On the Optimality and Interconnection of Valiant Load-Balancing Networks

Moshe Babaioff and John Chuang
School of Information
University of California at Berkeley
Berkeley, California 94720–4600
{moshe,chuang}@sims.berkeley.edu

Abstract— **The Valiant Load-Balancing (VLB) design has been proposed for a backbone network architecture that can efficiently provide predictable performance under changing traffic matrices [1]. In this paper we show that the VLB network has *optimal* performance when nodes can fail, in the sense that it can support the maximal homogeneous flow for any number of node failures. We generalize the VLB design to enable interconnection of multiple VLB networks, and study interconnection via bilateral peering agreements as well as transit agreements. We show that using VLB as a transit scheme yields the lowest possible network and interconnection capacities, while VLB peering can also achieve near-optimal use of capacity.**

I. Introduction

The Internet core consists of multiple interconnected backbone networks, with each backbone network independently provisioned, deployed, and administered by its owner. The interconnection regime evolves over time, as networks negotiate interconnection agreements with one another. It has become so complex and opaque that researchers have to devise various probing methods to infer the topology of the Internet. On top of this, the traffic matrices experienced by the backbone networks are becoming increasingly variable at both large and small timescales. This is due to a number of factors, including the popularity of new application classes characterized by dynamic overlay routing of large data flows. This makes the tasks of traffic engineering and network provision/upgrade extremely challenging for the network operators. Consequently, many backbone operators have resorted to over-provisioning by a factor of up to ten in order to maintain low latency in their networks.

This has led the networking community to revisit the design of the backbone network architecture. In particular, researchers at Stanford and Bell Labs have separately put forward two-phase load-balancing network designs that can provide predictable performance for highly variable traffic matrices [1], [2]. For example, the Valiant Load-Balancing (VLB) design from Stanford [1] imposes a specific topological structure (logical full mesh) on the backbone network, and routes data via exactly two hops over the full mesh using a simple load-balancing scheme (equal load-balancing on all nodes). This allows the network to efficiently support changing traffic matrices with robustness against failures. The two-phase routing scheme [2], [3] generalizes the above load-balancing scheme to any network topology, using non-equal load balancing on intermediate nodes.

The VLB two-phase load-balancing scheme has many advantages for the network operator. It handles the unpredictability of traffic, by being able to supports *any* traffic matrix as long as it falls within the capacity constraints of ingress-egress capacity of each node ("hose model"). Moreover, it avoids congestion without any dynamic or real-time configuration of the network. Routing is oblivious and independent of the specific traffic matrix. Finally, it is efficient in the sense that it has minimal total capacity provisioned [1].

As network operators begin to contemplate the VLB backbone design, the natural next question arises: how should multiple VLB networks interconnect with one another? This interconnection problem actually encompasses many specific open questions, including: how should the load-balanced routing algorithm be generalized across multiple VLB networks? How should the interconnection points be selected? Can the VLB design support different interconnection relationships, e.g., transit and peering? Are the efficiency and robustness properties of a single VLB network retained for multiple interconnected VLB networks? Different methods of interconnecting VLB networks are possible, and they should be compared against one another along dimensions such as efficiency, robustness, evolvability, and support for competition and innovation.

In this paper, we will focus on establishing the optimality of the VLB design for a standalone network as well as for various forms of interconnections. First, we will extend the analysis of [1], [4] and establish the universal optimality of the VLB design to node failures (Section III). Next, we will propose a generalization of the VLB network that facilitates interconnection, namely the *m-hubs* VLB network and the *m-hubs l-tolerant* VLB network (Section IV). This allows us to establish optimality results for the cases of transit and peering between multiple VLB networks (Section V).

A. Related Work

Our paper follows a line of research that aims to design networks under unpredictable traffic, thus replacing the assumption of a fixed traffic matrix by the "hose-model" [5] in which only a bound on the ingress and egress rates of each node in the network is known. The goal is to build efficient networks that support *any* traffic matrix that is consistent with the rate bounds. We next discuss some related literature on network design under the hose model.

Our paper is most closely related to the papers by Zhang-Shen and McKeown [1], [4] which consider the problem of Internet back-bone IP routing. They suggest to use a two-phase routing, which they call the Valiant Load-Balancing (VLB) scheme (following Valiant [6]), on a logical full mesh. Routing is done in two phases, first, each flow is equally split on all nodes, and then forwarded to its destination. This routing scheme is shown to be optimal (with respect to total capacity). We show that it has the best performance with respect to node failures, and that its generalizations can be used for efficient interconnection of networks.

Kodialam, Lakshman and Sengupta [2] suggest two-phase routing schemes that can be viewed as a generalization of the VLB scheme. They provide Linear Programming formulations for various goals [3], [7] and argue that it has many advantages over direct routing. The paper [8] considers the issue of resilience to a single router (node) failure with throughput as the optimization goal. Our universal optimality of VLB to node failures is, as far as we are aware of, the first result showing that the VLB scheme performs optimally with respect to any number of failures, with total capacity as the optimization goal.

Keslassy et al. [9] consider the use of a two-phase load-balancing scheme inside a switch. While the application is different, the basic model and goal is very similar to the model of a single network that we consider. The paper also considers routing under the homogeneous hose model on a full mesh. The objective of Keslassy et al. is to minimize the sum of capacities of all edges in the network including self edges, while our goal is to minimize the sum of capacities *excluding* self edges (as in [1], [4]), as this makes more sense for back-bone networks. [9] shows that optimizing for the sum of all edges capacities results with a unique optimal network that is biased (self edges have half the capacity of non-self edges), while for our optimization goal we observe that there are multiple optimal networks, a fact that we show useful in networks interconnection.

The literature on bandwidth provisioning for Virtual Private Network (VPN) under the hose model can be viewed as a generalization of the single network model considered in this paper. This generalization allows for heterogeneous rates at the nodes, capacity bounds on the edges and possibly different cost of unit capacity for different edges. Gupta et al. [10] consider routing along fixed routes and show that the problem of finding an optimal tree routing or single-path routing is NP-hard. Erlebach and Rüegg [11] allow for multi-path routing of splittable flows and show that the optimal provisioning problem can be solved in polynomial time using Linear Programming (LP). In contrast, our paper addresses resilience to failures and interconnection of networks, and offers a simple routing scheme with load-balancing splits that are independent of the source and destination.

II. Model

A. The Single Network Model

We begin by presenting the model of a single network with homogeneous access capacities. The network N consist of n nodes.The network can be represented as a directed graph, with set of vertexes of size n, and an edge between each pair of nodes. We assume that there are no constraints on the edges' capacities.

