Top Percentile Network Pricing and the Economics of Multi-Homing

Joseph Levy[1], Hanoch Levy[2*], Yaron Kahana[3]
[1]Comgates Ltd. Herzliya, Israel
[2] School of Computer Science, Tel-Aviv University
Tel-Aviv, Israel
Email: hanoch@cs.tau.ac.il
[3]Comgates Ltd. Herzliya, Israel

Abstract

Under *multi-homing* an Internet Service Provider (ISP) connects to the Internet via multiplicity of network providers. This allows the provider to receive proper quality of service when one of the networks fails. An ISP that uses multi-homing is subject to extra charges due to the use of multiple networks. Such extra costs can be very drastic under fixed cost pricing and non-meaningful under per-usage pricing. This work deals with the question of how large are these costs under *top-percentile pricing*, a relatively new and popular pricing regime. We provide a general formulation of this problem as well as its probabilistic analysis, and derive the expected cost faced by the ISP. We numerically examine several typical scenarios and demonstrate that despite the fact that this pricing aims at the peak traffic of the ISP (similarly to fixed cost), the bandwidth cost of multi-homing is not much higher than that of single-homing.

Keywords: Pricing, Traffic-Engineering, Multi-Homing, Top-Percentile pricing.

1 Introduction

Multi-homing [4] is an architecture used by *Internet Service Providers* (ISP's) to connect to the Internet via multiple *network providers* (backbones). This connectivity improves the network reliability and quality of the ISP, since when one of the networks fails and its quality degrades the ISP can use the alternate network. While multi-homing improves the ISP's experienced QOS, it increases its inflicted costs and thus may make it economically inefficient.

The objective of this work is to examine the bandwidth costs inflicted on an ISP and the economical viability of the multi-homing concept. This, however depends on the cost structure used by the network providers. The traditional cost structure, that has served as the basis for much of the research in network design and traffic engineering is the *fixed-cost,* in which the customer pays for a fixed capacity regardless of how much of it is being actually used. Under this cost structure, the cost of dual-homing (a special case of multi-homing with two network connections) is twice the cost of a single connection and thus may be too expensive. An alternative to the fixed-cost is the pure *per-usage cost* in which the customer pays exactly for the total bandwidth (measured in Mega-bytes and integrated over the month). Under this pricing there is hardly extra cost inflicted on the multi-homing architecture, since on each of the networks the customer pays only for the bytes transferred.

A new and increasingly popular pricing policy is the *top-percentile pricing* [2], [3]. To apply top-percentile charges, the provider measures the amount of data sent at fixed intervals (say 5 minutes or 15 minutes). It then evaluates these values for all the intervals over the charge period (8640=30x24x12 intervals per month, in the case of monthly charges and 5 minute intervals) and selects the "top 95%", i.e. the 95^{th} percentile (or *p%*) interval (the 8208^{th} lowest traffic interval, in the example) . The cost is then computed as a function of that number. Top-percentile pricing resembles somewhat the fixed price policy as it charges for one of the largest volume intervals, and thus if one does not use the network for 95% of the time, one still pays as if one used it for all the time. As such, there is a question whether multi-homing is economically viable under this pricing.

In this work we focus on analyzing the top-percentile pricing costs, in particular under the multi-homing environment. We develop a probabilistic model that models the stochastic nature of traffic. We provide the mathematical analysis of the model and derive its expected cost. We then use the model to examine several numerical examples. The examination reveals that the top-percentile pricing inflicts much lower costs than the fixed pricing. Accounting for bandwidth cost - the multi-homing cost inflicted on the ISP is higher from non-multi-homing cost only by several percents (as opposed to doubling it under fixed cost structure). As such we conclude that multi-homing is economically viable (unless the cost structure contains significant fixed price components). The analysis methodology developed in this work can be further used in more general traffic engineering and network planning frameworks.

---

[*] This work was done while the author was partially with Comgates Ltd., Herzliya, Israel.

