# Realizing Future Broadband Satellite Network Services

Sastri Kota
Lockheed Martin
160 East Tasman Avenue
San Jose, CA 95134
Phone: (408) 456-6300
e-mail: sastri.kota@lmco.com

Arjan Durresi, Raj Jain
The Ohio State University
2015 Neil Ave
Columbus, OH 43210

**New Address: Raj Jain, Washington University in Saint Louis, jain@cse.wustl.edu, http://www.cse.wustl.edu/~jain**

**ABSTRACT**

Future satellite communication systems proposed use geosynchronous (GEO) satellites, medium earth orbit (MEO), and low earth orbit (LEO) constellations. Most of the next generation satellite systems will use fast packet switching with onboard processing to provide full two-way services to and from earth stations. One of the major service drivers is a high data rate internet access carried over integrated satellite-fiber networks. Provisioning of quality of service (QoS) within the advanced satellite systems is the critical requirement.

In this paper, we present broadband LEO satellite network QoS model and simulated performance results. We discuss the TCP flow aggregates performance for their good behavior in the presence of competing UDP flow aggregates in the same assured forwarding. We identify several factors that affect the performance in the mixed environments and quantify their effects using a full factorial design of experiment methodology.

## 1.0 INTRODUCTION

The rapid globalization of the telecommunications industry and the exponential growth of the Internet is placing severe demands on global telecommunications. This demand is further increased by the convergence of computing and communications and by the increasing new applications such as Web surfing, desktop and video conferencing. Satisfying this requirement is one of the greatest challenges before telecommunications industry in the 21st century. Satellite communication networks can be an integral part of the newly emerging national and global information infrastructures (NII and GII).

In the past three years, interest in Ka-band satellite systems has dramatically increased, with over 450 satellite applications filed with the ITU. In the U.S., there are currently 13 Geostationary Satellite Orbit (GSO) civilian Ka-band systems licensed by the Federal Communications Commission (FCC), comprising a total of 73 satellites.

Two Non-Geostationary Orbit (NGSO) Ka-band systems, compromising another 351 satellites, have also been licensed. Eleven additional GSO, four NGSO, and one hybrid system Ka-band application for license and 16 Q/V-band applications have been filed with FCC [1].

The large delays in geostationary Earth orbit (GEO) systems and delay variations in low Earth orbit (LEO) systems affect both real-time and non-real-time applications. In an acknowledgement- and time-out-based congestion control mechanism (like TCP), performance is inherently related to the delay-bandwidth product of the connection. Moreover, TCP round-trip time (RTT) measurements are sensitive to delay variations that may cause false timeouts and retransmissions. As a result, the congestion control issues for broadband satellite networks are somewhat different from those of low-latency terrestrial networks

There has been an increased interest in developing Differentiated Services (DS) architecture for provisioning IP QoS over satellite networks. DS aims to provide scalable service differentiation in the Internet that can be used to permit differentiated pricing of Internet service [2]. This differentiation may either be quantitative or relative. DS is scalable as traffic classification and conditioning is performed only at network boundary nodes. The service to be received by a traffic is marked as a code point in the DS field in the IPv4 or IPv6 header. The DS code point in the header of an IP packet is used to determine the Per-Hop Behavior (PHB), i.e. the forwarding treatment it will receive at a network node. Currently, formal specification is available for two PHBs - Assured Forwarding [3] and Expedited Forwarding [4]. In Expedited Forwarding, a transit node uses policing and shaping mechanisms to ensure that the maximum arrival rate of a traffic aggregate is less than its minimum departure rate. At each transit node, the minimum departure rate of a traffic aggregate should be configurable and independent of other traffic at the node. Such a per-hop behavior results in minimum delay and jitter and can be used to provide an end-to-end `Virtual Leased Line' type of service.

In Assured Forwarding (AF), IP packets are classified as belonging to one of four traffic classes. IP packets assigned to different traffic classes are forwarded independent of each other. Each traffic class is assigned a minimum configurable amount of resources (link bandwidth and buffer space). Resources not being currently used by another PHB or an AF traffic class can optionally be used by remaining classes. Within a traffic class, a packet is assigned one of three levels of drop precedence (green, yellow, red). In case of congestion, an AF-compliant DS node drops low precedence (red) packets in preference to higher precedence (green, yellow) packets.

In this paper, we describe satellite network architectural options followed by a wide range of simulations, varying several factors to identify the significant ones influencing fair allocation of excess satellite network resources among congestion sensitive and insensitive flows. The factors that we studied in QoS Frame Work include *a)* number of drop precedence required (one, two, or three), *b)* percentage of reserved (highest drop precedence) traffic, *c)* buffer management (Tail drop or Random Early Drop with different parameters), and *d)* traffic types (TCP aggregates, UDP aggregates). We describe the simulation configuration and parameters and experimental design techniques. Analysis Of Variation (ANOVA) technique is described. Simulation results for TCP and UDP, for reserve rate utilization and fairness are also given. The study conclusions are summarized.

# 2.0 SATELLITE NETWORK AND QoS MODEL

## 2.1 Satellite Network Architectural Options

Global communications coverage by satellites has been available for many years using Geostationary Orbits (GSO) and large earth stations. Currently, this coverage is being extended to mobile small terminals. The three optional architectures include:

- Geostationary Orbits (GSOs)
- Non-geostationary orbit (NGSO)
  - Medium earth orbit (MEO)
  - Low earth orbit (LEO)
- *GSO Architectures:* The entire world except for the Polar Regions can be covered by only 3 satellites in equatorial orbits. The altitude of these satellites must be approximately 35,800 km above the surface of the earth. The satellites will appear to an observer on earth as being stationary.
- *MEO Architectures:* The altitude has to be selected between the inner and outer Van Allen Radiation belts, typically around 10.355 km above the surface of the earth. The orbit period in this case will be 6 hours. The world can be covered by 10-12 satellites in 2-3 planes, e.g., 5 satellites in each of 2 planes or 4 satellites in each of 3 planes.
- *LEO Architecture:* The satellite altitude is much lower, typically betwen700 to 2000 km. The orbit period will be somewhere between 100 and 120 minutes. Due to the lower altitude, the view of the world from the satellite is rather small and the number of planes to cover the entire globe will have to be increased to 6-8, and the number of satellites must be 6 per plane.

The main advantages and disadvantages of GSO versus non-GSO architectures include:

*a.* Three operational satellites are sufficient to provide full-coverage of the earth excluding the polar caps.

*b.* Not necessary to provide large earth station antennas with fast moving auto-tracking systems, as would be the case with LEO satellites.

*c.* An earth station antenna is required to work with only one GSO satellite for a continuous connection, whereas with LEO satellites it is necessary to use either an earth station antenna able to jump quickly from the setting satellite to a rising one, or to use two antennas for each earth station.

*d.* Onboard antenna of the GSO satellite may be very directive. For a LEO satellite, a small-gain antenna or fast moving tracking systems are required.

*e.* The propagation delay is much larger and may cause an echo with about 500-ms time difference.

