would like to acknowledge Kar-Ming Cheung, Carlyn-Ann Lee and Kristy Tran from NASA's Jet Propulsion Laboratory for their feedback and advice while conducting this research.

References

[1] G. S. F. Center, "Space network (sn) user's guide," National Aeronautics and Space Administration, Tech. Rep., 2007.

[2] J. S. Schier, J. J. Rush, W. D. Williams, and P. Vrotsos, "Space communication architecture supporting exploration and science: Plans and studies for 2010-2030," in *1st Space Exploration Conference: Continuing the Voyage of Discovery*, vol. 30, 2005, pp. 1–33.

[3] M. Sanchez, I. del Portillo, B. G. Cameron, and E. F. Crawley, "Architecting space communication networks under mission demand uncertainty," in *2015 IEEE Aerospace Conference*. Big Sky, Montana: Institute of Electrical and Electronics Engineers, March 7-14 2015.

[4] M. Sanchez, D. Selva, B. G. Cameron, E. F. Crawley, B. Seery, and A. Seas, "Results of the mit space communication and navigation architecture study," in *2014 IEEE Aerospace Conference*. Big Sky, Montana: Institute of Electrical and Electronics Engineers, March 1-8 2014.

[5] S. S. Board *et al.*, *Earth Science and Applications from Space:: National Imperatives for the Next Decade and Beyond*. National Academies Press, 2007.

[6] G. S. Wojcik, H. L. Szymczak, R. J. Alliss, R. P. Link, M. E. Craddock, and M. L. Mason, "Deep-space to ground laser communications in a cloudy world," in *Optics & Photonics 2005*. International Society for Optics and Photonics, 2005, pp. 589 203–589 203.

[7] R. Allis and B. Felton, "The mitigation of atmospheric impacts on free-space optical communications," *Proceedings of International Conference on Space Optical*

Systems and Applications, 2012.

[8] S. S. SSC, "Ground station networks vs. geo relay satellites for polar orbiting satellites data download," 2012.

[9] K.-M. Cheung and D. S. Abraham, "End-to-end traffic flow modeling of the integrated scan network," *the Interplanetary Network Progress Report*, vol. 42, p. 189, 2012.

[10] M. E. e. a. Brown, "Nasa earth science division study on data latency needs and requirements," Tech. Rep., 2013.

[11] "Edos high-rate data capture and delivery of low-latency hyspiri level-zero data," https://hyspiri.jpl.nasa.gov/downloads/2013_Symposium/day2_am/EDOS-HyspIRI_Worshop_Presentation_5-30-2013.pdf, accessed: 07/26/2015.

[12] D. Lester and H. Thronson, "Human space exploration and human spaceflight: Latency and the cognitive scale of the universe," *Space Policy*, vol. 27, no. 2, pp. 89–93, 2011.

[13] ESA-ESTEC, "Ground systems and operations telemetry and telecommand packet utilization," European Co-operation for Space Standardization, Tech. Rep., 2003.

[14] K. Bhasin and J. Hayden, "Developing architectures and technologies for an evolvable nasa space communication infrastructure," in *AIAA ICSSC*, 2004.

[15] T. Haidegger and Z. Benyo, "Surgical robotic support for long duration space missions," *Acta Astronautica*, vol. 63, no. 7, pp. 996–1005, 2008.

[16] G. A. Horsham, G. R. Schmidt, and J. H. Gilland, *Establishing a Robotic, LEO-to-GEO Satellite Servicing Infrastructure as an Economic Foundation for Exploration*. National Aeronautics and Space Administration, Glenn Research Center, 2010.

[17] W. S. Kim and A. K. Bejczy, "Demonstration of a high-fidelity predictive/preview display technique for telerobotic servicing in space," *Robotics and Automation, IEEE Transactions on*, vol. 9, no. 5, pp. 698–702, 1993.

[18] H. A. Thronson, D. Akin, J. Grunsfeld, and D. Lester, "The evolution and promise of robotic in-space servicing," in *Proceedings of the AIAA SPACE 2009 Conference & Exposition. California*, 2009, pp. 1–6.

[19] D. Lester and H. Thronson, "Low-latency lunar surface telerobotics from earth-moon libration points," *American Institute of Aeronautics and Astronautics Space*, 2011.

[20] K. M. Tsui, M. Desai, H. Yanco, C. Uhlik *et al.*, "Exploring use cases for telepresence robots," in *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*. IEEE, 2011, pp. 11–18.

[21] "Esd34. system architecture," http://ocw.mit.edu/courses/engineering-systems-division/esd-34-system-architecture-january-iap-2007/, accessed: 04/14/2015.

[22] I. O. A. Group, "The consultative committee for space data systems," Tech. Rep., 2014.

[23] L. Kleinrock, "The latency/bandwidth tradeoff in gigabit networks," *IEEE Communications Magazine*, vol. 30, no. 4, pp. 36–40, 1992.

[24] T. De Cola and M. Marchese, "Performance analysis of data transfer protocols over space communications," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 41, no. 4, pp. 1200–1223, 2005.

[25] "Communications service office," https://cso.nasa.gov/content/mission-routed-data, accessed: 10/17/2015.

[26] M. Sanchez Net, I. d. Portillo, B. Cameron, E. F. Crawley, and D. Selva, "Integrated tradespace analysis of space network architectures," *Journal of Aerospace Information Systems*, vol. 12, no. 8, pp. 564–578, 2015.

[27] D. Selva, "Rule-based system architecting of earth observation satellite systems," Ph.D. dissertation, Massachusetts Institute of Technology Department of Aeronautics and Astronautics, Cambridge, Massachusetts, June 2012.