2 Mathematical Formulation

2.1 Model and Assumptions

Let L_1,...L_T be the set of network providers. Assume that the charge period of a network provider is divided into T intervals of equal length for the purpose of top-percentile charge calculation (for simplicity, assume that T is the same for all providers). A network provider calculates the traffic shipped through the network during each of the T intervals, and the cost inflicted on the customer (normally a service provider) is determined, in the top percentile pricing by the volume of traffic shipped at the top percentile interval. For example, in a typical situation each of the intervals 15 minutes long and the number of intervals in the month is 2880. If top 95% pricing is used then the top 144th interval traffic forms the charge basis.

To account for a general model, we model the traffic demands in the following general form: A set of I traffic demands is a pair (X, π), where $X = \left\{X_i(t), i=1,\ldots,I, t=1,\ldots T\right\}$ is a collection of positive independent random variables, and $\pi = \left\{\pi_{ij}, i=1,\ldots,I, j=1,\ldots J\right\}$, such that $\pi_{ij} \geq 0$ for all i and j, and $\sum_{j=1}^{J} \pi_{ij} = 1$ for all i. The variables $X_i(t)$ represent the ith traffic demand at time interval t, and π_{ij} is interpreted as the probability (or proportion of time) that the ith traffic demand is routed via the jth network provider. The variables π_{ij} will be used as a general machinery to account for network failure or network quality degradation. Also assume, without loss of generality, that the set of T intervals is partitioned into K subsets, where each subset consists of T_k consecutive intervals and $\sum_{k=1}^{K} T_k = T$, and that $\left\{X_i(t), \quad t = T_{k-1}+1,\ldots,T_{k-1}+T_k\right\}$[1] have a common probability distribution function $F_i^k(\cdot)$. The latter assumption stems from practical considerations by which typically one is not equipped with different statistical information for each $X_i(t)$ (for each short time interval) but rather with more general statistics (e.g the amount of traffic between the hours 8-12 or 12-16).

2.2 The distribution of combined traffic flows

To define $D_j(t)$, the combined traffic demand on network provider j at time interval t, let $U_{ij}(t)$ be a random variable that takes the values 1 and 0 with probabilities π_{ij} and $1-\pi_{ij}$ respectively. Then the event $\{U_{ij}(t) = 1\}$ means that during time interval t traffic demand i is routed through service provider j. We assume that the random variables $U_{ij}(t)$ are independent of each other. Then $D_j(t) = \sum_{i=1}^{I} X_i(t)\cdot U_{ij}(t)$. Let $G_j^k(x) = P\big(D_j(t) \leq x\big)$, where $T_{k-1}+1 \leq t \leq T_{k-1}+T_k$, be the probability distribution function of $D_j(t)$.

Assertion 1:
For all $x \geq 0$,

$$G_j^k(x) = \sum_{u_{1j}=0}^{1} \cdots \sum_{u_{Ij}=0}^{1} \left[\prod_{i=1}^{I} \big((2\pi_{ij}-1)\cdot u_{ij} + 1 - \pi_{ij}\big)\right] \cdot \left(\mathop{*}_{i=1}^{I} u_{ij} F_i^k(x)\right) \tag{1}$$

where $*$ denotes the convolution operation.
Proof is straightforward, by conditioning on $U_{ij}(t)$.

Corollary 1:
Denote by $\Phi(x \mid \mu, \sigma^2)$ the Normal distribution with mean μ and variance σ^2. Assume that for all i and k, $F_i^k(x) = \Phi(x \mid \mu_{ik}, \sigma_{ki}^2)$. Then, knowing that a linear combination of independent Normal random variables is also Normally distributed:

[1] Where $T_0 \equiv 0$.

$$G_j^k(x) = \sum_{u_{1j}=0}^{1} \cdots \sum_{u_{Ij}=0}^{1} \left[\prod_{i=1}^{I} \left((2\pi_{ij} - 1) \cdot u_{ij} + 1 - \pi_{ij} \right) \right] \cdot \Phi\left(x \mid \sum_{i=1}^{I} u_{ij}\mu_{ik}, \ \sum_{i=1}^{I} u_{ij}\sigma_{ik}^2) \right) \tag{2}$$

2.3 The distribution of the traffic flow top-percentile

Our aim is at computing the *expected cost* of the customer. Thus, once the combined traffic demand function over a network provider is calculated, the expected value of its top percentile needs to be calculated.

