

A Piecewise Linear Approach to Overbooking

Feng Huang, Casey Deccio, Robert Ball, Mark Clement, Quinn Snell
3370 TMCB
Brigham Young University
Provo, Utah 84602
clement@cs.byu.edu

Abstract

Overbooking is frequently used to increase the revenue generated by a network infrastructure without incurring additional costs. If the overbooking factor is chosen appropriately, additional virtual circuits can be admitted without degrading quality of service for existing customers. Most implementations use a single factor to accept a linear fraction of traffic requests. If a piecewise linear approach is used in admissions, additional traffic can be accepted without causing proportional increases in loss rate and utilization. This additional accepted traffic can significantly improve the profit margin for network service providers.

1 Introduction

Overbooking can significantly increase the number of virtual circuits that can be supported in a network infrastructure. This occurs because customers routinely request and purchase more bandwidth than they actually use. It can be difficult to determine exact values for overbooking factors because it is difficult to gather detailed data linking virtual circuit admission to trunk bandwidth increases. Although linear approaches can be analyzed more easily using regression software, they do not fully take advantage of overbooking opportunities. A piecewise linear approach can lead to more liberal admissions without damaging quality of service.

1.1 Utilization Patterns in Large Scale Networks

Airlines have been successful in selling more seats on an airplane than are physically available because statistically, it is unlikely that all of the customers will use their tickets on a particular flight. Airline policy will dictate which percentage of additional bookings they will accept even when all seats are full in order to maximize profit. An optimal policy would be able to predict the exact number of passengers that will utilize their tickets on a given flight and then sell that many tickets. If the policy underestimates the number of passengers, they there will be some passengers that have paid for a ticket who will not be able to use that flight. Airlines will typically offer free flights at another time in order to compensate customers for the inconvenience of taking another flight. If this lack of “Quality of Service”

inflight scheduling becomes too severe, customers may select another airline and the profit margin will decrease because of decreased demand. If the policy overestimates the number of passengers on a given flight, then seats will go empty and revenue will be lost.

Network users often behave in ways that are similar to airline passengers. They will purchase bandwidth that they often don’t utilize. Network administrators can maximize profit by accurately predicting the actual utilization of customer Virtual Circuits (VCs) in order to specify an overbooking policy that will sell as much bandwidth as possible without causing degradation in network “Quality of Service” resulting in dropped packets and increased delay and jitter.

This research analyzed data from 475230 Frame Relay virtual circuits on a wide area network with approximately 2000 trunks on three selected days during a normal work week in November 2002. The topology of the network is shown in Figure 1. The network is similar in nature to backbones supported by many national providers.

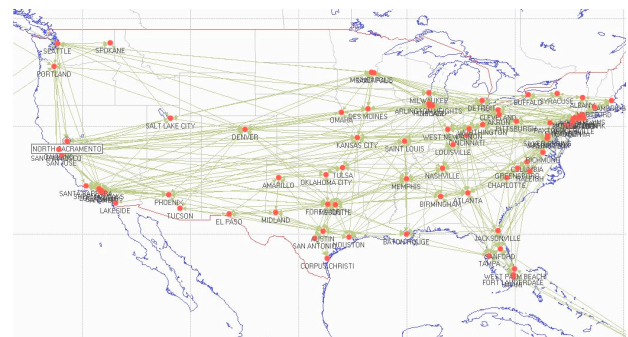


Figure 1. Network topology used in gathering overbooking statistics.

A summary of VC utilization is shown in Table 1. The Committed Information Rate (CIR) is the requested rate by the customer. The Peak Information Rate (PIR) is the peak rate that the customer can send at. The PIR may be much larger than the CIR and the customer should be able to send at rates between CIR and PIR for short

periods of time as long as the average bandwidth is no greater than the CIR. Packets in excess of the CIR will be marked for discard when network congestion occurs. The actual bandwidth utilization is shown in the table along with the number of virtual circuits that had that average utilization. The total values are weighted by the total bandwidth so that large VCs have a proportionally more significant impact on total averages than smaller VCs. The average utilization across all 400,000 virtual circuits was approximately 28%. This means that a provider could admit nearly four times as much traffic with an overbooking policy than he could with a strict admission policy. This overbooking should not result in lower quality of service since users are underutilizing their links by this factor.

