

Evaluating Network Information Models on Resource Efficiency and Application Performance in Lambda-Grids

Nut Taesombut

Department of Computer Science and Engineering
University of California, San Diego
9500 Gilman Dr.
La Jolla, CA 92093-0404
1-858-534-5486
nut@cs.ucsd.edu

Andrew A. Chien

Department of Computer Science and Engineering
University of California, San Diego
9500 Gilman Dr.
La Jolla, CA 92093-0404
1-858-822-2458
achien@ucsd.edu

ABSTRACT

A critical challenge for wide-area configurable networks is definition and widespread acceptance of Network Information Model (NIM). When a network comprises multiple domains, intelligent information sharing is required for a provider to maintain a competitive advantage and for customers to use a provider's network and make good resource selection decisions. We characterize the information that can be shared between domains and propose a spectrum of network information models. To evaluate the impact of the proposed models, we use a trace-driven simulation under a range of real providers' networks and assess how the available information affects applications' and providers' ability to utilize network resources. We find that domain topology information is crucial for achieving good resource efficiency, low application latency and network configuration cost, while domain link state information contributes to better resource utilization and system throughput. These results suggest that collaboration between service providers can provide better overall network productivity.

Categories and Subject Descriptors

C.2.3 [Computer-Communication Networks]: Network Operations – *network management*.

General Terms

Experimentation

Keywords

Lambda-Grids, configurable optical network, information model

1. INTRODUCTION

Continuing advances in optical transmission and network control plane are producing next-generation networks with dramatically lower cost-per-unit bandwidth, high-quality service and support for dynamic network provisioning. A key enabling technology is Dense Wavelength Division Multiplexing (DWDM) [1] which

allows large numbers of independent wavelengths (lambdas) to be carried over a single physical fiber and increases the aggregate throughput to several terabits per second. Recently, a large number of research efforts [2,3,4] are enabling dynamic provisioning of these lambdas. High-speed optical circuits (lambdas) can be dynamically configured to optimize application flows and provide guarantee over bounded communication latency and jitter [5]. Dynamic provisioning not only allows individual applications to obtain a "real private network" on-demand, but also enables efficient sharing of lambdas. Examples of advanced optical network facilities include OptIPuter [6], CANARIE's CA*net4 [7], GENI [8] and NLR [9].

To support the communication requirements of emerging large-scale scientific applications [5], a wide range of research in systems and applications is being pursued to develop Lambda-Grids [10]. Lambda-Grids are collections of geographically dispersed computing and storage resources that can be tightly interconnected by dedicated lambdas. Such a capability enables new, innovative applications [11,12] that compose large data collections, terabit-scale communication, and collaborative visualizations, only possible with the 10's to 100's of gigabits.

While configurable optical networks are gaining momentum, a significant challenge is definition and broad acceptance of Network Information Model (NIM). NIM provides information about network capabilities and resources (and possibly in the future, reliability, prices, etc.) to higher levels of the system and that information informs the selection and configuration process of a private network and Grid resources. Perhaps, the most critical tension is between service providers (e.g., AT&T, Sprint, and Verizon) who whilst being business competitors nonetheless share resources to promote overall network productivity. Because interworking between service providers raises numerous issues of security, trust and financial benefits, they don't share details of their internal networks [13]. In today's Internet, network providers only rely on the BGP protocol [14] as a basis for sharing interdomain reachability and route information. However, in configurable networks, information sharing is crucial for effective path computation across networks and for good resource selection decisions for distributed applications. To date, most Lambda-Grid systems have used simple NIM [2,4], consisting of general edge device connectivity speed or complete low-level control information (physical interconnection, switch ports, wavelengths, etc.). These models have clear complexity and commercial limitations, and represent only two extreme points on the spectrum of possible design.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and the copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SC07 November 10–16, 2007, Reno, Nevada, USA
(c) 2007 ACM 978-1-59593-764-3/07/0011...\$5.00

In this paper, we consider a range of approaches and tradeoffs for network information model. The primary contributions of this paper include:

- Identification of key motivations and difficult issues of information sharing for configurable optical networks that comprise multiple service providers,
- Classification and characterization of the basic types of network information that can be shared between service network providers and to Grid applications,
- Defining six network information models for configurable optical networks, and pointing out the underlying assumptions and characteristics of each, and
- Evaluation of the six information models on resource efficiency and application performance using trace-driven simulations across a range of real service providers' metropolitan, national and global networks.

Our key findings include:

- Domain topology information is crucial for good resource efficiency, low network transmission and configuration costs, while domain link state information contributes to better system throughput and utilization of lambdas.
- Full interdomain topology information alone provides little benefit over (BGP-like) connectivity information.
- When internal network information is not available, approximate domain connectivity information helps improve system throughput and application communication latency.
- An ISP network topology has strong impact on system lambda utilization and the utility of domain link state information.

Overall, these results give the first empirical data on the resource efficiencies of real service providers and the ability of an application to use these providers' networks with limited information sharing. Our results also suggest that cooperation between service providers can improve overall network efficiency and productivity.

The rest of the paper is organized as follows. In Section 2, we discuss information sharing challenges. In Section 3, we characterize the basic types of network information that can be shared and define six information models. In Section 4, we present our methodology to evaluate the proposed models. In Section 5, we present our evaluation results. Lastly, we discuss related work in Section 6, and present our summary and future work in Section 7.

2. INFORMATION SHARING CHALLENGES

While configurable optical networks provide intriguing opportunities for new application capabilities, they also present significant challenges in information sharing. Network information (including details of service providers' internal networks and their interconnection) is crucial for effective path computation across networks for distributed applications and enables efficient traffic engineering. Such traffic management allows the network resources (lambdas) to be utilized more efficiently and leads to better overall network productivity and net revenue for service providers. However, there are many reasons

why service providers are not inclined or able to share information of their internal networks:

- **Security:** Revealing sensitive details of service providers' internal networks (e.g., switch locations, core links without backup) makes them vulnerable to a range of security threats, including Denial-of-Service (DoS) attacks [15]. Because such threats potentially cause an infrastructural loss and/or service discontinuity, this information is often treated as confidential.
- **Financial benefits:** Publishing internal network information could point a way to other providers to gain competitive advantages in offering better network coverage, capacities and/or quality-of-service which can drive customers elsewhere. In a bandwidth broker model [16], exposing this information also makes network providers lose bargaining power over selling services to more profitable customers.
- **Internal network management:** By advertising detailed internal network information and letting an external entity manage path selection through their networks, service providers lose control over their own resource usage and management. This may cause their resources to be poorly utilized, thus making them unwilling to share information.
- **Interdomain routing policy enforcement:** Information hiding provides a means for service providers to enforce interdomain routing policies. Certain network providers may refuse to provide a transit service to all or a restricted set of other carriers by advertising to their neighboring peers only those routes that they use or allow [14].
- **Protocol heterogeneity:** Service providers may use diverse network management protocols (e.g., PNNI [17], OSPF with GMPLS-extension [18]) to obtain and propagate topology and resource information inside their networks. Incompatibility among these protocols may limit the nature and extent of network information that can be shared.

