Routing of VoIP traffic in multi-layered Satellite Networks

Durga Shankar Dash[a], Arjan Durresi*[b] and Raj Jain[b]
[a]Department of Electrical Engineering
[b]Department of Computer and Information Science
The Ohio State Unniversity
Columbus, Ohio, USA
*Durresi@cis.ohio-state.edu; phone 1 614 688 5610; fax: 1 614 292 2911

ABSTRACT

Broadband satellite constellation networks will be required to carry all types of IP traffic, real time interactive traffic as well as non-real time traffic, warranting the need for appropriate QoS for these different traffic flows. In this paper we investigate advantages of employing constraint-based routing using MPLS in a multilayered hierarchical satellite constellation. Bandwidth availability or residual bandwidth on a satellite link is taken into account when setting up routes for high priority real-time traffic e.g. VoIP, which is sensitive to delay and jitter. Also to protect the VoIP traffic from being swamped by bursty best-effort traffic we propose to have a separate queue for high priority traffic. The performance of the prioritized load balancing routing algorithm on a multi-layered satellite network is simulated and analyzed.

Keywords: VoIP routing; load balancing routing; QoS; HAPs; LEOs; GEO.

1. INTRODUCTION

Satellite networks with their potential for global coverage and high bandwidth availability are an attractive option for establishing an “internet in the sky”. In the absence of wired infrastructure, any individual host or an ad hoc network can access the rest of the wired network through satellites. Satellite networking has evolved from simple bent-pipe routing for geo-stationary orbit (GEO) satellite networks to on-board switching capabilities in low earth orbit (LEO) broadband satellite constellations and High Altitude Platforms (HAPs).

The next generation of networking will involve integration of terrestrial and space networks with High Altitude Platforms(HAPs) providing last mile connectivity to certain sensitive areas(e.g. disaster relief, battlefields) where high bandwidth and accessibility are necessary. With the advance of free space optical technology it is possible to have a network of these HAPs, which could be unmanned aerial vehicles, communicating with each other as well as with satellites through Inter Orbit Links (IOL). With radio access they make it possible for small hand-held terminals on the ground to access satellites with high speed connections [16,17].

Multi-layered satellite architectures with optical Inter orbit Links(IOLs) between layers of satellite constellations are of much interest as they yield much better performance than individual layers. In [12] a two layered architecture and a routing algorithm are proposed. A connectionless packet IP routing algorithm on a three layered satellite network consisting of GEOs, MEOs and LEOs is proposed in [11]. One of the main issues for satellite networks is to provide high levels of guaranteed Quality of Service (QoS) to priority traffic.

Given the nature of satellite constellations, static routing protocols based on topological properties of the network like minimum hop path or minimum delay will give rise to congestion at some points in the network (e.g. satellites that are visible over major cities) and a lot of unused bandwidth at other points in the network leading to under-utilization of network resources and degradation of service offered. The number of active sessions in such networks is unlimited and no kind of priority policy is implemented. Consequently, high priority calls are routed over paths heavily used by low priority calls.

Performance and Control of Next-Generation Communications Networks, Robert D. van der Mei,
Frank Huebner, Editors, Proceedings of SPIE Vol. 5244 (2003) © 2003 SPIE · 0277-786X/03/$15.00

MPLS has evolved as an IP based QoS architecture, though originally developed for IP over ATM integration. It combines the traditional datagram service with the virtual circuit approach. Basically a LSP is a virtual circuit and allows the setup of explicit routes for packets of a class, a critical capability for traffic engineering. Similar to the DiffServ approach the ingress LSP at the edge of the network classifies packet to classes and sets the initial label. MPLS also supports bandwidth reservation for classes, again enforced by a packet scheduler. MPLS along with constraint routing provides us with the option of routing over non-shortest paths subject to constraints of bandwidth and delay.

In this paper we propose a three layer satellite network architecture of GEOs, LEOs and HAPs with a prioritized load balancing algorithm. We identify traffic as either best-effort or high priority, with QoS guarantees for high priority traffic.

The rest of the paper is organized as follows; Section 2 discusses the motivation for MPLS and load balancing in layered satellite networks; Section 3 proposes the load balancing strategy; Section 4 details the multi layered architecture; Sections 5 and 6 gives the simulation details, results and analysis; and Section 7 concludes.