CIR Range (Kbps)	PIR Range (Kbps)					Totals:
	0 512	513 1536	1537 2212	2213 +		
0 5	36.48% (49330)	109.06% (2930)	99.46% (9830)	192.77% (440)	53.56% (62530)	
6 64	27.16% (193360)	39.06% (24090)	34.13% (56050)	27.02% (690)	29.24% (274190)	
65 128	27.77% (28390)	34.91% (10940)	28.10% (17040)	32.90% (290)	29.18% (56660)	
129 256	30.07% (4970)	30.98% (17570)	29.31% (20250)	72.24% (440)	30.61% (43230)	
257 512	42.21% (50)	28.74% (8910)	29.42% (12810)	30.40% (660)	29.16% (22430)	
513 1024	0.00% (0)	27.00% (1690)	27.34% (8520)	26.48% (670)	27.22% (10880)	
1025 +	0.00% (0)	14.44% (310)	20.67% (3550)	28.40% (1000)	23.62% (5310)	
Totals:	27.82% (276100)	29.97% (66440)	27.13% (128050)	29.40% (4190)	28.15% (475230)	

Table 1. Utilization values for ranges of CIR and PIR values. The numbers in parentheses are the total number of virtual circuits in that range.

Table 1 also shows that there are some users that significantly overutilize their VCs. Note that a significant number of VCS use 193% of the negotiated CIR. A successful overbooking policy must correctly estimate utilization for both small and large VCs. A piecewise linear approach is well suited to accomplish this utilization prediction.

1.2 Data Volume

Strict timing demands are placed on trunk networking equipment. Packets from a large number of high bandwidth lines must be switched with little delay in order to maintain network stability. Statistic gathering functions must impose low levels of overhead if they are

to be carried out by backbone routers. Although edge switches may be able to gather Per-VC statistics, High performance network switches do not have the storage space or time to perform per-VC accounting.

Even if per-VC accounting were possible, the The total volume of data is too large. The CISCO networking equipment we obtained statistics from required 14Mbytes for one sample of VC values. If this data were gathered at any reasonable frequency, it would quickly overwhelm router storage capacity.

Hourly average data is routinely collected in order to gauge the health of the network without causing unmanageable storage requirements. Routers collect data for each trunk containing the total CIR and average utilization for the hour. At the end of a 24 hour period, the maximum hourly average is saved for later use. Similar statistics are gathered at edge routers where hourly averages are collected for each virtual circuit. This data is used to detect trunks with above average utilization and perform traffic engineering to lower maximum hour utilization. The data also provides detailed information that can be used to predict utilization given CIR values.

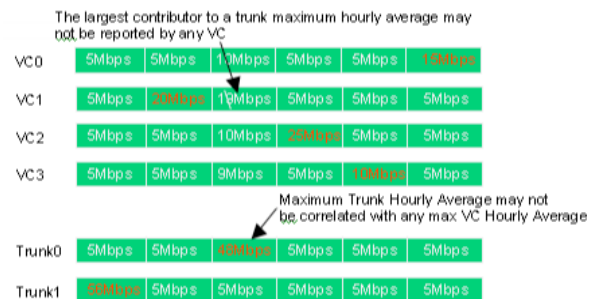


Figure 2: As time increases across the horizontal axis, example hourly averages for trunks and virtual circuits are shown. Saved maximum hourly averages may not be correlated, since they may be gathered from different hours during the day.

Initial research focused on ways of correlating virtual circuit data with trunk data in order to pinpoint the cause of congestion and perform global routing optimization. The uncorrelated data was difficult to use for these purposes, but can be used to assist in setting overbooking policy.

1.3 Previous research

Pazos and Gerla have performed research in overbooking in the use of ABR services on Internet backbones [6]. Their paper asserts that the use of ABR services permits better utilization of ATM resources than with CBR services, but that full utilization rarely occurs. They

propose an approach to improve resource utilization that relies on bandwidth overbooking.

Urgaonkar, et al. demonstrate the feasibility and benefits of overbooking resources in shared platforms [10]. Other research has also turned to overbooking as a reliable alternative to increasing link utilization [5,7,8].

