Does Over-Provisioning Become More or Less Efficient as Networks Grow Larger? *

Yaqing Huang Roch Guérin
Dept. of Elec. & Sys. Eng., U. of Pennsylvania
200 S. 33rd Street, Philadelphia, PA 19104, USA
{yaqing@seas, guerin@ee}.upenn.edu

Abstract

IP networks have seen tremendous growth in not only their size and speed, but also in the volume of traffic they carry. Over-provisioning is commonly used to protect network performance against traffic variations, be they caused by failures or transient surges. This paper investigates the influence that increasing network size has on the efficacy of over-provisioning in absorbing a certain range of traffic variations and preserving performance guarantees. For that purpose, we develop a general model that accounts for network topology, base offered traffic, and traffic variations, and allows us to explore how their combination behaves as the network and the traffic it carries grow. The model's generality enables us to investigate several representative scenarios and to identify critical thresholds in the relation between network and traffic growth, which delineate regions where a given amount of over-provisioning provides increasingly better protection against traffic variations. The results offer insight into how to grow IP networks in order to enhance their robustness.

1 Introduction

Over-provisioning is commonly used to protect network performance against traffic variations. It typically involves dimensioning links so that their bandwidth exceeds the expected traffic load by a certain margin, which is selected to ensure that the link can absorb both expected and unexpected traffic fluctuations. In other words, the bandwidth B_l of link l is chosen such that $B_l \geq (1+\beta)\widetilde{f}_l^b$, where $\widetilde{f}_l^b$ is the expected base offered traffic on link l, and $\beta > 0$ denotes the *over-provisioning factor*. The challenge is in determining what β is needed to offer a desired level of protection against a given range of traffic surges. As a result, and because of the need to accommodate the relatively large traffic fluctuations caused by link failures, values of $\beta \approx 5$ or even higher are not uncommon in large IP backbones [1, 18]. However, the emergence of high-speed applications and access links, and the ever greater heterogeneity of user traffic profiles, e.g., machine-to-machine, mean that there is considerable uncertainty regarding whether even such conservative over-provisioning factors can remain adequate as networks and the user population they serve continue growing.

It is, therefore, of interest to develop a better understanding of if and when the efficiency of over-provisioning in protecting network performance against traffic variations changes as networks grow larger. We define efficiency more precisely later, but it essentially amounts to identifying the minimum over-provisioning factor, which ensures that even in the presence of traffic variations, the network still maintains good performance, namely, the actual link load remains below link capacity with a certain (high) target probability. In other words, efficiency is a measure of the network's *robustness* against such traffic variations.

Focusing on the impact of network size is relatively natural, as the increased efficiency of large scale systems is a well documented phenomenon. For example, a property of the Erlang formula known as "trunking efficiency" tells us that as the number of circuits (trunks) grows large, the call blocking probability goes down to zero even as the system load approaches 100%. Similarly, the statistical multiplexing gain achievable on data links is known to increase (under certain assumptions) with the link bandwidth and the number of flows. The situation is, however, much less clear when it comes to networks, because of the many parameters involved, e.g., topology, routing, traffic model, etc., and their complex interactions. For example, as we shall see later, some related works that investigated the likely evolution of maximum link loads with network size, reached somewhat different conclusions simply because of their use of different models for routing and network traffic.

As a result, and because there is considerable uncer-

*This work was supported by NSF grants ANI-9902943 and ITR-0085930.

tainty regarding appropriate models for capturing network and traffic growth, the approach we take is centered around the development of a *parametric* model that enables us to investigate the efficiency of over-provisioning across a reasonable range of operating conditions and assumptions. As described in Section 5, using such a model we are able to identify the presence of *thresholds* in the relative growth of traffic volume compared to network size (measured in number of network nodes), which play a critical role in determining how the efficiency of over-provisioning changes with network size. The results provide insight into how to possibly grow backbone networks to improve their robustness against traffic variations.

The paper's contributions are two-fold. It formulates a number of scenarios that capture possible evolutions of network size and traffic volume, and for each assesses the efficiency of over-provisioning in absorbing traffic variations. However, the paper goes beyond investigating specific scenarios, it also develops a flexible model that can serve as a basis for further investigations using different traffic models. This is of interest, as the continued evolution of the network and its usage is expected to yield new and potentially different traffic patterns, whose impact will need to be assessed anew.

The rest of the paper is organized as follows. Section 2 introduces and motivates our models for both network topology and traffic. Section 3 discusses related works and reflects on the differences between these works and the present paper. Our main analytical results are presented in Section 4. In Section 5, we explore if and when over-provisioning becomes more efficient in larger networks by applying the tools of Section 4 to a number of representative traffic scenarios. Section 6 concludes the paper.

2 Network and Traffic Models

The ability to absorb traffic variations through over-provisioning depends on both the amount of spare bandwidth set aside for traffic surges, and the magnitude of these surges. The main parameters that influence these two quantities include the network topology, the routing in use in the network, the "base" traffic matrix, i.e., the anticipated "*average*" volume of traffic between different source-destination (SD) pairs, and the sources and destinations of traffic surges as well as their intensities. Specifically, network topology, routing, and the base traffic matrix together generate the expected load a link is designed to carry. The expected load and the over-provisioning factor β in turn determine the resulting link capacity, i.e., if the expected load on link l is f_l, the corresponding link bandwidth is $(1+\beta)f_l$ and the spare capacity available to absorb traffic surges is βf_l. Similarly, network topology, routing, and the traffic surge matrix that captures the intensity and distribution of traffic surges across SD pairs, determine which links will be affected and to what extent. As a result, exploring how network size influences the efficiency of over-provisioning calls for a model that incorporates all of the above parameters, and provides as much flexibility as possible in their specification. In this section, we describe the approach we take for developing such a model and the choices we make.

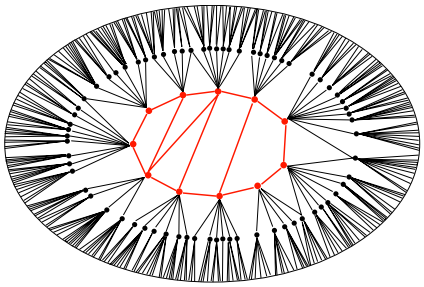


Figure 1. Abilene inspired topology.

Let us first consider what is involved in selecting a network model that is both representative of the network structure, and can be scaled to account for its growth. Developing models that reflect the key characteristics of today's IP networks, and can track their evolution has been an active research area over the past few years. Large IP networks, such as the Internet, have been observed to obey a power-law degree distribution [4, 8, 9]. However, Li *et. al.* [13] observed that while a power-law degree distribution indeed matches observations, it is by itself not sufficient as networks of totally different topological structure can obey the same power-law degree distribution. In particular, most methods that are used to generate power-law networks, e.g., the preferential attachment method [4], can fail to capture other structural properties of IP topologies such as mesh-connected backbones and tree-connected access networks. Li *et. al.* argued that more realistic topologies should reflect the different topological structures of each network level, while preserving an overall power-law degree distribution.

In this paper, we focus on the *backbone* component of networks where over-provisioning is most common, and therefore only model the topology of that component. Note that the topological properties of the backbone can be quite different from that of access layers as pointed out in [13]. A typical network topology [13], as shown in Fig. 1[1], consists of a backbone network (the red nodes and red links), which often exhibits a mesh-like topology, and several layers of access networks that exhibit tree-like structures. The backbone itself is usually comprised of one, possibly two, levels of hierarchy, especially when dealing with either very large backbones or the interconnection of multiple backbones. The number of backbone nodes (routers) is typically much smaller than the total number of nodes in the network, so that the degree distribution of backbone nodes

[1] See [1] for the topology of the Abilene network.