In brief, while detailed network information is important for effective network management, it is often unavailable due to numerous issues of security, economics and politics. This poses key challenges for intelligent information sharing that must not only enable effective path selection for Grid applications, but also preserve competitive advantages for individual providers.

3. NETWORK INFORMATION MODELS

Here, we describe a perspective and assumptions on the architecture of a configurable optical network, characterize network information, and define six information models.

3.1 Network Architecture Assumptions/ Information Categorization

A configurable optical network consists of a collection of optical switches interconnected by DWDM optical links. In such a network, a connection is created on-demand and formed by a set of optical switches which forward data along the established circuit path. In today's Internet, networks are partitioned into sub-networks which provide autonomous administrative domains for each Internet Service Provider (ISP), providing the autonomy and scalability of the Internet. Figure 1 depicts a simple example of a multi-domain, configurable optical network. The network consists of interconnected groups of optical switches managed independently by three ISPs.

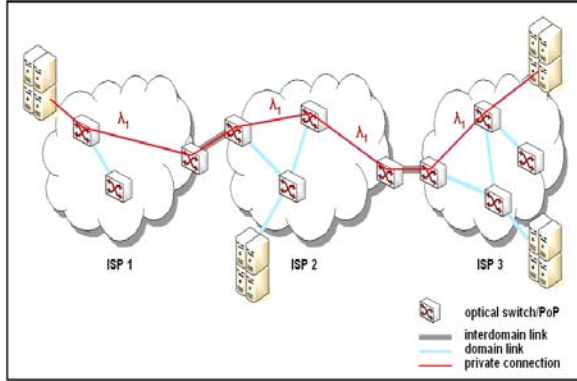


Figure 1. Physical architecture of a multi-carrier, optical circuit network.

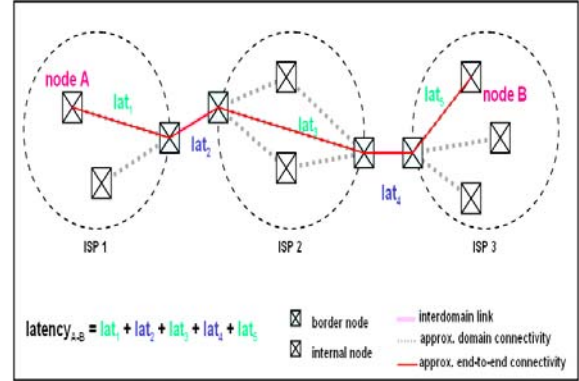


Figure 2. Approximating the latency of an end-to-end network path across domains using ConnDom

In general, we can classify network information into two main categories: *domain* and *interdomain* information.

- **Domain network information:** includes topology and link state information of a domain (i.e., each ISP’s network). *Domain topology information* specifies the interconnection between nodes and links within a domain, the latency of each link, and which end resources are attached to each node. The notion of ‘node’ can be generalized to an optical switch or Point of Presence (PoP) depending on whether the network is switch-level or PoP-level. *Domain link state information* specifies the capacity and usage of domain links (i.e., links between nodes within a domain).
- **Interdomain network information:** includes interdomain (domain-to-domain) topology and connectivity information. *Interdomain topology information* specifies the interconnection between domains, including the latency, capacity and usage of each interdomain link as well as their peering points. *Interdomain connectivity information* provides network reachability information among ISPs. It can be viewed as “distance vector” information similar to that of BGP [14] indicating at which domains and via which interdomain paths a particular domain can be reached.

3.2 Model Definition

We now present six different network information models.

1. **Open Interdomain, Open Domain (Open):** provides complete domain and interdomain network information. It assumes complete trust amongst ISPs (i.e., an open infrastructure) and allows an external agent to control the selection and configuration process of an entire network path for an application. This simple model is widely deployed in experimental Grid and advanced optical networks, such as OptiPuter [6], CANARIE’s CA*net 4 [7], and CHEETAH [4].
2. **Open Interdomain, Topology Domain (TopoDom):** includes all network information except domain link state information. The key idea here is that while ISPs can profitably share their domain topologies, they are unwilling to reveal information about internal resource capacity/usage as others can exploit it for their competitive advantages. An exemplary use of this model is an ISP with multi regional domains (e.g., AT&T US and Europe) which are operated by different business units with their own revenue targets [13].

3. **Open Interdomain, Connectivity Domain (ConnDom):** provides interdomain topology information and an approximation of domain link connectivity. The rationale here is while complete domain information cannot be shared, some abstraction of domain connectivity can be useful to enable more effective path selection [13]. In our specific approach, the model provides an approximate latency of connectivity between border nodes and between pairs of each border and internal node within a domain. Figure 2 gives an approximation of domain connectivity of the network in Figure 1. As can be seen, this approach is useful to estimate the total latency of a circuit path across domains while hiding physical domain topologies. In our implementation, we approximate this latency by computing the latency of the shortest physical path between two nodes assuming infinite lambdas on all links along the path.
4. **Topology Interdomain (TopoInter):** includes interdomain topology information. Due to numerous issues of economics and security, each ISP hides all details of its internal network. While this approach provides no domain information, it offers diverse interdomain paths.
5. **Connectivity Interdomain (ConnInter):** provides interdomain connectivity information. This approach reflects the philosophy of interdomain routing in today’s Internet which relies on the Border Gateway Protocol (BGP) [14] to disseminate interdomain reachability and route information. Being distance-vector-based, the model offers neither diverse interdomain path nor link state information.
6. **No Information (None):** provides no network information. A potential connection between two edge devices can only be inferred from their network interface card (NIC) speed.

