SPACOMM 2017

The Ninth International Conference on Advances in Satellite and Space Communications

ISBN: 978-1-61208-545-6

April 23 - 27, 2017

Venice, Italy

SPACOMM 2017 Editors

Timothy Pham, Jet Propulsion Laboratory/ California Institute of Technology, USA

SPACOMM 2017

Forward

The Ninth International Conference on Advances in Satellite and Space Communications (SPACOMM 2017), held between April 23-27, 2017 in Venice, Italy, continued a series of special events attempting to evaluate the state of the art in academia and industry on the satellite, radar, and antennas based communications, bringing together scientists and practitioners with challenging issues, achievements ,and lessons learnt.

Significant efforts have been allotted to design and deploy global navigation satellite communications systems. Satellite navigation technologies, applications, and services still experience challenges related to signal processing, security, performance, and accuracy. Theories and practices on system-in-package RF design techniques, filters, passive circuits, microwaves, frequency handling, radars, antennas, and radio communications and radio waves propagation have been implemented. Services based on their use are now available, especially those for global positioning and navigation. For example, it is critical to identify the location of targets or the direction of arrival of any signal for civilians or on-purpose applications; smarts antennas and advanced active filters are playing a crucial role. Also progress has been made for transmission strategies; multi antenna systems can be used to increase the transmission speed without need for more bandwidth or power. Special techniques and strategies have been developed and implemented in electronic warfare target location systems.

The conference had the following tracks:

- Space communications services
- Satellite and space communications

We take here the opportunity to warmly thank all the members of the SPACOMM 2017 technical program committee, as well as all the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to SPACOMM 2017. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We also gratefully thank the members of the SPACOMM 2017 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that SPACOMM 2017 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the area of satellite and space communications. We also hope that Venice, Italy provided a pleasant environment during the conference and everyone saved some time to enjoy the unique charm of the city.

SPACOMM 2017 Committee

SPACOMM Steering Committee
Timothy T. Pham, Jet Propulsion Laboratory - California Institute of Technology, USA
Stelios Papaharalabos, National Centre for Scientific Research "Demokritos", Greece
Cathryn Peoples, Queen Mary University of London, UK

SPACOMM Industry/Research Advisory Committee
Michael Sauer, Corning Cable Systems, USA
Vittorio Dainelli, Rheinmetall Italia S.p.A., Italy
Brian Niehoefer, TÜV Informationstechnik GmbH, Germany

SPACOMM 2017
Committee

SPACOMM Steering Committee

Timothy T. Pham, Jet Propulsion Laboratory - California Institute of Technology, USA
Stelios Papaharalabos, National Centre for Scientific Research "Demokritos", Greece
Cathryn Peoples, Queen Mary University of London, UK

SPACOMM Industry/Research Advisory Committee

Michael Sauer, Corning Cable Systems, USA
Vittorio Dainelli, Rheinmetall Italia S.p.A., Italy
Brian Niehoefer, TÜV Informationstechnik GmbH, Germany

SPACOMM 2017 Technical Program Committee

Ashfaq Ahmed, COMSATS Institute of Information Technology - Wah campus, Pakistan
Mohamed Al-Mosawi, University of Portsmouth, UK
Mark Bentum, University of Twente, Netherland
Yao-Yi Chiang, University of Southern California, USA
Vittorio Dainelli, Rheinmetall Italia S.p.A., Italy
Suleyman Eken, Kocaeli University, Turkey
Suk-Seung Hwang, Chosun University, Korea
Adil Hakeem Khan, Nation College of Engineering and Tech., Guna MP, India
Emmanouel T. Michailidis, University of Piraeus, Greece
Kresimir Malaric, University of Zagreb, Croatia
Brian Niehoefer, TUV Informationstechnik GmbH, Germany
Nele Noels, University of Gent, Belgium
Stelios Papaharalabos, National Centre for Scientific Research "Demokritos", Athens, Greece
Cathryn Peoples, Queen Mary University of London, UK
Timothy T. Pham, Jet Propulsion Laboratory - California Institute of Technology, USA
Vincent Roca, Inria research institute, France
Alexandru Rusu, University Politehnica of Bucharest, Romania
Michael Sauer, Corning Incorporated, USA
Cristian Stanciu, University Politehnica of Bucharest, Romania
Ouri Wolfson, University of Illinois, USA
Xiaozhou Yu, Northwestern Polytechnical University, China
Qiang Zhu, University of Michigan, USA

Copyright Information

Table of Contents

The 'C' Trinity

- Converged, Cross-layer and Cooperative Integration of Satellite Communication System

Su Gao, Yufei Shen, Ying Tao, Houtian Wang, Na Zhou, Zhicheng Zhou
Institute of Telecommunication Satellite,
China Academy of Space Technology,
Beijing, China
E-mail: 15712889970@163.com

***Abstract*—Application is the driving force of system evolvement. Satellite communication has seen its progression based on service provisioning capability and user experience. In this paper, we present an 'integrated' next-generation architecture, providing global users seamless connectivity, improving quality of service, and achieving synergy by robust network design. Satellite plays an important role in various aspects of our system. Techniques, such as powerful regenerative onboard processing, multi spotbeams with tunable structure, tailored frequency plan and flexible intersatellite links, etc., enable satellites with remarkable competence. Based on these, converged satellite-terrestrial network, cross-layer protocol stack and cooperative robust framework, as a trinity, together accomplish the integration of satellite communication system.**

Keywords-integration; satellite; communication; converge; cross-layer; cooperative.

I. INTRODUCTION

People appreciate the profound impact that modern wireless communication brings to us, while expecting more influential pervasion to all aspects of our daily life. Ubiquitous architecture distributing wireless services will become prominent in the next generation system [1]. It is important to design an interface for seamless connectivity while users change from terrestrial to satellite network, and vice versa. In addition, the system needs to consider and compensate deficiencies caused by various transmission environments. The protocol needs to be optimized to overcome deficiencies of the current separate structure, which is accomplished by cross-layer design. Furthermore, from the macrostructure point of view, the overall system performance needs to be promoted by applying suitable system engineering methods, on the basis of the above converged and Quality of Service (QoS)-based cross-layer design. To boost overall performance, cooperative strategies should be designed to combine information provided by different links in order to reduce data redundancies and improve transmission efficiency [2].

The remaining of this article is organized as follows. Section II portrays the system architecture we considered and illustrates converged, cross-layer and cooperative integration concepts. Section III describes integration challenges and solutions. Section IV concludes this article.

II. SYSTEM ARCHITECTURE

The system architecture is portrayed in Figure 1. Note this is a generalized system illustrating concepts in our discussion.

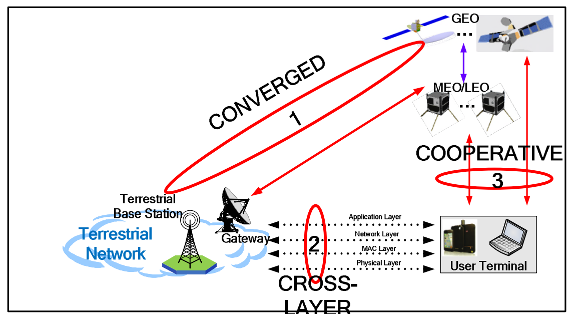


Figure 1. System Architecture

The system includes:

- Terrestrial network: well-established ground-based communication system. Base stations connect user terminals with traditional communication model. Here, the figure only shows a simplified layered framework including application layer, network layer, Media Access Control (MAC) layer and physical layer. These layers will be used in our discussion in the following sections.
- Satellite gateway: communicates with satellites via feeder links. In the figure, the gateway is shown within the terrestrial network but it does not have ground cell network communication functionality.
- Space segment: contains satellites. This framework does not target any particular satellite constellation type. The satellites are assumed to be bidirectional interactive, with transparent structure or OnBoard processing (OBP) capabilities. Satellites can provide mobile or fixed satellite communication service. Geosynchronous Earth Orbit (GEO) satellite can either carry deployable antenna for mobile users or generate C/Ku/Ka band multi-beams to produce broadband services. Medium Earth Orbit (MEO)/ Low Earth Orbit (LEO) satellite constellations may vary and have a certain design in the following sections. InterSatellite links (ISLs) are assumed.

 ISBN: 978-1-61208-545-6

- User terminals: are interactive. Terminals have access to satellite and terrestrial networks. Terminals may be fixed or mobile.

According to the analysis in the previous section, our system is designed as an integrated one. The integration is considered in three dimensions. First, to bridge the digital divide, converged integration is carried out in the interface dimension, providing seamless connectivity and effective merge between satellite and terrestrial networks. Second, to optimize QoS metrics, cross-layer integration is carried out in the protocol dimension, overcoming deficiencies of the current separate structure by joint modeling of different layers in the protocol stack. Third, to boost overall performance, cooperative integration is carried out in the link dimension, achieving synergy among transmission links for higher throughput and network robustness.

Each of the three integrations over three different dimensions solves one aspect of the system deficiencies according to user experience. They also reflect the sequence while optimizing the system: providing seamless connectivity for all users is the first step, followed by improving quality of service, and overall performance advancement brings all the users the best communication experience. These three integrations, which together form our design guidelines for satellite communication system, are a trinity.

III. Integration Challenges and Solutions

In this section, we discuss various integration issues in detail.

A. *Converged Integration*

This is the first guideline of unifying the system. The converged integration considers the interaction and interface design of satellite and terrestrial networks which is mainly reflected in radio interface design. It will be seen from our discussion that satellite design plays an important role as the satellite ability and performance affects various aspects of the converged interface.

1) *Radio Interface and Satellite Design*

To improve satellite communication capability and reduce system deployment cost, satellite design advancement plays a key role in aerospace systems. Therefore, it largely influences performance of the unified satellite-terrestrial system and its air interface design. An integrated solution should reflect satellite characteristics and modify the air interface design accordingly. For voice services, flexible transformation and routing onboard enabling the satellite to support single-hop communication among user terminals gives better service quality; for broadband access services, advanced antenna techniques can produce multiple and movable spotbeams providing users dynamic service management. All these satellite technology advancements inspire radio interface design to better satisfy user requirements.

In this subsection, we present an example that user and system requirement leads the satellite design, and how novel satellite ability affects radio interface architecture. We consider GEO mobile communication satellites carrying unfurlable antenna, which gives greater equivalent isotropic radiated power (EIRP) for ground users to have better receiving gain. This type of large diameter mesh antenna, together with the feed array, generates multiple spotbeams to achieve seamless coverage for a large number of users, where each beam can cover hundreds of square kilometers.

This new characteristic brings some design issues for mobile radio interface design. Due to variation of space environment and mechanical insensitiveness, the antenna oscillates, resulting in the satellite antenna pointing inaccurately during the oscillation lifecycle [3]. From the user experience point of view, this phenomenon mainly influences system performance and radio interface design in two aspects: synchronization and interruption. The first one affects communication system in various layers and is the basis in establishing the phone call for a single user. The second one guarantees the basic QoS during the phone call and usually measured in system level as interruption probability of the whole network. These two aspects are the most important aspects for users in the network and directly connected with users' communication experience, thus are essential in analyzing the effects of antenna oscillation to radio interface design.

2) *Interruption*

Radio link failure is the main reason of interruption. When errors occur, messages cannot be received successfully and certain system parameters are reset. Then, the radio link failure occurs and communication is interrupted [4].

Network interruption needs two conditions to happen, the location condition, where the user must be located at the edge of some beams, and the time condition, where the interruption must occur during the communication process. These two conditions are independent so can be analyzed separately and combined together as the whole interruption probability. To reduce interruption probability, two methods may be applied:

- Design channel allocation strategy according to time and frequency of antenna oscillation. The receiving C/N can not only be increased but also be more stable, thus guaranteeing sufficient and stable link budget.
- Modify terrestrial telecommunication protocol by increasing the number of re-transmissions of access channel and paging channel. This can reduce the probability of error as the probability that antenna oscillation causing the failure of multiple transmissions is reasonably low.

3) *Satellites in the Future Converged System*

It can be seen from previous discussion that radio interface design issues such as synchronization and interruption are closely related to satellite characteristics such as deployable antenna oscillation. It is important to investigate satellite antenna structure to know more about the oscillation process in various circumstances [1]. It is also important to design a tailored control strategy based on the satellite bus's ability to maintain position and attitude of the spacecraft to ensure satellite position is not changed by the oscillation.

Not all future satellites are equipped with deployable antenna. Some of them provide Ka broadband services and with multiple spotbeams. They allow tunable power level and dynamically changeable EIRP to reach balanced receiving power among users. Other characteristics include digital payload with onboard processing, advanced frequency reuse and optical intersatellite links, etc. These advanced techniques [1] make the convergence of satellite system to the mature terrestrial radio interface easier, but at the same time introduce new issues specifically applicable to satellite transmission such as channel environment and long transmission delay, which urges modifications to the terrestrial network.

B. Cross-Layer Integration

Cross-layer design is essential for wireless system design and has already been widely used in terrestrial communication systems [5]. It plays a key role in optimizing QoS metrics. For an integrated satellite communication system, the upper layer design uses the cross-layer concept. We mainly consider two aspects: cross-layer design of radio resource management and network layer routing. In the following, a generalized optimization model is given and the mechanisms to solve it are analyzed according to characteristics of relevant parameters. As an example, network layer routing optimization problem is simulated. It can be seen from the discussion that for this dimension of integration, more powerful satellites are required.

1) Optimization Model

To improve the robustness of satellite communication networks and realize better performance in a dynamically changing communication environment, cross-layer design should be carried out and models should be built to optimize the relevant cost function in the considered layer. Other layers' parameters should be included as constraints so that the optimization problem contains the whole system's impact, thus forming an integrated design solution.

The current trend toward the migration to satellite-terrestrial integrated services opens new opportunities for cross-layer system design [6]. Designing an efficient, reliable and flexible resource allocation or network routing mechanism in satellite networks has always been a big challenge. Cross-layer optimization can integrate the layers of the protocol stack into a comprehensive classification framework, which can meet the requirement of QoS in satellite networks. The cross-layer architecture proposed is shown in Figure 2.

2) Solvable Mechanism

It should be noted that balance should be found between computational complexity and optimality, i.e., in a multiple-access/network scheme with few parameters, few degrees of freedom are left to resource/routing optimization, whereas including lots of parameters and constraints may result into computationally unaffordable approaches. Therefore, the application of optimization model to resource management and network routing in satellite communication networks need to be reasonable and feasible. In addition, a computationally and numerically solvable mechanism of the model is essential. It depends on system characteristics and model properties.