(counting only connections between them) has little or no impact on the degree distribution of the overall network, which as mentioned above commonly obeys a power-law distribution. In this work, we represent the backbone using a two-level random graph model that captures its mesh-like topology and is amenable to analysis. The impact of access networks is accounted for by incorporating the traffic they source and sink into the traffic associated with the backbone node to which they are attached. In other words, the power-law degree distribution of the overall network is reflected in the distribution of the traffic intensities of backbone nodes. In Section 5, we investigate a gravity traffic pattern that is based on this approach.

The first or top level of the *backbone* topology is generated using a $\mathcal{G}(n_1, p_1)$ random graph model [6]. A random graph G generated by the $\mathcal{G}(n, p)$ model has n vertices and includes each potential edge (between any two nodes) with probability p. Specifically, the $\mathcal{G}(n, p)$ probability space consists of all graphs with vertex set $V = \{1, 2, \ldots, n\}$, and $Pr\{g\} = p^{e_g}(1-p)^{\binom{n}{2}-e_g}$ for all $g \in \mathcal{G}(n, p)$, in which e_g denotes the number of edges in graph g. For lack of a better notation, we call nodes in the top level "domains." We then expand each domain using a $\mathcal{G}(n_2, p_2)$ random graph model, where each node now represents a physical router. The end-points of "inter-domain" links in the top level are assigned to randomly selected nodes in the corresponding "domains." It is worth noting that while we limit ourselves in this paper to two levels, which we believe is a reasonable first step to capture the key topological features of IP backbones, the model can be extended to more than two levels to account for more complex topologies, e.g., multiple tiers of providers with different connectivity (degree p_i) at each tier. In terms of routing, we assume a hierarchical shortest path routing policy that first selects shortest paths at the inter-domain level, and expands them inside each domain using again shortest paths. When multiple equal cost paths are available, the traffic is either evenly split among them or sent on one of them chosen randomly. We believe that such a choice is a reasonable approximation of the routing policies used in IP networks.

Besides network topology and routing, the two other factors that a model needs to specify are the base offered traffic, and how to represent traffic surges.

The base offered traffic is the traffic that the network expects to carry under normal circumstances and for which it is, therefore, designed. Statistics on carried traffic are routinely collected by service providers, e.g., using SNMP [7] to poll counters that track the volume of data transmitted on links. These measurements are then used to build a (base) offered traffic matrix, which because of the inherent "noise" in the measurement process, represents only an *estimate* of the expected traffic offered to the network. Some of the factors that contribute to variations in the base traffic around these expected values include changes over time in the intensity and the geographical distribution of flows between end-users, e.g., from peak hour traffic to off-peak hour traffic, from corporate traffic to residential traffic, etc. Similarly, typical SNMP poll cycles are of the order of 5 minutes, and therefore only capture link loads averaged over such durations, which masks out traffic variations at *smaller time scales* [15]. See [10, 16, 17] for extensive discussions on this issue, but the sources of errors or fluctuations in the base traffic are well understood, so that their magnitude can be estimated and their impact incorporated in the dimensioning of the network, i.e., reflected in the choice of β. In our analysis, we account for possible variations in the base traffic by allowing the specification of an arbitrary general traffic matrix $\mathbf{X}$, whose entries are allowed to be random variables. The mean values of the traffic matrix entries together with routing are used to compute average link loads, and therefore determine the resulting link capacities.

Traffic surges on the other hand, represent *unexpected* variations in offered traffic in terms of either their location or magnitude. Such fluctuations have many causes and come in different forms. One major source is link failures, which result in traffic being rerouted from one path to another. Traffic variations can also arise because of Denial-of-Service (DoS) attacks, the sudden popularity of a web site, or as a result of external (BGP) routing changes. In order to provide as much flexibility as possible in specifying traffic surges, we model them through an independent general traffic surge matrix $\Delta\mathbf{X}$. As with $\mathbf{X}$, the elements of $\Delta\mathbf{X}$ are random variables, and represent traffic surges between pairs of backbone routers[2]. The generality of both $\mathbf{X}$ and $\Delta\mathbf{X}$ allows experimentation with a broad range of base and surge traffic models.

3 Related Works

Several previous works have studied the scaling properties of IP networks. Their focus has been on understanding how maximum link loads grow with network size, but they share some motivations and methodologies with our work. In this section, we briefly discuss two of the most relevant works, and comment on differences with our work.

Gkantsidis *et. al.* [11] showed that there exists an *optimal routing* policy such that the maximum link load is at most $O(n \log^2 n)$ (n is the number of nodes). The result is

[2]The impact of link failures can, for example, be captured by identifying the set of links over which traffic is re-routed after each failure and the corresponding load increases these link see, and then create an equivalent surge matrix with matching entries for the routers associated with the links experiencing those increases. Simulations of all single-link failure scenarios on a 500-node network indicate that failures could be modelled by a surge matrix with traffic loads between adjacent node pairs that are less than 10% of the original link loads. Large-scale failure events can be investigated similarly through simulations.

established assuming that for each pair of nodes with degree d_u and d_v, there are $O(d_u d_v)$ units of traffic demand.

Akella *et. al.* [2, 3] showed that when shortest path routing is used, the expected value of the maximum link load in a power law graph with exponent α grows as $\Omega(n^{1+1/\alpha})$ with n. This result is established under the assumption that there is a *unit* traffic demand between all pairs of nodes.

Although both works investigated how the maximum link load grows with the network size, their different choices in routing and traffic models led them to different conclusions. Specifically, the upper bound derived in [11] is asymptotically smaller than the lower bound of [2, 3]. This realization together with a different focus, namely, the investigation of the impact of network size on the efficiency of over-provisioning, is what led us to develop a parametric model that can accommodate different operating conditions and assumptions, especially when it comes to how traffic grows. As a result, both the approach used in this paper and its contributions differ from those of [2, 3, 11] in several important aspects. Specifically, we allow arbitrary traffic models, where each element is a random variable that captures intrinsic traffic variations. In addition, we allow the specification of both a base traffic matrix and a separate surge traffic matrix, where the base traffic matrix is assumed known and used to dimension the network, while the surge traffic matrix accounts for the unexpected traffic variations that the over-provisioning of link bandwidth is meant to absorb. Finally, using this model we derive a *closed-form* expression for the traffic load on a link, as opposed to a bound on the expected link load, and use this expression to estimate the probability that traffic variations result in a link load that exceeds the link capacity. This then allows us to explore, as a function of network size, the level of over-provisioning needed to ensure a certain robustness against traffic surges. As we shall see, we find that this is critically dependent on the assumed underlying traffic model, and in particular how traffic grows in relation to network size.

4 $\mathcal{G}(n,p)$ networks with general traffic matrices

This section introduces the main analytical results we rely on to evaluate the influence of network growth on the efficiency of over-provisioning. Due to space constraints, proofs and derivations are omitted and can be found in [12].