A *traffic matrix* Λ is an $n \times n$ matrix such that a flow (rate) of size $\lambda_{ij} \geq 0$ needs to be sent from node i to node j. We sometimes refer to λ_{ij} as the *stream* from i to j. As traffic is dynamic and changes over time, our goal is to build a network that can support a large set of traffic matrices. We adopt the "hose-model"([5]) in which each node in the network has an homogeneous bound r on its ingress and egress rates, and we wish to provision the network to support *any* traffic matrix that is consistent with the rate bounds. A *routing scheme* defines the way that traffic is routed in the network, and given a routing scheme, we can find the minimal capacity on each edge that is required to support all desired traffic matrices. We assume that the flows are splittable.

Formally, a traffic matrix is *legal with rate* r if for any $i \in N$, $\sum_{j\in N} \lambda_{ij} \leq r$, and for any $j \in N$, $\sum_{i\in N} \lambda_{ij} \leq r$. Given a network with an $n \times n$ capacity matrix C, we say that the network can *support* a traffic matrix Λ if there exists a solution to the multi-commodities flow problem [12] defined by the demands Λ and the capacity constraints C, on the directed graph of the network. A network with capacity C can *support homogeneous rate of* r if for any legal traffic matrix Λ with rate r, the network can support the traffic matrix Λ. We define the *capacity* of a network[1] with an $n \times n$ capacity matrix C, to be the sum of the edges' capacity, without self edges:

$$C = \sum_{i\in N} \sum_{\substack{j\in N \\ j\neq i}} c_{ij}$$

A network is *optimal* if it has minimal capacity over all networks that support all legal traffic matrices with rate r.

All the networks that we present will use local routing decisions (oblivious routing) and will not require any solving of multi-commodities flow problems (or any other Linear Programming), or the knowledge of the specific traffic matrix.

We also consider node failures. A failure of a node implies that it can no longer generate or receive any traffic, and all edges that incident on the node are not used for routing. Formally, given that the nodes $F \subset N$ failed, a legal traffic matrix Λ with rate r is *legal after* F *failed* if for any $i \in F$, $\sum_{j\in N} \lambda_{ij} = 0$, and for any $j \in F$, $\sum_{i\in N} \lambda_{ij} = 0$. A network with capacity matrix C *supports homogeneous rate of* r *after nodes* F *failed* if any Λ with rate r that is legal after F failed can be routed without violating the capacity constraints imposed by C, with flow of 0 on each edge that incident on F.

[1] We abuse notation and use C to denote the capacity of network with capacity matrix C.

B. The Multiple Networks Interconnection Model

We are interested in the interconnection of networks. Assume that there are q networks: $X = \{x_1, x_2, \ldots x_q\}$. Network $x \in X$ has n_x nodes, and has homogeneous rate of r_x at each node[2]. Similar to the case of a single network, we do not want to assume knowledge of a specific traffic matrix for the interconnection traffic. Rather, we adopt a similar "hose-model" for the interconnection traffic. We assume that there exists an homogeneous bound R^p on the ingress and egress rate of each network to the other networks. The network operators decide on R^p by negotiation.

Formally, we call an interconnection traffic matrix *legal* if it

- respects the constraint on local traffic: for any network $x \in X$ and for any node $i \in x$, $\sum_{j\in x} \lambda_{ij} \leq r_x$, and for any $j \in x$, $\sum_{i\in x} \lambda_{ij} \leq r_x$.
- respects the constraint on interconnection traffic: for any network $x \in X$ it holds that $\sum_{i\in x} \sum_{j\notin x} \lambda_{ij} \leq R^p$, and $\sum_{i\notin x} \sum_{j\in x} \lambda_{ij} \leq R^p$.

We further assume that each network is able to generate traffic of size R^p, that is $R^p \leq min_{x\in X}\{n_x \cdot r_x\}$.

While in the single network case we did not restrict the capacities of the edges inside the network, we impose natural restrictions on the capacities on edges between different networks. We assume that each node has some location, and the locations of two nodes in the same network are different. On the other hand, two nodes from different networks can share a location. We only allow such nodes that share a location to create a connection between them. Formally, let L be a set of locations. For each node i, let $l(i) \in L$ be the location of node i. For $x \in X$ and $i, j \in x$, $l(i) \neq l(j)$. For two different networks $x, y \in X$, $x \neq y$ and nodes $i \in x$, $j \in y$ it holds that if $l(i) \neq l(j)$ (they are not in the same location) it implies that there is no link between the nodes i and j ($c_{ij} = 0$).

Let $S_{xy} = \{i \in x | \exists j \in y \ s.t. \ l(i) = l(j)\}$ be the set of nodes in network x that can be peered to nodes in network y. That is, S_{xy} is the set of nodes from which traffic can be sent from network x to network y. We assume that these connections are bidirectional, that is, every node in S_{xy} is connected to a node in S_{yx} and vise versa.

For the case of multiple networks we care about the capacity of each network, that is for network $x \in X$, $C(x) = \sum_{i\in x} \sum_{\substack{j\in x\\ j\neq i}} c_{ij}$. We also care about the *interconnection capacity* (total capacity between the networks), defined to be $\sum_{x\in X} \sum_{i\in x} \sum_{j\notin x} c_{ij}$.

III. Universal Optimality of VLB to Node Failures

In this section we consider a single network using the VLB scheme. The VLB network suggested by Zhang-Shen and McKeown [4] routes each stream in two stages. First $1/n$-fraction of each stream is sent to each of the nodes in the network (each stream is *load balanced* on all nodes), and then each stream is sent to its destination. In [4] it is shown that for this routing scheme, capacity of $2r/n$ on each edge is sufficient, yielding total capacity of $2r(n-1)$ for the entire network. The paper also shows that total capacity of $2r(n-1)$ is necessary to support all traffic matrices. The following is a corollary of Theorem 1 of [4].

Lemma 1: The capacity of a network with $\tilde{n}$ nodes that supports homogeneous rate $\tilde{r}$ is at least $2\tilde{r}(\tilde{n}-1)$.

The proof of the above is based on the observation that in order to support the matrix in which each node $(1, \ldots, \tilde{n} - 1)$ is sending $\tilde{r}$ to the next node, total forward capacity $(\sum_{i\in N} \sum_{j\in N, j>i} c_{ij})$ of at least the size of the total forwarded flow of $\tilde{r}(\tilde{n}-1)$ is required (as each forwarded flow must travel forward at least once), and the same for the transpose matrix and backward capacity. Note that the capacity needed to support any traffic matrix is less than twice the capacity needed to support one specific matrix.