For resource allocation, slots allocated to the user terminal are integer in satellite networks. So the problem of cross-layer dynamic bandwidth resource allocation can be solved by transforming the optimization model into a nonlinear integer programming problem, which is more accessible.

For network routing, optimization is usually to provide a set of numbers, paths, destinations, etc. Integer programming is not suitable for this type of problem and dynamic programming needs to be applied. Take ant colony method [7] as an example. The process of ants looking for a path is similar to path discovery process from source to destination. There are two kinds of agents in this type of method: forward agent and backward agent. The basic idea is that the former travels through the satellite networks and collects routing information and the latter updates the routing table. When data packets are sent from the source satellite and reach an intermediate satellite, this satellite node will forward them according to the routing table. The process continues, until data packets reach the destination. The key idea is that a set of routes needs to be found during each step of the mechanism, so the optimization model is transformed into a dynamic programming problem.

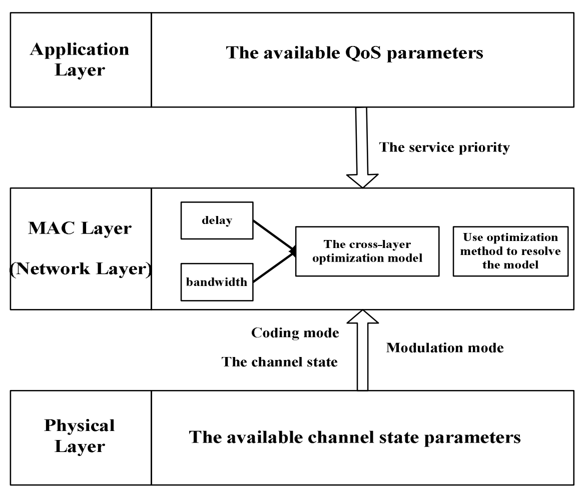


Figure 2. Cross-Layer Optimization

3) Simulation Results

In this part, we consider network routing as an example to explain the advantage that cross-layer design brings. The performance of a model designed based on the above idea, Cross-layer design and Ant-colony optimization based Load-balancing routing algorithm for LEO Satellite Networks (CAL-LSN) [8] is studied. Note that although the method is designed for LEO, it can be transferred into a network with hybrid constellation.

The simulation model is shown in Figure 3. An iridium-like satellite constellation is considered with two intra-plane ISLs (namely, links to the adjacent satellites in the same orbital plane) and two inter-plane ISLs (that is, links to the neighbouring satellites in the right-hand and left-hand orbital planes). While intra-plane ISLs are maintained for the whole satellite period, interplane ISLs are broken as satellites come close to the poles due to adverse pointing and tracking conditions, when satellites move to lower latitudes, inter-plane ISLs are re-established. Moreover, cross-seam ISLs, namely links between satellites in counter-rotating orbits, are not used.

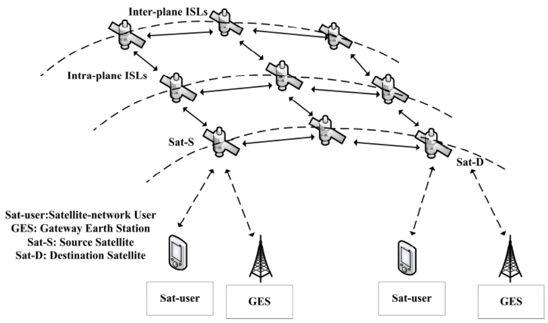
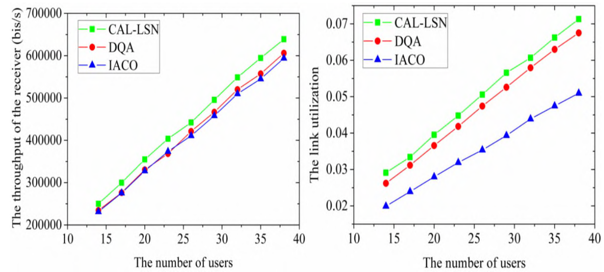


Figure 3. Cross-layer simulation model

The methods proposed in [9] and [10] are chosen to compare with CAL-LSN. The reason to choose these two algorithms is as follows. In [9], an adaptive routing algorithm based on an Improved Ant Colony System (IACO) was made use of in LEO satellite networks. The authors improve the original ant-colony algorithm in LEO satellite networks with its own cyclical and regular characteristics. In [10] the Distributed QoS-based Algorithm (DQA) was proposed. This algorithm is also based on ant-colony system. To meet QoS requirements minimum bandwidth constraints are also considered in DQA.

We first compare the performance of the receiver's throughput when CAL-LSN, DQA and IACO are utilized in the satellite networks. Figure 4. (a) shows the comparison. The reason why CAL-LSN has the highest throughput is that it can collect information from the physical layer. Comparing with DQA and IACO, the throughput is about 6.32% and 7.53% higher when CAL-LSN is used.

The average utilization of all inter- and intra-satellite links over the whole constellation is also compared between the above mentioned three strategies. We can see from Figure 4. (b) that although IACO improves the original ant-colony algorithm, it does not consider the influence of the residual bandwidth on QoS requirement. CAL-LSN has the same trends as DQA as the user number increases but has better performance. Comparing with the other two algorithms, the link utilization is the highest when CAL-LSN is used. It is about 6.52% and 29.24% higher comparing CAL-LSN with DQA and IACO respectively.



(a) (b)

Figure 4. Throughput and link utilization comparison

4) Satellite Design Issues

Cross-layer design brings a system with better performance in a dynamically changing communication environment. However, it also brings some challenging issues to satellite design. When the converged satellite-terrestrial system is built, layered design guideline and cross-layer optimization of terrestrial wireless network is introduced to satellites, making the space segment system more complicated [2].

First, the cross-layer design of satellite constellations requires ISLs with high transmission capacity while satellites needs a flexible pointing adjustment mechanism. Interplane ISLs are difficult to maintain when satellites come close to the poles due to adverse pointing and tracking conditions. Second, powerful onboard processing capability is required for computational complex tasks. However, signal processing onboard need to be flexible to some extent for dynamically changeable tasks, so advanced and flexible payload structure such as channelization and software defined radio (SDR) may be applied. Third, constellation design should be investigated to make network routing and resource management more efficient and realizable.

C. Cooperative Integration

Cooperative integration is the last dimension in our architecture and mainly aimed at increasing network robustness. It is on the basis of converged satellite-terrestrial and cross-layer QoS-based integration, which were discussed in the above sections. By applying cooperative strategies, data redundancies are reduced and transmission efficiencies are improved, thus allowing more users in the network with better QoS. In this section, cooperative diversity to use multi-path in receiver is first discussed followed by network coding which mixes different data flows to save resources. We again present satellite design issues as a final remark.

1) Cooperative Diversity

In satellite communication system, the satellite-to-ground channel is a typical fading channel. In order to assure the communication quality, efficient technology to defeat wireless fading is necessary. The advantage of diversity technology on improving the impact of channel fading gradually attracts attention. By sending the same signals through several independent paths and adopting a combination technology at the receiver, the power of the received signal is significantly enhanced.

Cooperative diversity technology in ground wireless communication system has been well studied. In satellite cooperative communication system, we investigate three cooperative strategies according to the processing scheme at relay satellite: amplify-and-forward (AF) scheme, decode-and-forward (DF) scheme and coding cooperative (CC) scheme.

We simulate the performance of the outage probability for the above satellite cooperative diversity scheme. Outage probabilities for satellite cooperative diversity are illustrated in Figure 5. Meanwhile, the performance of non-cooperative system is also provided for comparison. The readers may refer to [11] for theoretical modeling and deviation. The simulation and analytical results of outage probability are mostly coincident. We can also see that:

- Satellite cooperative diversity schemes lead to considerable improvement on the outage probability compared with the non-cooperative system. Among these three cooperative strategies, CC achieves the lowest outage probability.
- As SNR increases, the outage probability of AF and CC schemes decreases faster than that of the DF scheme. The DF scheme outperforms the AF scheme when the SNR is low, but becomes inferior at high SNR. This is because when SNR is high, transparent satellite structure in AF introduces less noise and the decoded scheme becomes less effective.
- The outage probability under urban scenario is greater than that under suburban one. However, the performance of DF scheme under these two scenarios seems similar.

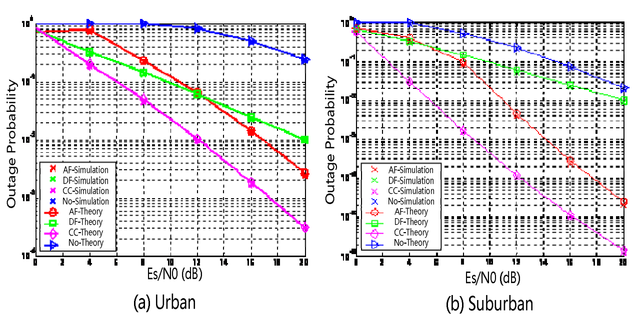


Figure 5. Cooperative diversity outage probability under urban and suburban scenarios

2) Network Coding

The basic idea of network coding is to allow conflict and combination of several data packets on the same time slot or the same frequency band. A network node receiving such a signal should broadcast the processed data, in order to assist the data recovery at other network nodes. We design network coding architecture for various applications:

- Transparent transmission satellite system: the satellite and the ground station are connected through a feeder link and the combination of them forms the relay node in the network. The satellite is equivalent to antenna of the relay node and ground station is responsible for signal processing.
- On-board processing satellite system: such complicated processing procedure is difficult to implement using today's satellite ability. However, complexity can be reduced if reliable coding techniques are applied with some loss of link performance.
- Relay satellite system: for an application where ground relay is used, the received signal can be transmitted to ground station regardless whether the satellite has the onboard processing ability or not. The signal can be encoded by the ground station and transmitted to user terminals. However, to complete the communication process, terminals communicating with each other must be located in the coverage area of the same ground station or the ground station can be interconnected through the network.
- Satellite systems with inter satellite links: a virtual relay channel can be formed by many satellites communicating with each other. In this case, design of resource scheduling algorithm is very important.

The above discussion assumes a three-point model. We now present design ideas of network with many points where many data streams are transmitted. In this case, corresponding routing and resource management algorithms are important. From the above discussion, we know that, if data transmitted between two nodes is equivalent, PNC technology can be used to save resources. In real communication system, there are many unidirectional links and many relay nodes serving different communication links. Moreover, even communication link is two-way, data transmitted in the two directions is not equal. In this case, virtual communication links that data transmitted in two directions can be constructed by designing corresponding scheduling algorithm, so that PNC technology can be used by this link to save resources [12].

3) Satellite in Cooperation

The cooperative strategy in our discussion above forms the third dimension of integration. Similar to the previous two, the cooperative integration ability in satellite communication systems is closely related to satellite performance [2].

- The satellite is always used as an active component instead of a transparent medium; thus, a regenerative structure is necessary in many scenarios. However as analyzed above, demodulation, decoding and parity information regeneration put heavy burden on satellite and raise challenges in on-board processing design.
- In advanced multi spotbeam satellite system, it was mentioned in [12] that the system does not perform well, as it requires the downlink beam to cover both satellite terminals. The likelihood of this situation is

lower for multi spotbeam satellites. However for other applications as multicast, channel independency is a key element for reaping the benefits of network coding multicast.

- From the simulation results, we can also see that cooperative strategy performs differently in various channel conditions. So, satellite payload that can adapt to channel conditions is essential for users to receive consistent service.
- Satellite mechanism enabling effective synchronization plays a key role in engineering realization in lots of cooperative scenarios.

IV. Concluding Remarks

Satellite communication has seen evolvement in recent years according to service provisioning requirements while global ubiquitous connectivity is on the way. Following this trend, our research considered system integration in three dimensions. Converged integration in interface dimension bridges the digital divide with satellite-terrestrial co-design providing seamless connectivity. Cross-layer integration in protocol dimension optimizes users' service quality with joint modeling of different layers in the protocol stack. Cooperative integration in link dimension promotes performance for all users by improving overall network robustness.

Our discussion revealed that satellite, in various aspects, plays an essential role [2] in the integrated system design.

- Powerful onboard processing capability based on advanced telemetry and command design is needed for computational complicated integration tasks, though a transparent structure can be used in some applications.
- Flexible regenerative structure should be applied to further reduce burden and adapt to variable channel conditions.
- Multi spotbeam techniques with dynamically allocation capability provide tunable structure for multiple users, which are useful in future broadband systems.
- Tailored frequency plan and frequency reuse schemes based on beam design should be applied to provide sufficient carrier to interference (C/I) ratio.
- Intersatellite link (ISL) giving intraplane and interplane connectivity needs flexible pointing adjustment mechanism, thus requires satellite mechanics innovation.
- Telemetry, command and other communication workforce onboard guarantees effective synchronization in various application scenarios.

In conclusion, future satellite communication system providing universal service requires satellite-based design throughout its construction [2]. It is thus possible to maximize the system capacity at present, make use of the benefits brought up by technical development and accelerate the accomplishment of newly integrated satellite communication system.

Acknowledgment

This paper is supported by the National Natural Science Foundation of China (Grant No. 91438205, 61675232).

References

[1] M. Sadek and S. Aissa, "Personal satellite communication: technologies and challenges," IEEE Wireless Communications, 2012, vol. 12, pp. 28-35.

[2] Z. Zhou and G. Qu, "System design and dynamics analysis of communication satellites," 2013.

[3] Z. Zhou and G. Qu, "Nonlinear finite element analysis of large mesh deployable antenna on satellite," Spacecraft Engineering, 2008, vol. 17, no. 6, pp. 33-38.

[4] R. B. Dybdal and K. M. SooHoo, "Narrow beamwidth satellite antenna pointing and tracking," IEEE International Symposium on Antennas and Propagation, 2011, pp. 2012-2015.

[5] L. Zhou, M. Chen, Z. Yu, Rodrigues J. and Chao H. C., "Cross-layer wireless video adaptation: Tradeoff between distortion and delay," Computer Communications, 2010, vol. 33, no. 14, pp. 1615-1622.

[6] R. Shen, "Some Thoughts of Chinese Integrated Space-Ground Network System," Engineering Science, 2006, vol. 8, no. 10, pp. 19-30.