As mentioned earlier, the network model is a two-level random network based on the $\mathcal{G}(n,p)$ model. Links are bidirectional, so that an edge between nodes i and j represents two directed links, namely, link $i \rightarrow j$ and link $j \rightarrow i$. Nodes correspond to backbone routers that can generate (source nodes), receive (destination nodes), and forward (transit nodes) traffic. Traffic consists of a base traffic matrix $\mathbf{X}$ and a surge matrix $\Delta\mathbf{X}$, defined as follows:

Definition 1 *In a network with node set V of cardinality n, let random variable X_{st} (ΔX_{st}) be the base traffic (surge) generated by source s to destination t ($s,t \in V$, $s \neq t$). Let $\mathbf{X}$ ($\Delta\mathbf{X}$) be the $n \times n$ base (surge) traffic matrix, in which the elements are X_{st} (ΔX_{st}), for $s \neq t$, and 0s on the diagonal.*

We first consider a single-level $\mathcal{G}(n,p)$ network with base traffic matrix $\mathbf{X}$, and derive the *base traffic load* $\widetilde{F}_{ij}(\mathbf{X})$ on link $i \rightarrow j$. We then extend the result to a two-level $\mathcal{G}(n,p)$ network, and obtain the *actual traffic load* $\widetilde{F}_{ij}(\mathbf{X}+\Delta\mathbf{X})$ on link $i \rightarrow j$, where the actual network traffic accounts for the contributions of both base and surge traffic. Based on the availability of expressions characterizing the link load, we then rely on Chebyshev's inequality to estimate the probability that the actual traffic stays below the link capacity, which indicates no congestion and good network performance. We finally use this expression to measure the efficiency of over-provisioning as a function of network size (as reflected in n) and traffic (as represented by $\mathbf{X}$ and $\Delta\mathbf{X}$), under the assumption that link capacities are chosen in proportion to the expected base traffic load.

4.1 Link traffic load in single-level $\mathcal{G}(n,p)$ random networks

Consider a single-level $\mathcal{G}(n,p)$ network with minimum hop count routing, i.e. shortest path routing with equal weight links. In case of multiple shortest paths, traffic is either evenly split among the shortest paths or sent on a randomly chosen path among them. In this setting, we introduce the following notation:

Definition 2 *For a given graph g, let $f_{ij}(g,s,t)$ be the amount of traffic on link $i \rightarrow j$ when there is one unit of traffic from source s to destination t ($s \neq t$). If there is no link $i \rightarrow j$ in graph g, then $f_{ij}(g,s,t) = 0$.*

Definition 3 *For a given graph g, let $f_{ij}(g,\mathbf{X})$ be the amount of traffic on link $i \rightarrow j$ when the traffic generated by SD pairs are represented by traffic matrix $\mathbf{X}$. If there is no link $i \rightarrow j$ in graph g, then $f_{ij}(g,\mathbf{X}) = 0$.*

Definition 4 *For a random graph G with a traffic matrix $\mathbf{X}$, let the unconditional mean traffic load on link $i \rightarrow j$ be $f_{ij} \triangleq Ef_{ij}(G,\mathbf{X}) = E_{\mathbf{X}}E_G f_{ij}(G,\mathbf{X})$. Let $\widetilde{F}_{ij}(\mathbf{X}) \triangleq E_G\{f_{ij}(G,\mathbf{X}) | \text{link } i \rightarrow j \text{ exists}\}$, namely the conditional expected traffic on link $i \rightarrow j$ with regard to the random graph and conditioned on the event that link $i \rightarrow j$ exists.*

Definition 5 *For any given graph g, $\forall s,t \in V, (s \neq t)$, we*

define the path length from node s to node t as

$$d(g,s,t) \triangleq \begin{cases} \text{hop counts of the shortest path from } s \text{ to } t & \\ \text{if } s,t \text{ are connected;} & \\ 0 \quad \text{if } s,t \text{ are not connected.} & \end{cases}$$

In the $\mathcal{G}(n,p)$ model, we denote the expected path length from s to t by $d(s,t) \triangleq E_G d(G,s,t)$, and the average expected path length over all SD pairs in the $\mathcal{G}(n,p)$ model by $d \triangleq \sum_{s \neq t} d(s,t)/n(n-1)$.

To obtain our first result (see Theorem 1) on $\widetilde{F}_{ij}(\mathbf{X})$, we introduce the following three lemmas.

Lemma 1 *For any given graph g and $\forall s,t \in V, (s \neq t)$, we have $d(g,s,t) = \sum_{i \neq j} f_{ij}(g,s,t)$. That is, the path length from s to t is equal to the sum of the traffic carried on all links when s sends one unit of traffic to t.*

Lemma 2 *In the $\mathcal{G}(n,p)$ model, we have $d(s,t) = d$, $\forall s,t \in V, (s \neq t)$. That is, in the $\mathcal{G}(n,p)$ model, the expected path length between any two nodes is the same, and thus equal to the average expected path length d.*

Lemma 3 *In the $\mathcal{G}(n,p)$ model, if every source node sends one unit of traffic to every other node, then $f_{ij} = d$. That is, the unconditional mean traffic load on link $i \to j$ is equal to the average expected path length d.*

The result of Lemma 3 assumed a simple, deterministic traffic matrix $\mathbf{X}$ with unit traffic exchanged between all SD pairs. Our goal is to allow the specification of arbitrary traffic matrices with different (and possibly random) traffic entries for different SD pairs. For that purpose, it is necessary to understand for a link $i \to j$, how much of the traffic exchanged between any given SD pair (s,t) actually traverses link $i \to j$. This depends on the position of s and t relative to i and j. For example, it is easy to see that if $s = i$ and $t = j$, then all of the traffic from s to t traverses link $i \to j$ as long as link $i \to j$ exists. Conversely, no traffic from s to t traverses link $i \to j$ if $s = j$ and $t = i$. In fact, for any given link $i \to j$, we can partition SD pairs into five subsets based on their relative position with respect to i and j, such that their expected contribution to the link load depends only on which subset they belong to.

Definition 6 *For any two node $i,j \in V$, $(i \neq j)$, we partition all source and destination pairs (s,t), $(\forall s,t \in V, s \neq t)$, with regard to their relative position to i,j as follows:*

$$\begin{aligned} S_{ij}^{(1)} &= \{(s,t) : s \neq i,j \text{ and } t \neq i,j\} \\ S_{ij}^{(2)} &= \{(s,t) : s = i, t \neq j \text{ or } s \neq i, t = j\} \\ S_{ij}^{(3)} &= \{(s,t) : s = i, t = j\} \\ S_{ij}^{(4)} &= \{(s,t) : s = j, t \neq i \text{ or } s \neq j, t = i\} \end{aligned}$$

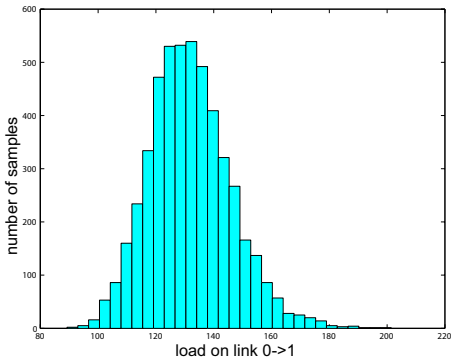
$$S_{ij}^{(5)} = \{(s,t) : s = j, t = i\}$$

We are now ready to state our first main result.