2. MOTIVATION FOR MPLS AND CONSTRAINT ROUTING IN LAYERED SATELLITE NETWORKS

2.1. Multi-layered architecture

We consider a three layer architecture of geo-stationary (GEO) satellites, low earth orbit(LEO) satellites and high altitude platforms(HAPs). GEOs orbiting in high altitude geo-stationary orbits (36000 Km) individually are unattractive for delay sensitive traffic because of large propagation delay, consequently making them unsuitable for real-time applications. GEOs do not provide coverage at high latitudes and require terrestrial users to have bulkier equipment.

All these issues motivated the deployment of low-earth orbit (LEO) satellites which orbit the Earth at a height of just 500 to 1,000 miles, which in turn necessitates the use of multiple satellites which constantly orbit around the earth in fixed planes, to provide constant service to any area. The LEO constellation can be viewed as a mobile network with fixed users and mobile nodes. The low altitude orbit makes them capable of providing smaller, more energy-efficient spot beams, and delivers latency potentially equal to (or better than) transcontinental fiber optic cable. Frequency reuse is also an important advantage considering the limited and costly frequency spectrum while increasing the system capacity. With the advent of multiple spot beams, inter-satellite links (ISLs) between satellites and on board switching and processing capabilities, these constellation of low-earth orbit(LEO) satellites along with their terrestrial gateway servers form Autonomous systems(AS). One of the distinct advantages of LEO satellite networks over GEO networks is the reduction in propagation delay making them an attractive option for routing real time traffic.

HAPs are unmanned aerial vehicles or balloons which cover areas of small radii(20 km) and hover at an altitude of 20-30 km.. These HAPs can be deployed in areas with heavy and sensitive traffic(e.g.battlefields, disaster relief) and provide access to the high-speed satellite network for terrestrial users with mobile and hand-held terminals. HAPs can easily form a high speed network among themselves and satellites in the higher layers [18] with optical links.

We propose a multi-layered satellite network with GEOs acting as the backbone routers, LEOs as the second layer and HAPs deployed in specific local regions. This combination provides high bandwidth access to all types of users and low latency to delay sensitive applications. Figure 1. shows our proposed multi-layer architecture. Such architecture is similar to the Transformational Communications architecture proposed in [20].

2.2. Prioritized load balancing

The need for QoS in satellite networks is fueled by several reasons. With an explosion of network traffic in terms of users and applications, ISPs want to offer different levels of service based on business priorities of the users or applications. With applications varying from real time interactive traffic (e.g. VoIP), real time non-interactive traffic (e.g. streaming video) to non-real time traffic (e.g. web traffic) it is necessary to differentiate in the levels of service provided. High speed networks should be able to support different degrees of Quality of Service (QoS) to different applications. For example, real-time traffic generated by multimedia applications has radically different requirements than best-effort traffic. So real-time applications require tight bounds on transfer delay (in the order of hundreds of milliseconds).