[7] D. Maniezzo, M. Dorigo and Colorni A., "Ant System: An Autocatalytic Optimizing Process," Technical Report 91-106, 1991, pp. 1-21

[8] H. Wang, Q. Zhang, X. Xin, Y. Tao and N. Liu, "Cross-Layer Design and Ant-Colony Optimization Based Routing Algorithm for Low Earth Orbit Satellite Networks," China Communications, 2013, vol. 10, no. 10, pp. 37-46.

[9] Z. Gao, Q. Guo and P. Wang, "An Adaptive Routing Based on an Improved Ant Colony Optimization in LEO Satellite Networks," Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 2007.

[10] H. Xu and S. Wu., "A Distributed QoS Routing Based on Ant Algorithm for LEO Satellite Networks," Chinese Journal of Computers, 2007, vol. 30, no. 3, pp. 361-367.

[11] A. Iqbal and K. M. Ahmed., "Outage probability analysis of multi-hop cooperative satellite-terrestrial network," The 8th Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology Association, 2011, pp. 256-259.

[12] Z. C. Chen, B. Y. Zheng, X. D. Ji and X. C. Xiao, "An Improved Joint Design of Physical Layer Network Coding and Channel Coding Based on Trellis Coded Modulation in Two-way Relay Channel," Journal of Electronics and Information Technology, 2011, vol. 33, no. 11, pp. 2594-2599.

Topologies for the Provision of Network-Coded Services via Shared Satellite Channels

Ulrich Speidel
The University of Auckland
Auckland, New Zealand
email:ulrich@cs.auckland.ac.nz

’Etuate Cocker
Spark Digital NZ Ltd.
Auckland, New Zealand
email:Etuate.Cocker@spark.co.nz

Muriel Médard
RLE, MIT
Cambridge, USA
email:medard@mit.edu

Péter Vingelmann, Janus Heide
Steinwurf ApS
Aalborg, Denmark
email:{peter|janus}@steinwurf.com

***Abstract*—Network traffic using the Transmission Control Protocol (TCP) across shared bottleneck satellite channels can suffer significant impairment due to TCP queue oscillation. Coding of such network traffic across multiple Internet Protocol (IP) packets allows packet loss to be masked from the senders, letting TCP senders sustain higher goodput rates. We argue that the concept of tunneling coded traffic across a satellite link is a flexible one and does not rely on a one-size-fits-all solution. This paper discusses a number of network topology options for the deployment of coding, from the perspective of satellite providers, Internet service providers, end users and third-party entities.**

Keywords–TCP/IP; network coding; network topologies; queue oscillation; tunneling.

I. Introduction

Consider the following scenario: An Internet Service Provider (ISP) on a small Pacific island receives its international connectivity via a geostationary (GEO) or medium earth orbit (MEO) satellite service. The capacity provisioned is in the range of several Mbps to several 100 Mbps, but always well below that of the networks connected at either end (assumed to be 1 Gbps or faster). The ISP services users on the island. The number of concurrently active client devices could be anywhere from a few dozen to a couple of thousand, and the ISP might observe up to a few thousand simultaneous TCP flows. For the purposes of this paper, a TCP flow is a set of TCP packets travelling in one direction and characterised by a unique combination of source and destination IP addresses and ports [1]. Each flow typically belongs to a single TCP connection (i.e., a connection typically consists of two flows in opposing directions).

The flows across the link will typically be a heavily skewed mix: Most flows on the link will contain at most a few hundred bytes and will be too small and short to have their rate controlled by TCP flow control (also known as congestion control). Long flows, which are subject to flow control, contribute the majority of bytes on the link, however.

Satellite links of this type present a significant challenge to TCP: The long latency bottleneck makes it difficult for TCP senders to find the correct congestion window [2]. Moreover, a large number of simultaneous connections face exactly the same congestion situation here. This causes the TCP senders involved to act in unison when adjusting their congestion windows, an effect known as *global synchronisation*. It can lead to *TCP queue oscillation*, where the input queue to the satellite link oscillates between empty and overflow, causing link underutilisation when the queue is empty. The resulting performance problem has been studied in the context of satellite links for over two decades (see, e.g., [3]) and remains essentially unsolved, despite the emergence of active queue management (AQM) techniques [4][5]. In large parts, this is due to senders overloading the queue based on feedback from the receivers that is already on its way when the queue shows signs of filling. Any feedback from explicit congestion notification or random early drops simply arrives too late to be useful.

Network coding [6] offers a potential part-remedy here: Error-correcting packet losses at the input queue can prevent premature back-off by the TCP senders, allowing some of the lost capacity to be reclaimed. The basic topology investigated in this paper is a tunnel, which operates across the link and both satellite input queues at either end. It accepts and delivers IP packets regardless of transport layer protocol involved, such that the end-to-end principle always remains intact. We have already demonstrated [7][8] that such a tunnel solution can improve goodput for individual TCP connections, even in the presence of a majority of legacy TCP traffic on the same link. Such tunnel-based solutions thus represent a potential alternative to and/or enhancement of performance-enhancing proxies (PEPs) [9], [10].

This paper investigates possible deployment scenarios for tunnel solutions of this kind. Section II explains the basic workings of the tunnel in a scenario where the ISP on the island provides the on-island tunnel endpoint and where the coded traffic between the tunnel endpoints travels in the payload of User Datagram Protocol (UDP) packets. On this basis, Section III discusses the question as to where the off-island tunnel endpoint could be located and presents a case for having multiple endpoints. Section IV describes a scenario in which a third-party entity operates the tunnel endpoint on the island. In Section V, we look at the advantages and drawbacks of offering coding-as-a-service tunnels to individual end users on an island. Section VI then looks at various options for non-UDP communication between the tunnel endpoints.

II. Coded Tunnels

We begin our tour of the basic tunnel model (Figure 1) by introducing our players and our components and following what happens during a TCP connection from an island end user client to a server off-island that the user wants to access:

The connection begins at an end user machine (client) on the island. In most scenarios, we will assume that we have no control over this machine, i.e., that we cannot assume anything beyond the existence of a TCP/IP stack and some TCP client program on the machine. It means in particular that we cannot

 ISBN: 978-1-61208-545-6

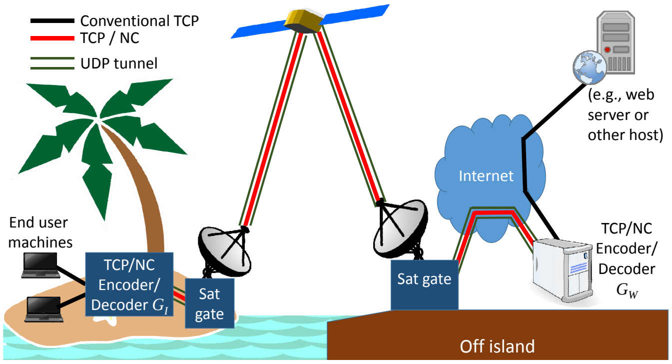


Figure 1. Network topology in a scenario where the on-island internet service provider operates the local encoder/decoder. The off-island encoder/decoder may be at an arbitrary off-island location on the internet.

install software on the machine or get the user to change machine settings. The client initiates the connection by sending a TCP packet to the off-island server, with the SYN flag set. Its TCP/IP stack encapsulates this packet in an IP packet whose source address is that of the end user machine. Its destination address is that of the off-island server.

The IP layer initially forwards this IP packet to a local gateway router on the island, and from there possibly along further gateway routers in the direction of the on-island satellite gateway. In our case, we replace one of these gateways by our on-island encoder/decoder G_I. Since our packet is heading off-island, we use G_I's encoder functionality here.

The encoding works as follows: G_I captures the original IP packet but does not forward it any further. Instead, G_I forms sets of n successive captured IP packets. Each such set is called a *generation* and n is the *generation size*. In random linear network coding (RLNC), G_I now creates $n + \omega$ byte-wise linear combinations of all packets p_j in the generation, using randomly (i.e., generally independently and uniformly) chosen coefficients c_{ij}. That is, the i'th linear combination r_i that G_I produces is given by:

$$c_{i1}p_1 + c_{i2}p_2 + \ldots + c_{in}p_n = r_i, \qquad (1)$$

The $n+\omega$ combinations thus produced form an overdetermined system of linear equations whose solution is the set of p_j. In doing so, G_I codes all bytes of the incoming packets, including the headers with the IP addresses of the island and off-island end hosts involved. G_I now communicates this system one equation at a time to G_W, located somewhere on the Internet on the off-island side of the satellite link. For this purpose, G_I sends $n + \omega$ UDP packets to G_I, with its own IP address as source and G_W's IP address as destination. Each UDP packet contains the equation for a particular i in the form of the c_{ij}'s and r_i.

These UDP packets now travel via the satellite link and the off-island Internet to G_W, which solves the system of linear equations. The solution consists of the p_j, of course, which G_W then forwards to their off-island destinations. Note that G_W generally only needs any n of the $n + \omega$ UDP packets in order to decode the p_j. The remaining ω UDP packets are surplus and may safely be dropped along the way – for example, at the input queue to the satellite gateway. The important point here is that we can leave it up to this queue to decide which packets to drop.

Our SYN packet has now arrived at its off-island server destination, and the server wishes to send a SYN+ACK in response. At this point, the network topology becomes critical: Most data flows *to* the island and it is important that we encode this direction in particular, so we need to ensure that the response packet finds its way to G_W. However, in most island scenarios, the island hosts including the satellite gateways at either end of the link belong to a single IP subnet. From the world's perspective, the off-island satellite gateway is also the IP gateway to this subnet, so the SYN+ACK response (and any subsequent packets) from the server would be routed straight to the off-island satellite gateway, entirely bypassing G_W. This is unacceptable, of course.

The solution is to split the subnet on the island: End hosts in the islands become part of a new subnet A (this could also be several subnets), whereas the on-island satellite gateway is placed in a disjoint subnet B. One then configures routing to make G_W becomes the gateway for traffic to A, and traffic to B is routed straight to the off-island satellite gateway.

In this scenario, the off-island server receives the SYN packet with a source from network A, and thus responds by forwarding the SYN+ACK to G_W. There, G_W encodes the packet in the same way G_I encodes packets in the opposite direction. It then forwards the coded packets inside UDP to G_I for decoding and release to the island end user machine. This completes the round-trip handshake. Further packets between the hosts follow the same path. That is, the packets travel trough a coded UDP *tunnel* between G_I and G_W and vice versa.

This scenario requires the ISP on the island to either operate G_W off the island, or contract an off-island entity to operate G_W on their behalf. In many cases, it will also be desirable to make at least network A an autonomous system (AS) for routing purposes, in which case G_W needs to be duplicated for redundancy. The current experimental software that we have been working with is capable of supporting two instances of G_W.

In the next sections, we will consider variations of this base scenario.

III. Tunnel Endpoints and Their Locations

Our basic tunnel scenario above assumes that G_W is located at an arbitrary location on the Internet. As long as the tunnel that it spans with G_I covers the satellite link, it can fulfill its purpose of masking packet loss at the satellite gateway input queues. There are however good reasons to consider the placement of G_W carefully. The following options may deserve consideration:

- G_W could be placed in the path between the Internet and the off-island satellite gateway (Figure 2), or even be a part of the satellite gateway hardware. In this case, network B could use private IP addresses, and G_W simply acts as a gateway for network A as in the previous scenario. That is, all machines on the island could be in the same subnet of an upstream provider's network, save the off-island facing interface of the on-island satellite gateway. For the ISP and/or their upstream provider, this removes the cost of

 ISBN: 978-1-61208-545-6

maintaining a separate block of public IP addresses or even a separate AS.

However, a placement at the satellite gateway requires the competent cooperation of whichever party controls the off-island satellite gateway: It needs to install and assist in commissioning G_W, or permit terminals with equivalent built-in functionality to be installed. In practice, one encounters a variety of scenarios, however: One ISP owns and controls both satellite gateways, another ISP owns and operates the island side only and contracts to a satellite provider and upstream ISP off-island, and yet another buys a turnkey solution from a satellite provider who also controls and services the on-island satellite gateway. We note in this context that especially in the latter case, satellite providers often already provide Wide Area Network (WAN) accelerators with network memory, parity packets and various other optimisation functions – inserting G_W or even G_I as part of such a solution would thus not be without precedent.

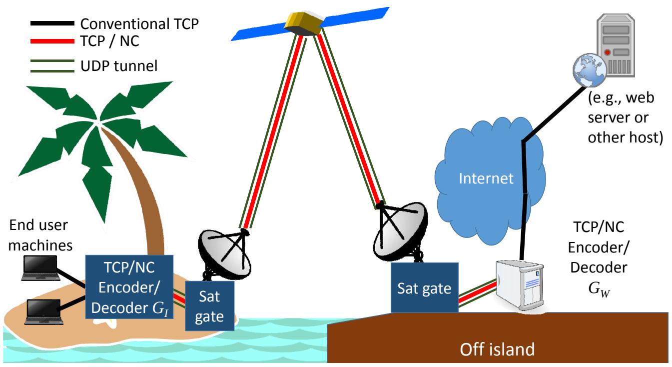


Figure 2. Network topology in a scenario where the on-island internet service provider operates the local encoder/decoder, and the off-island encoder/decoder is inserted in the path between off-island satellite gateway and the Internet.

- G_W (or several instances thereof, labelled G_{W1}, G_{W2}, etc.) could be placed close to the known primary sources of bulk data content sought by island clients (Figure 3). Such a placement would protect a longer portion of the paths between servers and clients by coding and would bridge other potential sources of loss. However, there is an obvious drawback: G_W is no ordinary server – as an AS gateway, it needs a significant amount of network configuration in its environment. Placing G_W in a site potentially far away from both ISP and satellite provider premises requires the cooperation of a third party able to host G_W and arrange for its routing needs. Such partners could potentially be difficult to recruit for an ISP based on a remote island.
- G_W could be placed at the premises of a specialised off-island provider, who could also own and operate the device and sell its encoding/decoding services to the ISP on the island. An obvious advantage of this model is that it allows a provider to specialise in this type of service and host the G_W for multiple island installations, achieving some economies of scale. A potential disadvantage is added latency: The latency between off-island satellite gateway, G_W and off-island data sources may be much higher than that between data sources and satellite gateway alone. This problem can be exacerbated by a failure to peer near the off-island satellite gateway. E.g., the authors are aware of a Pacific Island ISP whose off-island gateway is located in Hawaii. While the island has close cultural and economic links to New Zealand, lack of peering in Hawaii at the time of writing meant that all traffic between the island and New Zealand also has to travel between Hawaii and the U.S. mainland and back.