*f.* The free space attenuation is much larger and varies with the various frequency ranges.

*g.* GSOs may not provide coverage of big cities for land-mobile communications due to the shadows created by tall buildings.

## 2.2 Onboard Processing and Switching

One of the fundamental drivers of the next generation broadband satellite systems is the onboard processing and ATM fast packet/cell switching. Onboard processing involves demodulation and demultiplexing the received signal. The payload performs decoding and encoding, processing the header information, and routing the data, pointing the antennas, buffering, multiplexing, and retransmitting the data on downlink or inter-satellite link. The major reasons for onboard processing include separation of the uplink from the downlink, a gain of approximately 3 dB in performance, and provision of resources on demand. The advantages of onboard processing and switching include:

- Improved error rates by using effective encoding techniques
- Separation of uplink and downlink
- System efficiency can improve from 37% to nearly 99.5% with packet or cell switching
- Delay improvements
- Routing decisions onboard or via intersatellite links
- No end-to-end retransmissions
- Capacity improvements
- Multiple beams with dual polarization

Table 1 summarizes the architectural options and some of the technical challenges.

*Table 1. Architetcural Options and Technical Challenges*

| Architectural Option | On-board Processing and Switching | Issues |
|---|---|---|
| GSO | No on-board processing | Media Access protocol |
| NGSO<br>– MEO<br>– LEO | On-board processing<br>– Improves error rates<br>– Improves efficiency<br>– Delay improvements<br>On-board switching<br>– Circuit<br>– Packet<br>– Cell | Traffic management<br>– Call administration control<br>– Congestion control<br>– Policing of shaping<br>– Buffering/scheduling<br>QoS management for IP and ATM<br>Interoperability with legacy networks<br>Network control and management |

## 2.3 QoS Model

The key factors that affect the satellite network performance are those relating to bandwidth management, buffer management, traffic types and their treatment, and network configuration. Bandwidth management relates to the algorithms and parameters that affect service PHB given to a particular aggregate. In particular, the number of drop precedence (one, two, or three) and the level of reserved traffic were identified as the key factors in this analysis.

Buffer management relates to the method of selecting packets to be dropped when the buffers are full. Two commonly used methods are tail drop and random early drop (RED). Several variations of RED are possible in case of multiple drop precedence.

Two traffic types that we considered are TCP and UDP aggregates. TCP and UDP were separated out because of their different response to packet losses. In particular, we were concerned that if excess TCP and excess UDP were both given the same treatment, TCP flows will reduce their rates on packet drops while UDP flows will not change and get the entire excess bandwidth. The analysis shows that this is in fact the case and that it is important to give a better treatment to excess TCP than excess UDP.

In this paper, we used a simple network configuration which was chosen in consultation with other researchers interested in assured forwarding. This is a simple configuration, which we believe, provides most insight in to the issues and on the other hand will be typical of a GEO satellite network.

We have addressed the following QoS issues in our simulation study:

- Three-drop precedence (green, yellow, and red) help clearly distinguish between congestion sensitive and insensitive flows.
- The reserved bandwidth should not be overbooked, that is, the sum should be less than the bottleneck link capacity. If the network operates close to its capacity, three levels of drop precedence are redundant as there is not much excess bandwidth to be shared.
- The excess congestion sensitive (TCP) packets should be marked as yellow while the excess congestion insensitive (UDP) packets should be marked as red.
- The RED parameters have significant effect on the performance. The optimal setting of RED parameters is an area for further research.

**Buffer Management Classifications**

Buffer management techniques help identify which packets should be dropped when the queues exceed a certain threshold. It is possible to place packets in one queue or multiple queues depending upon their color or flow type. For the threshold, it is possible to keep a single threshold on packets in all queues or to keep multiple thresholds. Thus, the accounting (queues) could be single or multiple and the threshold could be single or multiple. These choices lead to four classes of buffer management techniques:

*1.* Single Accounting, Single Threshold (SAST)
*2.* Single Accounting, Multiple Threshold (SAMT)
*3.* Multiple Accounting, Single Threshold (MAST)
*4.* Multiple Accounting, Multiple Threshold (MAMT)

Random Early Discard (RED) is a well known and now commonly implemented packet drop policy. It has been shown that RED performs better and provides better fairness than the tail drop policy. In RED, the drop probability of a packet depends on the average queue length which is an exponential average of instantaneous queue length at the time of the packet's arrival [5]. The drop probability increases linearly from 0 to max_p as average queue length increases from min_th to max_th. With packets of multiple colors, one can calculate average queue length in many ways and have multiple sets of drop thresholds for packets of different colors. In general, with multiple colors, RED policy can be implemented as a variant of one of four general categories: SAST, SAMT, MAST, and MAMT.

Single Average Single Threshold RED has a single average queue length and same min_th and max_th thresholds for packets of all colors. Such a policy does not distinguish between packets of different colors and can also be called color blind RED. In Single Average Multiple Thresholds RED, average queue length is based on total number of packets in the queue irrespective of their color. However, packets of different colors have different drop thresholds. For example, if maximum queue size is 60 packets, the drop thresholds for green, yellow and red packets can be {40/60, 20/40, 0/10}. In these simulations, we use Single Average Multiple Thresholds RED.

In Multiple Average Single/Multiple Threshold RED, average queue length for packets of different colors is calculated differently. For example, average queue length for a color can be calculated using number of packets in the queue with same or better color [2]. In such a scheme, average queue length for green, yellow and red packets will be calculated using number of green, yellow + green, red + yellow + green packets in the queue respectively. Another possible scheme is where average queue length for a color is calculated using number of packets of that color in the queue [6]. In such a case, average queue length for green, yellow and red packets will be calculated using number of green, yellow and red packets in the queue respectively. Multiple Average Single Threshold RED will have same drop thresholds for packets of all colors whereas Multiple Average Multiple Threshold RED will have different drop thresholds for packets of different colors.

## 3.0 SIMULATION CONFIGURATION AND PARAMETERS

Figure 1 shows the network configuration for simulations. The configuration consists of customers 1 through 10 sending data over the link between Routers 1, 2 and using the same AF traffic class. Router 1 is located in a satellite ground station. Router 2 is located in a GEO (or LEO) satellite and Router 3 is located in destination ground station. Traffic is one-dimensional with only ACKs coming back from the other side. Customers 1 through 9 carry an aggregated traffic coming from 5 Reno TCP sources each. Customer 10 gets its traffic from a single UDP source sending data at a rate of 1.28 Mbps. Common configuration parameters are detailed in Table 1. All TCP and UDP packets are marked green at the source before being 'recolored' by a traffic conditioner at the customer site. The traffic conditioner consists of two 'leaky' buckets (green and yellow) that mark packets according to their token generation rates (called reserved/green and yellow rate). In two-color simulations, yellow rate of all customers is set to zero. Thus, in two-color simulations, both UDP and TCP packets will be colored either green or red. In three-color simulations, customer 10 (the UDP customer) always has a yellow rate of 0. Thus, in three-color simulations, TCP packets coming from customers 1 through 9 can be colored green, yellow or red and UDP packets coming from customer 10 will be colored green or red. All the traffic coming to Router 1 passes through a Random Early Drop (RED) queue. The RED policy implemented at Router 1 can be classified as Single Average Multiple Threshold RED as explained in the following paragraphs.