What if nodes can fail? How well does the VLB network perform when nodes might fail? In this section we show that VLB has the *best possible performance with respect to any number of failures*, and is the *only* network with this property. We show that for any given capacity of a network (total capacity of all non-self edges), the VLB network has the best resistance to failures over all networks with the same capacity, in the following sense. Assume that the capacity of the network is $C = 2 \cdot r \cdot (n-1)$ (minimal capacity to support a rate of r). For any $l \in \{1, \ldots, n\}$, after l failures (worst case failures, done by an adversary), the VLB network can support the maximal possible homogeneous rate of $\frac{n-l}{n} \cdot r$, and no other network has this property.

Consider the VLB network in which each edge has a capacity of $\frac{2r}{n}$, this network can support an homogeneous rate of r at each node. Now assume that there are l node failures - as the network is symmetric it does not matter which nodes have failed. We next show that after the failures the network can now support an homogeneous rate of $\frac{n-l}{n} \cdot r$ at each node. In the first stage, each node sends $\frac{1}{n-l}$ of each stream to each of the remaining nodes, and at the second stage all traffic is sent to its destination. On each edge, at each stage, there is a flow of at most $(\frac{n-l}{n} \cdot r) \cdot (\frac{1}{n-l}) = \frac{r}{n}$, thus there is enough capacity for this scheme. We note that the VLB network after any $l \in \{1, \ldots, n\}$ node failures has a capacity of $\frac{2r}{n} \cdot (n-l)(n-l-1)$.

Next we prove that for any network with capacity $C = 2 \cdot r \cdot (n-1)$, after l (worst case) failures, the network has at most the capacity of the VLB network after l failures.

Lemma 2: Given a network that has minimal capacity $2 \cdot r \cdot (n-1)$ needed to support homogeneous rate of r, there exists a set of nodes of size l, such that after all these nodes fail, the total remaining capacity is at most $\frac{2r}{n} \cdot (n-l)(n-l-1)$.

Proof: Let $\tilde{C}$ denote the capacity matrix of the network. Assume in contradiction that for every set of size l, the total remaining capacity is greater than $\frac{2r}{n} \cdot (n-l)(n-l-1)$. Let $\mathcal{S}$ denote the collection of all sets of size $n-l$. By our assumption, for any set $S \in \mathcal{S}$,

[2] All nodes at the same network has the same rate, but different networks might have different rates.

$$\sum_{i\in S}\sum_{\substack{j\in S\\ j\neq i}}\tilde{c}_{ij} > \frac{2r}{n}\cdot(n-l)(n-l-1)$$

As the size of $\mathcal{S}$ is $\binom{n}{l}$, by summing over all $S\in\mathcal{S}$ we get

$$\sum_{S\in\mathcal{S}}\sum_{i\in S}\sum_{\substack{j\in S\\ j\neq i}}\tilde{c}_{ij} > \binom{n}{l}\cdot\frac{2r}{n}\cdot(n-l)(n-l-1)$$

We use the symmetry between all nodes to figure out how many times each $\tilde{c}_{ij}$ is counted in the above summation. For any given pair $i<j$, there are $\binom{n-2}{l}$ ways to chose the l nodes to remove, out of all nodes but i and j (there are $n-2$ such nodes). Thus

$$\sum_{S\in\mathcal{S}}\sum_{i\in S}\sum_{\substack{j\in S\\ j\neq i}}\tilde{c}_{ij} = \binom{n-2}{l}\sum_{i\in N}\sum_{\substack{j\in N\\ j\neq i}}\tilde{c}_{ij}$$

We conclude that

$$\binom{n-2}{l}\sum_{i\in N}\sum_{\substack{j\in N\\ j\neq i}}\tilde{c}_{ij} > \binom{n}{l}\cdot\frac{2r}{n}\cdot(n-l)(n-l-1)$$

As $\binom{n-2}{l}=\binom{n}{l}\cdot\frac{(n-l)(n-l-1)}{n(n-1)}$ we derive

$$\sum_{i\in N}\sum_{\substack{j\in N\\ j\neq i}}\tilde{c}_{ij} > 2\cdot r\cdot(n-1)$$

which is a contradiction. ∎

Corollary 3: A network that has capacity $2\cdot r\cdot(n-1)$ cannot support homogeneous rate of more than $\frac{n-l}{n}\cdot r$ for every l nodes that fail.

Proof: By Lemma 2 for some set of l failing nodes the remaining capacity is at most $\frac{2r}{n}\cdot(n-l)(n-l-1)$. By Lemma 1, if a network with capacity $\tilde{C}$ has $\tilde{n}=n-l$ nodes, it can support homogeneous rate of at most $\frac{\tilde{C}}{2(\tilde{n}-1)}$. Thus if the network capacity is $\tilde{C}\leq\frac{2r}{n}\cdot(n-l)(n-l-1)$, the network can support homogeneous rate of at most $\frac{\frac{2r}{n}\cdot(n-l)(n-l-1)}{2(n-l-1)}=\frac{n-l}{n}\cdot r$. ∎

Given the corollary we can now define optimal performance after any failures.

Definition 4: A network that has minimal capacity $2\cdot r\cdot(n-1)$ needed to support homogeneous rate of r has *optimal l-failures performance* if for any set F of size l of nodes that fail, it can support the maximal homogeneous rate of $\frac{n-l}{n}\cdot r$ after F failed.

We next present the main result of the section.

Theorem 5: The VLB network has optimal l-failures performance for any $l\in\{1,\ldots,n\}$, and is the only network with this property (any network with this property has the same capacities on all edges).

Proof: The VLB network after any $l\in\{1,\ldots,n\}$ node failures has a capacity of $\frac{2r}{n}\cdot(n-l)(n-l-1)$, and as we have seen above, can support rate of $\frac{n-l}{n}\cdot r$ after any failures of up to l nodes. By Corollary 3 any network that has capacity $2\cdot r\cdot(n-1)$ cannot support homogeneous rate of more than $\frac{n-l}{n}\cdot r$ for every l nodes that fail, thus VLB has optimal l-failures performance.

Next we show that no other network has the property (support the same maximal flow as the VLB, for any l). If a network does not have exactly the same edge capacities as the VLB network, but has the same network capacity, this implies that there is an edge (i,j) such that $c_{ij}<\frac{2r}{n}$. Thus, for $l=n-2$, if all nodes other than i,j fail, the network can support less traffic than VLB. VLB can support a flow of $\frac{n-(n-2)}{n}\cdot r=\frac{2r}{n}$, while the other network cannot. ∎

There is another way to view the above results. Given a "capacity budget" C, the above results gives the optimal use of such capacity, if one wishes to build a network with maximal homogeneous rate for any number of node failures. VLB network will enable an homogeneous rate of $r=\frac{C}{2(n-1)}$ with no failures, and an homogeneous rate of $r\cdot\frac{n-l}{n}=\frac{C}{2(n-1)}\cdot\frac{n-l}{n}$, after l failures (for any l). No other network can use the capacity budget in a better way.