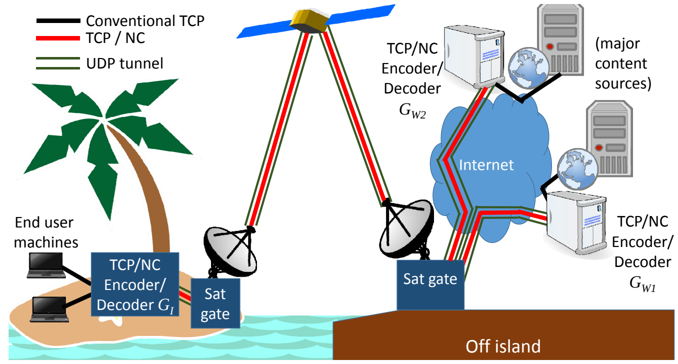


Figure 3. Network topology in a scenario where the on-island internet service provider operates the local encoder/decoder, and off-island encoders/decoders are placed close to the servers on the internet from which most of the download content originates.

IV. TUNNELS NOT INVOLVING ISPS

In all our scenarios so far, the island ISP has played a core role as the operator of G_I if not G_W. However, in principle there is no reason why G_I cannot be operated by another party on the island. Assuming for the moment that the ISP and satellite provider will pass UDP in both directions, any of the ISP customers within the island network can operate a G_I to tunnel to some G_W located off-island. This customer can then spawn their own network (Figure 4) or – in the case of individual rather than institutional customers – simply run G_I on their own host or local network address translator (NAT) box.

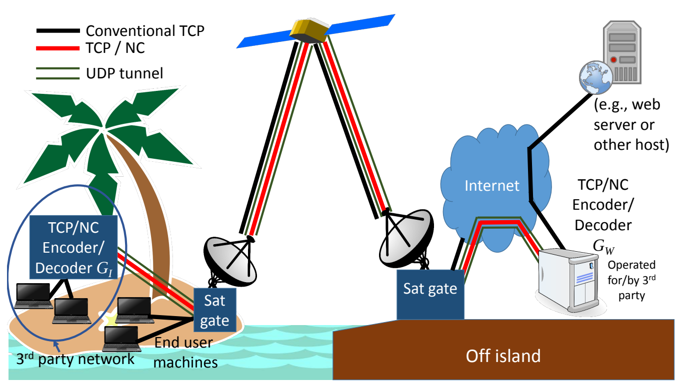


Figure 4. Network topology in a scenario where an on-island encoder/decoder is integrated into an end user machine on the island, and the off-island encoder/decoder is provided by a third party on the internet as a service, e.g., for a fee.

 ISBN: 978-1-61208-545-6

In the case of institutional customers, the corresponding G_W could be located at an organisation's off-island data centre or at the premises of a specialised off-island provider as discussed in the previous section. In the case of individuals, there could also be the option of G_W being provided on a subscription or pay-as-you-go basis by an off-island entity, as discussed in the next section.

Any such arrangement has a number of drawbacks, however. Firstly, it almost inevitably means that only some of the traffic on the link will be coded traffic, with the remainder being (mostly) conventional TCP. This residual uncoded traffic may still cause queue oscillation. While the coded traffic would be – at least to an extent – be protected from the associated packet loss and slow-down, the coding scheme involved would nevertheless have to provision sufficient overhead in order to cope with the potentially lengthy burst errors that queue oscillation causes. This would further increase the load on the link. However, this would be in parts offset by the fact that the link does not need to carry overhead for the uncoded part of the traffic.

Secondly, any overhead transmitted or received by a G_I under customer control increases that customer's data usage. In cases where the ISP on the island applies volume charges (a very common scenario in the Pacific), this results in additional cost for the customer. This may however be outweighed by data volume savings at the application layer as customers have to repeat fewer unsuccessful downloads.

V. Coding-As-A-Service Tunnels

Another possible scenario is to absorb G_I into a virtual network interface on the end user machine and provision G_W off-island on a subscription or pay-per-coded-volume basis (Figure 5). In this case, the end user would download an application which implements the client-side solution with G_I and interfaces with G_W off-island. The end user machine would then use two IP addresses: that assigned by the ISP, which appears in the header of the UDP packets between the machine and G_W, and an IP address assigned by the off-island provider of G_W, which belongs to the off-island provider's network and is not visible on the island to any host except G_I (which of course operates on the machine itself). This address is the source of IP packets departing G_W in the direction of off-island servers on behalf of the end user machine, and the destination of any packets that these servers send in response. In this respect, the service operates in a very similar fashion to a tunnelled VPN connection, except that the traffic across the tunnel is encoded rather than encrypted (it may of course also be encrypted in addition to the encoding).

An obvious advantage of this approach is that there is no need for dedicated on-island infrastructure, the ISP does not have to expend or upskill personnel resources (or even be aware of the tunnel operation), and there is no need for equipment or personnel to be sent to the island to install or support the system. These are significant factors as many Pacific islands with satellite connection are difficult to reach – air services may be infrequent or non-existent, and intervals between ship visits may be lengthy and freight is expensive. Similarly, many island ISPs struggle to hire and retain qualified personnel.

Naturally, there are also a number of drawbacks, which start with those discussed in the previous section. In addition,

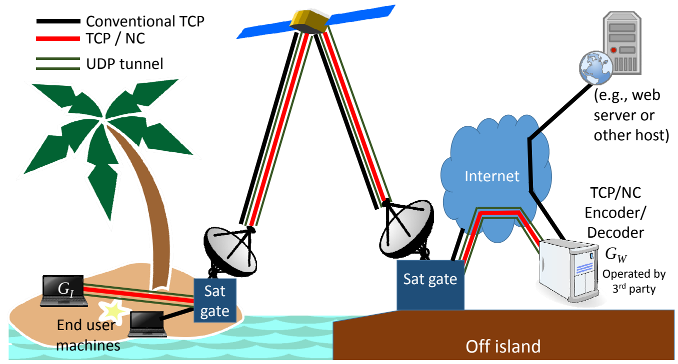


Figure 5. Network topology in a scenario where an on-island encoder/decoder is integrated into an end user machine on the island, and a third party provides the off-island encoder/decoder as a (potentially fee-paying) service on the internet.

there is now an additional challenge from the location perspective: As discussed in Section III, the latency between satellite gateway, G_W and data sources may be significantly higher than the latency between satellite gateway and data sources alone. If the specialised off-island provider of G_W implements a coding-as-a-service scenario at scale, it will inevitably find its client software used in multiple island locations, with satellite gateways in geographically dispersed locations: An island in the western Pacific can have its off-island GEO gateway in Canada, whereas an island in French Polynesia might opt for space segment terminating in Australia.

A further challenge is the diversity in consumer operating systems. To be able to serve a large majority of users, the off-island provider would need to supply the application implementing G_I on multiple popular operating systems such as Windows, MacOS, IOS and Android. This represents significant additional effort compared to a tunnel application on a single operating system of the implementor's choice. It also carries the risk of leaving the end user machine disconnected: The software needs to modify the network configuration of the machine. There could be unintended consequences if, in doing so, the software interferes with any of the myriad of network configuration managers, tools and utilities which commonly inhabit these ecosystems. Given that the off-island provider has no control over what else may be installed on the end user's machine, this risk could be substantial. Another question that arises in this context is how an island user would pay the off-island provider: Not every islander has access to a credit card.

VI. Connecting the Tunnel Endpoints

On many islands, ISPs and/or satellite providers block UDP to keep traffic off their satellite link that does not back off under congestion. This can backfire, however, as many applications that have higher bandwidth efficiency using UDP do sense congestion and will switch to less efficient TCP when UDP is blocked. It is worth noting in this context that the UDP carrying our coded packets *will* back off as well: If too many packets of a generation are lost, the generation as a whole will become undecodable and the TCP packets it contains are lost as well, causing the contributing TCP senders to back off, too. Conversely, the notion that TCP will always back off is also not true: Small TCP flows often complete before the first ACK reaches the sender, such that the ACKs do not get to influence packet transmission rates.

However, the communication between G_I and G_W need not rely on UDP. There are several options for this, two of which are discussed below:

A. Spoofing TCP

One option is to pass the coded combinations as TCP packets without actually running TCP at G_I or G_W. The only differences to the UDP variant are as follows:

- The packets carry a TCP header instead of a UDP header, with nominal sequence and acknowledgment numbers
- G_I and G_W acknowledge any packet received but do not attempt to retransmit any packets not received (and in fact ignore any ACK received)
- The first and second packet from G_I to G_W have their SYN and SYN+ACK flags set, respectively, and the first packet G_W to G_I correspondingly has SYN+ACK set.
- Either end ignores flags and ACK numbers upon receipt and concentrates on the packet payload instead.

To an outside observer, such flows are almost indistinguishable from genuine TCP and practically impossible to detect or block on a firewall with stateful inspection. Even in a real TCP connection, an observer somewhere along the path may not get to see all packets of the connection due to load balancing and asymmetric paths. One disadvantage of this approach is that it is a hack and, from the ISP's perspective, could be considered improper use. Another is that it adds encapsulation overhead, as the size of a TCP header is larger than that of a UDP header.

B. Multiple TCP Connections

Another option would be to open multiple TCP connections between G_I and G_W at the outset and communicate only a small number of linear combinations (or even just one) per generation as data across each connection.

In scenarios where G_I and G_W are the only significant users on the satellite link, this has the advantage of replacing what would otherwise be a mix of TCP flows of varying lengths by a fixed number of TCP flows with infinite length and more or less equal data rate. Since the arrival of each combination is now ensured by TCP, one could also set $\omega = 0$ and thus reduce overhead to zero. However, TCP also adds its own overhead. It is also possible to use TCP variants optimised for long latency networks, such as Hybla [12] or H-TCP [11].

One potentially significant problem occurs at G_W (and possibly G_I, too), however: In the UDP or spoofed TCP scenarios, data arrives at G_W at full Gbps network rates and leaves in the direction of the sat gate at the same high rate in encoded form. So G_W does not need to buffer or concern itself with keeping any form of state once the coded packets of a generation have left. If we connect G_W and G_I via TCP, we transfer at least a significant part of the sat link bottleneck and its associated queue to G_W. Since TCP sockets cannot queue drop, G_W would need to implement this functionality *before* the linear combinations are written to the TCP connections with G_I.

VII. Conclusion and Future Work

Network-coded tunnels carrying TCP/IP traffic in coded form across lossy bottlenecks in satellite networks have been shown to be able to improve goodput under TCP queue oscillation conditions even in the presence of a majority of flows using legacy TCP. The core insight that underpins the tunnel concept is that packet losses occur by queue drop at the input queue to the satellite link. As long as one can protect traffic against data loss at this location, the remaining system topology is a question of who will or can provide the tunnel service, and how cooperative the local ISP and satellite provider are. We have discussed a variety of potential topologies along with their advantages and drawbacks. All are feasible: Coding across satellite links does thus not rely on a single solution topology.

As a general rule, topologies in which ISP and/or satellite provider are not involved (or even actively oppose the use of coded tunnels) are bound to be less effective: The presence of legacy TCP connections forces coded traffic to use more overhead, so any parties on the island with coded traffic consume more data and bandwidth than necessary. Those not using coding are also put at a potential disadvantage as this may eat into their bandwidth as well. Active involvement of satellite providers and/or local ISPs thus seems advantageous.

At the time of writing, only experimental implementations of coded tunnels are available. These are based on a Debian/Ubuntu Linux kernel module. While they do not cater for the subscription model discussed in Section V at this point in time, they nevertheless represent a proof of concept for the remaining scenarios. Current work aims to demonstrate that the technology scales to whole-of-island coding.

Acknowledgments

The authors wish to thank the Information Society Innovation Fund (ISIF Asia/APNIC) and Internet NZ for their financial support for this work. We would also like to thank Brian Carpenter and Nevil Brownlee, among others, for their support and many helpful suggestions and discussions. We are also grateful for the support received from the Pacific Island Chapter of the Internet Society and a number of ISPs in the Pacific.

References

[1] Clark, D. D., “The design philosophy of the DARPA Internet protocols”, Proceedings of ACM SIGCOMM’88, pp.16–19, 1988.

[2] Jacobson, V. and Braden, R. *TCP Extensions for Long-Delay Paths*, RFC1072, 1988.

[3] Jouanigot, J. M. et al., “CHEOPS dataset protocol: an efficient protocol for large disk-based dataset transfer on the Olympus satellite”, International Conference on the Results of the Olympus Utilisation Programme, 20-22 April 1993, Sevilla, Spain, CERN CN/93/6.

[4] Braden, B. et al., *Recommendations on Queue Management and Congestion Avoidance in the Internet*, RFC 2309, 1998.

[5] Kim, J. and Yeom, I., “Reducing Queue Oscillation at a Congested Link”, IEEE Transactions on Parallel and Distributed Systems, 19(3), pp. 394–407, 2008.

[6] Sundararajan, J. K. et al., “Network Coding Meets TCP: Theory and Implementation”, Proc. IEEE, 99(3), pp. 490–512, 2011.

[7] Speidel, U. , Cocker, ’E., Vingelmann, P., Heide, J., and Médard, M., “Can network coding bridge the digital divide in the Pacific?”, International Symposium on Network Coding (NetCod), Sydney, pp. 86–90, 2015.

 ISBN: 978-1-61208-545-6

[8] Speidel, U. et al., “Can Network Coding Mitigate TCP-induced Queue Oscillation on Narrowband Satellite Links?”, International Conference on Wireless and Satellite Systems, pp. 301–314, Springer International Publishing, July 2015.

[9] Caini, C., Firrincieli, R., and Lacamera, D., “PEPsal: a Performance Enhancing Proxy designed for TCP satellite connections”, IEEE 63rd Vehicular Technology Conference, pp. 2607–2611, 2006.

[10] Delannoy, G., “Design and Implementation of a Performance-Enhancing Proxy for connections over 3G networks”, Dublin City University, May 27, 2013, https://github.com/GregoireDelannoy/TCPeP/blob/master/Final_Report.pdf [accessed: 2017-03-05].

[11] Leith, D., “H-TCP: TCP Congestion Control for High Bandwidth-Delay Product Paths”, Internet Draft, IETF, April 7, 2008, https://tools.ietf.org/html/draft-leith-tcp-htcp-06 [accessed: 2017-03-05].