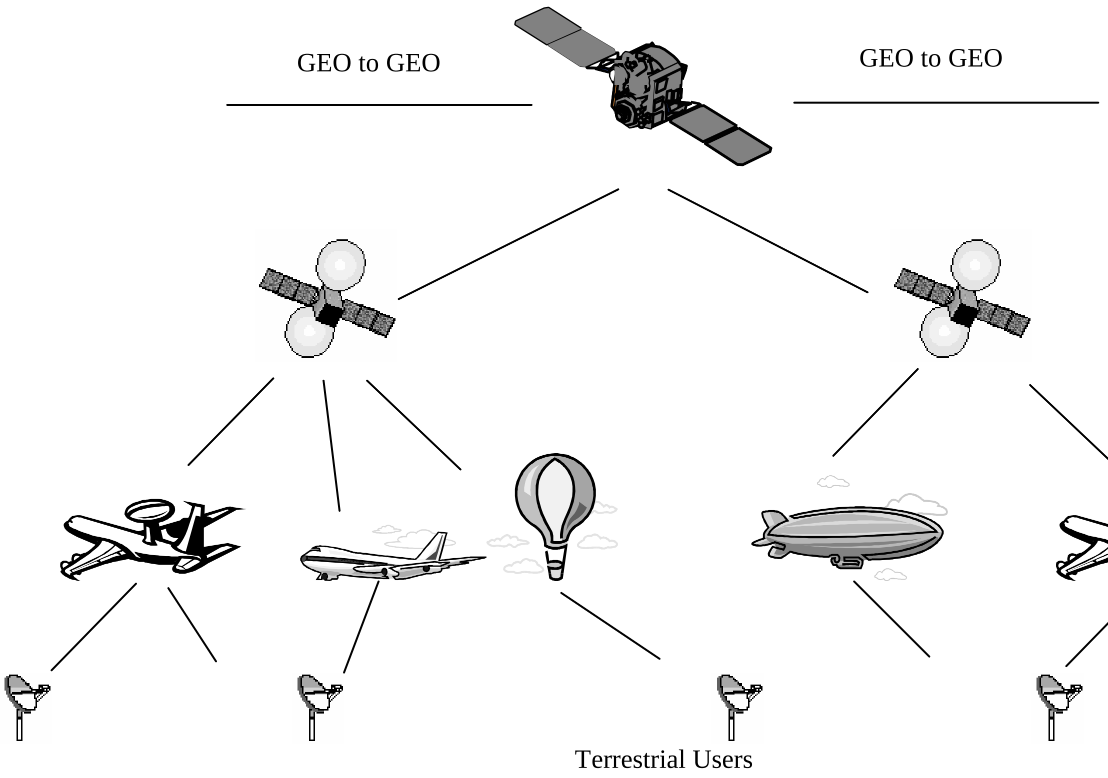

Figure 1. Satellite multi-layer architecture

Real-time applications such as VoIP and streaming video are susceptible to changes in the transmission characteristics of data networks. Voice over IP (VoIP) and real time traffic (VBR) are also susceptible to network behaviors, referred to as delay and jitter, which can degrade the voice application to the point of being unacceptable to the average user.

So it becomes essential to separate such high priority traffic from non-real time traffic and route them over explicit paths, which meet the desired QoS requirements. Satellite networks employing static routing; based on topological properties of the network lead to congestion in some parts of the network, which is heavily used by all classes of traffic even when there is a lot of leftover bandwidth in the network. By employing a load-balancing algorithm with delay as a constraint we manage to protect real-time flows from each other as well as from bursty best-effort traffic.

2.3. Need for MPLS

Due to the high-speed mobility of the nodes and ISL handoffs between satellites in the LEO/MEO layers there are several issues regarding routing of IP traffic over satellite networks [8]. Since MPLS operates independently of layer 3 and will use IP routing methods, standard IP QoS can be enforced during the LSP setup process. LSPs with specific bandwidth requirements, delays bounds can be setup using constraint-based routing and have labels associated with them. Consequently appropriate traffic can be routed along their desired QoS path.

3. PRIORITIZED LOAD BALANCING

In the following sections we propose the call admission and routing strategy for our scheme.

3.1. Call Admission Control (CAC) and Packet Scheduling

Given the costly nature of satellite resources over-provisioning is not an option in satellite networks, so we resort to a form of class based queuing. The overall bandwidth for each link is split, so that we have two queues, a fraction for high priority traffic (with real time constraints) and the rest for best-effort service. The call admission control scheme to which the high priority calls are subject is very simple; for each link, new calls are accepted if:

$$\sum i \quad Bp <= \mu B$$

Where i is the set of high-priority calls, Bp is the bandwidth of each call and C is the capacity of the link. The fraction μ, is the fraction of bandwidth on each link that is apportioned for high priority calls and can be assigned, and takes values from 0 to 1.

No CAC is implemented for best-effort calls, consequently if bandwidth normally reserved for high priority traffic is available, best-effort calls will be entertained on the link, but may be switched or pre-empted if a request for bandwidth comes from a high priority call.

This is done using Class Based Queuing (CBQ) to guarantee bandwidth allocations for all the different trunks at the routers. By doing this, we essentially separate the real-time VoIP flow from the TCP flows.
Thus, by isolating VoIP and TCP flows, we can guarantee a given Quality of Service to sources which are responsive to congestion control, although this is achieved at the overhead of maintaining separate queues for each of the trunks at each router. According to Li and Rekhter [1], the number of trunks within a given topology is within the limit of (N*(N-1)*C), where N is the number of routers in the topology and C is the number of traffic classes. This means the overhead is within a reasonable limit.

3.2. Constraint based Routing strategy

We introduce bandwidth availability and maximum hop count as additional constraints during the path selection process.
Let Ω be the generic cost function associated to a path p, related to the network topology; for example; Ω can be a function of the number of hops in p, or of the delay, or a function of both. Paths p1 and p2 are considered equivalent, provided they are between the same source and destination pair, when Ω (p1) = Ω (p2).