Many of the commercial simulation tools available do not support the QoS models as of today. Hence, we have used NS simulator version 2.1 [8] for these simulations. The code has been modified to implement the traffic conditioner and multi-color RED (RED_n).

## 3.1 Experimental Design

In this study, we perform full factorial simulations involving many factors:

- *Green Traffic Rates:* Green traffic rate is the token generation rate of green bucket in the traffic conditioner. We have experimented with green rates of 12.8, 25.6, 38.4 and 76.8 kbps per customer. These rates correspond to a total of 8.5%, 17.1%, 25.6% and 51.2% of network capacity (1.5 Mbps). In order to understand the effect of green traffic rate, we also conduct simulations with green rates of 102.4, 128, 153.6 and 179.2 kbps for two color cases. These rates correspond to 68.3%, 85.3%, 102.4% and 119.5% of network capacity respectively. Note that in last two cases, we have oversubscribed the available network bandwidth.

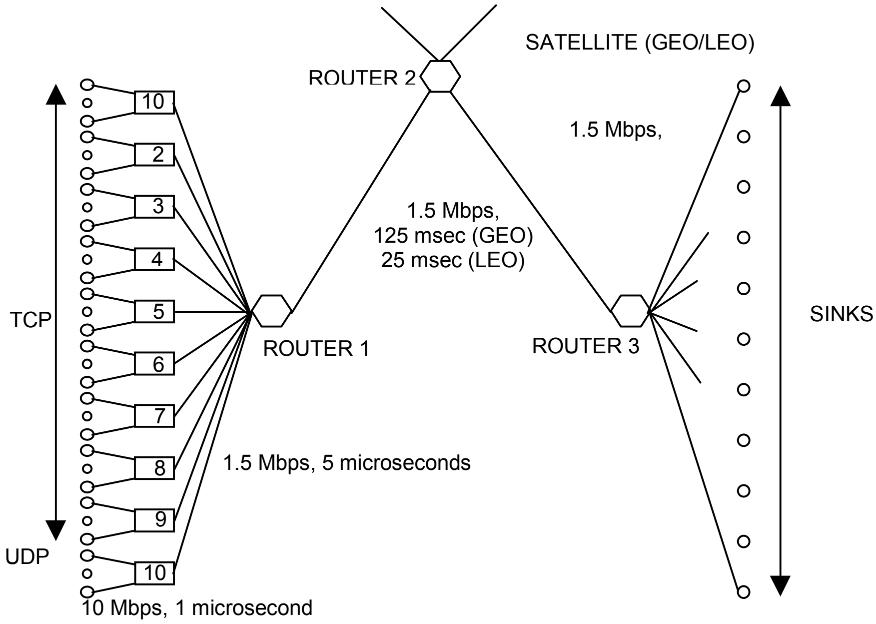


*Figure 1. Simulation Configuration*

- *Green Bucket Size:* 1, 2, 4, 8, 16 and 32 packets of 576 bytes each.
- *Yellow Traffic Rate* (only for three-color simulations): Yellow traffic rate is the token generation rate of yellow bucket in the traffic conditioner. We have experimented with yellow rates of 12.8 and 128 kbps per customer. These rates correspond to 7.7% and 77% of total capacity (1.5 Mbps) respectively. We used a high yellow rate of 128 kbps so that all excess (out of green rate) TCP packets are colored yellow and thus can be distinguished from excess UDP packets that are colored red.
- *Yellow Bucket Size* (only for three-color simulations): 1, 2, 4, 8, 16, 32 packets of 576 bytes each.
- *Maximum Drop Probability:* Maximum drop probability values used in the simulations are listed in Tables 2 and 3.

*Table 2: Two-Color Simulation Parameters*

| Simulation ID | Green Rate [kbps] | Max Drop Probability {Green, Red} | Drop Thresholds {Green, Red} | Green Bucket (in Packets) |
|---|---|---|---|---|
| 1-144 | 12.8 | {0.1, 0.1} | {40/60, 0/10} | 1 |
| 201-344 | 25.6 | {0.1, 0.5} | {40/60, 0/20} | 16 |
| 401-544 | 38.4 | {0.5, 0.5} | {40/60, 0/5} | 2 |
| 601-744 | 76.8 | {0.5, 1} | {40/60, 20/40} | 32 |
| 801-944 | 102.4 | {1, 1} | | 4 |
| 1001-1144 | 128 | | | 8 |
| 1201-1344 | 153.6 | | | |
| 1401-1544 | 179.2 | | | |

*Table 3: Three-Color Simulation Parameters*

| Simulation ID | Green Rate [kbps] | Max Drop Probability {Green, Yellow, Red} | Max Drop Probability {Green, Yellow, Red} | Yellow Rate [kbps] | Bucket Size (in packets) | |
|---|---|---|---|---|---|---|
| | | | | | Green | Yellow |
| 1-720 | 12.8 | {0.1, 0.5, 1} | {40/60, 20/40, 0/10} | 128 | 16 | 1 |
| 1001-1720 | 25.6 | {0.1, 1, 1} | {40/60, 20/40, 0/20} | 12.8 | 1 | 16 |
| 2001-2720 | 38.4 | {0.5, 0.5, 1} | | | 2 | 2 |
| 3001-3720 | 76.8 | {0.5, 1, 1} | | | 32 | 32 |
| | | {1, 1, 1} | | | 4 | 4 |
| | | | | | 8 | 8 |

- *Drop Thresholds* for red colored packets: The network resources allocated to red colored packets and hence the fairness results depend on the drop thresholds for red packets. We experiment with different values of drop thresholds for red colored packets so as to achieve close to best fairness possible. Drop thresholds for green packets have been fixed at {40,60} for both two and three color simulations. For three-color simulations, yellow packet drop thresholds are {20,40}.

In these simulations, size of all queues is 60 packets of 576 bytes each. The queue weight used to calculate RED average queue length is 0.002. For easy reference, we have given an identification number to each simulation as shown in Tables 2 and 3. The simulation results are analyzed using ANOVA techniques [9] briefly described in the following paragraphs.

## 3.2 Performance Metrics

Simulation results have been evaluated based on utilization of reserved rates by the customers and the fairness achieved in allocation of excess bandwidth among different customers.

Utilization of reserved rate by a customer is measured as the ratio of green throughput of the customer and the reserved rate. Green throughput of a customer is determined by the number of green colored packets received at the traffic destination(s). Since in these simulations, the drop thresholds for green packets are kept very high in the RED queue at Router 1, chances of a green packet getting dropped are minimal and ideally green throughput of a customer should equal its reserved rate.