[12] Caini, C., Firrincieli, R., “TCP Hybla: a TCP enhancement for heterogeneous networks”, Int. J. of Satellite Communications and Networks, 5(22), pp. 547-566, 2004.

 ISBN: 978-1-61208-545-6

Multi-Pool Based Resource Allocation Scheme in Broadband SATCOM System

Luoyu Qin*, Ying Tao†,Houtian Wang‡
China Academy of Space Technology, Beijing, China
Email: qinluoyu@icloud.com*,tao.ying@126.com†,1357953688@qq.com‡

***Abstract*—A Multi-Satellite Broadband Network (MSBN) appears to be a promising network for providing efficient and seamless multimedia services. To maximize the resource utility and optimize the Quality of Service (QoS) in MSBN, reasonable bandwidth allocation method is, indeed, important. State-of-the-art bandwidth algorithms, which are mostly based on the allocation of a single satellite-ground link, ignore limited traffic capacities in other links, such as inter-satellite links. In this paper, a Multi-Pool based Bandwidth Allocation (MPBA) scheme is proposed to solve this problem. Firstly, a typical multi-pool framework is built to manage bandwidth resources of all communication links instead of the traditional single satellite-ground link, and the corresponding resource allocation process is analyzed. Secondly, inspired by the non-cooperative game theory, the MPBA algorithm with a low computation complexity is designed to ensure different QoS demands of multimedia services. Extensive experiments are carried out, and results obtained demonstrate that our multi-pool framework can achieve improvement on the end-to-end delay and the network resource utility over existing bandwidth allocation methods in MSBN.**

***Keywords*–Resource Allocation; Satellite Communication; Multimedia services; Quality of Service.**

I. INTRODUCTION

Future broadband satellite communication (SATCOM) systems will be composed of multiple satellites and orbits to provide efficient multimedia services [1] [2], which is called multi-satellite broadband network. Notwithstanding the benefits stemming from this design approach, owing to the diversity of communication links, the system must be capable of managing all of the bandwidth resources in every communication link flexibly, with satellite-ground links and inter-satellite links included. There are three key points that must be considered in the bandwidth allocation scheme of MSBN:

- Guarantee different QoS demands for multimedia services.
- Maximize the network resource utility.
- Balance traffic loads of different links.

State-of-the-art broadband SATCOM system based bandwidth allocation algorithms have made great contributions on the first two points. A game-theoretic framework based on Nash bargaining solution from cooperative game theory for the bandwidth allocation of elastic services in high-speed networks is proposed in [3], which is the first to focus on different QoS requirements of multimedia services in bandwidth allocation. It provides the rate setting of users that are Pareto optimal from the point of view of the whole system. Based on the water-filling algorithm, a cross-layer framework for optimizing the dynamic bandwidth allocation of Digital Video Broadcasting - Return Channel via Satellite (DVB-RCS) system is proposed in [4]–[8], which takes the Media Access Control (MAC) layer into consideration and provides QoS requirement for multimedia services in dynamic satellite channels. Rain-fade attenuation is taken into consideration on the bandwidth allocation scheme in [9], which makes contribution on the real-time slot assignment. The resource allocation is modeled as a non-cooperative game in [10], and a fair equilibrium point is converged to improve the fairness.

However, [3]–[10] are based on the resource allocation of a single satellite-ground link, which means that they ignore the limited traffic capacity of other links in MSBN, such as inter-satellite links. Traditional bandwidth allocation schemes [3]–[10], without taken the third point into consideration, are actually local optimization methods, which could not achieve system optimization in MSBN.

In a typical scenario of MSBN, as is shown in Figure 1, for satellite node S_i with B spot beams, the resource pools can be divided into three types: $S_i^k(k = 1, ..., B)$, $S_i \to S_j$, $S_j \to S_i$. S_i^k refers to the resource pool of k-th satellite-ground link in satellite S_i, $S_i \to S_j$ is the resource pool of inter-satellite link from satellite S_i to S_j, and $S_j \to S_i$ is from S_j to S_i.

It can be observed from Figure 1, that $user_1$, accessing B_1-th spot beam of satellite S_1, initiates a communication request with $user_2$, with $user_2$ accessing B_2-th spot beam of S_4. Supposing the optimal route path of the communication from $user_1$ to $user_2$ is $S_1 \to S_2 \to S_4$, then bandwidth resources of resource pools $S_1^{B_1}$, $S_1 \to S_2$,$S_2 \to S_4$, and $S_4^{B_2}$ will be occupied by this communication.

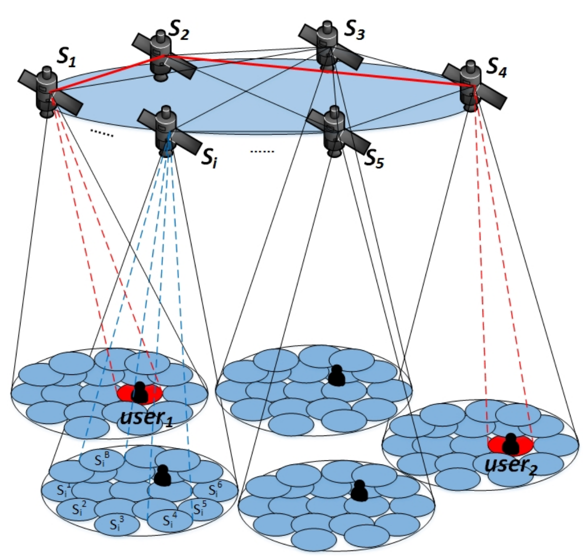


Figure 1. The resource framework in MSBN

Traditional researches [3]–[10] just take the bandwidth capacity of access link $S_1^{B_1}$ into consideration, as much service as $S_1^{B_1}$ can afford will be admitted to achieve higher local

 ISBN: 978-1-61208-545-6

resource utilization. However, with the lack of considering the limited capacity of inter-satellite links, there will be plenty of packets congestions in inter-satellite links, which will deteriorate the end-to-end delay and will increase the packet loss, especially when the system throughput rate is high. And lossless protocols, such as TCP, require retransmission of lost packets, which substantially increases transmission time and deteriorates congestion in inter-satellite links.

In this paper, we propose a multi-pool based bandwidth allocation scheme, which manages all links dynamically and minimizes network congestions in MSBN, to make up for the drawbacks of traditional methods above. By taking diversity QoS demands of multimedia services into account, our method can keep the resource utilization to maximum, while satisfying QoS demands of different multimedia services at the same time. By adopting the gradient descent and the non-cooperative game theories, a priority based bandwidth allocation algorithm is proposed to achieve network utility maximum (NUM) in our MPBA method, which features a low computational complexity and guarantees different QoS demands of multimedia services in MSBN.

The rest of this paper is organized as follows: In Section II, the multi-pool bandwidth optimization framework is constructed and the corresponding optimization problem is presented, which means to afford assurance of different QoS demands for multimedia services. In Section III, a non-cooperative theory based algorithm is proposed to solve the optimization problem and the complexity of this algorithm is analyzed. In Section IV, a simulation scenario is built based on OPNET, and comparison results of the MPBA and the traditional algorithm on the network end-to-end delay and the system resource utility are analyzed. The conclusion of this paper is given in Section V.

II. Proposed System Model

A. A Multi-pool based Resource Allocation Framework

As is shown in Figure 1, the traffic would occupy bandwidth resource of links within its route path. As a result, the traffic carrying capability of all these related links, not just access link, should be taken into consideration in the decision of whether a traffic could be accepted. Giving this situation, a MPBA framework is proposed in this paper. The process of the MPBA framework can be divided into four steps:

1. A bandwidth request is initiated by $user_i$ to the system Radio Resource Manager (RRM), with the destination node, the user type, the service type, the minimum bandwidth requirement $a_i$bps, and the current bandwidth requirement $b_i$bps included.

2. The RRM asks for the optimal route path information of $user_i$ from the network route manager.

3. The RRM obtains remaining bandwidth informations of all relevant satellite-ground and inter-satellite links within the optimal route path of $user_i$.

4. The RRM checks if all related links can meet the bandwidth demand of $user_i$. The bandwidth request of $user_i$ will be successful only when all links can meet the requirement. Otherwise, a bandwidth allocation failure signaling would be sent to $user_i$.

Different from traditional model [3]–[10], the traffic capacities of every link in system, rather than only the access link, are considered in the MPBA scheme, which is the key that global optimism could be achieved in MSBN.

There are different types of multimedia services in MSBN. The key of bandwidth allocation is to meet different QoS demands of multimedia services while maximizing the system resource utility. The MSBN based optimization problem is analyzed below.

B. Problem Formulation

Define a link vector $\mathbf{L} = (l_1, l_2, ..., l_M)$, where l_i is the i-th link in MSBN and M is the number of links in MSBN. Let the link resource vector $\mathbf{C} = (C_1, C_2, ..., C_M)$ be the current remaining bandwidth of links in $\mathbf{L}$, set $Q = \{user_1, user_2, ..., user_K\}$ consists of users to be allocated bandwidth resources. For multimedia $user_i(i = 1, 2, ..., K)$, let a_i be the minimum bandwidth demand and b_i be the current requested bandwidth.

The utility function is a tool of measuring the cost and the benefit in the bandwidth allocation. The user utility function $u_i(x_i)$ of $user_i$ is a classical model [11] [12], which meets

$$u_i(x_i) = P_i * \ln(x_i + 1) \tag{1}$$

where x_i denotes the allocated bandwidth resource, P_i is the traffic priority, the system utility function $U = \sum_{i=1}^{K} u(x_i)$ is the sum of all user utility values.

Then, we model the optimization problem of the bandwidth allocation in MSBN as follows:

$$\mathbf{P}: \qquad \max \sum_{i=1}^{K} u_i(x_i) \tag{2}$$

$$\text{s.t.} \qquad \mathbf{Ax} \leq \mathbf{C} \tag{3}$$

$$(\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}) \cup ((\mathbf{x} = \mathbf{0}) \cap (\mathbf{x} \in \mathbf{N})) \tag{4}$$

Equation (4) constraints that the allocated resource for the $user_i(user_i \in Q)$ should meet the minimum bandwidth request $a_i(a_i \in \mathbf{a})$ and less than the maximum demand $b_i(b_i \in \mathbf{b})$, and, at the same time, the allocated resource $x_i(x_i \in \mathbf{x})$ should be an integer. if it cannot meet the minimum demand, the system would not allocate resources for the user, which means that $x_i = 0$. Matrix $\mathbf{A}$ represents the connection between $l_j(l_j \in \mathbf{L})$ and $user_i(user_i \in Q)$, which meets

$$\mathbf{A} = \begin{cases} \mathrm{A}_{ji} = 1, & \text{the route path of } user_i \text{ through } l_j \\ \mathrm{A}_{ji} = 0, & \text{otherwise} \end{cases} \tag{5}$$

where $\mathbf{A} \in \Re^{M \times K}$ and the route path of $user_i$ is computed by Dijkstra algorithm [13].

As is shown in (3), the link bandwidth constraint condition is multi-dimensional, which is different from the one-dimensional constraint condition in the traditional model [3]–[10]. Notwithstanding the benefits stemming from this design, owing to the multi-dimensional of link capacity constraint in (3) and discontinuous constraint in (4), this optimization problem can not be solved with the traditional method. Inspired by non-cooperative game theory, a multi-pool resource allocation algorithm is proposed to solve this optimization problem.

III. The Multi-pool Resource Allocation Algorithm

Considering that the constraints of the original optimization in (2) is complex, we separate the optimization into two sub-problems to iteratively obtain the global optimization of the original problem.

A. Obtain Lagrange Approximate Solution

Given the optimization problem in (2) is a strictly increasing and convex function, if just the continuity constraint condition in (3) is considered, according to the Lagrangian duality theory [14] and the convex programming theory [15], this problem can be solved with Lagrangian multiplication. The Lagrange expression of optimization problem in (2) is

$$\mathcal{L}(\mathbf{x},\lambda)=\sum_{i=1}^{K}P_i\times\ln(x_i+1)-\lambda^{\mathrm{T}}(\mathbf{Ax}-\mathbf{C})$$
$$\text{s.t.}\qquad \lambda\geq\mathbf{0},\lambda\in\Re^{M\times1} \tag{6}$$

The derivative of (6) is as follows:

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}x_i}=\frac{P_i}{x_i+1}-\sum_{j=1}^{M}\lambda_jA_{ji}=0\quad(i=1,2,...,K) \tag{7}$$

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\lambda}=\mathbf{Ax}-\mathbf{C}=\mathbf{0} \tag{8}$$

Then we have:

$$x_i=\frac{P_i}{\sum_{j=1}^{M}\lambda_j\mathrm{A}_{ji}}-1\qquad(i=1,2,...,K) \tag{9}$$

$$\sum_{i=1}^{K}\mathrm{A}_{ji}x_i=C_j\qquad(j=1,2,...,M) \tag{10}$$

Inspired by the gradient descent method [16], we set the iterative equation as follows:

$$x_i^{n+1}=\max(0,\min(\frac{P_i}{\sum_{j=1}^{M}\lambda_j^n\mathrm{A}_{ji}}-1,\min(\mathrm{A}_{ji}C_j,C_j\in C)))$$
$$(i=1,2,...,K) \tag{11}$$

$$\lambda^{\mathbf{n+1}}=\max(\mathbf{0},\lambda^{\mathbf{n}}-\mathbf{r^n}\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\lambda^{\mathbf{n}}}) \tag{12}$$

where $\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\lambda^{\mathbf{n}}}=\mathbf{Ax^n}-\mathbf{1}$, and $\min(\mathrm{A}_{ji}C_j,C_j\in C)$ in (11) denotes the minimum remaining bandwidth resources in related links, which guarantees the convergence of the iteration. The initial step value of the iteration step is $\mathbf{r^0}$, $\mathbf{r^{n+1}}=\mathbf{r^n}*\mathbf{T}(\mathbf{n}=\mathbf{0},\mathbf{1},...)$, where $\mathbf{T}$ is the attenuation factor, which determines the iterative rate.

The initial iteration value $\begin{pmatrix}\mathbf{x^0}\\ \lambda^{\mathbf{0}}\end{pmatrix}$, which is the key of convergence rate, is analyzed in this paper. Let N_i be the number of links occupied by $user_i$, which obeys $N_i\sim U(1,M)$. According to the probability density function theory, the average number of links is $\frac{M+1}{2}$, thus we can set $x_i^0=\frac{2*\sum_{j=1}^{M}C_j}{(M+1)K}(I=1,2,...,K)$. Since the traffic priority P_i can be normalized to $P_i\sim U(0,1)$, we get $\lambda_j^0=\frac{K}{(M+1)K+2*\sum_{j=1}^{M}C_j)}(j=1,2,...,M)$ from (9).