The routing function first determines the set of paths satisfying the generic cost requirements i.e. the delay bound for the specific traffic.For each set of equivalent paths, we then associate a dynamic cost based on the residual bandwidth of the path. For each link (l) along the path we have the available bandwidth B_l, the residual bandwidth of the path is:

$$R = \min (B_l) \text{ where link } l \text{ belongs to path P}$$

For the high priority calls the residual bandwidth is in terms of B_p(bandwidth assigned for high priority traffic) assigned for that link, whereas for best-effort traffic it is in terms of the available bandwidth on the link.
The overall cost Ψ to the path p is then assigned:

$$\Psi(p) = \Omega(p) / R$$

The paths are then sorted according to Ψ, the first path in the set being the primary path and the other being the secondary paths.

4. MULTI-LAYERED SATELLITE NETWORK

4.1. Interconnections

The satellites maintain three types of links:

1. *User Data Links(UDLs)* between terrestrial users and the satellite network. A user can have UDLs to multiple satellites in each of the layers and vice-versa.

2. *Inter orbit links(IOLs)* exist between the layers of the satellite hierarchy. Each satellite is linked by IOLs to the satellites under its coverage as well as to the satellites in the upper layer that cover it.

3 .*Inter Satellite Links(ISLs)* exist in the LEO layer which can be interplane ISLs, intraplane ISLs or cross-seam ISLs. The HAPs also communicate with each other through line of sight optical links.

4.2. Domains and grids

Three GEOs divide the earth surface into three fixed domains, the size of which correspond to the footprint size of the GEO satellites, since the GEOs will always be over their respective positions with respect to the earth surface. Furthermore, each domain is subdivided into sub-domains, the size of which corresponds to the footprint of the LEO satellites, assuming earth-fixed footprints for LEO satellites. Depending on whether HAPs are deployed, the sub-domain could have further division into smaller cells. Assuming rectangular grids for the purpose of easy representation, Figure 2 below shows the domain structure.

Figure 2. above represents a single footprint from GEO A, divided by six LEOs (I-VI) and the footprint of LEO I is subdivided by the presence of HAPs in the coverage of I.

Addressing for the purpose of deciding the egress node is also derived from the grid structure. In the above figure a user in grid U, where $U \subset I \subset A$, can be reached by all three: HAP(U),LEO(I) and GEO(A). Consequently, all users in U can have an address prefix (e.g. A.I.U) which identifies the egress nodes through which the user can be reached.

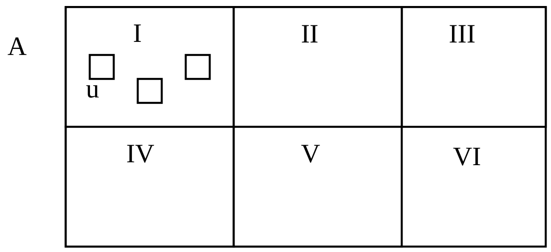

Figure 2. Domains And Grids

We assume earth-fixed footprints for the LEO satellites, so that handovers between LEO satellites of ground to satellite links (GSLs) are always synchronous and periodic. In Figure 2 when the LEO over grid I moves over to grid IV, and LEO over grid IV is moving out, the LEO IV forwards its routing table with the active connections and their next hops to LEO I. So that an active connection is always routed over the same logical topology, though the LEO satellites keep changing. We reduce the dynamic LEO constellation into a logically fixed topology [11].

4.3. Hierarchical Routing Architecture

GEOs being at the highest altitude act as the “eyes” for the satellite constellation, and keep track of the LEO and HAP movement and interconnections under its coverage. Most of the interconnections (e.g. the ISLs and cross-seam links between LEO satellites) are periodic and can be predicted.

4.3.1. Static Connection Matrix

Given the regular topology of the satellite networks, our logical grid structure for the LEO and the assumption that HAPs can be considered stationary with respect to LEOs and GEOs, a static connection matrix can be easily maintained and updated at regular intervals or manually (in case HAPs are deployed or satellite/link failures).

Figure 3. shows a representation of the connection matrix. The shaded portion refers to the cross-seam plane across which the ISLs will be switched off much rapidly because of counter-rotating orbits. The GEOs are connected to each other through line of sight links. Adjacent grids on the horizontal plane in the LEO layer are interconnected by interplane ISLs at all times except; cross-seams where links are handed over very fast, at poles where they are switched off. The link delay on interplane ISLs changes with latitude, with link delay reducing as the latitude increases, but since this behavior is also periodic the link delays between neighbors can be predicted. Adjacent grids on the vertical plane in the LEO layer are interconnected by intra-plane ISLs, which are never switched off and have constant delay. Whenever HAPs are deployed over a region, their IOL delays to both the LEO satellites and GEO are uploaded to the matrix.