The fairness in allocation of excess bandwidth among n customers sharing a link can be computed using the following formula [9]:

$$Fairness \quad Index = \frac{\left(\sum x_i\right)^2}{n \times \sum \left(x_i^2\right)}$$

Where $x_i$ is the excess throughput of the ith customer. Excess throughput of a customer is determined by the number of yellow and red packets received at the traffic destination(s).

## 4.0 SIMULATION RESULTS

Simulation results of two and three color simulations are shown in Figure 2. In this figure, a simulation is identified by its Simulation ID listed in Tables 2 and 3. Figures 2a and 2b show the fairness achieved in allocation of excess bandwidth among ten customers for each of the two-colors in GEO and LEO architectures. Figures 3a and 3b show simulation results of fairness achieved for GEO and LEO networks for three colors with different reserved rates. It is clear from Figures 2 and 3 that fairness is not good in two-color simulations. With three colors, there is a wide variation in fairness results with best results being close to 1. Note that fairness is zero in some of the two color simulations. In these simulations, total reserved traffic uses all the bandwidth and there is no excess bandwidth available to share. As shown in Figures 2a and 2b, 3a and 3b, there is a wide variation in reserved rate utilization by customers in two and three color simulations.

*Table 4: Main Factors Influencing Reserved Rate Utilization Results*

| Factor/Interaction | Allocation of Variation (in %age) | | | |
|---|---|---|---|---|
| | 2 Colors | | 3 Colors | |
| | TCP | UDP | TCP | UDP |
| Green Rate | 1.60% | 15.65% | 2.21% | 20.40% |
| Green Bucket Size | 97.51% | 69.13% | 95.24% | 62.45% |
| Green Rate - Green Bucket Size | 0.59% | 13.45% | 1.96% | 17.11% |

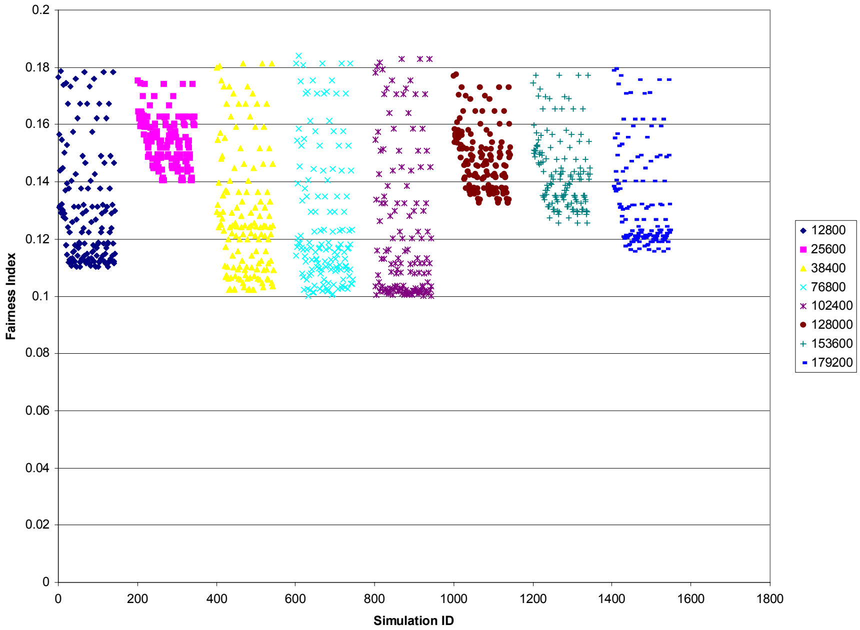


*Figure 2(a). GEO Simulation Results: Fairness Achieved in Two-Color Simulations with Different Reserved Rates*

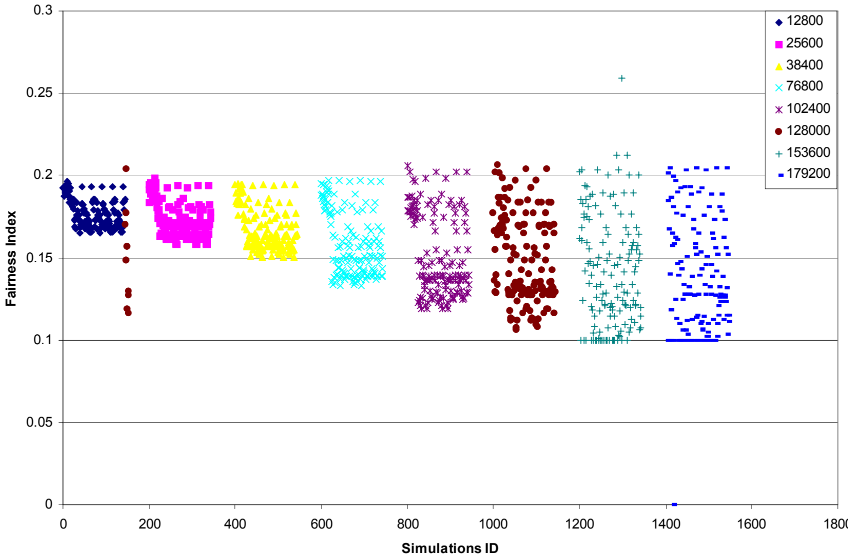


*Figure 2(b). LEO Simulation Results: Fairness Achieved in Two-Color Simulations with Different Reserved Rates*

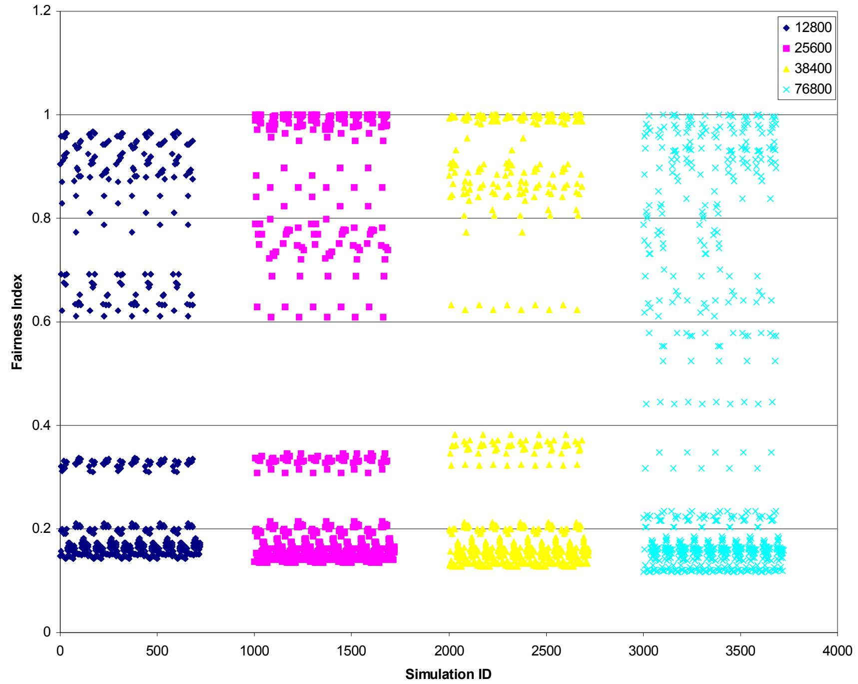


*Figure 3(a). GEO Simulation Results: Fairness Achieved in Three-Color Simulations with Different Reserved Rates*