4. METHODOLOGY

We describe our methodology for evaluating the proposed network information models. In brief, we used trace-driven simulations across a range of metropolitan, national and global networks, and used a Commercial Content Distribution (CCD) application as our workload. We chose to study CCD because it shares many aspects with “scientific data distribution and sharing” [11,12], which features large distributed collections of data objects and on-demand communication and is a dominant large-scale scientific application targeting Lambda-Grids today. In the following, we describe the details of our simulation model, resource selection strategies as well as evaluation metrics.

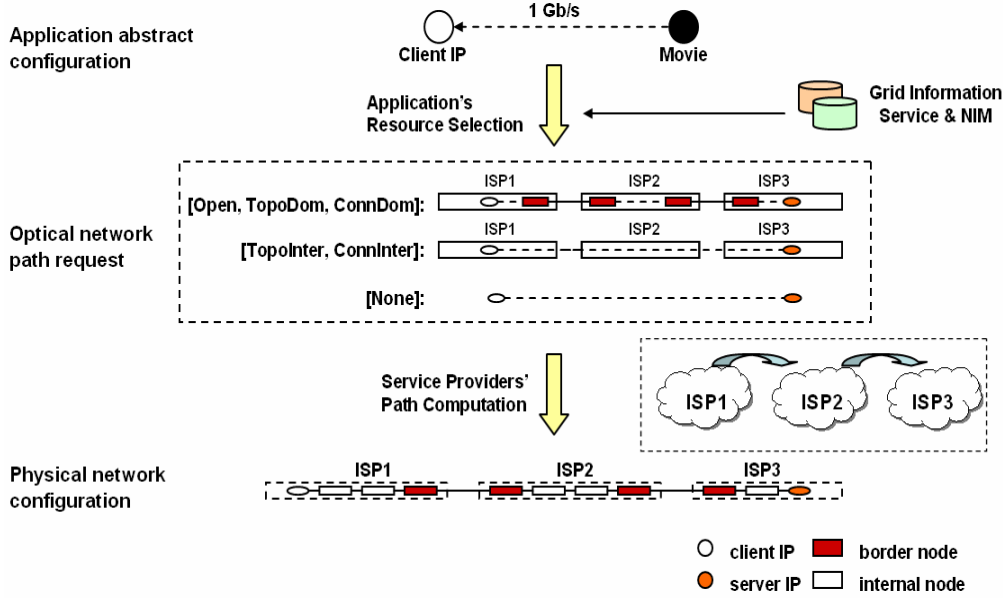


Figure 3. Resource selection and network path computation architecture for a content distribution application

4.1 Approximating ISP Fiber Topologies

To make our evaluation of NIMs broadly useful, we consider a range of realistic ISP optical network topologies. Ideally, the studied networks must be diverse in size, network design, complexity and geographical presence so that we can explore the impact of these factors on the utility of different information models. Unfortunately, due to numerous issues of security and economics, ISPs often regard their physical fiber topologies as confidential. To evaluate the proposed NIMs, our strategy is to utilize real ISP fiber network topologies wherever possible and also use ISP PoP-level network topologies for approximating their physical fiber maps.

Our metropolitan network topology models were derived from AboveNet’s metro-area fiber maps [19]. While a few ISPs publish information about their metropolitan networks, AboveNet provides the most comprehensive network maps. We chose to use eight of the AboveNet metro-area networks (all in major cities) and decoded them manually from the published fiber map images. The details of these networks are summarized in Table A.1.

Our wide-area network topology models were derived from Rocketfuel’s router-level, ISP backbone network map collection [20, 21]. These maps were originally extracted from the “traceroute” data generated by 300 traceroute web servers across the world. We carefully selected eight ISP network maps – AT&T, Ebone, Exodus, Level3, Sprint, Telstra, Tiscali and Verio, which are large and diverse enough for meaningful study. Using these ISPs, we reduced their router-level topologies to PoP-level (city-level) topologies. Specifically, we grouped routers by their geographical locations which were inferred from their DNS names [22]. This reduction simplifies our analysis while preserving its validity because ISP’s traffic engineering decisions are usually made at the PoP-level [23, 24]. The details of these ISP networks are summarized in Table A.2.

To study the impact of various inter-domain factors, we used a realistic map of the multi-carrier, Internet backbone network

consisting of most large ISPs – AT&T, BT, Cogent, Global Crossing, Level3, NTT/Verio, Qwest, Sprint, Time Warner and Verizon. To obtain such network map, we first derived the current network map of individual ISPs from their company websites and then inferred their peering points from the traceroute data [20]. To reduce the number of traces to look up, we used the AS-peering relationships information published by CAIDA [25, 26], to select only those traceroutes that traverse pairs of the peering ISPs. The details of the derived multi-ISP backbone network are given in Table A.3. While we consider only a small number of ISPs, these ten ISPs are dominant network service providers in the world (9 tier-1 ISPs and 1 high-degree tier-2 ISP [27]) and altogether account for a large fraction of today’s optical fiber infrastructures. Note that, here, we didn’t use Rocketfuel’s ISP network topologies (which we used to study intra-domain factors above) because only five of these ISPs are directly peered and not sufficient for meaningful study.

Once the network topologies were derived, we assigned lambdas and latency for each link in these networks. For each link we assigned 20 lambdas, each at 1 Gb/s. To obtain the latency of a link between two PoPs, we first determined the latitude and longitude of their geographical presence, calculated their distance using the great circle method [28], and lastly computed the latency using this distance divided by the speed of light. Using this method, we made the assumption that all links are laid along the shortest path between two cities (PoPs).

4.2 Generating Grid Resources

Using a statistical Grid resource generator [29], we generated end resources (cluster and host information) with the distribution matching that of the currently deployed Grid infrastructures, such as TeraGrid [30] and iVDGL [31]. These resources were given unique IP addresses and randomly assigned to PoPs of each network topology model in Section 4.1. Each PoP consists of 270 end resources on average, and each resource has a 10 Gbps uplink to the core network.

4.3 Content Distribution Applications

To evaluate the proposed NIMs, our workloads are synthetic traces of movie content delivery requests. Each request is as shown in the application abstract configuration model on the top of Figure 3, which specifies a client’s IP address (chosen randomly), a requested movie, and a private network path (1 Gbps). We assume each of the derived networks in Section 4.1 contains a set of replica servers, each maintaining a collection of movie contents. For each request, the goal is to find the server with the movie replica closest to the client. The notion of “closest” is determined by the minimum lambda distance (or latency) between the client and the chosen server. If the request cannot be satisfied (e.g., no available optical path), it will be placed in the system queue and re-evaluated the next time some lambdas in the system are released.