Zhang-Shen and McKeown [1] considered the problem of designing a network that can support homogeneous rate of r after l node failures (an l-tolerant VLB network). The paper suggests to use capacity of $\frac{2r}{n-l}$ on each edge, and load balance each stream on all surviving nodes, this gives total capacity of $\frac{2r}{n-l}n(n-1)$. [1] pointed out that the function $\frac{2r}{n-l}$ is very flat for small values of l, but gave no proof for the optimality of this scheme. Our result shows that the network capacity of their l-tolerant VLB network is actually *optimal*.

Corollary 6: Any network that can support any legal traffic matrix with homogeneous rate of r *after* any l nodes failures, has capacity of at least $\frac{2r}{n-l}n(n-1)$.

Proof: We have seen that a network with n nodes and rate r', can support homogeneous rate of at most $r'\cdot\frac{n-l}{n}$ after l failures. If we like to support rate of r after the l failures, then $r=r'\cdot\frac{n-l}{n}$ thus $r'=r\cdot\frac{n}{n-l}$ and by Lemma 1 the necessary capacity to support this rate is $2r'(n-1)=2r\frac{n}{n-l}(n-1)$ as required. ∎

IV. Generalizations of the VLB Network

We now generalize the VLB scheme of [4] and show that if each stream is load balanced on m nodes (hubs) instead of all n nodes (this can be viewed as a special case of the generalization of [2]), the total capacity of the network does not change. This "m-hubs VLB network" would be useful later in designing optimal interconnection network.

Additionally, we discuss a generalization of this network that can support up to l node failures. Our "m-hubs l-tolerant VLB network" can be viewed as a generalization of the scheme of Zhang-Shen and McKeown [1] which handles node failures but load balances each stream on all nodes. Our design load balances each stream only on m nodes, and this implies some increase in capacity. Nevertheless, this scheme is also useful in designing optimal interconnection networks in the presence of node failures.

A. The m-hubs VLB Network

The m-hubs VLB network load balances each stream on m hubs.

Definition 7: The *m-hubs VLB network* is an n nodes network where m nodes serve as hubs.

Routing: For a given legal traffic matrix Λ, it load balances each source-target stream λ_{ij} on each of the m hubs: At the first stage each node i sends $\frac{1}{m}$ of each stream λ_{ij} to each of the m hubs, and at the second stage each stream is forwarded to its destination.

Capacities: Let H be the set of hubs, $|H| = m$. The capacity of the edge (i, j) is

- 0 if $i \notin H$, $j \notin H$.
- $\frac{r}{m}$ if $i \notin H$, $j \in H$ (for the first stage).
- $\frac{r}{m}$ if $i \in H$, $j \notin H$ (for the second stage).
- $\frac{2r}{m}$ if $i \in H$, $j \in H$ ($\frac{r}{m}$ for each of the two stages).

The capacity of the network is $2r(n-1)$ as

$$C = \sum_{i\in N}\sum_{\substack{j\in N\\ j\neq i}} c_{ij} = 2(n-m)m\frac{r}{m} + m(m-1)\frac{2r}{m} = 2r(n-1)$$

Note that a star network is a special case where $m = 1$, and the VLB scheme is the special case where $m = n$. Observe that the capacity is optimal and *independent* of m !

Observation 8: For any m, the m-hubs VLB network supports homogeneous rate of r. It has the same capacity of $2r(n-1)$, and this capacity is necessary to support any legal traffic matrix.

Proof: At the first stage, each node i send $1/m$ fraction of each stream λ_{ij} originating from i, to each node $k \in H$. As $\sum_{j\in N} \lambda_{ij} \leq r$, capacity of r/m is sufficient for the traffic on the edge from i to $k \in H$.

At the second stage, as each node $k \in H$ received $1/m$ fraction of each stream with destination node j, capacity of $\sum_{i\in N} \frac{\lambda_{ij}}{m} \leq \frac{r}{m}$ is sufficient from node k to node j for the second stage.

The network has capacity of $2r(n-1)$, and this capacity is shown to be necessary in [4]. ■

We use the m-hubs VLB network to build optimal interconnection of networks by peering agreements. In Section V-B.1 we show that if all networks has $m > 0$ shared location, no extra capacity in the networks is needed to support peering traffic. Each network runs an m-hubs VLB network on the set of the m shared locations. Routing is done by first load-balancing traffic on the hubs of the source network, then peering interconnection traffic to the destination network, and finally sending all traffic to its destination node. Not only does this scheme has optimal capacity in each network, it also has optimal interconnection (peering) capacity.

We note that the m-hubs VLB network can also be useful in cases where one wishes to reduce the number of hubs to be managed, as well as when there is economics of scale with respect to edge capacities (as now each non-zero capacity edge has larger capacity). On the other hand, lower number of hubs reduces the tolerance to failures.

Next we consider node failures, and generalize the above definition.

B. The m-hubs l-tolerant VLB Network

We begin by defining tolerance to at most l failures.

Definition 9: A network is *l-tolerant* if it can support any legal traffic matrix after F failed, for any $F \subset N$ of size at most l.

The m-hubs l-tolerant VLB network load balances each stream on m hubs, and is l-tolerant.

Definition 10: The *m-hubs l-tolerant VLB network* is an n nodes network which has m hubs. Let H be the set of hubs, $|H| = m$. Assume that the set F of nodes failed, and that $|F| \leq l$.

Routing: For any legal traffic matrix Λ, it load balances each source-target stream λ_{ij} on each of the hubs that are not in F. That is, fraction $\frac{1}{|H\setminus F|} \leq \frac{1}{m-l}$ of λ_{ij} is sent from node i to each node $k \in H \setminus F$ in the first stage, and forwarded to the destination in the second stage.

Capacities: The capacity of the edge (i, j) is

- 0 if $i \notin H$, $j \notin H$.
- $\frac{r}{m-l}$ if $i \notin H$, $j \in H$ (for the first stage).
- $\frac{r}{m-l}$ if $i \in H$, $j \notin H$ (for the second stage).
- $\frac{2r}{m-l}$ if $i \in H$, $j \in H$ (for the two stages).

The capacity of the network is $C = 2r(n-1)\frac{m}{m-l}$. Lemma 11 shows that the m-hubs l-tolerant VLB network can indeed support up to l node failures.