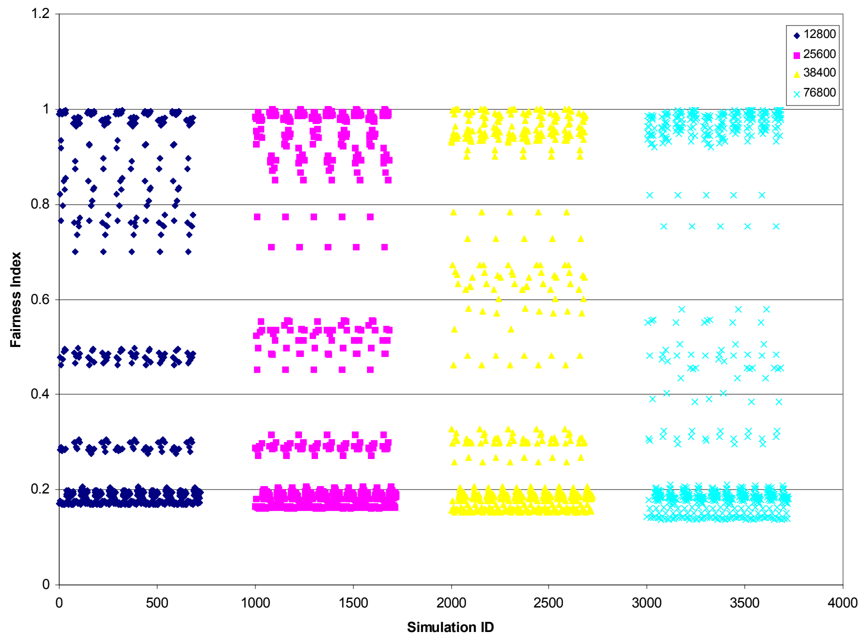


*Figure 3(b). LEO Simulation Results: Fairness Achieved in Three-Color Simulations with Different Reserved Rates*

Figures 4(a) and 4(b) show reserved rate utilization by TCP customers in two color for GEO and LEO networks. Figures 6(a) and 6(b) show reserved rate utilization for TCP customers in three colors, for

GEO and LEO, respectively. For TCP customers, we have plotted the average reserved rate utilization in each simulation. Note that in some cases, reserved rate utilization is slightly more than one. This is because token buckets are initially full which results in all packets getting green color in the beginning. Figures 5a and 5b, and 7a and 7b, show that UDP customers have good reserved rate utilization in almost all cases. In contrast, TCP customers show a wide variation in reserved rate utilization.

In order to determine the influence of different simulation factors on the reserved rate utilization and fairness achieved in excess bandwidth distribution, we analyze simulation results statistically using Analysis of Variation (ANOVA) technique. A brief introduction to ANOVA technique used in the analysis is provided. In later paragraphs, we present the results of statistical analysis of two- and three-color simulations.

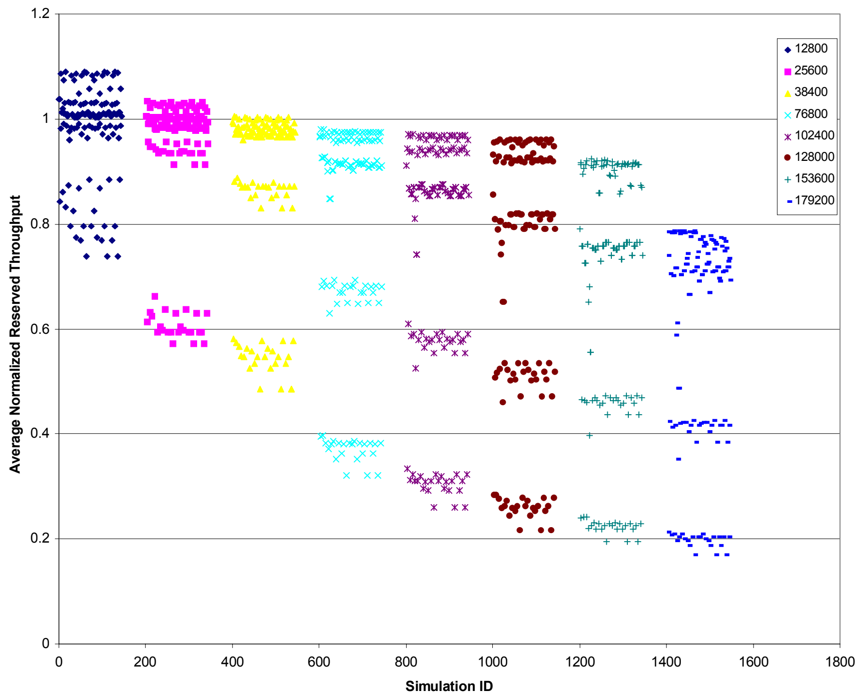


*Figure 4 (a). GEO Reserved Rate Utilizat ion by TCP Customers in Two Color Simulations*

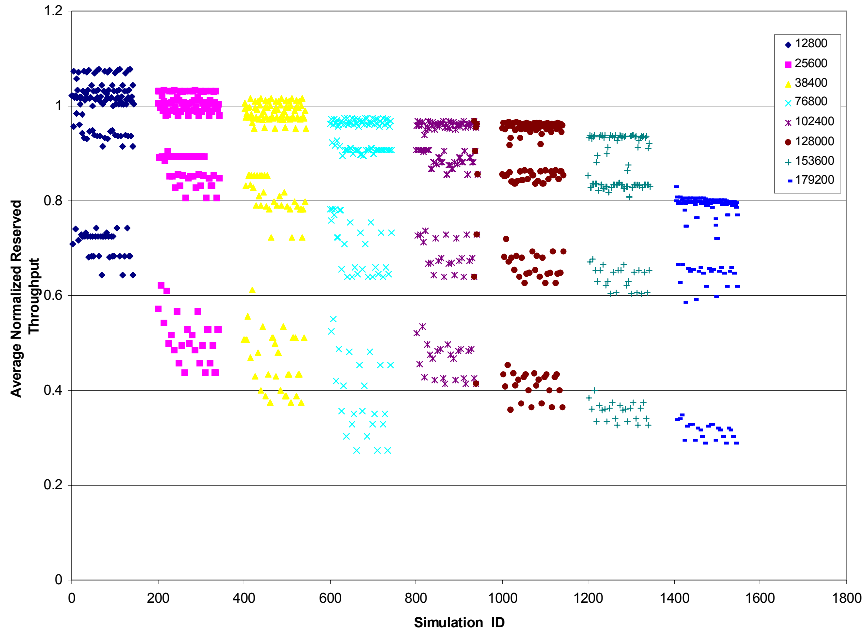


*Figure 4 (b). LEO Reserved Rate Utilization by TCP Customers in Two-Color Simulations*