Overbooking can be an effective tool for increasing link utilization for many different types of applications. Much of the existing research has focused on ways of choosing linear factors to use in overbooking. This research proposes a piecewise linear approach that promises to admit more traffic (and thus increase profits) while maintaining reasonable quality of service.

2 Overbooking Techniques

Several techniques have been explored in research to determine accurate overbooking techniques. The most obvious prediction technique is to use the CIR (the customers best estimate of bandwidth usage) multiplied by some constant to allocate bandwidth. Intuitively, it may seem likely that the PIR could also be combined in a linear system with two variables. Hardware vendors have recently incorporated additional overbooking hardware in core switches to allow for an increased linear factor to be applied on a trunk by trunk basis. It seems reasonable to believe that large trunks may be able to benefit from increased statistical multiplexing so that a more aggressive overbooking factor could be used. Although these schemes show some promise, utilization appears to be a non-linear function that does not fit well to this kind of analysis.

A statistical regression analysis was performed to determine the correlation between CIR and utilization. Results shown in Table 2 indicate that CIR is significant in predicting VC utilization. The linear model also indicates a overbooking factor of 20.4% based on the linear fit to data. The PIR is also correlated to utilization, but explains less of the variance in the data than CIR. The mean squared error did not decrease significantly when using PIR and CIR. This also shows that the inclusion of PIR in an overbooking technique may not lead to more accurate predictions of utilization.

Figures 3 and 4 show a more detailed view of the relationship between CIR, PIR and utilization. Although increasing PIR values are positively correlated with larger utilization values, this impact is greatest for smaller values of PIR. The negative correlation between CIR and utilization shown in Figure 4 is also non-linear. An overbooking factor that was piecewise linear provides a better fit for the data. For example, an overbooking factor based on CIR could have a negative coefficient for CIR values less than 64 and a small positive value for larger CIR values.

Estimate		MS Error	t value	Pr > t		
CIR alone						
0.2115		9398	421	<.0001		
CIR and PIR						
CIR	PIR		CIR	PIR	CIR	PIR
0.204	0.004	9368	304	39	<.0001	<.0001

Table 2. GLM regression analysis of CIR and PIR correlation with utilization. The first set of results are from a regression using only CIR. The second set results from a multiple linear regression with both PIR and CIR.

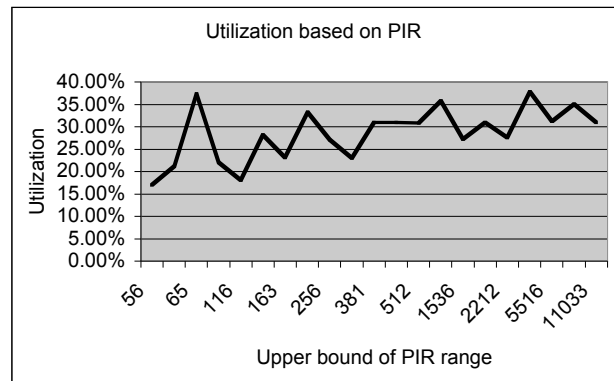


Figure 3. Although there is an upward trend in higher utilizations for larger PIR values, the impact of PIR on utilization is small

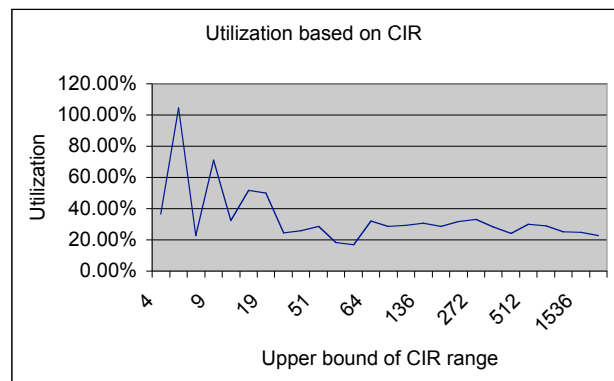


Figure 4. There is a negative correlation between CIR and utilization for small CIR values. For large CIR values utilization is invariant with respect to CIR.