Lemma 11: The m-hubs l-tolerant VLB network supports homogeneous rate of r after F failed, for any set $F \subset N$ with $|F| \leq l$.

Proof: After the set F of nodes failed ($|F| \leq l$) the flow sent on the edges is as follows. At the first stage, if $i, k \notin F$ and $k \in H$, i sends to k a flow of size $\frac{\lambda_{ij}}{m-|F\cap H|}$ out of the flow λ_{ij}, for any $j \notin F$ (as it sends $\frac{1}{m-|F\cap H|}$ fraction of any flow from $i \notin F$ to $j \notin F$ through any node $k \in H \setminus F$). This implies that on the edges from i to k, the flow that is sent is of size

$$\sum_{j\notin F} \frac{\lambda_{ij}}{m-|F\cap H|} = \frac{1}{m-|F\cap H|} \cdot \sum_{j\notin F} \lambda_{ij} \leq$$
$$\frac{r}{m-|F\cap H|} \leq \frac{r}{m-l}$$

Thus on the edge i to k, a capacity of $\frac{r}{m-l}$ is enough for the first stage.

At the second stage, if $j, k \notin F$ and $k \in H$, k sends to j all the flow it has received in the first stage, that is destinated to j. k has received from node $i \notin F$ the flow $\frac{\lambda_{ij}}{m-|F\cap H|}$ that is destinated to j. This implies that k has received at most

$$\sum_{i\notin F} \frac{\lambda_{ij}}{m-|F\cap H|} \leq \frac{r}{m-l}$$

Thus on the edge k to j, a capacity of $\frac{r}{m-l}$ is enough for the second stage. We conclude that capacity of the m-hubs l-tolerant VLB network is enough to support both stages. ■

Definition 12: The *l-tolerant VLB network* is defined to be the n-hubs l-tolerant VLB network.

Note that the function $\frac{m}{m-l}$ is monotonically decreasing with m, thus load balancing on all nodes (using the l-tolerant VLB network for which $m = n$) minimizes the network's capacity. The "m-star" network is the m-hubs l-tolerant VLB network. The above implies that the l-tolerant VLB scheme ($m = n$) is better than the "m-star" for any $m < n$.

Observation 13: The l-tolerant VLB network (with n-hubs) has lower capacity than the m-hubs l-tolerant VLB network (the "m-star" network), for $m < n$.

Moreover, Corollary 6 shows that the l-tolerant VLB network is optimal, that is it has minimal capacity over all networks that support homogeneous rate of r after l node failures.

V. Interconnection of VLB Networks

We are now ready to present interconnection schemes for multiple VLB networks based on transit and peering agreements. In addition to quantifying the capacity requirements for different schemes, we will also discuss the implications on the design of interconnection networks.

Consistent with established terminology, when a VLB network plays the role of a transit network, it may carry traffic that neither originates nor terminates within itself. With peering, non-local traffic with origin in network $x \in X$ and destination in network $y \in X$ s.t. $y \neq x$ cannot go through any other network $z \in X$ s.t. $z \neq x, y$. In this section (except in Section V-B.1) we consider that networks only have pair-wise shared locations, i.e., no location is shared by more than two networks.

A. Transit

We first consider a single transit network to which multiple stub networks are connected, and show that using the VLB scheme is optimal. This architecture may be appropriate for a national utility model or a regulated monopoly model, as the single transit network exercises monopolistic power over the stub networks. This scheme may also be appropriate for a single network domain distributed over a large geographic area with low traffic volumes between regions, as it reduces the latency of traffic that is local to a region. We also consider the case of two transit networks, such that each network alone can support any legal interconnection traffic matrix. This scheme ensures that no transit network exercises monopolistic power over the stub networks. We can also view this scheme as robust against failure of one of the transit networks.

1) A Single Transit Network: Assume that there are q networks: $X = \{x_1, x_2, \ldots x_q\}$ and let $z = x_q$ be the transit network. Recall that S_{xy} denotes the nodes of network x that share common locations with nodes of network y.

Definition 14: The *interconnection network by VLB transit* is an internetwork consisting of a transit network z and $q-1$ stub networks $X \setminus \{z\} = \{x_1, x_2, \ldots x_{q-1}\}$. On each stub network x we build a $|S_{xz}|$-hubs VLB network, using the nodes of S_{xz} as the hubs.

Let $S_z = \cup_{x \in X} S_{zx}$ be the set of location in the transit network that are shared with the stub networks. In the transit network z we build a $|S_z|$-hubs VLB network using the nodes of S_z as the hubs.

Routing: The VLB scheme in the transit network runs between the two stages of the VLB scheme of the stub networks. Stream λ_{ij} going from node i in network x to node j in network y is routed as follows:

1) The stream is load balanced on the hubs of network x. If $x = z$ go to step 4.
2) the stream is peered to the transit network z, from S_{xz} to S_{zx}.
3) It is load balanced from S_{zx} on all hubs S_z of z (without load balancing from S_{zx} to S_{zx})[3]
4) If $y = z$ go to step 7.
5) The stream is load balanced from S_z on the peering nodes S_{zy} with network y (without load balancing from S_{zy} to S_{zy}).
6) It is peered to the stub network y, from S_{zy} to S_{yz}, the hubs of y.
7) The load balanced traffic on the hubs of y is sent to the destination j.

Capacities: In network z capacity of $2 \cdot r_z \cdot (n_z - 1)$ (defined by the $|S_z|$-hubs VLB network capacities) is needed to support stages 1 and 7. Additional to the capacities required for local traffic, in order to support stage 3 we add capacity of $R^p/(|S_{zx}| \cdot |S_z|)$ from each node $i \in S_{zx}$ to each node $j \in S_z \setminus S_{zx}$. To support stage 5 we add capacity of $R^p/(|S_{zy}| \cdot |S_z|)$ from each node $i \in S_z \setminus S_{zy}$ to each node $j \in S_{zy}$. The capacity of the transit network z is $2 \cdot r_z \cdot (n_z - 1) + 2 \cdot R^p \cdot (q-2)$.

The capacity of each stub network x is $2 \cdot r_x \cdot (n_x - 1)$. The interconnection capacity is $2 \cdot R^p \cdot (q-1)$.

Note that given an "interconnection network by VLB transit" with $q \geq 2$ networks, adding a stub network causes an increase in capacity of $2 \cdot R^p$ in the transit network (this holds for any stub network other than the first), and additional interconnection capacity of $2 \cdot R^p$ is needed between the new stub and the transit network. The transit network operator can charge this extra cost to the stub network operator.