[28] D. P. Bertsekas, R. G. Gallager, and P. Humblet, *Data networks*. Prentice-Hall International New Jersey, 1992, vol. 2.

[29] J.-Y. Le Boudec and P. Thiran, *Network calculus: a theory of deterministic queuing systems for the internet*. Springer Science & Business Media, 2001, vol. 2050.

[30] A. K. Sinha, *Methodologies and analyses of broadband access network traffic*. ProQuest, 2007.

[31] J. P. Laboratory, "Dsn telecommunications link design handbook," California Institute of Technology, Tech. Rep., 2012.

[32] G. S. F. Center, "Near earth network (nen) user's guide," National Aeronautics and Space Administration, Tech. Rep., 2010.

[33] S. Communications and N. Program, "Space communications and navigation program (scan) service catalog," National Aeronautics and Astronautics Administration, NASA Headquarters. Washington, D.C., Tech. Rep. 1, 2011.

[34] K. J. Murphy, "Land and atmosphere near-real-time capability for earth observing system," 2011.

[35] K. Hogie, E. Criscuolo, A. Dissanayake, B. Flanders, H. Safavi, and J. Lubelczyk, "Tdrss demand access system augmentation," in *Aerospace Conference, 2015 IEEE*. IEEE, 2015, pp. 1–10.

[36] F. Flentge, "Study on cfdp and dtn architectures for esa space missions," in *The Third International Conference on Advances in Satellite and Space Communications*. Houston, Texas: Space Commerce Conference and Exposition, 2011.

[37] K.-M. Cheung, E. H. Jennings, and J. S. Sergui, "Coarse-grain bandwidth estimation scheme for large-scale network," 2013.

[38] R. G. Lyons, *Understanding digital signal processing*. Pearson Education, 2010.

[39] S. Gadde, J. Chase, and A. Vahdat, "Coarse-grained network simulation for wide-area distributed systems," in *Communication Networks and Distributed Systems Modeling and Simulation Conference*, 2002.

[40] "Archnet documentation," http://www.mit.edu/~msnet/ArchNet/index.html, accessed: 01/06/2015.

[41] J. M. Zelle, *Python programming: an introduction to computer science*. Franklin, Beedle & Associates, Inc., 2004.

[42] N. Matloff, "Introduction to discrete-event simulation and the simpy language," *Davis, CA. Dept of Computer Science. University of California at Davis. Retrieved on August*, vol. 2, p. 2009, 2008.

[43] M. Summerfield, *Rapid GUI programming with Python and Qt: the definitive guide to PyQt programming*. Pearson Education, 2007.

[44] D. Megginson, *Structuring XML documents*. Prentice-Hall, Inc., 1998.

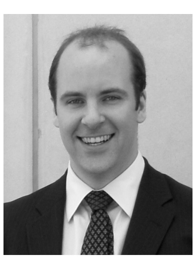
[45] M. Sanchez, "Architecting space communications networks," Master's thesis, Massachusetts Institute of Technology Department of Aeronautics and Astronautics, Cambridge, Massachusetts, June 2014.

[46] J. L. Duda, J. Mulligan, J. Valenti, and M. Wenkel, "Latency features of safetynet ground systems architecture for the national polar-orbiting operational environmental satellite system (npoess)," in *Fourth International Asia-Pacific Environmental Remote Sensing Symposium 2004: Remote Sensing of the Atmosphere, Ocean, Environment, and Space*. International Society for Optics and Photonics, 2005, pp. 233–241.

BIOGRAPHY

Marc Sanchez Net *is currently a second year M.S. student in the department of Aeronautics and Astronautics at MIT. His research interests include machine learning algorithms and rule-based expert systems, and their suitability to the fields of system engineering and space communication networks. Prior to his work at MIT, Marc interned at Sener Ingenieria y Sistemas as a part of the team that develops and maintains FORAN, a CAD/CAM/CAE commercial software for shipbuilding. Marc received his degrees in both Industrial engineering and Telecommunications engineering in 2012 from Universitat Politecnica de Catalunya, Barcelona.*

Iñigo del Portillo *is a graduate student in the department of Aeronautics and Astronautics at MIT. His research interests include optical communications for space-based networks and small satellites communications. Iñigo received his degrees in Industrial Engineering, Electronics Engineering and Telecommunications Engineering in 2014 from Universitat Politecnica de Catalunya, Barcelona.*



Dr. Bruce Cameron *is a Lecturer in Engineering Systems at MIT and a consultant on platform strategies. At MIT, Dr. Cameron ran the MIT Commonality study, a 16 firm investigation of platforming returns. Dr. Cameron's current clients include Fortune 500 firms in high tech, aerospace, transportation, and consumer goods. Prior to MIT, Bruce worked as an engagement manager at a management consultancy and as a system engineer at MDA Space Systems, and has built hardware currently in orbit. Dr. Cameron received his undergraduate degree from the University of Toronto, and graduate degrees from MIT.*

Dr. Edward F. Crawley *received an Sc.D. in Aerospace Structures from MIT in 1981. His early research interests centered on structural dynamics, aeroelasticity, and the development of actively controlled and intelligent structures. Recently, Dr. Crawleys research has focused on the domain of the architecture and design of complex systems. From 1996 to 2003 he served as the Department Head of Aeronautics and Astronautics at MIT, leading the strategic realignment of the department. Dr. Crawley is a Fellow of the AIAA and the Royal Aeronautical Society (UK), and is a member of three national academies of engineering. He is the author of numerous journal publications in the AIAA Journal, the ASME Journal, the Journal of Composite Materials, and Acta Astronautica. He received the NASA Public Service Medal. Recently, Prof Crawley was one of the ten members of the presidential committee led by Norman Augustine to study the future of human spaceflight in the US.*