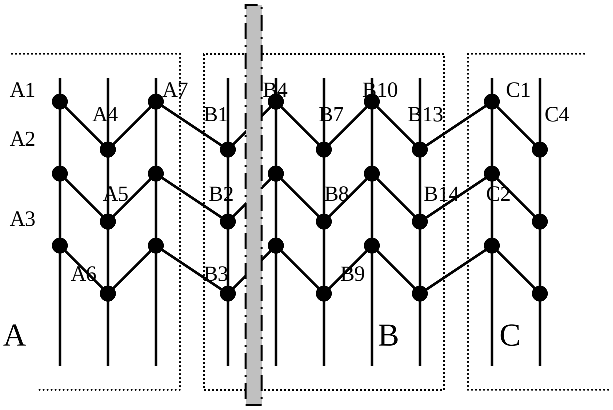

Figure 3. Connection Matrix

x	B	B13	B14	B15
B				
B1				
B2				
B3				

Figure 4. Aggregated routing table for domain B

4.3.2. Connection matrix with current traffic load

From the above connection matrix and link propagation delays we have a static snapshot of the network with no knowledge about the traffic load.

The LEOs and HAPs under a GEO measure their output buffers on all their outgoing links (UDL, ISL and IOL) to determine the residual bandwidth on the links and upload it to the GEO at regular intervals. Since the GEOs have limited onboard processing capabilities, but are always visible to the fixed terrestrial gateway server in its domain this link state information can be downloaded onto the gateway for processing. The intra-domain routing table for each node in the domain is formed. Also an aggregated routing table for the domain is formed which includes the maximum residual bandwidth paths from each border node at one end of the domain to every other node at the other end of the domain and vice-versa. Figure 4. represents the information in the aggregated routing table for domain B under the satellite GEO B.

The intra-domain routing and the aggregated routing tables are uploaded to the GEO, which then uploads the intra-domain routing table to the LEOs and HAPs under its coverage, and the aggregated routing table is sent to its peer GEOs. Each GEO upon receiving the aggregated routing table from its peer GEO floods it to the LEOs and HAPs under its coverage.

4.3.3. Routing Strategy

We outline the steps to setting up the LSP, the example below elucidates.

- S looks up the static connection matrix to find the optimal set of paths to D.
- Using the intra-domain routing table and the aggregated routing table for domain B it determines optimal border nodes in B to reach domain C.
- The paths are ranked according steps outlined in Section III.
- Routing from the border node in C to the egress node D is done intra-domain in domain C during the LSP setup process.

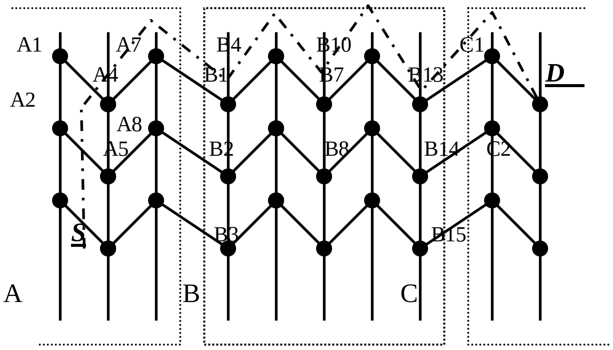

Figure 5. LSP setup for S-D pair

5. SIMULATION CONFIGURATION AND RESULTS

5.1. Simulation configuration

In all simulations we use the Iridium configuration [9], with 66 satellites, as the LEO layer. Furthermore, we assume three equally spaced GEOs on the equatorial orbit. VoIP can adequately represent high priority traffic, as it is interactive and has strict delay and jitter requirements. A high priority call is interpreted as an aggregate of several VoIP calls of the order of Megabits/s(Mb/s) where each VoIP call is an exponential on-off source at 8 Kb/s according to the G.729 standard [19]. Background web traffic was simulated as TCP flows with infinite ftp sources. In our study we analyze 3 real-time VoIP calls (RT1, RT2, RT3) originating from the terrestrial ground station S and terminating at terrestrial ground station D. The queue was split with μ = 0.8 (fraction of queue reserved for real time traffic). Round robin packet scheduling was used among the CBQ queues. To increase ISL utilization without having to increase traffic several folds, the ISLs at the LEO layer are 2.3 Mb/s links.