Theorem 1 *In single-level $\mathcal{G}(n,p)$ networks with a traffic matrix $\mathbf{X}$, $\forall i,j \in V, (i \neq j)$, the conditional expected traffic on link $i \to j$ given the existence of link $i \to j$ is:*

$$\begin{aligned} \widetilde{F}_{ij}(\mathbf{X}) &= \sum_{(s,t) \in S_{ij}^{(1)}} X_{st} \frac{d + p - 2\theta}{p(n-2)(n-3)} \\ &\quad + \sum_{(s,t) \in S_{ij}^{(2)}} X_{st} \frac{\theta - p}{p(n-2)} + \sum_{(s,t) \in S_{ij}^{(3)}} X_{st} \quad (1) \end{aligned}$$

in which $\theta = \Pr\{\text{there is a path from } s \text{ to } t\}$, and $c \leq \theta \leq 1$, c is a constant related to the average node degree np.



$n = 1000$, $p = 0.02$, 5000 samples of load on link $0 \to 1$.

Figure 2. Histogram of link loads with uniform traffic matrix.

Theorem 1 gives the conditional (given that the link exists) expected (over all graphs) traffic on a link, as a weighted sum of the traffic generated by all SD pairs based on their locations relative to the link. The weight associated with the traffic from a particular SD pair depends only on which of the five subsets of SD pairs it belongs to.

$\widetilde{F}_{ij}(\mathbf{X})$ is essentially the link load averaged over all graphs in $\mathcal{G}(n,p)$ that contain this link. We illustrate by simulation that such an averaged value is representative of the actual link load in a graph randomly generated by the $\mathcal{G}(n,p)$ model. Fig. 2 shows the histogram of link loads for link $0 \to 1$ with data samples collected from 5000 random graph generations of a 1000-node network. A uniform traffic matrix with 1 unit of traffic between every SD pairs is used in the simulation. The histogram has a reasonably narrow bell shape with mean 131.84 and standard deviation 14.16. The $\widetilde{F}_{ij}(\mathbf{X})$ in this case can be calculated from Eq. (1) with $X_{st} = 1$ for all $s \neq t$. The resulting value $\widetilde{F}_{ij}(\mathbf{X}) = \frac{d}{p} = 131.95$ (the average path length from the 5000 randomly generated graphs is 2.639), is very close to the sampled mean of 131.84. Extensive simulations with different topology parameters show similar results.

4.2 Link traffic load in two-level $\mathcal{G}(n,p)$ random networks

We now consider a two-level hierarchical random network, which consists of a base level, i.e., the router-level; and a top level or "domain-level." We generate the two-level network as described in Section 2. There are, therefore, n_1 domains, and each inter-domain edge is chosen independently with probability p_1; each domain contains n_2 nodes, and each intra-domain edge is chosen independently with probability p_2. Connected node pairs are randomly chosen within domains to serve as gateways, and inter-domain edges are anchored to these end nodes. In total, the network consists of $n_1 n_2$ nodes and $n_1 n_2(n_1 n_2 - 1)$ SD pairs.

Routing is as described earlier, namely, inter-domain routing is determined by considering only inter-domain links and choosing the path(s) with the smallest domain hop count. In case of multiple shortest paths, traffic is evenly split among them or sent on a randomly selected one. For every domain in an inter-domain path, the path(s) from the entry gateway to the exit gateway is a minimum hop count path(s) between the two nodes, considering only intra-domain links. Again, when multiple equal cost paths exist, traffic is evenly split among them or sent on a randomly selected one. This amounts to shortest path routing where all intra-domain links have the same weight, and all inter-domain links have the same but a much higher weight.

$\forall u \in V$, $A(u)$ denotes the domain that node u belongs to, and we extend Definition 6 as follows:

Definition 7 *Given any two domains $A_i, A_j, (i \neq j)$, we partition all $n_1(n_1-1)$ domain pairs (with regard to A_i and A_j) into sets $S^{(k)}_{A_i A_j}$, $k = 1, \ldots, 5$, as in Definition 6, except that domains are considered instead of nodes.*

For any two node i and j $(i \neq j)$ in the same domain (i.e. $A(i) = A(j)$), we partition all nodes in $A(i)$ (with regard to i and j) into sets $S^{(k)}_{ij}$, $k = 1, \ldots, 5$, as in Definition 6, except that only nodes in domain $A(i)$ are considered.

The next theorem is a natural extension of the single-level result of Theorem 1:

Theorem 2 *In two-level hierarchical $\mathcal{G}(n,p)$ networks with a general traffic matrix $\mathbf{X}$, the conditional expected traffic on inter-domain link $i \to j$ between domains A_i and A_j $(i \neq j)$ is given by*

$$\widetilde{F}_{A_i A_j}(\mathbf{X}) = \sum_{(A(s),A(t)) \in S^{(1)}_{A_i A_j}} X_{st} \theta_3^2 \frac{d_1 + p_1 - 2\theta_1}{p_1(n_1-2)(n_1-3)} \quad (2)$$

$$+ \sum_{(A(s),A(t)) \in S^{(2)}_{A_i A_j}} X_{st} \theta_3^2 \frac{\theta_1 - p_1}{p_1(n_1-2)} + \sum_{(A(s),A(t)) \in S^{(3)}_{A_i A_j}} X_{st} \theta_3^2$$

Similarly, the conditional expected traffic on intra-domain link $i \to j$ between nodes i and j $(i \neq j)$ is given by

$$\begin{aligned} \widetilde{F}_{ij}(\mathbf{X}) &= \sum_{(s,t) \in S^{(1)}_{ij}} X_{st} \frac{d_2 + p_2 - 2\theta_2}{p_2(n_2-2)(n_2-3)} \quad (3) \\ &+ \sum_{(s,t) \in S^{(2)}_{ij}} X_{st} \frac{\theta_2 - p_2}{p_2(n_2-2)} + \sum_{(s,t) \in S^{(3)}_{ij}} X_{st} \\ &+ \theta_1 \theta_3 \left\{ \sum_{\substack{t \notin A(i), \\ s \in A(i), s \neq i,j}} \frac{X_{st}(d_2 - \theta_2)}{p_2 n_2 (n_2-2)} + \sum_{t \notin A(i)} \frac{X_{it}\theta_2}{p_2 n_2} \right\} \\ &+ \theta_1 \theta_3 \left\{ \sum_{\substack{s \notin A(i), \\ t \in A(i), t \neq i,j}} \frac{X_{st}(d_2 - \theta_2)}{p_2 n_2 (n_2-2)} + \sum_{s \notin A(i)} \frac{X_{sj}\theta_2}{p_2 n_2} \right\} \\ &+ \frac{\theta_1 \theta_3^2}{\theta_2} \sum_{A(s) \neq A(t) \neq A(i)} \frac{X_{st} d_2 (d_1 - 1)}{p_2(n_1-2) n_2 (n_2-1)} \end{aligned}$$

in which d_1 and d_2 are the average expected path length in $\mathcal{G}(n_1,p_1)$ and $\mathcal{G}(n_2,p_2)$, respectively.
θ_1, θ_2 are $\Pr\{$there is a path from s to $t\}$ in $\mathcal{G}(n_1,p_1)$ and $\mathcal{G}(n_2,p_2)$, respectively. $\theta_3 = [\frac{1}{n_2} + (1 - \frac{1}{n_2})\theta_2]$. $c \leq \theta_1, \theta_2, \theta_3 \leq 1$, c is a constant related to $n_1 p_1$ and $n_2 p_2$.

Theorem 2 gives the conditional expected traffic on any link in a two-level $\mathcal{G}(n,p)$ random network as a function of network topology and traffic. We investigate next the robustness of such a network, i.e., measure the likelihood that the total (base + surge) traffic on a link is below its capacity, assuming that the link capacity is $(1+\beta)E\widetilde{F}_{ij}(\mathbf{X})$, where $E\widetilde{F}_{ij}(\mathbf{X})$ is computed using Eqs. (2) and (3).