Assertion 2:

Let $D_{(r)}$ be the r-largest of $D(t), \quad t = 1, \ldots, T$ [2], and let $B(\cdot \mid n, p)$ be the cumulative distribution function of a Binomial random variable with parameters n and p. Then for all $y \geq 0$,

$$P\left(D_{(r)} \leq y\right) = 1 - \mathop{*}_{k=1}^{K} B\left(r - 1 \mid T_k, G^k(y)\right).$$

Proof:

Fix $y \geq 0$ and for each $1 \leq k \leq K$, define $N_k^y = \#\{D(t) \leq y \ , t = T_{k-1} + 1, \ldots, T_{k-1} + T_k\}$. First observe that $D_{(r)} \leq y$ if and only if $\sum_{k=1}^{K} N_k^y \geq r$. Secondly, note that N_k^y is a Binomial random variable with parameters T_k and $G^k(y)$. Then

$$P\left(D_{(r)} \leq y\right) = P\left(\sum_{k=1}^{K} N_k^y \geq r\right) = 1 - P\left(\sum_{k=1}^{K} N_k^y \leq r - 1\right) = 1 - \mathop{*}_{k=1}^{K} B_k\left(r - 1 \mid T_k, G^k(y)\right) \tag{3}$$

QED.

Now, the expected value of $D_{(r)}$ is given by:

$$E[D_{(r)}] = \int_0^\infty y dP\left(D_{(r)} \leq y\right) \tag{4}$$

2.4 Implementation Considerations and Computational Complexity

These derivations can be done in an efficient implementation whose overall complexity is given by $O(2^I \mid X \mid + K \mid Y \mid)$. Generally we have $\mid X \mid = \mid Y \mid$, and thus the complexity is $O((2^I + K) \mid X \mid)$. The detailed complexity analysis is provided in [1].

3 Numerical Examples

Below we use a typical scenario in order to evaluate the economical viability of operating in the multi-homing mode. We are considering a service provider (customer) which has to place two traffic demands (the customer can easily form two demands by classifying its traffic into two classes) over one or two networks, provided by different providers. The customer may use a single provider, in which case when the network fails the customer is subject to severe quality degradation. Alternatively, the customer may purchase service at two network providers, and use one network as a primary and one network as an alternate (to be used when the primary fails). We assume that the network providers divide the month to intervals of 15 minutes length (that is about 3000 intervals) and the charges are set as function of the load on the top 5% interval. We consider a traffic requirement faced by the customer to consist of random variables that depend on time. For example the traffic volume in the morning is a random variable whose mean is much larger than the random variable of the traffic at a night hour.

For the sake of the examples we assume that we are given 6 traffic representatives, representing say, the traffic volume of 9AM-1PM, 1PM-5PM, 5PM-9PM, 9PM-1AM, 1AM-5AM and 5AM-9AM. For each of these representatives there are 496 random variables (16 per day, for 31 days) all 496 are mutually independent and identically distributed. For ease of presentation we will assume that the number of random variables is 500 (all together 3000 per month).

---

[2] For ease of reading, the network index is omitted.

We now consider 2 traffic demands, X_1, X_2 $X_i(t)$ is a random variable denoting the traffic demand of source *i* at time *t,* and two networks L_1, L_2 where we assume that network L_i fails with probability p_i. We will evaluate the cost of various policies for placing the demands on the networks.

3.1 The traffic demands

We will consider a sample traffic demand that represents differences between day and night. The demand is given by: $X_1(1),\ldots,X_1(500)$ [3]is *uniform* with M=20, S=5.77, where M is the mean and S is the standard deviation. All other demands are uniform: $X_1(501),\ldots,X_1(1000)$ with M=30, S=5.77, $X_1(1001),\ldots,X_1(1500)$ with M=50, S=11.6, $X_1(1501),\ldots,X_1(2000)$ with M=70, S=11.6, $X_1(2001),\ldots,X_1(2500)$ with M=100, S=17.3, $X_1(2501),\ldots,X_1(3000)$ with M=120, S=17.3. An alternative traffic assumption used is the *normal distribution* in which we have: $X_1(1),\ldots,X_1(500)$ with M=20, S=3.33, $X_1(501),\ldots,X_1(1000)$ with M=30, S= 3.33, $X_1(1001),\ldots,X_1(1500)$ with M=50, S=6.66, $X_1(1501),\ldots,X_1(2000)$ with M=70, S=6.66, $X_1(2001),\ldots,X_1(2500)$ with M=10, S=10, $X_1(2501),\ldots,X_1(3000)$ with =120 , S=10.