5.2. Scenario 1: Without Load Balancing

Table 1 below shows the simulation results with normal static routing:

Table 1. With static routing

Traffic Source	Packet Loss(%)	Mean Delay(ms)	Jitter(ms)
RT 1	11.89	172.477	54.0916
RT 2	22.46	187.028	39.526
RT 3	36.34	201.222	22.362

Average delay = 186.54 ms

Average jitter = 39.66   ms

With the static routing all the traffic flows are routed along the shortest path, leading to heavy packet loss and jitter, while there is a lot of unused bandwidth in the network.

5.3. Scenario 2: With Load Balancing

Table 2 below shows the simulation results with the load balancing algorithm implemented.

Table 2. with QoS routing

Traffic source	**Packet Loss(%)**	**Mean Delay (ms)**	**Jitter(ms)**
RT 1	0.67	65.0703	4.697
RT 2	0.71	60.4791	1.486
RT 3	1.0	80.9157	1.58593

Average Delay = 68.81 ms

Average jitter   = 2.59 ms

The average constraint-based LSP setup time was 0.06 seconds.

Both the average delay and jitter values have gone down considerably as now the real time flows are established along separate paths and have been separated from best effort traffic. RT1 experiences some jitter since all of the best effort traffic flows along the path used by RT1.

The optimum network requirements for acceptable VoIP traffic;

One way delay < 100ms

Jitter < 30 ms

Packet Loss < 3%

The next three figures Fig. 6, Fig. 7 and Fig. 8 are graphs of average one-way delay, average jitter and average packet loss respectively for the three real-time flows as the background web traffic is continually increased.