Theorem 15: The "interconnection network by VLB transit" can support any legal interconnection traffic matrix. Additionally, any interconnection network that can support any legal interconnection traffic matrix and uses a transit network has at least the same capacity in each network and at least the same interconnection capacity.

Proof: First observe that capacity of $r_z(n_z - 1)$ is indeed sufficient for each of the steps 1 and 7, to handle traffic that its origin or destination is network z.

In order to support stage 3, a capacity of $R^p/(|S_{zx}| \cdot |S_z|)$ from each node $i \in S_{zx}$ to each node $j \in S_z \setminus S_{zx}$ is sufficient. This is true because, after the load balancing of step 1, each node $i \in S_{zx}$ has $1/|S_{zx}|$ of at most R^p of interconnection traffic with origin at x, and it needs to send $1/|S_z|$ of it to $j \in S_z \setminus S_{zx}$. A similar argument holds for the capacity needed to support stage 5. We are left to show that the capacity allocated

[3]This means that each node in S_{zx} sends $1/|S_z|$ fraction of each stream to each node in $S_z \setminus S_{zx}$. As each stream is already load balanced on S_{zx}, there is no need for load balancing traffic between nodes in S_{zx}: it uses capacity but at the end of the stage each node will have the same fraction of each stream as in the absence of this load balancing.

in the transit network z to support each of these two stages sums to $R^p \cdot (q-2)$. Indeed, for stage 3 it holds that

$$\sum_{x\in X\setminus\{z\}} \sum_{i\in S_{zx}} \sum_{j\in S_z\setminus S_{zx}} R^p/(|S_{zx}|\cdot|S_z|) =$$

$$\sum_{x\in X\setminus\{z\}} |S_{zx}|\cdot|S_z\setminus S_{zx}|\cdot R^p/(|S_{zx}|\cdot|S_z|) =$$

$$R^p \cdot \sum_{x\in X\setminus\{z\}} (1-|S_{zx}|/|S_z|) =$$

$$R^p\left(q-1-\sum_{x\in X\setminus\{z\}} |S_{zx}|/|S_z|\right) = R^p(q-2).$$

Similar calculations give a capacity of $R^p \cdot (q-2)$ for stage 5.

As the traffic is load balanced in each stub network, a capacity of R^p from the stub to the transit network is clearly sufficient for stage 2. The same capacity is sufficient for stage 6. As there are $q-1$ stub networks, the total interconnection capacity is $2\cdot R^p(q-1)$. ■

The next lemma presents lower bounds on the capacity of each network, and the interconnection capacity. The lower bounds match the capacities that we achieve using the VLB scheme.

Lemma 16: Given q networks such that no more than two networks share a common location. If we wish to support all legal interconnection traffic matrices using a single transit network, then it is necessary to allocate capacity of at least

- $2\cdot r_z\cdot(n_z-1)+2\cdot R^p\cdot(q-2)$ in the transit network z.
- $2\cdot r_x\cdot(n_x-1)$ in each stub network x.
- $2\cdot R^p\cdot(q-1)$ for interconnection.

Proof: By Lemma 1, in order to support the local traffic in network x, we need capacity of at least $2\cdot r_x\cdot(n_x-1)$. Additionally, each of the $q-1$ stub networks needs a capacity of at least R^p to and from the transit network to support the interconnection traffic to and from this network. This gives the lower bound of $2\cdot R^p\cdot(q-1)$ on the interconnection capacity. Finally, we consider the capacity of the transit network. Consider some ordering over the nodes of z such that a node connected to network x precede all nodes connected to network y, whenever $x<y$. For any legal interconnection traffic matrix, the total forward capacity of the transit network, defined to be $\sum_{i\in z}\sum_{j\in z, j>i} c_{ij}$ must be at least the size of the total forward flow in z, plus the total forward interconnection flow between the stub networks, i.e.,

$$\sum_{i\in z}\sum_{\substack{j\in z\\ j>i}} c_{ij} \geq \sum_{i\in z}\sum_{\substack{j\in z\\ j>i}} \lambda_{ij} + \sum_{x\neq z}\sum_{\substack{y\neq z\\ y>x}}\sum_{i\in x}\sum_{j\in y}\lambda_{ij}$$

This holds in particular for the matrix in which each of the $q-2$ first stub networks sends a combined interconnection flow of R^p to the next stub network, and each of the n_z-1 first nodes in z sends r_z to the next node in z. The total forward flow of this matrix is $r_z(n_z-1)+R^p(q-2)$. The same bound holds for the transpose matrix and backward capacity, and when combined we conclude that the capacity of the transit network must be at least $2r_z(n_z-1)+2R^p(q-2)$. ■

2) Two Transit Networks: It is possible to generalize the construction presented in the previous subsection to build an interconnection network with multiple transit networks [4]. Assume that we would like to create an interconnection network with two transit networks and $q-1$ stub networks (the number of stub networks is the same as in the previous section), such that even if one of the transit networks fails, the other transit network could support any legal interconnection traffic matrix.

Assume that no traffic from one transit network is sent to the other transit network (removing this assumption will cause minor changes in the following observations). In each of the transit networks the same capacity as in the "interconnection network by VLB transit" network is necessary and sufficient. Additionally, the interconnection capacity will double (again, it is necessary and sufficient).

At each stub network x, if x has a shared location with both transit networks, we could build its hubs on the set of common locations to the three networks, and no extra capacity will be needed at the stub network. If there are no locations shared by all three networks, we can consider either one hop or two hops schemes for routing of interconnection traffic inside network x. In case of one hop, capacity of $2\cdot(2r_x(n_x-1))$ is sufficient (by allocating capacities as if the hubs are on the peering nodes with one of the transit and with the other).

In case of two hops, we can use the following scheme. We build the hubs of x on a set of peering nodes with the two transit networks, with equal number of peering nodes with each of the two. That is, let S_{xz_1} and S_{xz_2} be the set of peering nodes with transit network z_1 and z_2, respectively. Assume w.l.o.g. that $|S_{xz_1}|\leq|S_{xz_2}|$, then we build the hubs on S_{xz_1} and a set of nodes of size $|S_{xz_1}|$ from the nodes of S_{xz_2}. If at some time transit network z_i is used (for $i\in\{1,2\}$ and $j\neq i$), then routing is done as follows. First traffic is load balanced on all hubs, then from each hub that belongs to j, we send the interconnection traffic to a hub that belong to i (using one-to-one matching between the two sets of hubs). This requires capacity of $R^p/(2|S_{xz_1}|)$ on each edge, and total capacity of $R^p/2$. When interconnection traffic is received from z_j, we do the same in reverse order. The total capacity in x that is sufficient for this scheme is $2r_x(n_x-1)+R^p$ (if network j is used for transit, the usage of capacities between the hubs will be in reverse order, but the same capacities will be sufficient). This is less than the capacity needed for one hop routing, since instead of adding capacity of $2r_x(n_x-1)$ we are adding $R^p\leq r_xn_x$ (which is smaller for any $n_x\geq 2$).