The outcome of resource selection for each application request above is an optical path request to network service providers. As shown in the middle of Figure 3, resource selection using different NIMs results in three types of an optical path request. For all NIMs, the optical path request specifies the IP addresses of the client and chosen server together with a ‘loose’ network path between them. For TopoInter and ConnInter, the loose network path is an interdomain path, a result of path computation and selection using interdomain topology and connectivity information. For Open, TopoDom and ConnDom, this interdomain path is also decorated with information about which border nodes to be used to connect between providers’ networks. Such border nodes are derived from the chosen end-to-end network path from resource selection using domain network information provided by Open, TopoDom and ConnDom. For None, the loose network path is merely an abstract link and has no route information. Subsequently, the optical path request is given to the corresponding ISPs in an order as specified in the chosen path. As the request crosses different ISPs, a representative entity of each ISP uses its full internal network information to compute a specific intra-domain path through its network. Given these intra-domain paths, a final end-to-end physical network configuration is derived.

For each network model, the replica servers were randomly chosen from its end resource pool. In order to compare the results across network topologies, we used the same ratio of servers to PoPs; the number of servers is four times that of PoPs. Each server has 2 TBytes of disk space. We generated 500 movie objects for each metro-area network model and 5000 objects for each of the rest. We assume these movie contents are of 2K Digital Cinema resolution (up to 2160x1080) with a stream rate of 250 Mbit/s [32]. The average size of these movies is 200 GB which runs for approximately 1 hour and 50 minutes. Our decisions for replicating movie objects to replica servers are based on the popularity replication heuristic algorithm [33]. The popularity of these movie objects follows a Zipf-like distribution. Using this distribution, each replica server picked and stored as many objects as its storage constraint allows. In order to observe the system under different loads, we scaled the inter-arrival time between subsequent requests (or conversely the request rate). For each request rate and network model, we used 15 traces, each with 50,000 applications. This is high enough to ensure that our metrics are measured over simulations that spend greater than 90 percent of their time in a steady state.

4.4 Resource Selection & Network Path Computation

Here, we describe the different resource selection algorithms (applied in the case of different network information models) to select a replica server with the content replica requested by a client.

- **Open:** the algorithm combines all interdomain and domain information to construct complete, flat network information (including all nodes, interdomain and domain links, and available capacity on each link). It uses this information to compute the minimum-latency path (with sufficient capacity) between the requesting client to each replica server candidate. It selects the candidate with the shortest path from the client.
- **TopoDom:** the algorithm combines interdomain and domain topology information to construct flat network topology information, and assumes infinite capacity on each domain link. It uses this information to compute the minimum-latency path between the client and each candidate and chooses the candidate with the shortest path.
- **ConnDom:** for each candidate, the algorithm uses interdomain topology and approximate domain connectivity information to approximate the minimum-latency path to the client (see an example in Figure 2). During this path computation, it assumes infinite capacity on all links inside a domain. It chooses the candidate with the shortest approximate path.
- **TopoInter:** for each server candidate, the algorithm uses interdomain topology information to compute the path with the minimum number of transit domains to the client. It chooses the candidate with the minimum hop count.
- **ConnInter:** for each candidate, the algorithm uses the provided interdomain path in the interdomain connectivity information to determine the number of transit domains to the client. It chooses the candidate with the minimum hop count.
- **None:** the algorithm assumes all candidates have an equal network transmission cost and randomly selects one candidate.

In our simulations, we determined all server candidates with the requested movie replica using information provided by Grid Information Service (GIS). To compute the minimum-latency path (or minimum-domain-hop-count path), we employed Dijkstra’s shortest path algorithm.

Given an optical path request (in the second step in Figure 3), each ISP computes an intra-domain path through its network. To compute a minimum-latency, intra-domain path, we used all domain topology and link state information and employed Dijkstra’s shortest path algorithm. It should be noted that a pair of ISPs may have multiple peering points. Unless specific border nodes are provided, we implemented the “early exit” peering policy for an upstream ISP to select an intra-domain path to a downstream ISP. Specifically, the upstream ISP uses the peering point closest to the source (or ingress border node) as destination for path computation and selection. In [24], Spring et al. discovered that “early exit” is the most common policy accounting for 20-30% of all ISP pairs in the Internet.