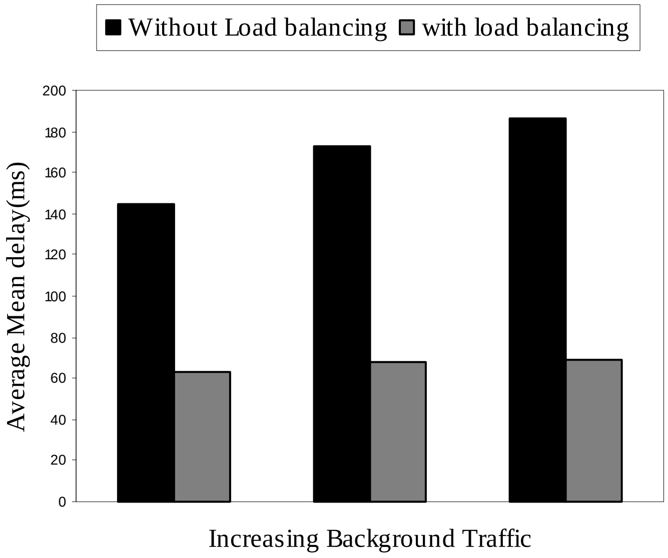

Figure 6. Average mean delay

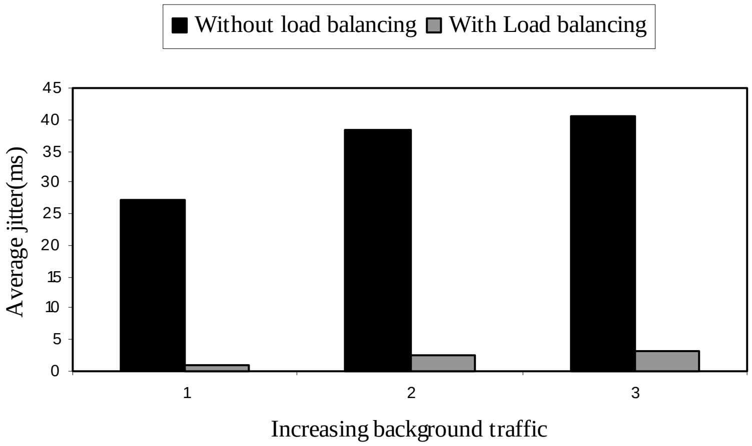

Figure 7.Average jitter

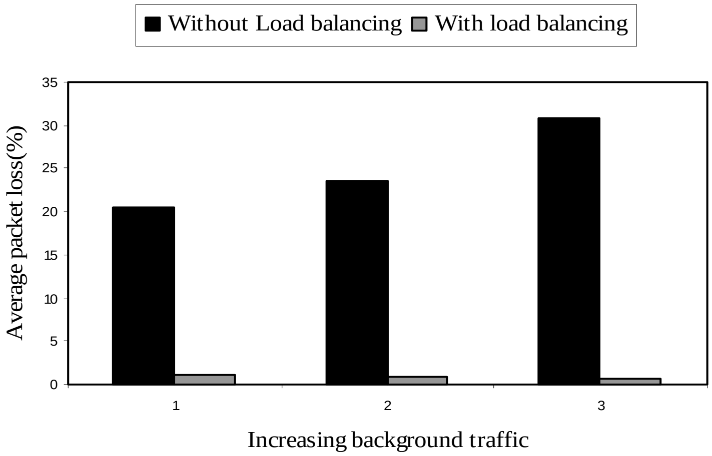

Figure 8. Average Packet Loss

The values for delay, jitter and packet loss almost remain constant for the real-time flows even if the background traffic is increased with the multipath algorithm in effect, whereas for the static routing case the values go beyond acceptable VoIP standards.

6. SIMULATION ANALYSIS

Figure 9 is a plot between increasing back ground traffic and VoIP goodput keeping the bandwidth of the bottleneck links constant i.e. 2.3 Mb. As we can infer from the plot with static routing the real-time rate has to be decreased in order to have acceptable delay and jitter values for the VoIP flow, whereas with the load balancing scheme the VoIP goodput remains constant even when background traffic is increased.

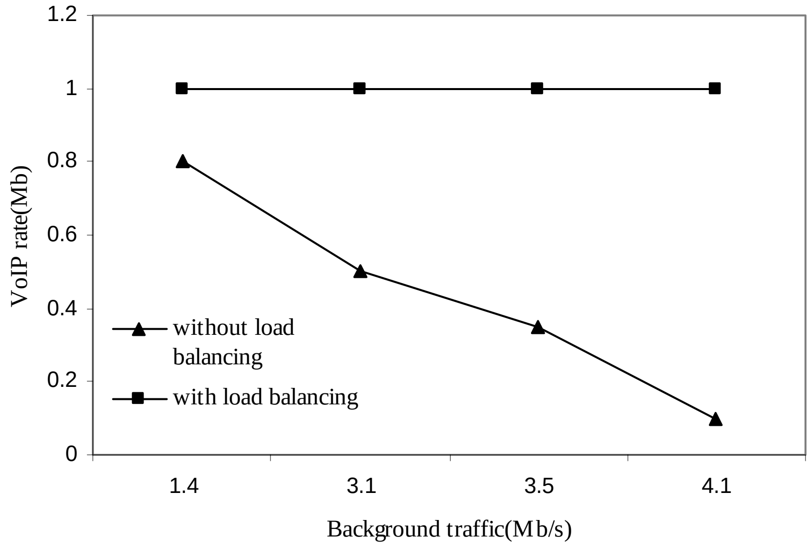

Figure 9. Goodput vs. Background Traffic

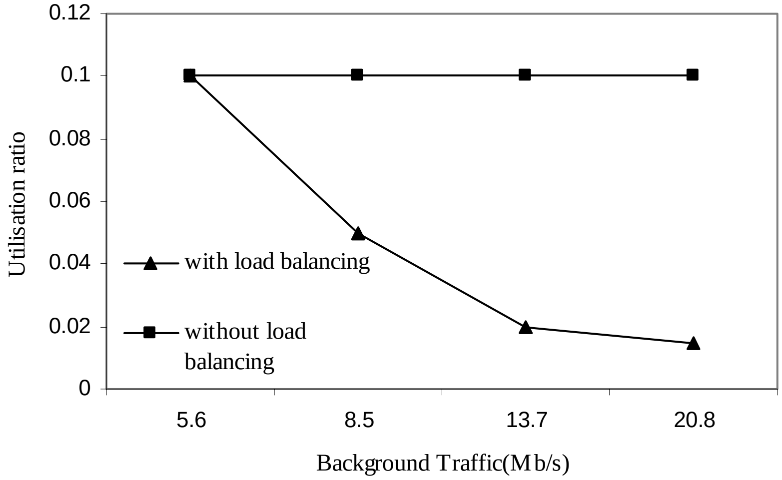

Figure 10. Utilization ratio vs. Background traffic

Next we analyze the utilization of the bottleneck link in terms of the real-time goodput flowing through it. We define a term *utilization ratio* where;

Utilization ratio = real-time goodput / Link bandwidth

Figure 10 above is a plot between the utilization ratio and increasing background traffic, without load balancing keeping the real-time rate constant we have to increase the bandwidth of the bottleneck link to maintain necessary QoS for VoIP flow.