An additional investigation was made to determine the effect of increasing the overbooking factor on trunks with higher bandwidth values. Experiments were performed with DS3 and OC3c connections. Figure 5 shows that as the overbooking factor was varied, the utilization remained approximately the same for DS3 and OC3c links. If the utilization would have been significantly lower for the OC3c connection then an additional trunk overbooking value could be used. Since the values were nearly the same, this approach will

probably not yield increased admissions with little additional utilization.

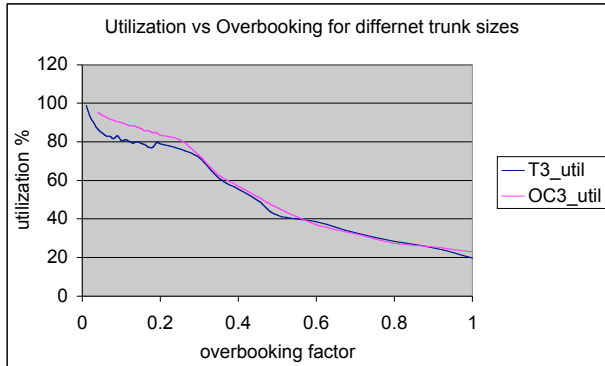


Figure 5. DS3 trunks do not appear significantly different from OC3c trunks in terms of the impact of particular overbooking values.

3 Experimental Results

Although log data is useful in determining patterns, simulations can help in detailing the impact of changes on performance. Several experiments were performed to determine the impact of changes in the overbooking policy on utilization and packet loss.

3.1 Simulation Infrastructure

The NS network simulator was used to create a simple topology where overbooking could be explored. A dogbone topology was configured with a large number of sources communicating through two routers. Figure 5 shows the overall simulation configuration.

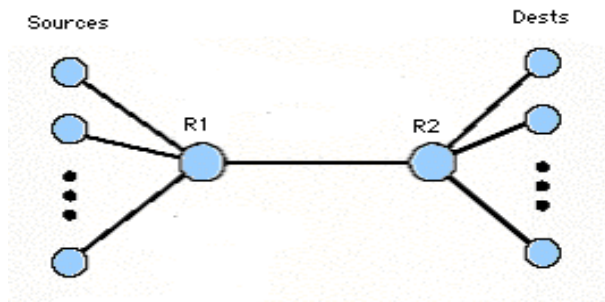


Figure 4. Simulation architecture. Sources are admitted until the trunk bandwidth has been filled.

For linear CIR based admissions, a random source was chosen out of the tracefile with 400,000 virtual circuits represented. The remaining bandwidth for the trunk was then set to be:

$$\text{Available Bandwidth}(i) = \text{Available Bandwidth}(i-1) - \text{CIR}[j] * \text{overbooking factor}$$

This process continued until the available trunk bandwidth reached zero. Smaller overbooking factors allow for larger amounts of data to be admitted into the simulation. In practice, values as low as 5% are commonly used for overbooking. For this reason, all of these values were included in the experiment.

Exponential sources were then configured in the ns simulator with average bandwidth determined by the utilization value for each source that had been randomly chosen from the trace file. Experiments were performed with each overbooking factor to determine the impact of overbooking policy on utilization and packet loss.

For the piecewise linear based admissions, available bandwidth was calculated in the following way.

$$\text{If}(\text{CIR} < \text{threshold}) \\ \text{Available Bandwidth}(i) = \text{Available Bandwidth}(i-1) - \text{CIR}[j] * \text{overbooking factor1}$$

$$\text{If}(\text{CIR} > \text{threshold}) \\ \text{Available Bandwidth}(i) = \text{Available Bandwidth}(i-1) - \text{CIR}[j] * \text{overbooking factor2}$$

Since virtual circuits with larger CIR values result in lower utilization values, it is possible to admit as much as 25% more traffic without increasing the utilization or packet loss for the system. Figure 5 and 6 show utilization and packet loss for the linear CIR based overbooking. Figure 7 and 8 show the results of a piecewise linear system. The piecewise linear system was able to admit nearly 25% more circuits without significantly increasing utilization and packet loss.