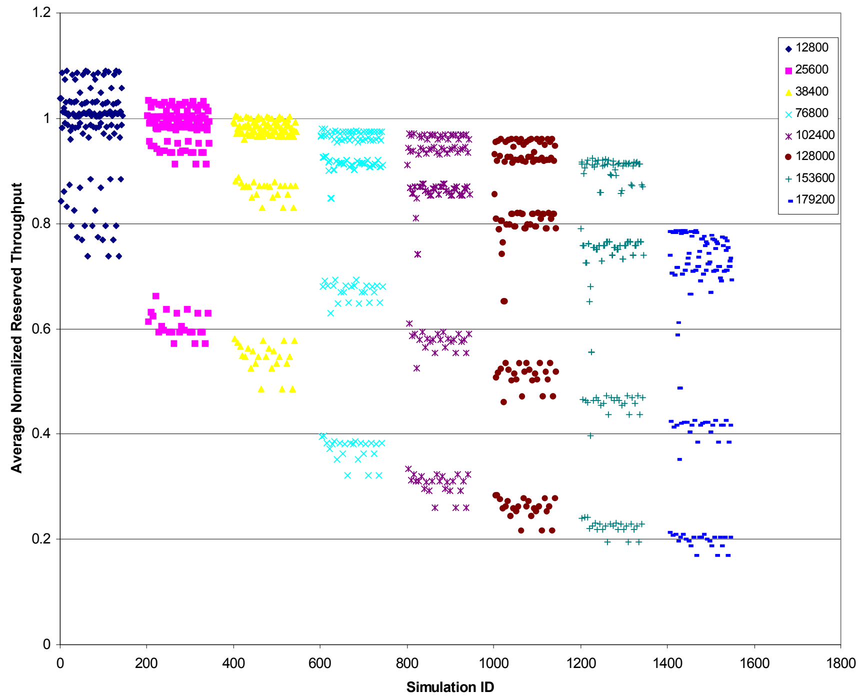


*Figure 5 (a). GEO Reserved Rate Utilization by UDP Customers in Three-Color Simulations*

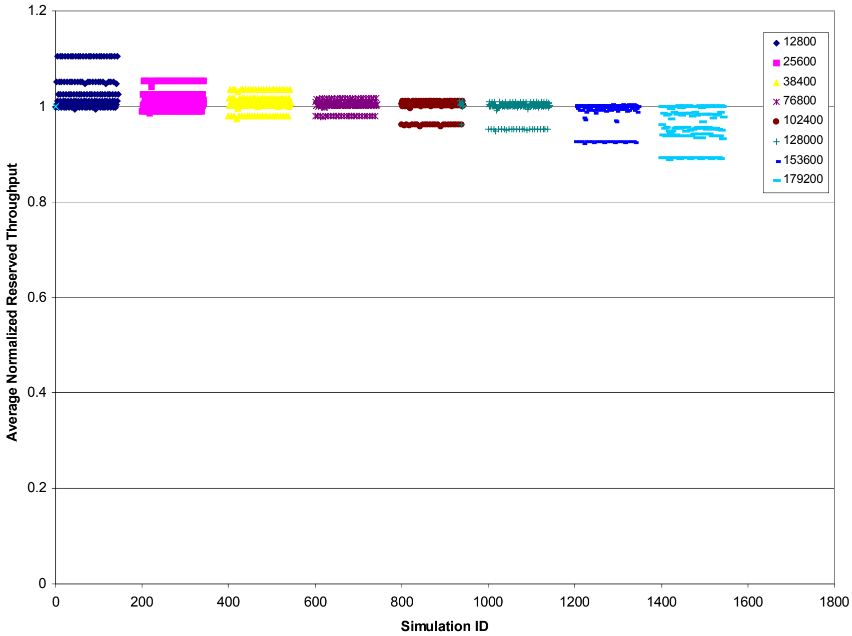


*Figure 5 (b). LEO Reserved Rate Utilization by UDP Customers in Three-Color Simulations*

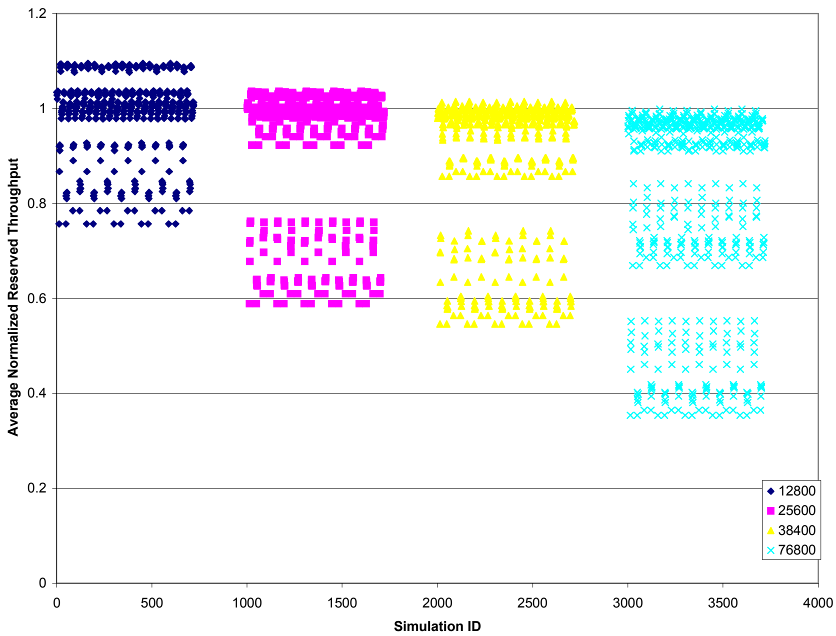


*Figure 6 (a). GEO Reserved Rate Utilization by TCP Customers in Three-Color Simulations*

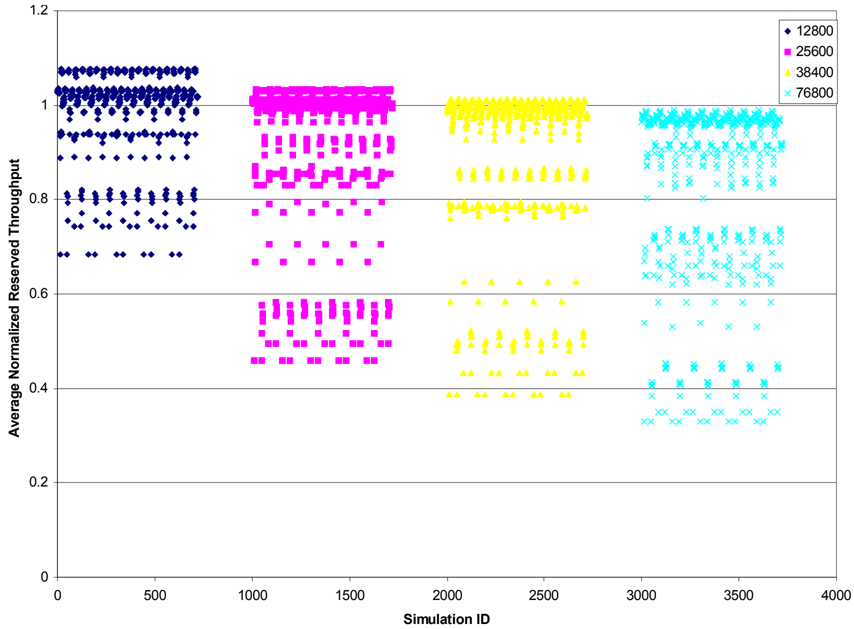


*Figure 6 (b). LEO Reserved Rate Utilization by TCP Customers in Three-Color Simulations*