4.3 Robustness in $\mathcal{G}(n,p)$ networks with general surges

Theorem 2 allows us to derive an explicit expression for the probability that the link capacity is larger than the actual load, as a function of the statistical properties of both the base and surge traffic. For simplicity, we rely on Chebyshev's inequality to express this probability as a function of just the mean and variance of the total offered traffic. This yields the following relationship between link bandwidth and the actual offered traffic:

Proposition 1 *Let B_{ij} be the bandwidth on link $i \to j$. For any $\epsilon > 0$, if B_{ij} satisfies $B_{ij} \geq E\widetilde{F}_{ij}(\mathbf{X} + \Delta\mathbf{X}) + \epsilon$, then*

$$\Pr\left\{ B_{ij} > \widetilde{F}_{ij}(\mathbf{X} + \Delta\mathbf{X}) \right\} \geq 1 - Var\widetilde{F}_{ij}(\mathbf{X} + \Delta\mathbf{X})/\epsilon^2$$

Proposition 1 gives a lower bound on the probability, $\psi \triangleq Pr\{B_{ij} \geq \widetilde{F}_{ij}(\mathbf{X} + \Delta\mathbf{X})\}$ that a link can accommodate its actual load. We refer to ψ as the target network *tolerance probability*. As mentioned before, link bandwidth is provisioned proportionally to the expected base offered

load $\widetilde{f}_{ij}^b \triangleq E\widetilde{F}_{ij}(\mathbf{X})$, i.e., $B_{ij} = (1+\beta)\widetilde{f}_{ij}^b$. Using Proposition 1, we can find the appropriate β for any target value of ψ by setting ϵ to be $\sqrt{Var\widetilde{F}_{ij}(\mathbf{X}+\Delta\mathbf{X})/(1-\psi)}$.

Definition 8 *For any constant $0 < \psi < 1$, we define the minimum over-provisioning factor β^* as*

$$\beta^*(\psi) = \min_{\beta>0}\left\{(1+\beta)E\widetilde{F}_{ij}(\mathbf{X}) \geq E\widetilde{F}_{ij}(\mathbf{X}+\Delta\mathbf{X}) + \sqrt{Var\widetilde{F}_{ij}(\mathbf{X}+\Delta\mathbf{X})/(1-\psi)}\right\}$$

β^* determines the lowest over-provisioning factor that meets a required tolerance probability of ψ. Using Theorem 2, we can express B_{ij}, $E\widetilde{F}_{ij}(\mathbf{X}+\Delta\mathbf{X})$, and $Var\widetilde{F}_{ij}(\mathbf{X}+\Delta\mathbf{X})$ as functions of n_1, p_1, d_1, n_2, p_2, d_2, $\mathbf{X}$ and $\Delta\mathbf{X}$ for both inter-domain and intra-domain links. Therefore, for a given target ψ, we can identify the required over-provisioning factor β^*, as a function of those parameters.

4.4 Average expected path length d in $\mathcal{G}(n,p)$

From Theorem 2, $\widetilde{F}_{ij}(\mathbf{X})$ depends on n, p, and d. n and p are input parameters to the model, but d still needs to be determined. The next proposition provides an expression for the order of d as a function of n and p.

Proposition 2 *In the $G(n,p)$ model, if $np \geq c$ for some constant $c > 1$, but $np = O(\log^\alpha n)$ for some constant $\alpha > 0$, then $d = \Theta(\log n/\log np)$.*

Combining Propositions 1 and 2, we can now identify β^* as an explicit function of the base and surge matrices, and the size of the network, namely, n.

5 Representative results

In this section, we assume a given tolerance probability ψ and use the analytical results of the previous section to evaluate for a selected set of base and surge traffic models, how network growth affects the minimum over-provisioning factor β^*.

Network growth is obviously driven to a large extent by traffic growth. However, there is considerable uncertainty regarding what are appropriate models not only for the rate of traffic growth, but also for how it is to be distributed between SD pairs. This uncertainty applies to the base traffic, and possibly even more so to traffic surges. The models we have developed so far are capable of handling this generality, as they can accommodate *any type of base and surge patterns*. However, exploring how network growth might affect the minimum over-provisioning factor β^* calls for introducing some structure into how one expects the network and the traffic to grow. As more data points become available to better characterize how network and traffic are growing, specific models will likely emerge and can then be "plugged" into the equations of Theorem 2, but in the absence of such definite answers, it is useful to introduce strawman traffic scenarios in order to explore possible trends in the evolution of β^*. For that purpose, we rely on three types of traffic patterns, which we believe are not only representative of possible evolutions of traffic patterns, but also have enough structure that we can use them to develop meaningful insight. In the remainder of this section, we use these three patterns to produce three distinct combinations of base and surge traffic for which we explore the evolution of β^* as the network grows. Investigations of additional traffic combinations can be found in [12].

The first pattern we consider corresponds to a traffic matrix where each entry (except for the diagonal terms that are zero) is an i.i.d. random variable. The use of random variables allows us to incorporate temporal fluctuations in the traffic exchanged between pairs of nodes. The main limitation of this model is obviously the i.i.d. constraint, which in particular assumes that *every* node sends equally to *all* other nodes in the networ, irrespective of the network size. However, an i.i.d. pattern, in spite of its limitations, is a good first step in capturing the variable nature of network traffic and allows us control this variability in a systematic manner.

The second traffic pattern we consider is a random destination selection pattern. It consists of a traffic matrix where each row is limited to having only k ($0 \leq k < n$) non-zero entries, where the locations of those entries (columns) are randomly chosen. This allows us to account for scenarios where at any point in time a node sends traffic to only a subset of possible destinations, rather than to all of them. For simplicity, we assume that the amount of traffic sent to each selected destination is constant and equal to a units of traffic. We use this traffic pattern to explore the impact of having a node distribute its traffic across a variable set of possible destinations, which does induce traffic variations.

The third traffic pattern we consider is one that allows us to incorporate the impact of access networks on the traffic exchanged between backbone nodes. For this purpose, we use a gravity traffic pattern in which the base traffic between two backbone nodes, is a function of the nodes' rank, where the rank of a backbone node is a function of the number of access networks attached to it, and therefore of its aggregate traffic generating (or receiving) capacity. Specifically, the traffic between nodes s and t, X_{st}, is proportional to $r_s r_t$, where r_s and r_t are the "ranks" of s and t respectively. In other words, traffic from a given source node is "fanned out" to destination nodes in proportion to their ranks. In [5], it was observed that backbone nodes could be ranked roughly into three categories (large, medium and small), where the "large" category contributes the majority of POP-level traf-

Table 1. Efficiency of over-provisioning in two-level $\mathcal{G}(n,p)$-based random networks