B. Peering

We now consider the interconnection of multiple VLB networks using only peering agreements. We first consider the case where there exists at least one location that is shared by all the networks. In this case, we show that no extra capacity is needed in each network to support the peering traffic. We

[4]In the interest of space we only present the basic ideas and observations, details can be found in the extended version of the paper [13].

then consider the case where each location is shared by at most two networks.

1) Peering with Universally Shared Locations:

Definition 17: Assume that S, the set of location that are shared by all q networks, is not empty. The *peering on universally shared locations VLB network* is an interconnection network in which each of the q networks run a $|S|$-hubs VLB networks with hubs on the nodes of S. The q networks peer at all nodes of S.

Routing: Routing in this network is done in three stages. In the first stage each stream is load balanced on all hubs of the originating network, then any peering traffic is sent on the peering edges to the destination network, and in the final stage all the traffic is forwarded to its destination.

Capacities: Each network x has a capacity of $2r_x(n_x - 1)$. There is interconnection (peering) capacity or R^p between any of the $q(q-1)$ ordered pairs of networks, thus the total interconnection capacity is $R^p \cdot q(q-1)$.

Theorem 18: The "peering on universally shared locations VLB network" can support any legal interconnection traffic matrix by peering. Additionally, any network that can support any legal interconnection traffic matrix by peering, has at least the same capacity in each network and at least the same interconnection capacity.

Proof: The same arguments as the ones presented in the proof of Observation 8 also can be used to prove that the "peering on universally shared locations VLB network" can support any legal interconnection traffic matrix. (It makes no difference if the traffic with destination node j is local or not. The important observation is that the rate of all the traffic with destination node j, from all origins, is at most r).

By Observation 8, a capacity of $2r_x(n_x - 1)$ is necessary in network x to support any *local* legal traffic matrix, even without any peering traffic. Thus this amount is clearly necessary. Additionally, recall that we assume that each network can generate and receive interconnection traffic at the rate of R^p. Thus for any of the $q(q-1)$ pairs of networks $x \neq y$, a peering capacity of R^p is necessary between x and y. ■

We conclude that if universally shared locations exist peering using the m-hubs VLB scheme on the shared locations is optimal. Moreover, an increase in R^p only results in an increase in the interconnection capacity, but not in the capacities of the individual networks.

With some additional assumptions, the above result can be generalized to the case of node failures[5]. If we like to support $l < m$ node failures, each network could use the m-hubs l-tolerant VLB scheme with hubs on the peering nodes S ($|S| = m$). Now traffic is load balanced on the hubs that survived failure. The capacity of each network, as well as the interconnection capacity, grow by a factor of $\frac{m}{m-l}$. Under the stronger assumption that R^p is no larger than the post-failure rate of each network, i.e., $R^p \leq min_{x \in X}\{r_x(n_x - l)\}$, by arguments similar to those presented in the proof of Lemma 2 one can show that such peering (interconnection) capacity is necessary to support failures of up to l nodes (Lemma 22 in the Appendix). Moreover, if all networks have the same number of nodes n, and each of the n locations is shared by all networks, each network could run the l-tolerant VLB network which has optimal network capacity (by Corollary 6).

2) Peering with Pair-Wise Shared Locations: Above we observed that if there is at least one interconnection location universally shared by all the networks, no extra capacity is needed to support the peering traffic. We next consider the other extreme, in which each location is shared by at most two networks. We leave the intermediate cases for future research.

Definition 19: Assume that any location is shared by at most two networks. The *pair-wise peering VLB network* is an internetwork in which each of the q networks runs a $|S_x|$-hubs VLB networks with hubs on the nodes of $S_x = \cup_{y \neq x} S_{xy}$, the set of locations that x share with other networks. Networks x any y peer at the set of common locations S_{xy} and S_{yx}, respectively.

Routing: After load balancing all traffic on the hubs, all peering traffic is sent to the nodes that are peering with the destination network. The traffic is then handed off to the corresponding peering nodes, load balanced on the hubs of the destination network, and sent to the final destination. Formally, stream λ_{ij} from node i in network x to node j in network y is routed as follows:

1) The stream λ_{ij} is load balanced on the hubs of network x. If $x = y$ go to step 5.
2) The stream is forwarded from the hubs of x to S_{xy}, the peering nodes with y.
3) It is handed off to network y, from S_{xy} to S_{yx}.
4) It is load balanced from S_{yx} on all hubs of network y.
5) It is delivered from the hubs of y to the final destination j.

Capacities: In network x, a capacity of $2 \cdot r_x \cdot (n_x - 1)$ is sufficient to support stages 1 and 5. In order to support stage 2 we add capacity of $R^p/(|S_{xy}| \cdot |S_x|)$ from each node $i \in S_{xy}$ to each node $j \in S_x \setminus S_{xy}$. To support stage 4 we add capacity of $R^p/(|S_{yx}| \cdot |S_y|)$ from each node $i \in S_y \setminus S_{yx}$ to each node $j \in S_{yx}$. Each network x has a capacity of $2r_x(n_x - 1) + 2R^p(q - 2)$, and the interconnection capacity is of size $R^p \cdot q(q-1)$.

While we cannot prove that this peering scheme yields optimal capacity in each network, we can prove that it is within a constant factor of the optimal. This constant is independent of q, the number of networks that are peering.

Theorem 20: The "pair-wise peering VLB network" can support any legal interconnection traffic matrix by peering. Additionally, any network that can support any legal interconnection traffic matrix by peering, has at least $1/5$ of the capacity in each network and at least the same interconnection capacity.

Proof: Clearly the routing scheme can route any legal interconnection traffic matrix, and the capacities are sufficient for this routing scheme. We only need to verify that the capacity of each network x is indeed $2r_x(n_x - 1) + 2R^p(q - 2)$.

[5]In the interest of space we only present the basic ideas and observations, details can be found in the extended version of the paper [13].

The proof for this is very similar to the proof of the transit network capacity of Theorem 15, and is omitted in the interest of space.