3.2 A single demand over a single provider

Here we consider the cost of traffic demand X_1 when applied on a single network, say L_1. This is given by the top 95% of the traffic of a single source on a single network.

3.3 No Multi-Homing: Two demands over a single provider

Here we consider the cost of running both X_1 and X_2 on a single network, say L_1. As in the previous sub-section this is a non multi-homing system, and thus the traffic may be subject to non-recoverable quality degradation. This evaluation is added for the sake of comparison.

3.4. Multi Homing: Two demands over two providers:

3.4.1 Identical primary networks

In this setting we consider a situation in which we are faced with two traffic demands, X_1 and X_2 which are statistically identical to each other. We assume that both demands are placed on L_1 as primary and on L_2 as secondary. We evaluate the cost of this solution as a function of π_1 the probability that L_1 fails.

3.4.2 Crossed-over primary networks

This setting is identical to that of Section 4.4.1 but where we assume that one demand is placed on L_1 as primary and on L_2 as secondary and the other demand is placed on L_2 as primary and on L_1 as secondary. For simplicity of presentation we also assume that the failure probabilities obey $\pi_1 = \pi_2$ and evaluate the cost as a function of π_1.

3.5 Results

The results provided below are based on the analysis provided in Section 2, which in some of the cases is combined with an approximation. The results for the normal distribution are provided in Figure 1. In the figure we depict the cost of running the demands, under various configurations, as a function of the probability of network failure. The solid (blue) line represents the cost of a single demand times two (3.2). This is given as a reference. The dashed (green) line represents the cost of running the two demands on a single network, under the assumption that there is no secondary network (no multi-homing, (3.3)). The dotted (red) line represents a multi-homing scenario where the two demands are

[3] The reader should note that the indexing of the internals (1 through 3000) is done for mathematical convenience. The index is not necessarily related to the specific time of the interval 9during the month or the day).

placed on an identical primary (and they also share the same secondary network, (3.4.1)). The dashed (light blue) curve represents the placement of the two demands on different primary networks where each of these networks serves as the secondary for the other demand (3.4.2).

We can observe the following properties: The cost of running two demands on the same network (no multi-homing) is less than double the cost of single demand. This is due to statistical multiplexing. The difference is in the order of 1.5%. The cost of running two demands on the same primary network and using a second network for alternate (dotted) is very low for low failure rates (up 4% failure rate) but then increases quite sharply for high failure rates. The cost of running the two demands on different primary networks (and use the other networks for alternate) is somewhat higher (than running them on the same network) for low failure rates but grows more modestly for high failure rates. Overall – the extra cost due to applying the multi-homing mechanism is in the order of several percents.

Figure 2 demonstrates quite similar results for the uniform distribution traffic.

We examined a variety of other cases [1], including scenarios consisting of more than 2 traffic demands and other percentile costs. We found out that in all these cases the extra bandwidth cost is in the order of several percents.