Traffic combination	Inter-domain links	Intra-domain links
Sce 1: i.i.d base traffic, i.i.d surges	$\widetilde{f}^b_{A_iA_j} = \Theta(y_1 N^{1-\lambda}\log N)$ $\beta^* = \Theta\left(\frac{\Delta y_1}{y_1} + \frac{\sqrt{\sigma_1^2+\Delta\sigma_1^2}}{y_1 N^{-\lambda/2}\log N\sqrt{1-\psi}}\right)$	$\widetilde{f}^b_{ij} = \Theta\left((y_1\log N + y_2)\log N\right)$ $\beta^* = \Theta\left(\frac{\Delta y_1 \log N+\Delta y_2}{y_1\log N+y_2} + \frac{\sqrt{(\sigma_1^2+\Delta\sigma_1^2)N^\lambda+(\sigma_2^2+\Delta\sigma_2^2)}}{(y_1\log N+y_2)N^{\frac{\lambda-1}{2}}\log N\sqrt{1-\psi}}\right)$
Sce 2: i.i.d base traffic, Random dest surges $(\Delta a_1,\Delta k_1;\Delta a_2,\Delta k_2)$	$\widetilde{f}^b_{A_iA_j} = \Theta(y_1 N^{1-\lambda}\log N)$ $\beta^* = \Theta\left(\frac{\Delta y_1}{y_1} + \frac{\sqrt{\sigma_1^2 N^{2-\lambda}+\Delta y_1^2(\frac{N^{1-\lambda}}{\Delta k_1}-\frac{1}{N^\lambda})}}{y_1 N^{1-\lambda}\log N\sqrt{1-\psi}}\right)$	$\widetilde{f}^b_{ij} = \Theta((y_1\log N + y_2)\log N)$ $\beta^* = \Theta\left(\frac{\Delta y_1\log N+\Delta y_2}{y_1\log N+y_2} + \frac{\sqrt{\sigma_1^2 N+\sigma_2^2 N^{1-\lambda}+\Delta y_1^2(\frac{1}{\Delta k_1}-\frac{1}{N})+\Delta y_2^2(\frac{1}{\Delta k_2}-\frac{1}{N^{1-\lambda}})}}{(y_1\log N+y_2)\log N\sqrt{1-\psi}}\right)$
Sce 3: Gravity base traffic, i.i.d surges	minimum loaded link: $\widetilde{f}^b_{A_iA_j} = \Theta(y_1 N^{1-\lambda}\log N)$ $\beta^* = \Theta\left(\frac{\Delta y_1}{y_1} + \frac{\sqrt{\Delta\sigma_1^2/(1-\psi)}}{y_1 N^{-\lambda/2}\log N}\right)$ maximum loaded link: $\widetilde{f}^b_{A_iA_j} = \Theta\left(y_1 N^{1-\lambda}(\log N + l^{-1})\right)$ $\beta^* = \Theta\left(\frac{\Delta y_1\log N}{y_1(\log N+l^{-1})} + \frac{\sqrt{\Delta\sigma_1^2/(1-\psi)}}{y_1 N^{-\lambda/2}(\log N+l^{-1})}\right)$	minimum loaded link: $\widetilde{f}^b_{ij} = \Theta((y_1\log N + y_2 l)\log N)$ $\beta^* = \Theta\left(\frac{\Delta y_1\log N+\Delta y_2}{y_1\log N+y_2 l} + \frac{\sqrt{N^\lambda\Delta\sigma_1^2+\Delta\sigma_2^2}/\sqrt{1-\psi}}{(y_1\log N+y_2 l)N^{\frac{\lambda-1}{2}}\log N}\right)$ maximum loaded link: $\widetilde{f}^b_{ij} = \Theta\left((y_1(\log N + l^{-1}) + y_2 l^{-1})\log N\right)$ $\beta^* = \Theta\left(\frac{\Delta y_1\log N+\Delta y_2}{y_1(\log N+l^{-1})+y_2 l^{-1}} + \frac{\sqrt{N^\lambda\Delta\sigma_1^2+\Delta\sigma_2^2}/\sqrt{1-\psi}}{(y_1(\log N+l^{-1})+y_2 l^{-1})N^{\frac{\lambda-1}{2}}\log N}\right)$

fic but comprises only a small number of nodes, while the majority of nodes are in the "medium" and "small" categories which contribute only a small fraction of the total POP-level traffic. Such an observation is consistent with the heavy-tail degree distribution and uneven geographical distribution of user population that have been reported for large-IP networks. The use of such a gravity traffic pattern, allows us to investigate how the disparate rank of backbone nodes affects not only the expected base load on backbone links, but also their ability to absorb traffic variations.

We now explore three scenarios involving different combinations of the above three traffic patterns, as summarized in Table 1. The network we consider is a two-level $\mathcal{G}(n,p)$ network as previously described, where the average inter-domain and intra-domain degrees are taken to be constant[3] and equal to $n_1p_1 = c_1$ and $n_2p_2 = c_2$, respectively. We further assume that the total size of the two-level network is $N = n_1n_2$ with $n_1 = N^\lambda$ and $n_2 = N^{1-\lambda}$ $(0 < \lambda < 1)$. The parameter λ provides some flexibility in deciding how network growth translates into growth in either the size or the number of domains. Our findings are summarized in Table 1 (see [12] for the underlying derivations), which displays for each scenario the expected base offered load $\widetilde{f}^b_{ij}$ and the minimum over-provisioning factor β^*. The results give the order of those quantities as functions of the total number of nodes, and the characteristics of the base and surge traffic of each scenario.

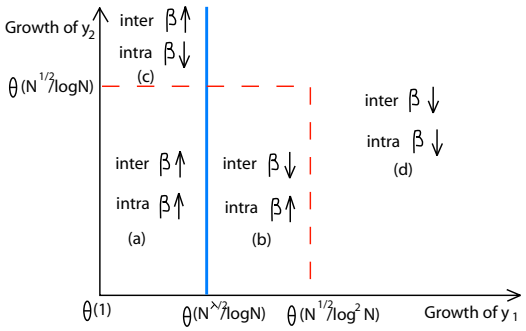
When interpreting these results, we make the following two general assumptions. First, we assume that the expected *total* base intra-domain and inter-domain traffic sourced by a *single* node, denoted by y_1 and y_2 respectively, are non-decreasing functions of N. We believe this to be a reasonable assumption that is consistent with current observations and predictions of future traffic growth, e.g., as in [14]. Note from Table 1 that although y_1 and y_2 may differ in each scenario, the expressions for $\widetilde{f}^b_{ij}$, as functions of y_1 and y_2, are similar in all three. They all point to increasing link bandwidth as the network grows under the assumption that y_1 and y_2 are non-decreasing. Our second assumption is that the ratio of the average intensity of traffic surges and the base traffic remains approximately constant, namely $\frac{\Delta y_1}{y_1} = \frac{\Delta y_2}{y_2}$ [4], and is not a function of N. We also believe this to be a reasonable assumption, since both types of traffic are influenced by the same parameters such as link and host speeds, application characteristics, etc.

5.1 i.i.d normal traffic and i.i.d surges

In this scenario, the base traffic between any two nodes in different domains (inter-domain traffic) is i.i.d. with mean x_1 and variance σ_1^2, and the base traffic between any two nodes in the same domain (intra-domain traffic) is i.i.d. with mean x_2 and variance σ_2^2. We also assume i.i.d. inter-domain traffic surges with mean Δx_1 and variance $\Delta\sigma_1^2$, and i.i.d. intra-domain traffic surges with mean Δx_2 and variance $\Delta\sigma_2^2$. This yields $y_1 = \Theta(x_1 N)$, $y_2 = \Theta(x_2 N^{1-\lambda})$, $\Delta y_1 = \Theta(\Delta x_1 N)$, and $\Delta y_2 = \Theta(\Delta x_2 N^{1-\lambda})$.

[3] This restriction can be readily removed, and was introduced to limit the number of variables to consider.

[4] Δy_1 and Δy_2 are the *total* inter-domain and intra-domain traffic surges sourced by a node, respectively.