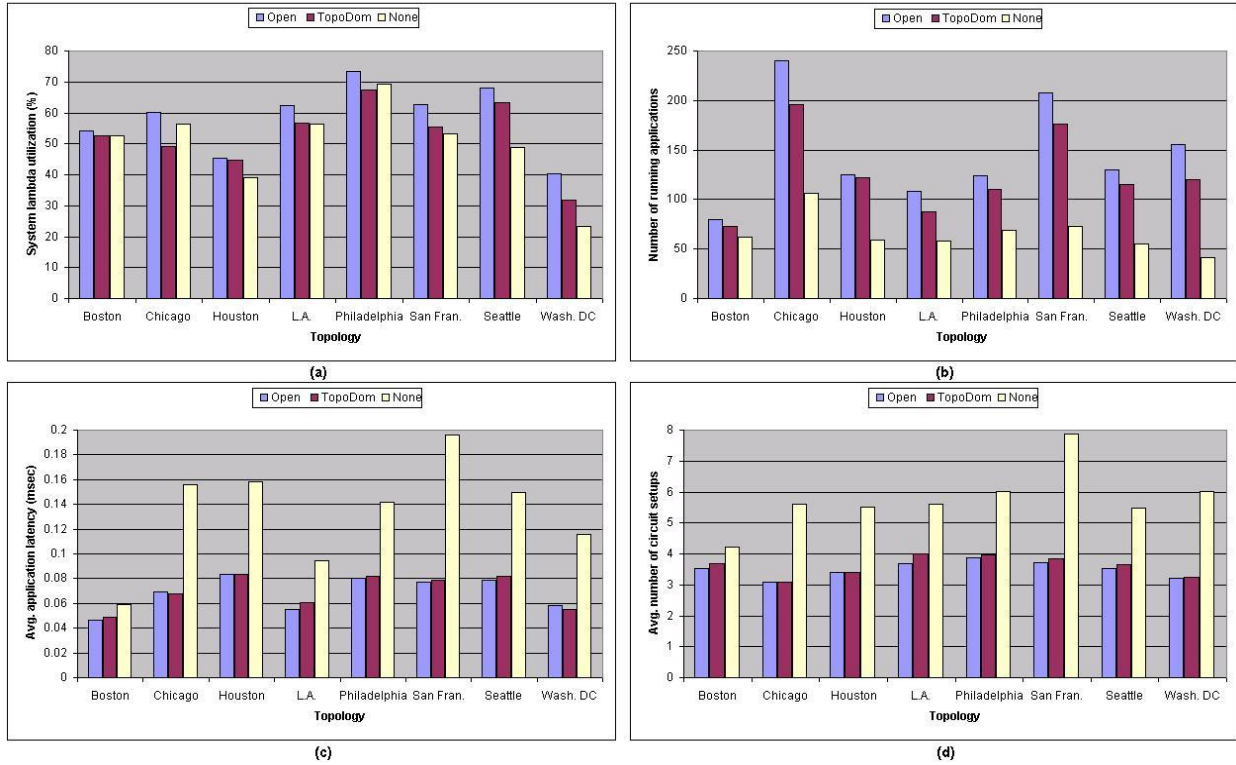


Figure 4. Evaluating the intra-domain impact of network information models using AboveNet metro-area networks: a) system lambda utilization; b) system throughput; c) average application latency; and d) network configuration cost

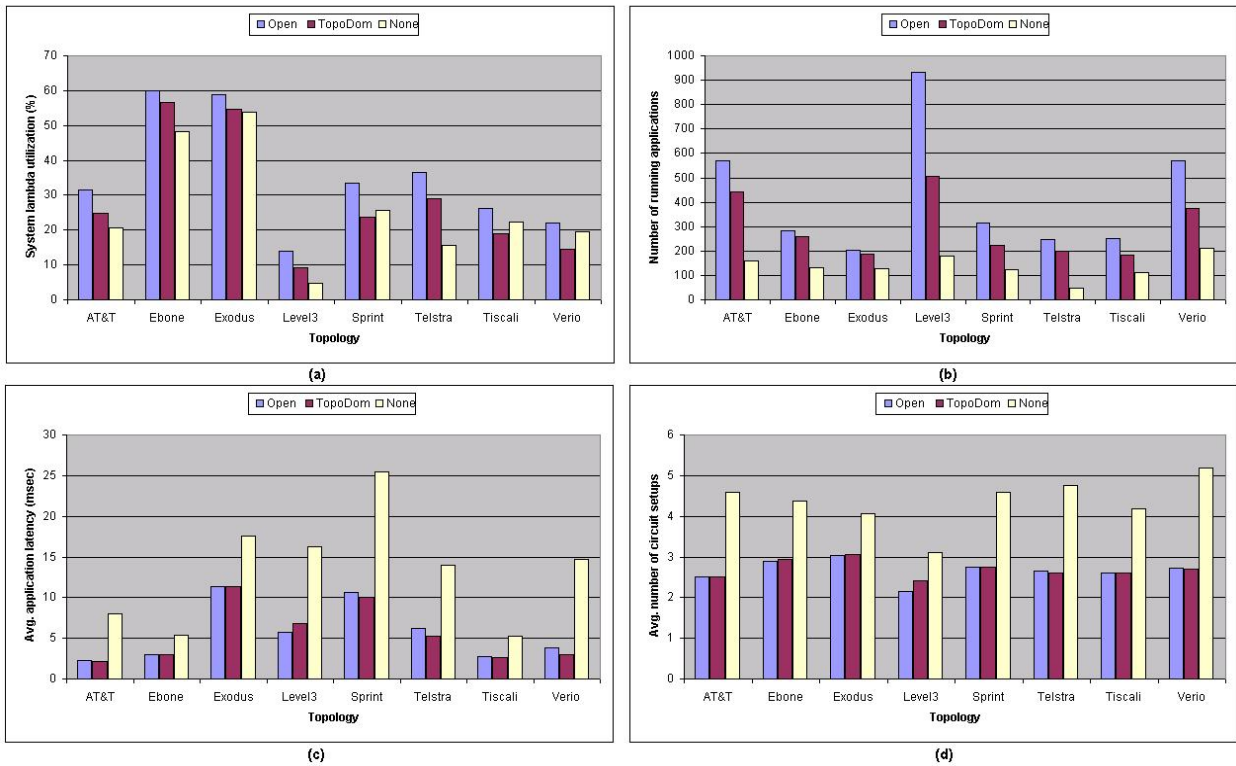


Figure 5. Evaluating the intra-domain impact of network information models using ISP backbone networks: a) system lambda utilization; b) system throughput; c) average application latency; and d) network configuration cost

4.5 Evaluation Metrics

To evaluate the proposed information models, we use the following four metrics:

1. **System lambda utilization:** the average fraction of available lambdas in the system that are allocated for use.
2. **System throughput:** the average number of running applications in the system
3. **Application latency:** the average lambda distance (or latency) of the network path allocated for each application
4. **Network setup cost:** the average number of optical circuit setups that must be configured for each application

It should be noted that high system lambda utilization alone doesn't necessarily mean the network resources are efficiently utilized. Good network efficiency should be determined altogether by two metrics: high system lambda utilization and high system throughput.

5. EVALUATION

5.1 Intra-domain Factors

We now evaluate the impact of various intra-domain factors on the usefulness of the proposed network information models across a range of ISP metropolitan, national and global networks. Here, we only consider the three models (Open, TopoDom and None) to investigate the utility of domain topology and link state information.

5.1.1 Metropolitan Network

We used eight AduveNet metropolitan networks (see Table A.1) to evaluate and compare the three models. The results below were reported using the request rate of 10 requests/min which is high enough for all metrics to be measured at their saturation regions.

Figure 4(a) compares the average system lambda utilization of Open, TopoDom and None. For all cases Open achieves the highest lambda utilization. This is because Open provides full domain information, allowing us to exploit diverse domain paths and always select solutions with lowest application latency. As a result, Open makes more efficient use of lambdas and can admit more applications into the system on average. On the other hand, we see no clear superiority between TopoDom and None. As explained below, while TopoDom produces higher system throughput, it allocates fewer lambdas per applications.

As shown in Figure 4(b), for all topologies Open's system throughput is higher than that of TopoDom, and both drastically outperform None. These results imply that domain topology information is crucial for achieving good system throughput, while link state information also has positive impact. Without domain topology information, many solutions with long-latency path are chosen, but they cannot be realized due to their high demand of lambdas. We also find that the size and topological structure of a metropolitan network have impact on the advantage of Open over TopoDom. The superiority of Open becomes more evident in larger and denser networks such as Chicago and Washington DC. This is because these networks offer more diverse domain paths, and link state information in Open allows us to exploit these paths when the network becomes congested.