Let the iterative error factor be $\Delta=\|\lambda^{i+1}-\lambda^i\|_2$ and the iterative precision be I_M, the iterative process ends until $\Delta\leq I_M$. After that, the Lagrangian approximation solution $\tilde{\mathbf{x}}=(\tilde{x}_1,\tilde{x}_2,...,\tilde{x}_K)^{\mathrm{T}}$ is obtained, and according to the strong duality theory in the convex optimization, the approximate optimal solution is unique.

B. Remove Maximum and Minimum Bandwidth Constraint

Considering the discontinuous constraint condition in (4), a non-cooperative theory based method is used to achieve the global optimal solution in this paper.

Firstly, assume that the set of users who meet $x_i=0$ is G_1, the set of game users is G_2, the set of users who game success is G_3, and initialize $G_1=\emptyset$, $G_2=Q$, $G_3=\emptyset$.

Definition 1. (Remove Maximum Bandwidth Constraint)

Obtain the Lagrangian approximate solution of all users in G_2 is $\tilde{\mathbf{x}}=(\tilde{x}_1,\tilde{x}_2,...,\tilde{x}_N)^{\mathrm{T}}$, for every $user_i(user_i\in G_2)$, if $\tilde{x}_i>b_i$, then remove $user_i$ from G_2 to G_3, and set the allocated bandwidth for $user_i$ be $x_i=b_i$, and update the remaining bandwidth of every resource pool $C_j(C_j\in\mathbf{C})$ be $C_j=(C_j-\mathrm{A}_{ij}b_i)$.

We define the process of removing all game success users from G_2 to G_3 above as a subroutine. Continue iteratively through this subroutine, until the Lagrangian approximate solution of every user in set G_2 meets $\tilde{x}_i\leq b_i(\tilde{x}_i\in\tilde{\mathbf{x}})$.

Definition 2. (Remove Minimum Bandwidth Constraint)

Execute the process of Remove Maximum Bandwidth Constraint in Definition 1, then obtain the Lagrangian approximate solution of all users in G_2 is $\tilde{\mathbf{x}}=(\tilde{x}_1,\tilde{x}_2,...,\tilde{x}_N)^{\mathrm{T}}$. Define a set $G_4=\emptyset$, for every $user_i(user_i\in G_2)$, if $\tilde{x}_i<a_i$, then copy $user_i$ from G_2 to G_4. Define the utility rate $R_i=\frac{u(a_i)-u(\tilde{x}_i)}{a_i-\tilde{x}_i}$, and the $user_m(user_m\in G_4)$ whose utility rate R_m is minimum is removed from G_2 to G_1, then set the allocated bandwidth for $user_m$ be $x_m=0$.

We define the process of removing the user with minimum utility rate from G_2 to G_1 above as a subroutine. Continue iteratively through this subroutine, until the Lagrangian approximate solution of every user in set G_2 meets $\tilde{x}_i\geq a_i(\tilde{x}_i\in\tilde{\mathbf{x}})$.

Theorem 1. The process of Remove Maximum Bandwidth Constraint in Definition 1 is the optimal solution for users in G_3.

Proof. According to the reduction to absurdity, we take a hypothesis as follows:

After the process of Remove Maximum Bandwidth Constraint, there are $user_i\in G_3$ and $user_j\in G_2$, which meet

$$u_i(b_i-1)+u_j(\tilde{x}_j+1)>u_i(b_i)+u_j(\tilde{x}_j) \tag{13}$$

The hypothesis in (13) could be reduced to

$$\frac{u_i(b_i)-u_i(b_i-1)}{b_i-(b_i-1)}<\frac{u_j(\tilde{x}_j+1)-u_j(\tilde{x}_j)}{(\tilde{x}_j+1)-\tilde{x}_j} \tag{14}$$

Since the utility function $u_i(x_i)=P_i*ln(x_i+1)$ is logarithmic, its derivative function is strictly monotonically decreasing. According to the Lagrange Mean Value Theorem, $\exists\ b_i>\xi_i>(b_i-1),(\tilde{x}_j+1>\xi_j>\tilde{x}_j)$, which meets:

$$u_i'(\xi_i)=\frac{u_i(b_i)-u_i(b_i-1)}{b_i-(b_i-1)} \tag{15}$$

 ISBN: 978-1-61208-545-6

$$u'_j(\xi_j) = \frac{u_j(\tilde{x}_j + 1) - u_j(\tilde{x}_j)}{(\tilde{x}_j + 1) - \tilde{x}_j} \tag{16}$$

From inequality (14), we get

$$u'_i(\xi_i) < u'_j(\xi_j) \tag{17}$$

However, with the process of Remove Maximum Bandwidth Constraint in Definition 1, the optimal solution obtained by the Lagrangian multiplication method in set G_2 meets

$$\frac{\mathbf{d}\mathcal{L}}{\mathbf{d}\tilde{x}_i} = u'_i(\tilde{x}_i) - \sum_{l=1}^{M} \lambda_l \mathrm{A}_{li} = 0 \tag{18}$$

We get

$$u'_i(x_i) = \sum_{l=1}^{M} \lambda_l \mathrm{A}_{li} = \sum_{l=1}^{M} \lambda_l \mathrm{A}_{lj} = u'_j(x_j) \tag{19}$$

Thus, due to the decreasing property of $u'_i(x)$ and $u'_i(x)$, then

$$u'_i(x_i) > u'_i(b_i) \geq u'_i(x_i) = u'_j(x_j) \geq u'_j(x_j) \tag{20}$$

Since the contradiction of (17) and (20), the hypothesis in (13) is false. Thus, we verify the validity of Theron 1.

Theorem 2. The process of Remove Minimum Bandwidth Constraint in Definition 2 is the optimal solution for users in G_1 and G_3.

Proof. Because the subroutine which finds the user with minimum utility rate R_i is a greedy algorithm [3], the optimal solution could be achieved for users in G_2 with the execution of subroutine.

We define the system utility of i-th subroutine in Definition 2 as U^i_{opt} and the iterative time as n.

When $n = 1$, the system utility U^1_{opt} could achieve maximum after the execution of the first subroutine. And we have $U^1_{opt} > U^0_{opt}$.

When $n = k$, the optimal solution could be achieved for users in G_3 with the process of Remove Maximum Bandwidth Constraint, and for users in G_2 with the process of finding the minimum utility rate R_i. Thus, the maximum value of the system utility U^k_{opt} could be achieved with the process of k-th subroutine, and we have $U^k_{opt} > U^{k-1}_{opt}$.

The iteration of this subroutine is completed until all users in G_2 satisfy $a_i \leq \tilde{x}_i \leq b_i$. The iterative time is denoted as N, and the system utility is U^N_{opt}, which meets

$$U^N_{opt} > U^{N-1}_{opt} > ... > U^k_{opt} > ...U^1_{opt} > U^0_{opt} \tag{21}$$

Since the iteration of the subroutine is completed, with the conclusion in (21), thus, we verify the validity of Theorem 2.

With the theoretical analysis above, the algorithm of solving the optimization problem in (2) could be expressed in Algorithm 1.

Algorithm 1 The Multi-Pool based Resource Allocation algorithm

Input: The parameter of set Q for users; the maximum iterative time I in Definition 2.
Output: The allocated bandwidth $\mathbf{x}$ for users in set Q.

Set $G_1 = G_3 = \emptyset$, $G_2 = Q$;
$f_{min} = 0$;
for $i = 1 \rightarrow I$ **do**
  if $f_{min} = 0$ **then**
    Execute the process of Removing Maximum Bandwidth Constraint in Definition 1.
    Obtain the Lagrangian approximate solution in set G_2.
    Set $G_4 = \emptyset$, and copy all users who meet $\tilde{x}_i < a_i$ from G_2 to G_4.
    if $G_4 \neq \emptyset$ **then**
      Find the $user_m$ with the minimum value of the utility rate R_m in G_4, and remove $user_m$ from G_2 to G_1, set the allocated bandwidth for $user_m$ be $x_m = 0$.
    else
      $f_{min} = 1$;
    end if
  else
    break;
  end if
end for
Set the allocated bandwidth for users in G_2 as $x_i = \lfloor \tilde{x}_i \rfloor$, where $\lfloor \tilde{x}_i \rfloor$ is the rounded down value of $\tilde{x}_i$.

C. Complexity Analysis

Let K be the maximum number of iterations in the gradient descent process, so the complexity in solving the Lagrangian multiplication is $O(K)$. The complexity of the process of Remove Maximum Bandwidth Constraint in Definition 1 is $O(KN_Q \log_2 N_Q)$. As the subroutine of Remove Minimum Bandwidth Constraint in Definition 2 contains the process of Remove Maximum Bandwidth Constraint in Definition 1, the complexity of Remove Minimum Bandwidth Constraint in Definition 2 is $O(KN_Q \log_2 N_Q \times N_Q \log_2 N_Q) = O(K(N_Q \log_2 N_Q)^2)$.

In summary, the complexity of the proposed MPBA algorithm is $O(K(N_Q \log_2 N_Q)^2 + N_Q N_L)$, which satisfies the requirement of the low computational complexity and the real-time bandwidth allocation.

IV. Simulation Analysis

A. Simulation Scenarios and Parameter Settings

As is shown in Figure 2, an OPNET based simulation scenario of MSBN is built to verify the MPBA scheme in this paper.

The constellation of this scenario is composed of six MEO satellites, with the height of 12,800km and 0° inclination, which means that the propagation delay of satellite-ground link is $56ms$ and inter-satellite link is $64ms$. For each satellite, they contain 4 satellite-ground links and 2 inter-satellite links. For every satellite-ground link, there are 20 users to be allocated bandwidth resources. Traffic parameters of these users are shown in Table I. The conversation and streaming service

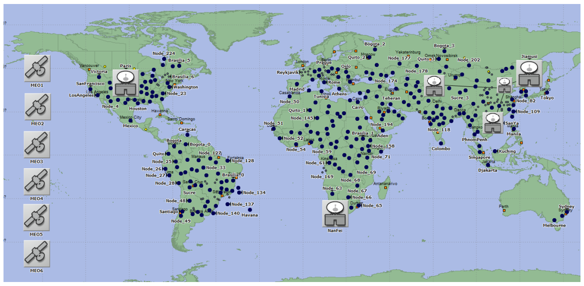


Figure 2. The Simulation Scenario of MSBN

are based on IP protocol, while interactive and background service are based on TCP protocol. We set that the number of inter-satellite route hops for every user is 3, so the inherent propagation delay for every user is $56 \times 2 + 64 \times 3 = 304ms$.

In the process of solving the Lagrangian relaxation solution, the initial iteration step of the projection gradient descent method $\mathbf{r^0 = 0.5}$, the attenuation factor $\mathbf{T = 0.75}$.

TABLE I. THE TRAFFIC PARAMETERS OF USERS

User ID	Traffic Type	Priority	Requests(bps)	Minimums(bps)
1-5	Conversation	4	25,600	25,600
6-10	Streaming	3	25,600	12,800
11-15	Interactive	2	19,200	9,600
15-20	Background	1	6,400	3,200

B. Simulation Results

The performance of the end-to-end delay and the system utility in the traditional single resource pool framework based bandwidth allocation algorithm [4] and our MPBA algorithm are analyzed in this section.

Let C_{s-g} and C_{s-s} be the bandwidth capacity of the satellite-ground link and the inter-satellite link for every satellite.

For $C_{s-g} = 307,200bps$, which means that bandwidth resources of satellite-ground links are adequate for bandwidth requests of 20 multimedia users. Take the bandwidth capacity of inter-satellite C_{s-s} as a variable, which changes from $7 \times C_{s-g}$ to $13 \times C_{s-g}$, the performance of the end-to-end delay is shown in Figure 3. When $C_{s-s} \geq 12 \times C_{s-g}$, the resource capacity of inter-satellite links are sufficient to meet maximum bandwidth requests of 20 users, so the end-to-end delay is equal to the inherent propagation delay $(304ms)$ for both the traditional algorithm and the MPBA algorithm. However, when $C_{s-s} < 12 \times C_{s-g}$, it shows that the MPBA improves significantly on the end-to-end delay over the traditional algorithm since limited resource capacities of inter-satellite links are taken into consideration in the process of the bandwidth allocation. Moreover, the more scarce the resource of the inter-satellite link is, the most significant the end-to-end delay improvement could be in the MPBA scheme.

The performance of the end-to-end delay for $C_{s-g} = 204,800bps$, in which the resource capacity of every satellite-ground link meets minimum bandwidth requests of 20 users, is shown in Figure 4. And $C_{s-g} = 153,600bps$, in which resource capacities of satellite-ground links could not meet the minimum bandwidth request of every user, is shown in Figure

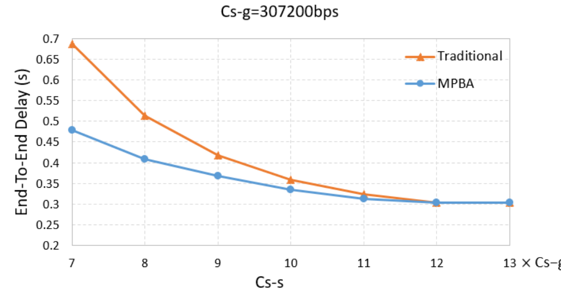


Figure 3. The End-to-End Delay for $C_{s-g} = 307,200bps$

5. Both Figure 4 and Figure 5 verify that the MPBA scheme outperforms over the traditional algorithm on the end-to-end delay for MSBN system.