↑: increases asymptotically, ↓: decreases asymptotically.
Scenario I: i.i.d base traffic and i.i.d traffic surges.
Scenario II: i.i.d base traffic and random destination selection base surges.

Figure 3. Efficiency of over-provisioning for scenarios 1 and 2.

Based on Table 1, the minimum over-provisioning factor β^* exhibits different behaviors depending on how fast nodal traffic grows with N. In order to facilitate the discussion, we introduce the additional constraint that the variances of the base and surge traffic for any SD pair, σ_1^2, σ_2^2, $\Delta\sigma_1^2$ and $\Delta\sigma_2^2$, remain constant, and not be functions of N.

Incorporating those constraints into Table 1, we find that on inter-domain links $\beta^* = \Theta(y_1^{-1} N^{\lambda/2} \log^{-1} N)$ and on intra-domain links $\beta^* = \Theta((y_1 \log N + y_2)^{-1} N^{\frac{1}{2}} \log^{-1} N)$. As a result, for inter-domain links β^* decreases asymptotically as N grows, i.e., over-provisioning becomes more efficient, if and only if y_1 grows faster than $\Theta(N^{\lambda/2}/\log N)$. This states that the growth in the expected base load needs to exceed a certain threshold to ensure that the bandwidth margin grows faster than the variations induced by surges. A similar but different threshold exists for intra-domain links, for which β^* decreases as N grows, if and only if either y_1 grows faster than $\Theta(\sqrt{N}/\log^2 N)$ or y_2 grows faster than $\Theta(\sqrt{N}/\log N)$. Fig. 3 displays graphically the four corresponding regions. The results for this specific scenario highlight a theme common to all three, namely, the existence of thresholds in the growth of traffic sourced by a node that define regions in which over-provisioning is either more or less efficient as networks grow.

5.2 i.i.d base traffic and random destination selection surges

In this scenario, the base traffic is the same i.i.d as that of scenario 1, but the traffic pattern used for surges is different and based on the random selection of a certain number of destinations. Specifically, each source s generates Δa_1 units of surge to each of Δk_1 destinations uniformly chosen among all possible $(n_1 - 1)n_2$ nodes not in its domain; and generates Δa_2 units of surge to each of Δk_2 destinations uniformly chosen among all possible $n_2 - 1$ nodes within its own domain (s does not send traffic to itself). Traffic sent by different source nodes are independent of each other, and so are inter-domain and intra-domain traffic.

In this scenario, we have $y_1 = \Theta(x_1 N)$, $y_2 = \Theta(x_2 N^{1-\lambda})$, $\Delta y_1 = \Theta(\Delta a_1 \Delta k_1)$, and $\Delta y_2 = \Theta(\Delta a_2 \Delta k_2)$. As in scenario 1, we assume that σ_1^2 and σ_2^2 are constant. We also assume that $\Delta k_1 = \Theta(N^\xi)$ $(0 \le \xi < 1)$, $\Delta k_2 = \Theta(N^{(1-\lambda)\zeta})$ $(0 \le \zeta < 1)$, indicating that the number of nodes that a source node sends traffic to grows, but not as fast as the network size.

Incorporating these assumptions into Table 1 and focusing on only the higher order terms, we find that on inter-domain links there exists a transition point at $\Theta(N^{\lambda/2}/\log N)$, such that β^* decreases asymptotically as the network grows, if and only if y_1 grows faster than this value. For intra-domain links, we find that β^* also experiences a transition and decreases asymptotically as the network grows, if and only if either y_1 grows faster than $\Theta(\sqrt{N}/\log^2 N)$or y_2 grows faster than $\Theta(\sqrt{N}/\log N)$. In summary, the behavior of β^* for both inter-domain and intra-domain links is essentially the same as that of the previous scenario as illustrated in Fig. 3.

5.3 Gravity base traffic and i.i.d traffic surges

In this third scenario, we consider a combination of a gravity base traffic and i.i.d traffic surges. The gravity traffic pattern translates into a base traffic matrix with entries that are proportional to the ranks of the corresponding source and destination nodes. Specifically, we have $X_{st} = x_1 r_s r_t$ for inter-domain traffic and $X_{st} = x_2 r_s r_t$ for intra-domain traffic. Backbone nodes are classified according to their rank into three rank categories "large", "medium" or "small", denoted by sets L, M, and S, respectively. Based on the observations of [5], we assume that the rank categories obey the following two rules: (1) Each category generates twice as much traffic as the next lower category, namely, $4\sum_{u\in S} r_u = 2\sum_{u\in M} r_u = \sum_{u\in L} r_u$. (2) The size of these three categories are: $|S| = sN$, $|M| = mN$, and $|L| = lN$, in which $s = 0.5 - o(1)$, $m = 0.5 - o(1)$, $l = o(1) > \frac{1}{n_1}$, and $N = n_1 n_2$ is again the total size of the backbone network. We assume that n_1 is large enough, so that sn_1, mn_1 and ln_1 are integers. Note that l reflects the extent to which the majority of backbone traffic is concentrated on only a few nodes with high rank.

Note that the gravity traffic pattern itself does not contain variations. It only determines the base offered loads on links, which when scaled by β determines their *spare* capacity. Traffic variations in this scenario are solely contributed by the i.i.d. traffic surge matrix, for which we assume i.i.d. inter-domain surges with mean Δx_1 and variance $\Delta\sigma_1^2$, and i.i.d. intra-domain surges with mean Δx_2

and variance $\Delta\sigma_2^2$. Results from scenario 1 show that under such a surge model, traffic variations are uniform across all links. As a result, the minimum over-provisioning factor is solely determined by the link that has the smallest amount of spare capacity, i.e., the link with the lowest base load. The base load $\widetilde{f}_{ij}^b$ on the "thinnest" link depends on how nodes of different ranks are distributed across domains. The next proposition establishes that when nodes are assigned to domains according to a "strict ordering" of their rank, this results in a network with min-min link loads (see [12] for a proof). Furthermore, this particular distribution of nodes also provides a lower bound on the difference between base link loads, i.e., the difference between the "fattest" and the "thinnest" links, which provides yet another perspective on the impact that such an imbalanced base traffic on has the efficiency of over-provisioning.

Proposition 3 *In two-level $\mathcal{G}(n,p)$ networks with gravity-type traffic matrices, the minimum loaded link achieves its min-min value when nodes are assigned to domains according to a "strict ordering" of their ranks:*

Domain A_i contains n_2 nodes ranging from the $[(i-1)n_2+1]^{th}$ smallest rank to the $[in_2]^{th}$ smallest rank, $i = 1,\ldots,n_1$.

Based on the strict ordering domain formation, $\widetilde{f}_{ij}^b$ and β^* can be computed for both minimum loaded links and maximum loaded links (see [12] for details) as shown in Table 1. Focusing first on inter-domain links, we see that for minimum loaded links $\widetilde{f}_{ij}^b$ and β^* are essentially similar in their order as that of the scenario 1. Therefore, on inter-domain links β^* exhibits the same behavior and thresholds as in this earlier scenario. Considering next maximum loaded inter-domain links illustrates that the gravity traffic model can introduce very substantial differences in how over-provisioning performs on different links. Specifically, when l is not too small, i.e., $l = \omega(\log^{-1} N)$, the base load on the minimum loaded and maximum loaded links are of the same order, and therefore correspond to a similar β^*. However, when l is small, indicating a highly concentrated base traffic pattern, the difference in the base offered load on maximum loaded and minimum loaded links can be substantial. This indicates that a much smaller level of over-provisioning is needed on those maximum loaded links. This is in a sense "good news" as the very high capacity of these links may make it technically difficult to over-dimension them by the same amount as lower bandwidth links. If we focus next on intra-domain links, we see that the impact of traffic imbalance is even more pronounced on those links. This is due to the added contribution of intra-domain traffic on the link, which shows a difference of the order of l^2 between maximum loaded and minimum loaded links. However, as with inter-domain links, the difference is in the "right" direction, with higher capacity links requiring a much lower level of over-provisioning.