Figure 4(c) and 4(d) illustrate the average application communication latency and network setup cost for the three

models. We see the comparable results between Open and TopoDom, while they both achieve much lower application latency and network setup cost than None. This demonstrates the utility of domain topology information for these performance dimensions, while link state information has little impact. Domain topology information is needed to compute the latency of a network path which is essential for comparing the quality of solutions and making good path selection decisions.

5.1.2 ISP Backbone Network

We now analyze the utility of the three information models (Open, TopoDom and None) using real ISP backbone networks (see Table A.2). In the results presented below, we used the request rate of 40 requests/min which is high enough for all metrics to be measured at their saturation regions.

The chart in Figure 5(a) illustrates the average system lambda utilization for the three models. We find that Open always achieves the highest system lambda utilization. Depending on ISPs, Open improves the lambda utilization over TopoDom and None by 3.1-9.8 percent and 2.8-21.0 percent, respectively. This is because Open provides complete network information which allows computation of diverse domain paths and enables the system to admit more applications during high resource contention.

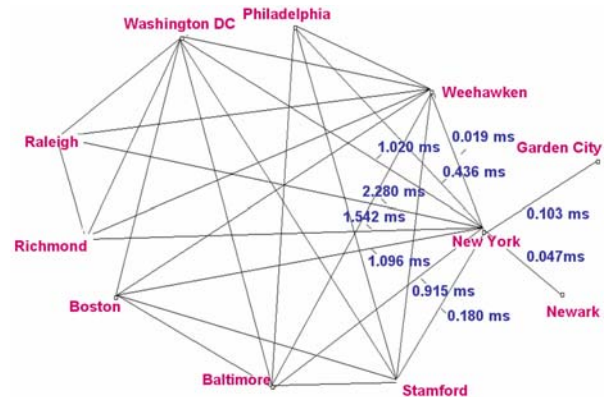


Figure 6. Fiber map of Level3 backbone network in the Northeastern US.

Further, our simulations show a strong influence of network design on system lambda utilization. The most common network design among the studied ISPs is "hub-and-spoke". These ISPs, including AT&T, Telstra, Tiscali and Verio, have hubs in major cities and spokes that fan out connections to smaller cities. For such ISPs, the bottlenecks are the links connecting major hubs and we observe their system lambda utilization using Open in the range of 22.2 to 36.5 percent. On the other hand, Level3 includes PoPs in major cities in US and Europe. While these PoPs are highly connected, the bottlenecks are the links crossing the two continents. Because our workloads contain a fair number of requests to these links (applications and data replica are in different continents), it causes other links to be relatively underutilized and thus leads to low system lambda utilization even with Open (13.9 percent). Exodus and Ebone represent another network design paradigm where their network links are more uniformly distributed among nodes. These two networks have no true bottleneck link, and we can achieve higher system utilization for all the three models (48-60 percent).

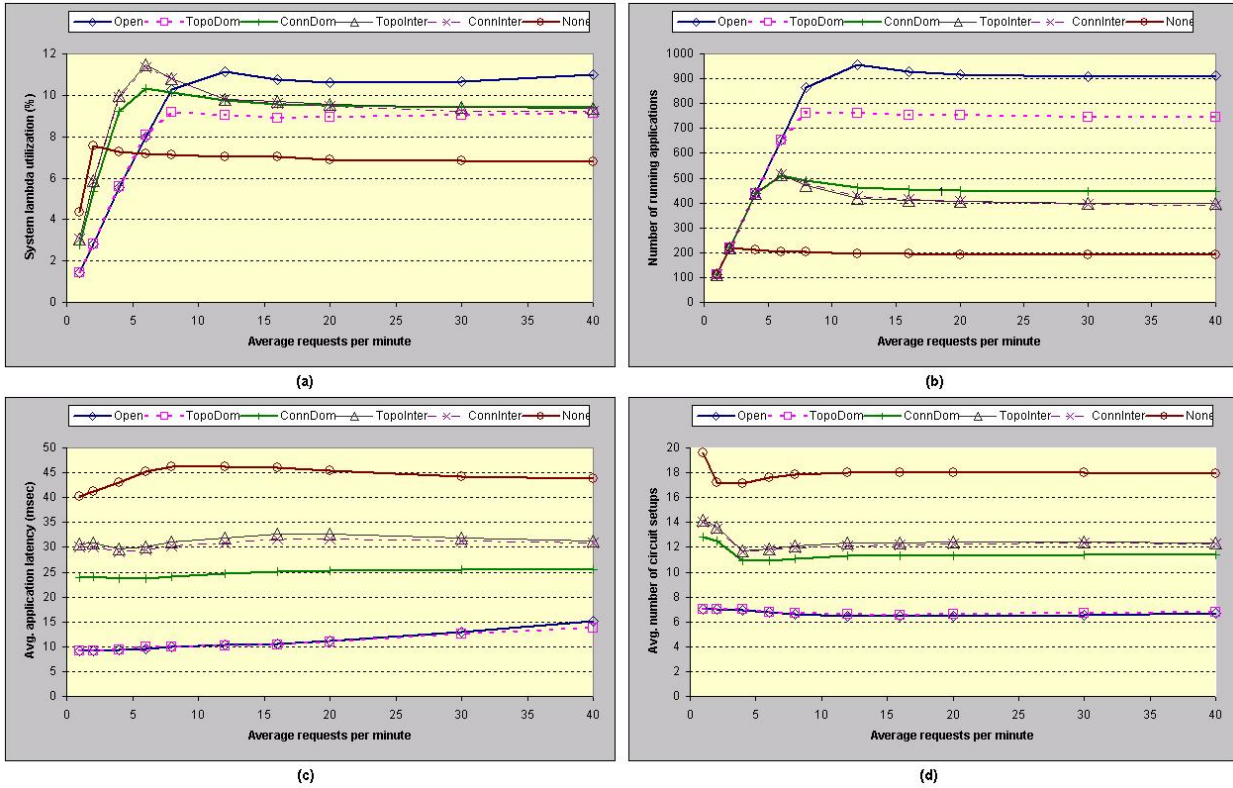


Figure 7. Evaluating the impact of network information models using a multi-domain network with top-tier ISPs: a) system lambda utilization; b) system throughput; c) average application latency; and d) network configuration cost

As shown in Figure 5(b), for all ISPs Open achieves higher system throughput than TopoDom and they both outperform None. These results confirm our findings above that both domain topology and link state information contribute to better system throughput. Depending on the studied ISPs, we see varying degrees of advantage of Open over TopoDom. This implies the impact of network topology of an ISP on the utility of link state information. Among the studied ISPs, we see its most benefit using Level3. As an example, Figure 6 illustrates the portion of the Level3 network in the Northeastern USA. While most cities are highly-connected, Newark and Graden City each has one link to New York. Because these two links have one of the lower latency (0.047 and 0.103 ms), they are highly utilized (often picked by our selector). At high load when these links become congested, if the resource selector doesn't know about the link usage, it will continue to choose solutions that include these links and fail. With link state information, it can avoid these congested links and select other feasible candidates. While observing similar effects across ISPs, we find its most impact on Level3 because their PoPs are highly-connected and hence a large fraction of their links are blocked from being utilized.