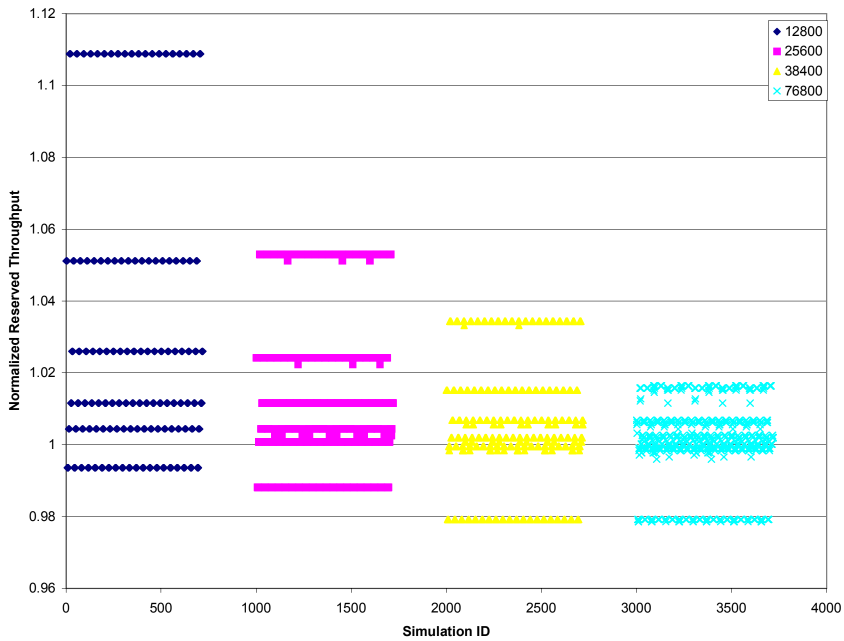


*Figure 7 (a). GEO Reserved Rate Utilization by UDP Customers in Three-Color Simulations*

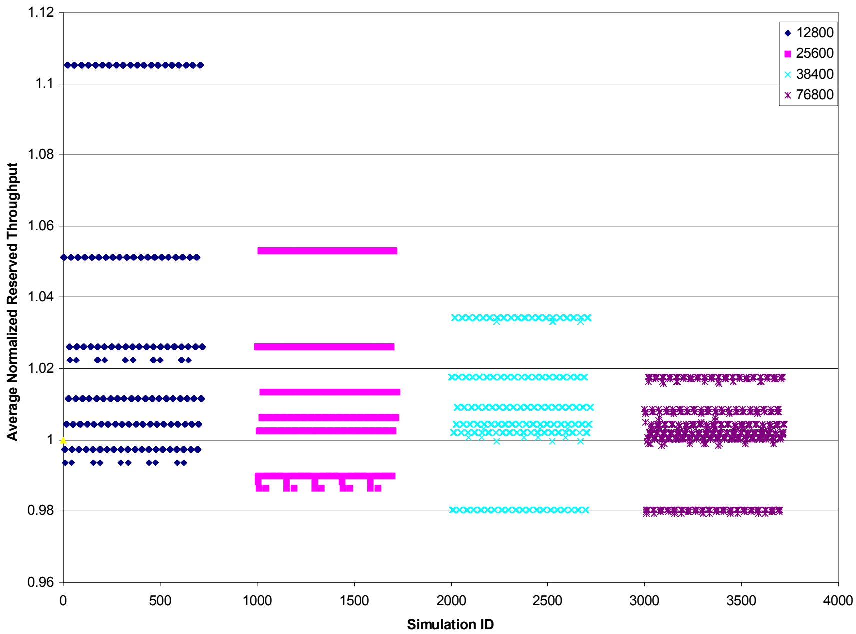


*Figure 7 (b). LEO Reserved Rate Utilization by UDP Customers in Three-Color* Simulations

## 4.1 Analysis Of Variation (ANOVA) Technique

The results of a simulation are affected by the values (or levels) of simulation factors (e.g. green rate) and the interactions between levels of different factors (e.g. green rate and green bucket size). The simulation factors and their levels used in this simulation study are listed in Tables 3 and 4. Analysis of Variation of simulation results is a statistical technique used to quantify these effects. In this section, we present a brief account of Analysis of Variation technique. More details can be found in [9].

Analysis of Variation involves calculating the Total Variation in simulation results around the Overall Mean and doing Allocation of Variation to contributing factors and their interactions. Following steps describe the calculations:

1. Calculate the *Overall Mean* of all the values.
2. Calculate the individual effect of each level *a* of factor *A*, called the *Main Effect* of *a*:
   Main $\text{Effect}_a = \text{Mean}_a$ - Overall Mean
   where, Main $\text{Effect}_a$ is the main effect of level *a* of factor *A*, $\text{Mean}_a$ is the mean of all results with *a* as the value for factor *A*.
   The main effects are calculated for each level of each factor.
3. Calculate the *First Order Interaction* between levels *a* and *b* of two factors *A* and *B* respectively for all such pairs:
   $\text{Interaction}_{a,b} = \text{Mean}_{a,b}$ - (Overall Mean + Main $\text{Effect}_a$ + Main $\text{Effect}_b$)

where, $\text{Interaction}_{a,b}$ is the interaction between levels *a* and *b* of factors *A* and *B* respectively, $\text{Mean}_{a,b}$ is mean of all results with *a* and *b* as values for factors *A* and *B*, Main $\text{Effect}_a$ and Main $\text{Effect}_b$ are main effects of levels *a* and *b* respectively.

4. Calculate the *Total Variation* as shown below:
   Total Variation = $\sum(\text{result}^2)$ - (Num_Sims) $\times$ (Overall $\text{Mean}^2$)
   where, $\sum(\text{result}^2)$ is the sum of squares of all individual results and Num_Sims is total number of simulations.
5. The next step is the *Allocation of Variation* to individual main effects and first order interactions. To calculate the variation caused by a factor *A*, we take the sum of squares of the main effects of all levels of *A* and multiply this sum with the number of experiments conducted with each level of *A*. To calculate the variation caused by first order interaction between two factors *A* and *B*, we take the sum of squares of all the first-order interactions between levels of *A* and *B* and multiply this sum with the number of experiments conducted with each combination of levels of *A* and *B*. We calculate the allocation of variation for each factor and first order interaction between every pair of factors.