In summary, our initial investigation revealed that in many cases, the level of over-provisioning required to achieve a given tolerance probability decreases as long as the rate of (base) traffic growth exceeds that of network growth by a certain factor. This conclusion is obviously predicated on the specific traffic models used, but it indicates that under a reasonably broad range of conditions, over-provisioning can become more efficient as the network grows. In the next section, we briefly touch on possible guidelines to promote such an outcome.

5.4 Implications for network design and provisioning

There are many decisions that network providers face when designing and provisioning networks, which affect network performance in various ways. For example, should one use many low capacity routers or fewer high capacity ones, or is it better to structure a network into multiple small domains or fewer larger ones? The analytical tools developed in this work can help explore the implications of such decisions, at least in terms of network robustness to traffic variations.

Our analysis of various traffic scenarios indicates that the relative growth of traffic volume versus network size plays a critical role in determining whether or not over-provisioning is more efficient in larger networks. From a practical standpoint, this means that we need to be concerned with how growth in router capacity[5] compares with the increase in the number of routers in the network. In particular, our results indicate that as long as router capacity grows faster than the number of routers by a certain ratio, the robustness of the network against traffic variations (assuming a given level of over-provisioning) will keep increasing. Obviously, this conclusion needs to be tempered by the fact that not all routers in a network are of the same type and capacity, but one can argue that routers in backbone networks, which are the focus of this paper, are reasonably homogeneous. As a result, this gives some guidelines on how to best grow such networks. Specifically, the rate at which backbone routers are upgraded to higher capacity versions should exceed the rate at which new routers are deployed. When looking back, it appears that we might have been heading in the right direction. Specifically, today's largest routers boast capacities of roughly 10 Terabits (10^{13} bits/sec), in comparison to a capacity of about 10 T1 links (10^7 bits/sec) for some of the early NSFNet routers[6]. This translates into a growth

[5] Assuming that router capacity is reasonably correlated with the traffic sourced by the router.

[6] See for http://moat.nlanr.net/INFRA/NSFNET.html for a perspective on the evolution of the NSFNet.

ratio of 10^6, and while the backbones of large Internet Service Providers have also grown in the mean time, the corresponding growth ratio has been more of the order of 10^2 or maybe 10^3 (the early NSFNet had of the order of 20 routers, while the backbones of tier 1 providers typically consist of several hundred routers). There is, therefore, hope that if router capacity continues to grow at a similar pace, the efficiency of over-provisioning will continue improving.

The influence of domain sizes can also be studied using our models by varying the parameter λ. A smaller λ corresponds to fewer but larger domains. Under the assumptions of our three traffic scenarios, the cost of over-provisioning on intra-domain links is independent of domain size, as the threshold in the relative growth of traffic generated by a single node compared to the network size is not affected by λ. However, the threshold for the inter-domain links does depend on λ, and a smaller λ translates into a lower threshold for the relative traffic growth, which is thus easier to meet. As a result, fewer domains with larger domain sizes are preferred when it comes to the efficiency of over-provisioning on inter-domain links. This being said, there are clearly other factors that affect this choice, e.g., routing complexity and stability, and this is a decision that should be made only after accounting for these parameters.

6 Conclusion

This paper has investigated the extent to which the size of the network can play a positive role in its ability to absorb traffic variations through over-provisioning. Given the many parameters that have the potential to affect the answer and the complex interactions that exist between them, the approach taken was to develop a model that could be adjusted to account for different scenarios in terms of network and traffic growth. The first contribution of the paper is, therefore, in developing such a model that can be used under a broad range of conditions. The second contribution of the paper is in using the model to explore a specific set of traffic scenarios that are representative of possible traffic growth models. The investigation identified the ratio of network size and traffic growth rates as a key parameter, and pointed to thresholds separating regions associated with different behaviors. The results provide some insight into how to possibly grow networks as a function of the underlying traffic growth and the available technology.

References

[1] Abilene Network. Abilene weather map - real time traffic. http://abilene.internet2.edu/.

[2] A. Akella, S. Chawla, A. Kannan, and S. Seshan. Scaling properties of the Internet graph. In *Proc. ACM PODC*, Boston, MA, July 2003.

[3] A. Akella, S. Chawla, A. Kannan, and S. Seshan. On the scaling of congestion in the Internet graph. *ACM SIGCOMM Computer Communication Review, Special Issue on Science of Network Design*, 34(3):43–56, July 2004.

[4] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.

[5] S. Bhattacharyya, C. Diot, J. Jetcheva, and N. Taft. POP-level and access-link-level traffic dynamics in a Tier-1 POP. In *Proc. ACM Internet Measurement Workshop (IMW)*, San Francisco, CA, November 2001.

[6] B. Bollobas. *Random Graphs*. Cambridge University Press, 2001.

[7] J. Case, M. Fedor, M. Schoffstall, and J. Davin. *Simple Network Management Protocol (SNMP)*. RFC1157, May 1990.

[8] F. Chung and L. Lu. The average distance in a random graph with given expected degrees. *Proc. National Academy of Sciences of the United States of America*, 99(25):15879–15882, December 2002.

[9] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. In *Proc. ACM SIGCOMM*, pages 251–262, Boston, MA, August 1999.

[10] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, J. Rexford, and F. True. Deriving traffic demands for operational IP networks: Methodology and experience. *IEEE/ACM Trans. Netw.*, 9(3):265–279, June 2001.

[11] C. Gkantsidis, M. Mihail, and A. Saberi. Conductance and congestion in power law graphs. In *Proc. ACM SIGMETRICS*, pages 148–159, San Diego, CA, June 2003.

[12] Y. Huang. *Understanding how to provide service guarantees in IP-based networks*. PhD thesis, Dept. of Electrical and Systems Engineering, University of Pennsylvania, May 2005.

[13] L. Li, D. Alderson, W. Willinger, J. Doyle, R. Tanaka, and S. Low. A first-principles approach to understanding the Internet's router-level topology. In *Proc. ACM SIGCOMM*, Portland, OR, September 2004.

[14] A. M. Odlyzko. Internet traffic growth: Source and implications. In *Proc. SPIE*, volume 5247, pages 1–15, August 2003.

[15] K. Papagiannaki, R. Cruz, and C. Diot. Network performance monitoring at small time scales. In *Proc. ACM Internet Measurement Conference (IMC)*, Miami, FL, October 2003.

[16] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg. Fast accurate computation of large-scale traffic matrices from link loads. In *Proc. ACM SIGMETRICS*, San Diego, CA, June 2003.

[17] Y. Zhang, M. Roughan, C. Lund, and D. Donoho. An information-theoretic approach to traffic matrix estimation. In *Proc. ACM SIGCOMM*, Karlsruhe, Germany, August 2003.

[18] R. Zhang-Shen and N. McKeown. Designing a predictable Internet backbone network. In *Proc. HotNet-III*, San Diego, CA, November 2004.