Figure 5(c) and 5(d) compares the average application latency and network setup cost for the three models. We see the comparable results of Open and TopoDom, while both achieve much lower application latency and network setup cost than None. These results confirm our findings above that while domain topology information is crucial for achieving lower application latency and network setup cost, link state information has minimal impact on

these dimensions of performance. This is because domain topology information allows us to always use a shortest available path.

5.2 Interdomain Factors

We now analyze the impact of various inter-domain factors on the usefulness of network information models using the real multi-ISP, global network (see Table A.3). We consider all the six network information models (Open, TopoDom, ConnDom, TopoInter, ConnInter and None) to investigate the utility of different interdomain and domain network information. In order to observe the system under different loads, we present the results with varying application request rates.

Figure 7(a) compares average system utilization for the six information models as a function of a load. At low system load, a model with better network efficiency can be observed by its slower growth rate of the utilization with higher load. We see the growth rate of Open and TopoDom is slower than that of ConnDom, TopoInter and ConnInter, whereas the growth rate of all these models is slower than that of None. These results show that the two significant factors for high resource efficiency are interdomain connectivity and domain topology information. Interdomain connectivity information allows selection of network paths with the minimum interdomain hop count, while domain topology information permits computation of a shortest path through or within an ISP's network.

We also see that as the request rate continues to increase and the network resources become congested, the growth rate of each

model moves from linear increase to a flattened saturation region. Open’s saturation region is the highest at ~10.7 percent, that of None is the lowest at ~6.8 percent, and the rest has the saturation region at ~9.1-9.2 percent. These results show the benefit of domain link state information on improving overall system lambda utilization. At high system load, this information allows us to avoid congested links and take advantage of diverse paths.

The chart in Figure 7(b) compares system throughput for the proposed models. At low load, for all models the average number of running applications increases linearly with higher load and at approximately the same rate. This is because without resource contention almost all new application requests can be satisfied and admitted. We find that at high load (>20 requests/min) the system throughput for each model reaches a saturation point. The saturation points for Open, TopoDom, ConnDom, TopoInter, ConnInter and None are at 910, 746, 446, 396, 391 and 190 applications, respectively. These results show that while interdomain topology and connectivity information is useful, domain topology and link state information has greater impact on improving system throughput. Closer investigation reveals that because the studied network contains a small number of large ISPs, we see more influence of domain information over interdomain information. Further, since most pairs of ISPs are directly peered in the network, we see little benefit of interdomain topology information over connectivity information. While topology information offers diverse interdomain paths, these paths are harder to allocate because they span multiple large networks. Lastly, we see some advantage of ConnDom over TopoInter, implying the utility of approximate domain connectivity information in ConnDom.

As shown in Figure 7(c), we observe the differences in average application latency achieved by different network information models. We find that Open and TopoDom produce the lowest and comparable application latency. This confirms our findings in Section 5.1 that while domain topology information is essential for low application latency, link state information has little impact on it. We also find that TopoInter’s and ConnInter’s average application latency is much lower than that of None. This implies the usefulness of interdomain topology and connectivity information. However, the latency of TopoInter is slightly higher than that of ConnInter. This is because ConnInter limits us to use only the shortest interdomain path (given in the interdomain connectivity information) which leads to lower application latency on average. Lastly, in terms of average application latency we see some benefit of approximate domain connectivity information in ConnDom. This information allows estimation of the latency of a network path across domains, which is useful for comparing the quality of solution candidates and making good path selection decisions.

Figure 7(d) compares the average number of circuit setups for the six information models as a function of a load. For all load, both Open and TopoDom require the lowest network setup cost, while ConnDom, TopoInter and ConnInter all produce a higher and comparable cost. These results imply the need for domain topology information to achieve low network setup cost, while interdomain network information also has some positive impact.

6. RELATED WORK

To the best of our knowledge, our work is the first to address both the definition and evaluation of network information models

(NIMs) for configurable optical networks. Our work share some similarities with [13] in that both point out the information sharing problems and characterize network information that can be shared. However, they don’t define any NIM and only discuss several use cases of network information at high level. In contrast, we present the first empirical data comparing NIMs using a wide range of realistic ISP networks.

There is a body of work on NIM for Grids. The Network Measurement Working Group (NM-WG) of the Global Grid Forum (GGF) focus on the formal classification of network characteristics [34], but they don’t provide any model for sharing private information, a critical tension likely to occur between commercial providers. Their goal is to address the portability issues between network measurements across Grid sites, in contrast to our work that addresses the information sharing problems.

Tools, such as Meridian [35], TopoMon [36] and iPlane [37], are measurement infrastructures estimating network performance characteristics for distributed applications. Furthermore, several research efforts have investigated a wide range of inference techniques [21, 24, 38] to approximate ISP and Internet topologies at different granularities, such as router-level and AS-level. However, they don’t solve the information sharing problems because physical optical network infrastructures and configurable resources cannot be inferred from measurements.

Study of lambda scheduling and resource allocation problems for configurable optical networks has received considerable attentions [39,40]. While these efforts present transfer-level simulation studies similar to ours, they focus on the algorithm design and assume a simple network information model with complete low-level information. In contrast, we also take into account different NIMs in network path selection and scheduling. Our previous work [5] explores several models of use for configuration networks which focus on intelligent network, file transfer and system abstractions. This work built on the system abstraction model (called Distributed Virtual Computer [41]) where an application explicitly describes and acquires a set of Grid resources and private network on-demand. However, our previous work focuses on the comparison of service models and assumes complete knowledge of network information. This is in contrast to this work that evaluates the impact of different NIMs on network resource efficiencies and application capabilities.