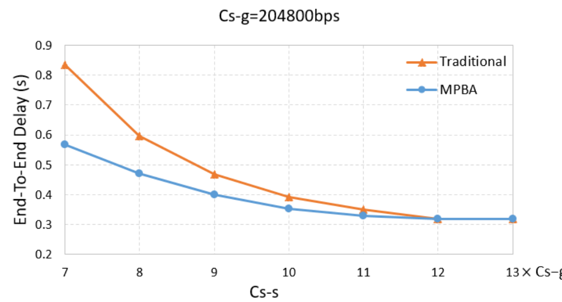


Figure 4. The End-to-End Delay for $C_{s-g} = 204,800bps$

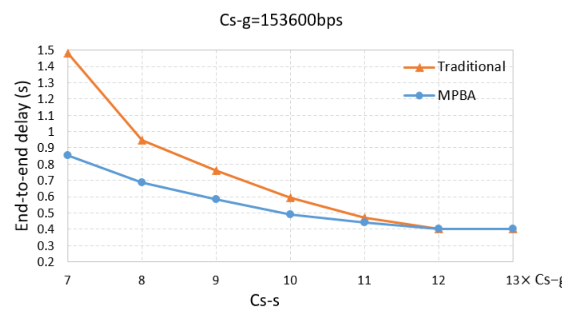


Figure 5. The End-to-End Delay for $C_{s-g} = 153,600bps$

The system utility, which is computed in (1), for $C_{s-g} = 307,200bps$ is shown in Figure 6. Thanks to the consideration of the limited resource capacity in every link, the resource manager would not accept too much network traffics in the MPBA scheme. However, the traditional algorithm would accept as much traffics as the satellite-ground link could afford, with the ignorance of traffic carrying capacities on inter-satellite links, which could cause congestion in the system. For lossless protocol based traffics, there would be much retransmission in the traditional algorithm, which results in the repeated occupancy of system resources and the aggravation of congestion. As a result, the system utility of the traditional algorithm would be lower than our MPBA scheme. As is shown in Figure 6, the less resource capacities of inter-satellite links, the more improvement on the system utility of the MPBA algorithm than the traditional one.

 ISBN: 978-1-61208-545-6

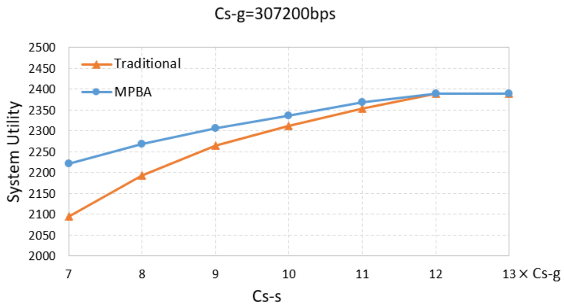


Figure 6. The System Utility for $C_{s-g} = 307,200bps$

V. Conclusion

The contributions of this paper are a multi-pool framework for MSBN system along with the game theory based bandwidth allocation algorithm that takes different QoS bandwidth demands of multimedia services into account.

Unlike traditional single satellite-ground link based approaches, the dynamic bandwidth capacities of whole links are taken into consideration in the MPBA scheme, with satellite-ground links and inter-satellite links included. This results in the effective improvement in the end-to-end delay, increased the robustness to the dynamic change of bandwidth resources in different links. Then, depending on the proposed multi-pool framework, the game theory based bandwidth allocation algorithm takes the different bandwidth demands of multimedia services into consideration, which meets the requirement of the low computational complexity of the on-board process.

Comparing with the traditional algorithm, simulation results show that a good usage of the system resource and a significant improvement of the end-to-end delay could be achieved. Moreover, the more unbalanced the resources of inter-satellite links and satellite-ground links are, the most significant the improvement could be in the MPBA scheme.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No.91438205, No.61271281).

References

[1] R. Radhakrishnan, W. Edmonson, F. Afghah, R. Rodriguez-Osorio, and F. Pinto, "Survey of inter-satellite communication for small satellite systems: Physical layer to network layer view," arXiv preprint arXiv:1609.08583, 2016.

[2] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on smart grid communication infrastructures: Motivations, requirements and challenges," IEEE communications surveys & tutorials, vol. 15, no. 1, 2013, pp. 5–20.

[3] H. Yaïche, R. R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," IEEE/ACM Transactions on Networking (TON), vol. 8, no. 5, 2000, pp. 667–678.

[4] A. Morell, G. Seco-Granados, and M. Á. Vázquez-Castro, "Cross-layer design of dynamic bandwidth allocation in dvb-rcs," IEEE Systems Journal, vol. 2, no. 1, 2008, pp. 62–73.

[5] X. Sun, Q. Zhang, X. Xin, and C. Yu, "Cross-layer dynamic bandwidth allocation based on fairness and system utility," in Instrumentation, Measurement, Computer, Communication and Control (IMCCC), 2012 Second International Conference on. IEEE, 2012, pp. 1322–1325.

[6] A. Morell, G. Seco-Granados, and M. A. Vazquez-Castro, "Computationally efficient cross-layer algorithm for fair dynamic bandwidth allocation," in Computer Communications and Networks, 2007. ICCCN 2007. Proceedings of 16th International Conference on. IEEE, 2007, pp. 13–18.

[7] B. de la Cuesta, L. Albiol, J. M. Aguiar, C. Baladrón, and B. Carro, "Innovative dama algorithm for multimedia dvb-rcs system," EURASIP Journal on Wireless Communications and Networking, vol. 2013, no. 1, 2013, pp. 1–14.

[8] H. Feng, Y. Liu, H. Hu, D. Yuan, and J. Ran, "A hierarchical resource allocation algorithm for satellite networks based on mf-tdma," in Mechatronics, Materials, Chemistry and Computer Engineering (ICMMCCE), 2015 4th International Conference on. Atlantis Press, 2015, pp. 2819–2824.

[9] I. Bisio, S. Delucchi, F. Lavagetto, and M. Marchese, "Comparison among resource allocation methods with packet loss and power metrics in geostationary satellite scenarios," in 2013 IEEE International Conference on Communications (ICC). IEEE, 2013, pp. 4271–4275.

[10] S. Vassaki, A. D. Panagopoulos, and P. Constantinou, "Game-theoretic approach of fair bandwidth allocation in dvb-rcs networks," in Satellite and Space Communications, 2009. IWSSC 2009. International Workshop on. IEEE, 2009, pp. 321–325.

[11] I. Bisio and M. Marchese, "The concept of fairness: definitions and use in bandwidth allocation applied to satellite environment," IEEE Aerospace and Electronic Systems Magazine, vol. 29, no. 3, 2014, pp. 8–14.

[12] J. E. Xu, R. Torres, J. Border, and Y. Li, "Load balancing for return satellite channels with multiple traffic classes," in MILCOM 2013-2013 IEEE Military Communications Conference. IEEE, 2013, pp. 1069–1074.

[13] S. Skiena, "Dijkstras algorithm," Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica, Reading, MA: Addison-Wesley, 1990, pp. 225–227.

[14] Y. Yi and S. Shakkottai, "Hop-by-hop congestion control over a wireless multi-hop network," IEEE/ACM Transactions On Networking, vol. 15, no. 1, 2007, pp. 133–144.

[15] S. Boyd and L. Vandenberghe, Convex optimization. Cambridge university press, 2004.

[16] S. H. Low and D. E. Lapsley, "Optimization flow controli: basic algorithm and convergence," IEEE/ACM Transactions on Networking (TON), vol. 7, no. 6, 1999, pp. 861–874.

 ISBN: 978-1-61208-545-6

DSN Wide Area Network Architecture, Capacity and Performance

Timothy Pham, Roger Cortez, Jason Liao
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA
e-mails: Timothy.Pham@jpl.nasa.gov, Roger.Cortez@jpl.nasa.gov, Jason.Liao@jpl.nasa.gov

***Abstract* - This paper discusses the architecture of the wide area network that connects key communications facilities within the National Aeronautic and Space Administration (NASA) Deep Space Network (DSN). Several considerations are given to the design of this wide area network to ensure a timely, reliable, and secure data delivery between the mission users and their spacecraft. The network star configuration simplifies data delivery to users and minimizes operational cost. The dual-path connections maximize the system reliability, with geographical diversity in data routing to avoid single point of failure. Data encryption enhances the protection of mission users' data. The system bandwidth is determined by balancing the needs to minimize the operating bandwidth cost and to have sufficient bandwidth to be able to deliver data to all users within the required. The DSN uses a class base weighted fair queuing (CBWFQ) method in its data delivery. This scheme guarantees a minimum bandwidth to each class of users and allows users to also access any unused bandwidth by other groups. The paper will also show the performance of system reliability and bandwidth margin.**

Keywords - DSN; network architecture; network design

I. INTRODUCTION

The Deep Space Network (DSN), an infrastructure of the National Aeronautics and Space Administration (NASA), is a global network that enables communications with far-flung spacecraft exploring the outer space. Its three main tracking facilities – Goldstone, Canberra and Madrid - are spread equidistance across the Earth, at both northern and southern hemisphere, to enable constant contact between Earth's mission controllers and their spacecraft. Antennas at each of these facilities enable communications with 35-plus spacecraft currently in operations. Data are exchanged in both directions - in the forward links where commands are sent to spacecraft and in the return links where spacecraft's housekeeping data and collected scientific data are sent back to Earth. In addition to spacecraft command and telemetry, radiometric measurements of spacecraft position and velocity, i.e., ranging and Doppler data also flow through this wide area network to mission navigation teams for orbit determination. Data collected at the tracking facilities also include observations/measurements for radio astronomy, planetary radar, and radio science. As shown in Figure 1, the three tracking facilities are connected to the Deep Space Operations Center (DSOC) at the Jet Propulsion Laboratory (JPL) in Pasadena, California via a wide area network, which is the focus of this paper. The DSOC facility in turn connects to mission operations centers (MOC) and science investigators.

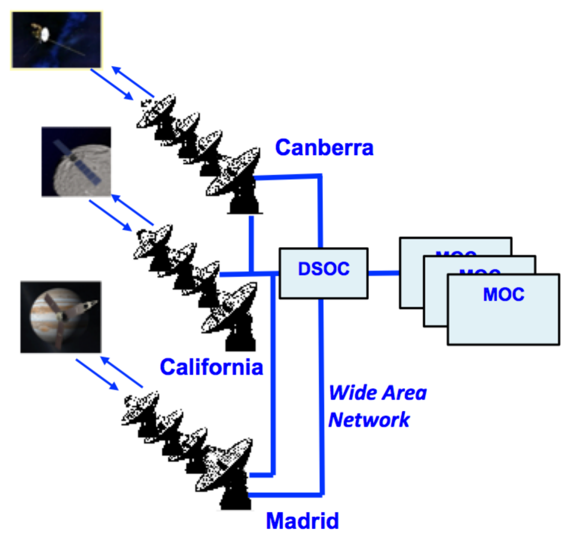


Figure 1. Topography of the DSN Wide area Network

Section II will discuss the architecture of the DSN wide area network. Several design considerations are important to maximize the system resiliency against possible outages and cyber security. The network capacity, driven by the needs of mission users, science users and DSN operations, are discussed in Section III. A scheme for fair allocation of bandwidth among users is described in Section IV. Section V provides a gleam of system performance in terms of bandwidth usage and margin, and connection reliability.

II. ARCHITECTURE

DSN engineers working with the NASA Communications Services Office (CSO) design the DSN wide area network architecture. The network connections are implemented, maintained and operated by a telecommunications company who has contract with the NASA Communications Service Office. The US-based prime contractor, in turn, works with domestic providers who have circuit access to Goldstone, and with international providers who have access to the other two overseas facilities at Madrid and Canberra. Past architecture relied on point-to-point dedicated circuits, based on Time Division Multiplex (TDM) technology. In the current architecture, the connections are done through the provider's IP VPN

 ISBN: 978-1-61208-545-6

(Internet-protocol virtual private network) solutions that run over their multi-protocol label switching (MPLS) backbone. Each site (DSOC, Canberra, Madrid and Goldstone) is connected to the carrier's MPLS backbone via dual access circuits. With all sites being connected to the MPLS backbone, virtual point-to-point circuits are created for connections between the three tracking facilities and DSOC.

This architecture offers a greater redundancy inherent in the network backbone, more flexibility to set virtual circuit for any-to-any node connection, no physical change required when there is a need to increase the capacity, and a lower cost for equivalent bandwidth compared to traditional TDM services. The drawbacks include the loss of end-to-end visibility and monitoring. The physical route across the MPLS backbone is not determinable, making it hard to verify the route diversity within the cloud. Note that although with the cloud, the realizable bandwidth is still limited to the physical capacity of the access circuits that connect the tracking stations to the cloud (i.e., last mile).

Star Configuration - The topology of the DSN wide area network resembles that of a star configuration where the three tracking facilities (Goldstone, Canberra, Madrid) are connected to a central DSOC node at the Jet Propulsion Laboratory, Pasadena, California. From DSOC, data are then delivered to external mission operations centers. This star configuration has both benefits and drawbacks. On the benefits, it provides a single-node connection to the mission users, regardless of where the tracking pass takes place. It simplifies the functions normally done at DSOC such as packet extraction and data interfacing with mission users. With a centralized configuration, the equipment only needs to be deployed once and operated by one team, reducing the hardware and operation costs. On the drawbacks, it makes the wide area network becomes a critical element to DSN operations and thus, requires funding to sustain it. Nowadays with Internet access being widely available and cheaply, a non-star configuration may offer a possible lower cost alternative. In such scenario, the DSN would deliver data to users directly from the tracking facilities. The users would connect to those tracking facilities. The cost of such connection would be born by the users, but not the DSN. The overall cost for NASA, however, may not change much since NASA pays for both mission operations cost and DSN operations cost. The second drawback, albeit minimal impact, is that data access and delivery are less direct. For example, a mission operation team for one of European missions, instead of being able to get to its mission data as soon as the data are received at Madrid, would need to wait for data to travel from Madrid to California before coming back to Europe. This delay impact is rather insignificant – an extra ~200 millisecond delay is quite small compare to the delay of several minutes/hours associated with signal travel in space before reaching Earth. System reliability is also a bit lower because there is more transferring nodes in the data delivery; however, in practice, the difference in reliability is quite small due to the built-in redundancy of the DSN wide area network, which is discussed next.

Redundancy - The tracking facilities are typically located in remote areas to minimize potential radio frequency interference (RFI) with terrestrial sources from nearby cities. As a result, the connections from the tracking site to the telecommunication provider's nearby point of presence are often of long distance, making it being both expensive and limited in choices of routing. In the past, even though the dual connections from the site to the nearby city were routed via two distinct fiber optic cables, both cables often ran through the same fiber bundle that extended over many tens of kilometers. There were instances where the roadwork or power crews accidentally cut both fiber optic links during construction and maintenance activities. Since these links are in remote area, recovery from outages often took many hours, and at times days. To avoid these problems, it is critical that the routing of signals from the tracking facility to DSOC is geographically distinct. The insulation between the two links starts at the tracking facilities where there are two equipment racks provided by vendor. The signals are routed via two nearby towns to two different metropolises, and in the case of Canberra and Madrid, traversing on two different undersea cables prior to reaching the United States. The VPN circuits arrive to JPL at two different entry points and terminate in two different buildings. All of these precautions help to ensure that even if there were a large-scale natural or man-made disruption affecting one of the cities or metropolis that one of the links traverses, the other link would remain operational. As shown in Figure 2, the Madrid tracking facility has one connection through the nearby village (Node 1) onto Madrid. The other connection is through a different village (Node 2) and onto Barcelona. Similarly, Canberra links are routed such that one going through Sydney and the other through Melbourne. The current dual paths from Goldstone have one common point of presence in a nearby town Victorville; however, with the effort spent over the past few years, another fully diverse route bypassing Victorville will become operational in 2017.