The low link utilization is a result of the bursty nature of best-effort traffic, which results in increased jitter for real-time traffic. By separating real-time from the bursty traffic we ensure low values of jitter while increasing link utilization.

7. CONCLUSION

Dividing the satellite network into domains and pushing most of the computation to the terrestrial gateways reduces the messaging and computational overheads. The use of HAPs and LEOs allows terrestrial users with hand-held terminals to connect to the high speed satellite network with high bandwidth availability. Given the geography sensitive satellite network where a satellite over a major city might experience heavy traffic, whereas a neighboring satellite over an ocean is under-utilized, the load balancing algorithm ensures that traffic from neighboring satellites does not further clog the already loaded satellite. Congestion sensitive applications e.g. VoIP are protected from best-effort traffic increasing the attractiveness of using satellites for routing such traffic.

REFERENCES

1. Li, T., Rekhter, Y. "A Provider Architecture for Differentiated Services and Traffic Engineering (PASTE)", RFC 2430, October 1998.
2. Awduche, D.O., Malcolm, J., O'DEll, M., McManus, J. "Requirements for Traffic Engineering Over MPLS", draft-ietf-mpls-traffic-eng-00.txt, October 1998
3. UCB, LBNL, VINnT Network Simulator - ns2.http://www-mash.cs.berkeley.edu/ns/ns.html
4. Rosen, E.C., Vishwanathan, A., Callon, R. "Multiprotocol Label Switching Architecture", draft-ietf-mpls-arch-04.txt, Feb. 1999 Eric C. Rosen, Y. Rekhter, "MPLS Label Stack Encoding", draft-ietf-mpls-label-encaps-03.txt, Sep. 1998
5. Sally Floyd, Van Jacobson, "Link-sharing and Resource Management Models for Packet Networks", Feb. 1995
6. Praveen Bhaniramka, Wei Sun, Raj Jain "Quality of Service using Traffic Engineering over MPLS: An Analysis," Proc. 25th Annual IEEE Conference on Local Computer Networks (LCN 2000), Tampa, Florida USA, November 8 - 10, 2000, pp. 238-241.
7. T. R. Henderson and R. H. Katz, Network Simulation for LEO Satellite Networks, 18th AIAA International Communications Satellite Systems Conference (ICSSC), Oakland, California, April 2000.
8. L. Wood, A. Clerget, I. Andrikopoulos, G. Pavlou, and W. Dabbous, "*IP Routing Issues in Satellite Constellation Networks*," International Journal of Satellite Communications, vol. 19, no. 1, pp. 69--92, January/February 2001.
9. http://www.iridium.com/
10. Yun Sik Kim, Young-ho Bae, Youngjae Kim, Chul Hye Park. “Traffic Load Balancing in Low earth orbit Satellite Networks” International Conference on Computer Communications and Networks, October 12-15 1998.
11. Ian F. Akyilidig, Eylem Ekici, Michael D Bender. “MLSR: A Novel Routing Algorithm for Multilayered Satellite IP Networks” IEEE/ACM Transactions on Networking, Vol. 10,No. 3,June 2002.
12. J. Lee and S.Kang. “Satellite over Satellite (SOS) network: A novel architecture for satellite network.” In Proc. IEEE INFOCOM, vol. 1, March 2000, pp. 315-321.
13. http://www.iridium.com/corp/iri_corp-news.asp?newsid=62
14. Mauro de Sanctis, Ernestina Cianca, Marina Ruggieri.“IP based routing algorithms for LEO satellite networks in Near Polar Orbits”
15. T.R. Henderson “Networking over next generation Satellite Systems”,PhD thesis, University of California at Berkeley.
16. http://www.bakom.ch/imperia/md/content/english/funk/forschungundentwicklung/studien/HAPS.pdf
17. http://www.elec.york.ac.uk/comms/haps-main.html
18. M. Antonini, E. Cianca, A. De Luise, M. Pratesi, M. Ruiggieri, “Stratospheric Relays: Potentialities of New Satellite-High Altitude Platforms Integrated scenarios”, IEEAC paper #1096.
19. International Telecommunication Union. http://www.itu.int/rec/recommendation.asp?type=folders&lang=e&parent=T-REC-G.729
20. Christine M. Anderson, Transformational Communications, presented at GSAW 2002, available online at: http://sunset.usc.edu/GSAW/gsaw2002/s8/canderson.pdf