## 4.2 ANOVA Analysis for Reserved Rate Utilization

Table 4 shows the Allocation of Variation to contributing factors for reserved rate utilization. As shown in Figures 5 and 7, reserved rate utilization of UDP customers is almost always good for both two and three color simulations. However, in spite of very low probability of a green packet getting dropped in the network, TCP customers are not able to fully utilize their reserved rate in all cases. The little variation in reserved rate utilization for UDP customers is explained largely by bucket size. Large bucket size means that more packets will get green color in the beginning of the simulation when green bucket is full. Green rate and interaction between green rate and bucket size explain a substantial part of the variation. This is because the definition of rate utilization metric has reserved rate in denominator. Thus, the part of the utilization coming from initially full bucket gets more weight for low reserved rate than for high reserved rates. Also, in two color simulations for reserved rates 153.6 kbps and 179.2 kbps, the network is oversubscribed and hence in some cases UDP customer has a reserved rate utilization lower than one. For TCP customers, green bucket size is the main factor in determining reserved rate utilization. TCP traffic because of its bursty nature is not able to fully utilize its reserved rate unless bucket size is sufficiently high. In our simulations, UDP customer sends data at a uniform rate of 1.28 Mbps and hence is able to fully utilize its reserved rate even when bucket size is low. However, TCP customers can have very poor utilization of reserved rate if bucket size is not sufficient. The minimum size of the leaky bucket required to fully utilize the token generation rate depends on the burstiness of the traffic.

## 4.3 ANOVA Analysis for Fairness

Fairness results shown in Figures 2a and 2b indicate that fairness in allocation of excess network bandwidth is very poor in two color simulations. With two colors, excess traffic of TCP as well as UDP customers is marked red and hence is given same treatment in the network. Congestion sensitive TCP flows reduce their data rate in response to congestion created by UDP flow. However, UDP flow keeps on sending data at the same rate as before. Thus, UDP flow gets most of the excess bandwidth and the fairness is poor. In three-color simulations, fairness results vary widely with fairness being good in many cases. Table 5 shows the important factors influencing fairness in three-color simulations as determined by ANOVA analysis. Yellow rate is the most important factor in determining fairness in three-color simulations. With three colors, excess TCP traffic can be colored yellow and thus distinguished from excess UDP traffic which is colored red. Network can protect congestion sensitive TCP traffic from congestion insensitive UDP traffic by giving better treatment to yellow packets than to red packets. Treatment given to yellow and red packets in the RED queues depends on RED parameters (drop thresholds and max drop probability values) for yellow and red packets. Fairness can be achieved by coloring excess TCP packets as yellow and setting the RED parameter values for packets of different colors correctly. In these simulations, we experiment with yellow rates of 12.8 kbps and 128 kbps. With a yellow rate of 12.8 kbps, only a fraction of excess TCP packets can be colored yellow at the traffic conditioner and thus resulting fairness in excess bandwidth distribution is not good. However with a yellow rate of 128 kbps, all excess TCP packets are colored yellow and good fairness is achieved with correct setting of RED parameters. Yellow bucket size also explains a substantial portion of variation in fairness results for three color simulations. This is because bursty TCP traffic can fully utilize its yellow rate only if yellow bucket size is sufficiently high. The interaction between yellow rate and yellow bucket size for three color fairness results is because of the fact that minimum size of the yellow bucket required for fully utilizing the yellow rate increases with yellow rate.

*Table 5: Main Factors Influencing Fairness Results in Three-Color Simulations*

| Factor/Interaction | Allocation of Variation (in %age) |
|---|---|
| Yellow Rate | 41.36 |
| Yellow Bucket Size | 28.95 |
| Interaction between Yellow Rate and Yellow Bucket Size | 26.49 |

It is evident that three colors are required to enable TCP flows get a fairshare of excess network resources. Excess TCP and UDP packets should be colored differently and network should treat them in such a manner so as to achieve fairness. Also, size of token buckets should be sufficiently high so that bursty TCP traffic can fully utilize the token generation rates.

## 5.0 CONCLUSIONS

One of the goals of deploying multiple drop precedence levels in an Assured Forwarding traffic class on a satellite network, either in GEO or LEO architecture, is to ensure that all customers achieve their reserved rate and a fair share of excess bandwidth. In this paper, we analyzed the impact of various factors affecting the performance of assured forwarding. The key conclusions are:

- The key performance parameter is the level of green (reserved) traffic. The combined reserved rate for all customers should be less than the network capacity. Network should be configured in such a manner so that in-profile traffic (colored green) does not suffer any packet loss and is successfully delivered to the destination.
- If the reserved traffic is overbooked, so that there is little excess capacity, two drop precedence give the same performance as three.
- The fair allocation of excess network bandwidth can be achieved only by giving different treatment to out-of-profile traffic of congestion sensitive and insensitive flows. The reason is that congestion sensitive flows reduce their data rate on detecting congestion however congestion insensitive flows keep on sending data as before. Thus, in order to prevent congestion insensitive flows from taking advantage of reduced data rate of congestion sensitive flows in case of congestion, excess congestion insensitive traffic should get much harsher treatment from the network than excess congestion sensitive traffic. Hence, it is important that excess congestion sensitive and insensitive traffic is colored differently so that network can distinguish between them. Clearly, three colors or levels of drop precedence are required for this purpose and is independent of the orbital selection.
- Classifiers have to distinguish between TCP and UDP packets in order to meaningfully utilize the three drop precedence.
- RED parameters and implementations have significant impact on the performance. Further work is required for recommendations on proper setting of RED parameters.

## 6.0 REFERENCES

[1] Sastri Kota, "Multimedia Satellite Networks: Issues and Challenges," Proc. SPIE International Symposium on Voice, Video, and Data Communications, Boston, Nov 1-5, 1998.

[2] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, An Architecture for Differentiated Services, RFC 2475, December 1998.

[3] J. Heinanen, F. Baker, W. Weiss, J. Wroclawski, Assured Forwarding PHB Group, RFC 2597, June 1999.

[4] V. Jacobson, K. Nichols, K. Poduri, An Expedited Forwarding PHB, RFC 2598, June 1999.

[5] S. Floyd, V. Jacobson, Random Early Detection Gateways for Congestion Avoidance, IEEE/ACM Transactions on Networking, 1(4): 397-413, August 1993.

[6] D. Clark, W. Fang, Explicit Allocation of Best Effort Packet Delivery Service, IEEE/ACM Transactions on Networking, August 1998.

[7] N. Seddigh, B. Nandy, P. Pieda, Study of TCP and UDP Interactions for the AF PHB, Internet Draft - Work in Progress, draft-nsbnpp-diffserv-tcpudpaf-00.pdf, June 1999.

[8] NS Simulator, Version 2.1, Available from http://www-mash.cs.berkeley.edu/ns.

[9] R. Jain, The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Simulation and Modeling, New York, John Wiley and Sons Inc., 1991.