Interconnection capacity of $R^p \cdot q(q-1)$ is proved to be necessary for interconnection by peering in Theorem 18. We next consider the capacity in each of the networks. By Lemma 21 below, if we like to support all legal interconnection traffic matrices using peering, then it is necessary to allocate capacity of at least $max\{2r_x(n_x-1), R^p(q-2)/2\}$ in each network x. The capacity of network x in the "pair-wise peering VLB network" is $2r_x(n_x-1) + 2R^p(q-2)$. As $4 \cdot max\{2r_x(n_x-1), R^p(q-2)/2\} \geq 2 \cdot R^p(q-2)$:

$$5 \cdot max\{2r_x(n_x-1), R^p(q-2)/2\} \geq 2r_x(n_x-1) + 2R^p(q-2)$$

which implies that in any interconnection network by peering, each network has at least $1/5$ of the capacity of network x in the "pair-wise peering VLB network". ∎

Lemma 21: Given $q \geq 2$ networks such that no more than 2 networks share a location. If we like to support all legal interconnection traffic matrices using peering then it is necessary to allocate capacity of at least $max\{2 \cdot r_x \cdot (n_x-1), R^p \cdot (q-2)/2\}$ in each network x.

Proof: By Observation 8, a capacity of $2r_x(n_x-1)$ is necessary in network x to support any *local* legal traffic matrix.

If there exists network $y \neq x$ such that the peering nodes of x with y can generate at least half of R^p ($r_x \cdot |S_{xy}| \geq R^p/2$) then for each of the $q-2$ networks $w \neq y$, a capacity of at least $r_x \cdot |S_{xy}| \geq R^p/2$ must enter S_{xw}. Thus all the capacities that enter all the nodes must sum to at least $(q-2) \cdot R^p/2$.

If for any network $y \neq x$, $r_x \cdot |S_{xy}| < R^p/2$, then for each y, the nodes that are not in S_{xy} can generate at least $R^p/2$ (as we assume that all nodes can generate R^p). Thus for each of the $q-1$ networks $y \neq x$, a capacity of at least $R^p/2$ must enter S_{xy}. Therefore, all the capacities that enter all the nodes must sum to at least $(q-1) \cdot R^p/2$.

We conclude that in any case the capacity of network x is at least $(q-2) \cdot R^p/2$. ∎

Note that if we assumed that for each network y, the nodes that are not in S_{xy} could generate a rate of R^p, then we could improve the lower bound to $max\{2 \cdot r_x \cdot (n_x-1), R^p \cdot (q-1)\}$. Using this stronger assumption, similar arguments show that the capacity of each network in the "pair-wise peering VLB network" has at most 3 times the optimal capacity. By extending the LP formulation of [11] we believe that the optimal provisioning can be calculated in polynomial time, but we leave this for future research.

Finally, we note that one might also consider peering with one hop in each network (instead of 2 hops as in the above scheme). Under the mild assumptions that all interconnection traffic cannot be generated by a single node ($R^p \geq r_x$) and the interconnection traffic is small enough ($R^p < \frac{r_x \cdot n_x}{2} - \frac{r_x \cdot (n_x-1)}{q-2}$), this scheme causes an increase in capacity (which might be significant if R^p is much smaller then the above expression). If $R^p \geq r_x$ this implies that from each node we need capacity of r_x to each set of peering nodes with at least $q-2$ networks (excluding x and possibly the network that this node share location with). Thus capacity of at least $r_x n_x (q-2)$ is necessary for this scheme, and if R^p is small enough this capacity is much larger than the capacity of the network when using 2 hops in each network.

VI. Conclusions

In this paper we have established the optimal resilience of the Valiant Load-Balancing network to node failures, and its usefulness as a building block for interconnected networks. In particular, building a transit-based interconnection network using the VLB scheme yields optimal capacity for each network as well as optimal interconnection capacity.

This work can be extended in the future by considering heterogeneous rates, edge capacity constraints, as well as heterogeneous edge cost structures (possibly by extending the LP formulation of [11]). It would also be important to consider the resilience of the design to edge failures in addition to node failures. Additionally, networks may be interconnected using a combination of transit and peering agreements. Therefore we should extend, in future work, our understanding of the possible use of the VLB scheme in such a hybrid environment.

References

[1] R. Zhang-Shen and N. McKeown, "Designing a predictable internet backbone network." in *HotNets III*, 2004.

[2] M. Kodialam, T. V. Lakshman, and S. Sengupta, "Efficient and robust routing of highly variable traffic," in *HotNets III*, 2004.

[3] ——, "Guaranteeing predictable performance to unpredictable traffic," in *43rd Annual Allerton Conference on Communication, Control, and Computing*, 2005.

[4] R. Zhang-Shen and N. McKeown, "Designing a predictable internet backbone with valiant load-balancing." in *IWQoS*, 2005, pp. 178–192.

[5] N. G. Duffield, P. Goyal, A. Greenberg, P. Mishra, K. K. Ramakrishnan, and J. E. van der Merive, "A flexible model for resource management in virtual private networks," *SIGCOMM Comput. Commun. Rev.*, vol. 29, no. 4, pp. 95–108, 1999.

[6] L. G. Valiant, "A scheme for fast parallel communication," *SIAM Journal on Computing.*, vol. 11, no. 2, pp. 350–361, 1982.

[7] M. Kodialam, T. V. Lakshman, and S. Sengupta, "Maximum throughput routing of traffic in the hose model," in *IEEE INFOCOM*, 2006.

[8] M. Kodialam, T. V. Lakshman, J. B. Orlin, and S. Sengupta, "Preconfiguring ip-over-optical networks to handle router failures and unpredictable traffic," in *IEEE INFOCOM*, 2006.

[9] I. Keslassy, C.-S. Chang, N. McKeown, and D.-S. Lee, "Optimal load-balancing," in *INFOCOM*, 2005.

[10] A. Gupta, J. M. Kleinberg, A. Kumar, R. Rastogi, and B. Yener, "Provisioning a virtual private network: a network design problem for multicommodity flow," in *ACM Symposium on Theory of Computing (STOC)*, 2001, pp. 389–398.

[11] T. Erlebach and M. Rüegg, "Optimal bandwidth reservation in hose-model vpns with multi-path routing." in *INFOCOM*, 2004.

[12] R. Ahuja, T. Magnanti, and J. Orlin, *Network Flows: Theory, Algorithms, and Applications.* Prentice Hall, 1993.

[13] M. Babaioff and J. Chuang, "On the optimality and interconnection of valiant load-balancing networks," 2006, extended version, on the authors' web sites.

Appendix

Lemma 22: Assume that $R^p \leq min_{x \in X}\{r_x(n_x-l)\}$. If two networks are peering at m nodes and support any legal interconnection traffic matrix after F failed by peering, where F is any set of nodes with $|F| \leq l$, then the peering capacity between the two networks is at least $2R^p \cdot \frac{m}{m-l}$.

Proof: In [13]. ∎