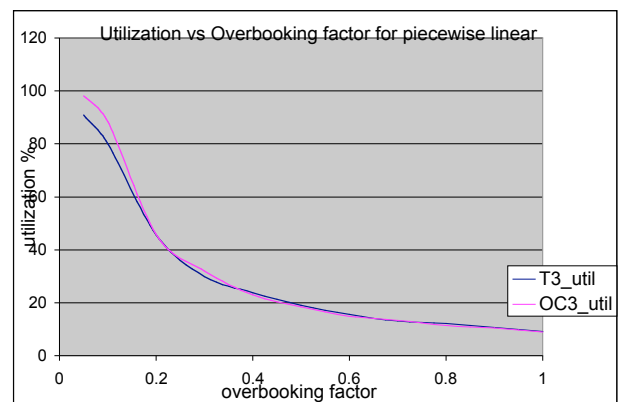
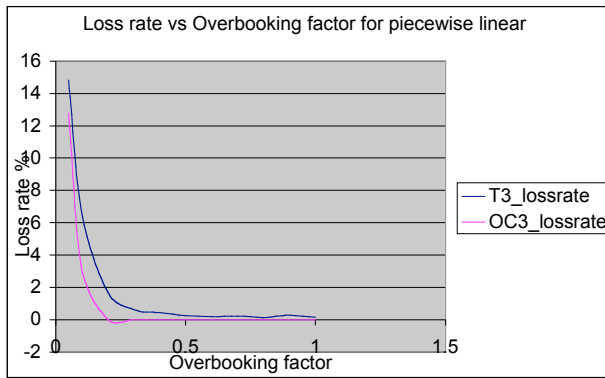


Figure 7. Utilization for piecewise linear overbooking policy.



4 Conclusions

Overbooking can significantly increase the profit margin for network providers. If an overbooking policy is implemented with a piecewise linear prediction model, more accurate measures can be performed and additional traffic can be accepted.

Traditional overbooking with CIR and PIR are inefficient because they do not account for the non-linear nature of the CIR to utilization relationship. Piecewise linear approaches can model this relationship more accurately and allow additional virtual circuits to be admitted without degrading performance.

5 Bibliography

[1] Cisco Systems, Incorporated. Series Universal Broadband Routers Chapter 3: Configuring Cable Modem Interface Features. July 14, 2003. http://www.cisco.com/en/US/products/hw/cable/ps2217/products_configuration_guide_chapter09186a00801170dd.html.

[2] Cisco Systems, Incorporated. Series Switches Traffic Management. July 14, 2003. <http://www.cisco.com/en/US/products/hw/switches/ps52>

5/products_configuration_guide_chapter09186a0080086fee.html

[3] Kweon, S. and Shin, K. Real-time transport of MPEG video with a statistically guaranteed loss ratio of ATM networks. *IEEE Transactions on Parallel and Distributed Systems*, 12(4):387-403, April 2001.

[4] Parulkar G., Schmidt, D., Kraemer, E., Turner J., and Kantawala A., "An Architecture for Monitoring, Visualization, and Control of Gigabit Networks". *IEEE Network*, vol. 11, September/October 1997.

[5] Pazos, C. and Gerla. M. "A Rate Based Backpressure Flow Control for the Internet", In *Proceedings of HPN'98*. Vienna, Austria - 1998.

[6] Pazos, C. and Gerla. M. "Improving Internet Traffic Transport over ABR Backbones Through Bandwidth Overbooking", In *Proceedings of Globecom '98*. Sidney, Australia, 1998.

[7] Pazos, C. and Gerla. M. "Improving IP over ATM Efficiency Using Bandwidth Overbooking". In *Proceedings of LANMAN'98*. Alberta, Canada - 1998.

[8] Turner, Jonathan S. "Maintaining High Throughput During Overload in ATM Switches". *Proceeding of IEEE INFOCOM 1996*, San Francisco, California, April 1996, pp. 287 – 295.

[9] Younger, Mary Sue. "A Handbook for Linear Regression". Duxbury Press, Belmont California. 1979, p. 1.

[10] Uргаonkar, B., Shenoy P., and Roscoe T.. "Resource Overbooking and Application Profiling in Shared Hosting Platforms". *USENIX Symposium on Operating Systems Design and Implementation (OSDI 2002)*, Dec. 2002.