Between the two links, the prime path is normally the direct connection between Canberra/Madrid and DSOC. The backup path is routed via Goldstone. The primary driver for having the backup Canberra/Madrid connection via Goldstone is for the DSN Emergency Control Center. The ECC will be activated if DSOC is not available, and having direct links between Goldstone and Canberra/Madrid sites means that these two sites can maintain communication with the Emergency Control Center even if the DSOC node is down. As a result, the Goldstone-DSOC link is expected to carry Goldstone-own traffic and either Canberra or Madrid traffic in the event of failure of the prime connection at one of these two sites. As seen in Figure 2, the bandwidth capacity of Goldstone facility is greater than that of Canberra and Madrid. The capacity for Goldstone link is scoped to accommodate a failure in the prime link at either Canberra or Madrid, but not both. Such limitation is a tradeoff of bandwidth subscription cost versus the low probability of simultaneous failures at Canberra and Madrid.

Security - As indicated earlier, the information being sent to and received from spacecraft traverses through the provider's network backbone. It is important that the data

 ISBN: 978-1-61208-545-6

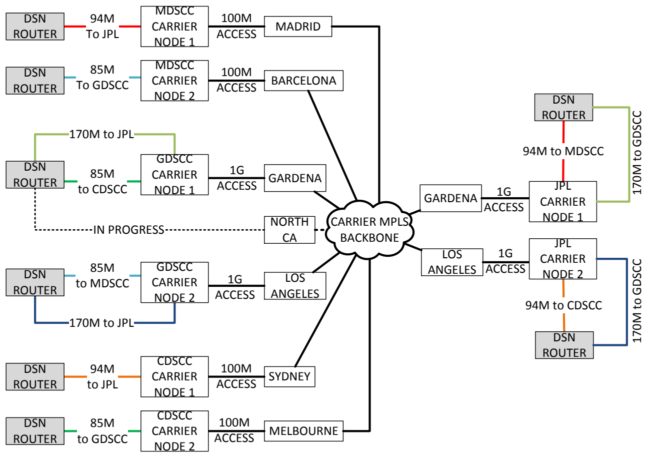


Figure 2. Dual path connections from tracking facilities to DSOC

confidentiality and integrity are not compromised. To further protect this information, beyond the vendor's IP VPN, encryption is applied to the data traverses over the DSN wide area network. IPSec tunnels are created between the sites, using the NSA Suite B Algorithms. In addition, the TCP maximum segment size (set to 1330 bytes) is being interjected by DSN routers to ensure that all frames for TCP connections stay under 1500 bytes to prevent fragmentation. This was required to accommodate the IPSec overhead. The encryption nominally adds about 6% bandwidth overhead.

III. BANDWIDTH CAPACITY

The consideration on bandwidth capacity is determined by two factors: sufficiently meet the need of users and incur minimum cost. Because of the remote location of the tracking facilities and the high availability requirement (99.95%) to meet mission critical need, the bandwidth cost for DSN wide area network is much higher than what is typically encountered in commercial business or residential rates. As such, it is important to correctly size the required bandwidth so that NASA would not have to pay the extra cost for idle bandwidth. Reference [2] provided a detailed description on the method of estimating the required bandwidth. In this paper, the estimated bandwidth reflects the greater demand from some recent missions with higher data rates.

There are many users of the DSN wide area network. In general, they can be divided into three groups – mission users, science users and DSN operations.

Mission Users – Telemetry data return from missions' spacecraft drives the required bandwidth because of the low latency requirement for proper mission operations. Command data to spacecraft, which flow in reverse direction from telemetry, and radiometric data needed for mission navigation consume much lower bandwidth. Most missions are tracked daily; thus, data that are downlinked on a given day need to be delivered within 24 hours so that the DSN would be cleared and ready for next day track. For some missions, the required latency is much shorter – in a few hours – so that mission operations team can determine the activity for the next day or for the follow-on tracking pass at the next tracking facility. Prior to 2017, the highest data rate among the current missions being supported by the DSN is 6 Mbps from the Mars Reconnaissance Orbiter. The aggregated downlink data rate for all missions tracked by various antennas within a tracking facility is typically 8 Mbps or less. In late 2017, the data rate will be significantly increased with the upcoming Terrestrial Exoplanet Survey Satellite (TESS) mission at 125 Mbps [3]. TESS tracking schedule is however infrequent, occurring roughly once every two weeks. The tracking passes for TESS is also rather short, ~2.5 hrs a day. The data collected over 2.5 hrs of TESS downlink pass can be delivered over a 24 hrs window at a much lower rate of 13 Mbps. In late 2018, another high rate mission - the James Webb Space Telescope (JWST) - will be launched. Its 28 Mbps downlink, although smaller compared to TESS, require a real-time data delivery [4]. The bandwidth driven by JWST is 28 Mbps. Looking forward into the future years in early or mid 2020's, there are likely missions with data rate upward of 150 Mbps, and even 300 Mbps. These missions will then drive a need for greater capacity in the DSN wide area network.

Science Users - Besides supporting standard telemetry, tracking and command functions in communications with scientific spacecraft, the DSN also serves as a science instrument platform for planetary radar science, radio astronomy, space geodesy via Very Long Baseline Interferometry (VLBI), and radio science. These observations typically require high-rate sampling of the wide-band received signal to extract information in the received signal spectrum. As a result, these science observations generate very high data volume. Fortunately, these observations are done less frequent (every few days or few weeks), the latency requirement is not as stringent as telemetry data. Currently, radio science and radar science recordings could be relayed via the network, but data from VLBI and radar science observations need to be delivered off-line via shipping of recorded disks. A better operation mode is to deliver data via the network. It would increase the timeliness of data access and reduce the burden of manual shipping of recorded data. In the current capacity, 15 Mbps is booked as reserved for science data transfer.

DSN Operations – Beside data delivery to missions and science users, there are other traffic flows necessary for DSN operations. Monitor data reflecting the performance and configuration of the ground equipment, and conditions of the received signals need to be delivered to the DSN Operations at DSOC and passed onto mission control centers.

Up to now, each tracking facility is responsible for the operations of equipment at its site. Toward the end of 2017 the DSN will be moving to a new operational concept called follow-the-sun operation in which one tracking facility (Goldstone, Canberra, or Madrid) would take control of the entire DSN operations on a rotating basis [5]. To support such operation, video streams of antennas that used to be flown internal within the facility will need to be routed through the wide area network to another facility in order to

 ISBN: 978-1-61208-545-6

provide assurance of proper operations of the remote antennas thousands of kilometers away. This monitoring traffic is estimated at 20 Mbps.

In addition to monitor data and antenna video streams, a small portion of the bandwidth is needed for voice-over-IP traffic that is essential to maintain communications among operation teams at different facilities, computer network monitoring (via simple network management protocol), and occasional remote logins for system testing and diagnostic. This type of traffic is estimated about 5 Mbps.

In total, the near-term bandwidth need is estimated to be ~100 Mbps. This capacity has been recently implemented; enables the DSN to support upcoming TESS and JWST missions.

IV. BANDWIDTH ALLOCATION AND ACCESS

Bandwidth allocation for DSN WAN traffic is done with a class base weighted fair queuing (CBWFQ) method [1]. Of the total bandwidth available, each class of users is given a minimum guaranteed allocation. The minimum allocation guarantees that the users within a specific class will have that bandwidth available to them, should they need it. If data flows within one class are below the allocation, the unused bandwidth then becomes available to users in other classes. Conversely, if the bandwidth demand within one class is more than the minimum allocation and there is unused bandwidth elsewhere, the users would be able to access it. Certain data traffics that are sensitive to delay (e.g., voice over IP) could be placed in a low-latency queue, which lets the data be sent out first.

Out of the current 100 Mbps capacity, telemetry allocation is guaranteed about 60 Mbps. Science users are given about 15 Mbps and DSN operations about 25 Mbps. Since science users typically desire to have access to greater bandwidth, an option to maximize the WAN resource is to flow science data via the available backup link that is normally idle.

V. PERFORMANCE

To ensure that system achieves its performance as required, monitoring capability is built into the system to collect necessary metrics for assessing the connectivity and bandwidth utilization.

Figure 3 shows how network monitoring is done in the DSN. As telemetry and science data comes to JPL from Goldstone, Canberra and Madrid, a copy of data packet is routed to the Network Monitor server for traffic loading analysis. Based on the frequency and volume of incoming data packets for each class of traffic, the Network Monitor creates the network loading profile.

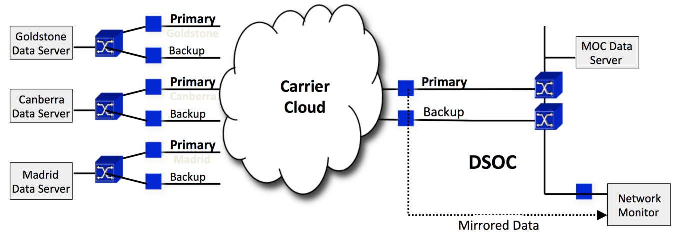


Figure 3. Peak flow of different data traffic

To monitor the status of link connection, the Network Monitor server pings the routers at Goldstone, Canberra and Madrid every 15 second. If there were a failure in the current active signal path, the router would automatically switch the data routing to the redundant path. Thus, the ping over the failed link would not be successful. Figure 4 shows a sample of IP connectivity over one month period for one of the Goldstone-JPL links. The blue line shows the actual availability, which is nearly 100% available. The red line shows the required 99.5% availability. Because of the dual-path redundancy, even if one of the paths is down, the overall service availability remains high at 100% level.

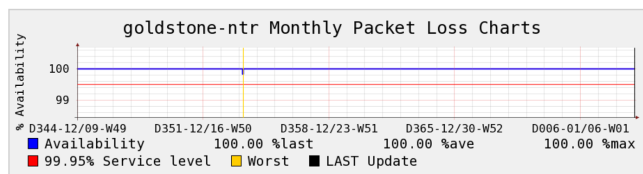


Figure 4. Service availability of a sample Goldstone link

On bandwidth utilization, Figure 5 shows the current average loading of ~15 Mbps (prior to the launch of the high-rate TESS and JWST missions). That represents a 15% usage or 85% margin of the link capacity. We expect however that the bandwidth usage will be significantly increased toward the end of 2017 when TESS mission is launched, followed by another increase with JWST launch in 2018.

One may also notice in Fig. 5 that different classes of traffic have different peak flow. Telemetry data return, shown in blue, is typically constrained by the spacecraft downlink data rates. Currently, the maximum downlink rate for a single mission is 6 Mbps, which is quite small compared to the link capacity. Toward the end of 2017, TESS will enter its mission operational phase and will downlink data at 125 Mbps. At that rate, telemetry data will flow at the maximum link capacity during the tracks and a few hours at the end of the pass. TESS tracking occurs infrequent, typically every 14 days. In between the two TESS passes, the maximum telemetry data flow - including JWST -would be at a lower peak, around 30 – 35 Mbps.

 ISBN: 978-1-61208-545-6

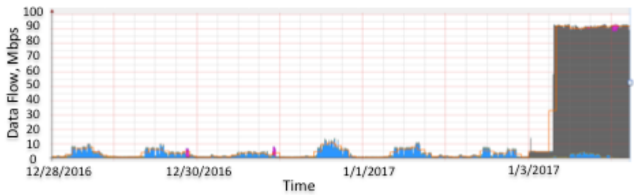


Figure 5. Peak flow of different data traffic

Science data, shown in dark gray curve, always reaches the maximum bandwidth of the wide area network. It is because science observations for planetary radar and radio astronomy studies typically involve the high-rate recording of wideband RF spectrum. The sampling data usually exceeds the capacity of the WAN.

VI. CONCLUSION

In summary, this paper discusses the characteristics of the wide area network of the Deep Space Network. Delivery of spacecraft telemetry to mission users is done via a single node at JPL. Features such as geographically distinct routing between the prime and backup path help to avoid a single point of failure and improve system reliability. Encryption helps to protect the information integrity, particularly relevant to the command data that are critical to spacecraft operations. The capacity of the WAN is set based on the balance between being able to deliver data to users within the required latency and minimizing the WAN cost. Bandwidth is allocated among users via class base weighted fair queuing scheme. The system has been recently upgraded to 100 Mbps capacity in preparation for upcoming support to a few high rate missions such as TESS and JWST. The current average loading at 15 Mbps shows that there is good margin of system capacity. The margin reflects DSN readiness to support upcoming missions in the next few years.

ACKNOWLEDGMENT

The work described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

[1] S. Floyd and V. Jacobson, “Link-sharing and Resource Management Models for Packet Networks,” IEEE/ACM Transactions on Networking, Vol. 3 No. 4, August 1995, pp. 365-386.

[2] T. Pham, “A Cost Effective Modelling of Near-Term Mission Bandwidth Demand”, 6th International Symposium on Reducing Cost of Space Ground Communications Systems Operations Symposium, Darmstadt, Germany, September 2005

[3] TESS – Transiting Exoplanet Survey Satellite, NASAfacts, National Aeronautics and Space Administration, NASAfacts, https://tess.gsfc.nasa.gov/documents/tess_factsheet_06-05-15.pdf, accessed March 15, 2017

[4] JWST Mission Operations Concept Document, Space Telescope Science Institute, January 13, 2006, http://spacese.spacegrant.org/JWST_Mission_Operations_Concept_Document.pdf , accessed March 15, 2017

[5] M. Johnston, M. Levesque, S. Malhotra, D. Tran, R. Verma, S. Zendejas, “DSN Automation Improvements in the Follow-the-Sun Era”, International Joint Conference on Artificial Intelligence Workshop on Artificial Intelligence in Space (AI Space, IJCAI 2015), Buenos Aires, Argentina, July 2015

 ISBN: 978-1-61208-545-6