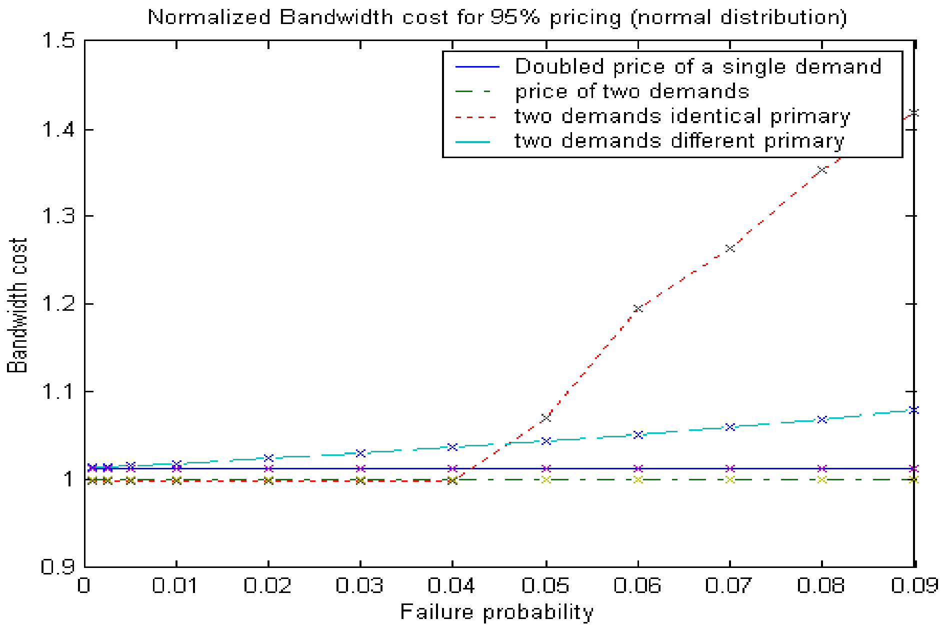

Figure 1: The relative cost of multi-homing for two traffic demands (normal distribution).

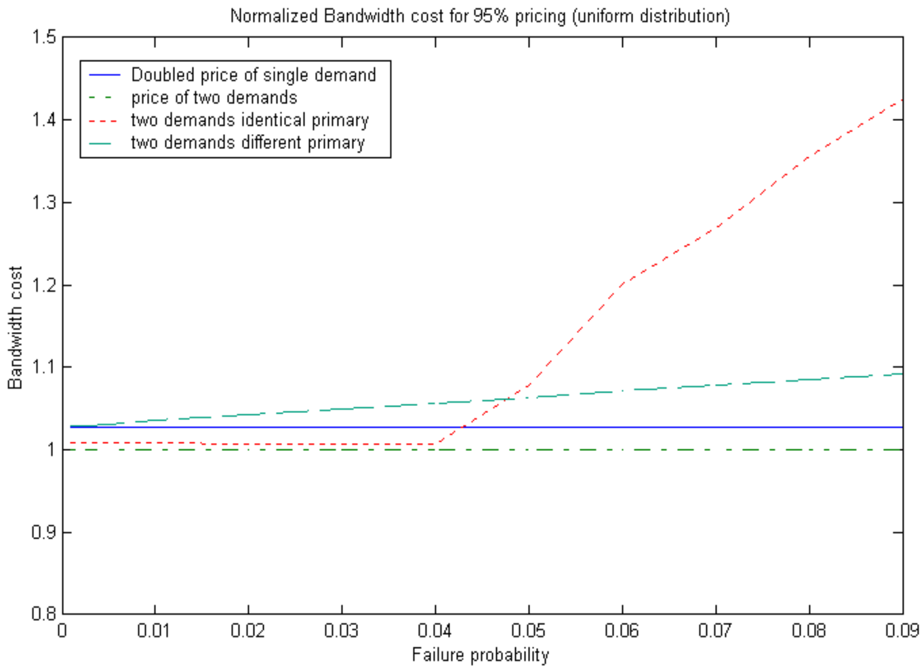

Figure 2: The relative cost of multi-homing for two traffic demands (uniform distribution)

References

[1] J. Levy, Handling Top-Percentile Cost Functions, Technical Report, COMGATES Ltd, October 2002.

[2] T. Monk, Inter-domain Traffic Engineering: Applications in Complex Networks, Ixia, presented at NFOEC2002, Dallas, TX, September 15-19, 2002. http://www.ixiacom.com/support/techinfo/presentations/Paper_IDTE_NFOEC2002.pdf.

[3] A. M. Odlyzko, Internet pricing and the history of communications, *Computer Networks* 36 (2001), pp. 493-517.

[4] A. Orda and R. Rom, Multi-homing in Computer Networks – A Topology Design Approach, *Computer Networks and ISDN Systems*, Vol. 18, pp. 133-141, February 1990.