Table 1. Summary of usefulness of different network information on the studied metrics

Network information	Usefulness on the metric			
	System utilization	System throughput	Average application latency	Average network setup cost
Interdomain connectivity	Medium	Medium	Medium	Medium
Interdomain topology	No	Minimal	No	No
Appox. domain connectivity	Low	Low	Medium	Low
Domain topology	-	High	High	High
Domain link State	Medium	Medium	Minimal	Minimal

7. SUMMARY AND FUTURE WORK

Emerging configurable optical networks provide superior network service at the cost of more heavy-weight, user-controlled resource management. A key challenge is intelligent information sharing that not only enables effective resource selection for Grid applications, but also maintains competitive advantages of individual service providers.

In this paper, we defined and evaluated six network information models (Open, TopoDom, ConnDom, TopoInter, ConnInter and None) with varying degrees of abstraction in terms of domain-level and interdomain-level information that they provide. Table 1 summarizes the usefulness of different network information as determined by its improvement over the previous information factor for a given metric. Among the studied models, Open and TopoDom produce the lowest application latency and network setup cost because they provide domain topology information which is a key for efficient path selection. Further, Open always leads to the highest system utilization and throughput, a result of it offering domain link state information which is essential for identifying congested links and computing diverse paths. When domain information cannot be shared, we find that an approximation of domain connectivity is useful, allowing the latency of network paths to be more accurately estimated and producing better system throughput and application latency. In addition, we find that interdomain topology information provides minimal improvement (sometimes negative impact) over connectivity information. This is because the studied Internet backbone network has a small number of large ISPs and the efficiency of path selection is highly influenced by intra-domain factors. We expect the benefits of interdomain topology information to be more apparent in the network with a larger number of smaller ISPs.

Further, we find that network topology of an ISP does impact system lambda utilization and throughput. Level3 achieves the lowest utilization due to its highly connected structure and the bottleneck links connecting cities on different continents, while Exodus and Ebone produce the highest utilization as their links are more uniformly distributed among cities. Additionally, we find that domain link state information have greater benefit on larger and denser networks as it allows us to exploit available lambdas on diverse paths to improve system throughput.

In summary, our key contributions are two-folded: 1) our results demonstrate the impact of limited network information on the ability of an application to use real ISPs' networks; and 2) these results encourage cooperation between ISPs to share their network information for better overall network efficiency and productivity. We believe that such collaboration is possible with a trustworthy, third-party external agent such as the DVC [41], that manages resource configuration planning for applications, while maintaining the confidentiality of ISPs' sensitive information.

To evaluate the proposed network information models, this paper used a Commercial Content Distribution (CCD) application as workloads. However, for more comprehensive evaluation, we are currently extending our simulations to include a broader range of realistic application models. These include high-performance distributed computing (HPDC) and collaborative data visualization applications; both of which are among the most popular applications targeted for Lambda-Grids today. It would

be interesting to see how different workloads affect the utility of different types of network information.

Furthermore, the current simulations use only simple (greedy) resource selection algorithms to evaluate all the network information models. While this evaluation approach is fair for comparison between different models, there are opportunities to improve resource efficiency and application performance by designing specialized selection algorithms for individual information models. We are exploring and evaluating new resource selection strategies and will present the results in our future work.

8. ACKNOWLEDGEMENTS

Supported in part by the National Science Foundation under awards NSF EIA-99-75020 Grads and NSF Cooperative Agreement ANI-0225642 (OptIPuter), NSF CCR-0331645 (VGrADS), NSF NGS-0305390, and NSF Research Infrastructure Grant EIA-0303622. Support from Hewlett-Packard, BigBangwidth, Microsoft, and Intel is also gratefully acknowledged.

9. REFERENCES

- [1] Laude, J. DWDM Fundamental, Components and Applications. Artech House, January 2002.
- [2] Yu, O., Li, A., et al. Multi-domain lambda grid data portal for collaborative grid applications. *Journal of Future Generation Computer Systems*, Vol. 22(8), October 2006.
- [3] Lehman, T., Sobieski, J., and Jabbari, B. DRAGON: A framework for service provisioning in heterogeneous grid networks. *IEEE Communications Magazine*, March 2006.
- [4] Verraraghavan, M., Zheng, X., et al. CHEETAH: circuit-switched high-speed end-to-end transport architecture. In *Proceedings of the 4th Annual Optical Networking and Communications Conference (OptiComm'03)*, October 2003.
- [5] Taesombut, N., Uyeda, F., et al. The OptIPuter : high-performance QoS-guaranteed network service for emerging e-science applications. *IEEE Communication Magazine*, May 2006.
- [6] Smarr, L. L., Chien, A. A., et al. The OptIPuter. *Communication of the ACM*, Vol. 46(11), November 2003.
- [7] CANARIE Inc
<http://www.canarie.ca>
- [8] GENI: Global Environment for Network Investigations
<http://www.geni.net>
- [9] National LambdaRail
<http://www.nlr.net>
- [10] DeFanti, T., Latt, C. D., et al. TransLight: a global-scale LambdaGrid for e-science. *Communications of the ACM*, Vol. 46(11), November 2003.
- [11] Astakhov, V., Gupta, A., et al. Data integration in the Biomedical Informatics Research (BIRN). In *Proceedings of the 2nd International Workshop on Data Integration in Life Sciences (DILS'05)*, July 2005.

- [12] Newman, H. B., Ellisman, M. H., and Orcutt, J. A. Data-intensive e-science frontier research. *Communications of the ACM*, Vol. 46(11), November 2003.
- [13] Bernstein, G. M., Sharma, V., and Ong, L. Interdomain optical routing. *Journal of Optical Networking*, Vol. 1(2), February 2002.
- [14] Rekhter, Y., and Gross, P. *Application of the Border Gateway Protocol in the Internet*. RFC 1772, T. J. Watson Research Center, IBM Corporation, MCI, March 1995.
- [15] Mirkovic, J., Dietrich, S., et al. *Internet Denial of Service: Attack and Defense Mechanisms*. Prentice Hall, 2005.
- [16] Zhang, Z. L., Duan, Z., and Hou, Y. T. On scalable network resource management using bandwidth brokers. In *Proceedings of the 8th Network Operations and Management Symposium (NOMS'02)*, April 2002.
- [17] The ATM Forum Technical Committee. *Private Network-Network Specification Interface v.1 (PNNI 1.0)*. af-pnni-0055.000, March 1996.
- [18] Liu, H., Pendarakis, D., et al. GMPLS-based control plane for optical networks: early implementation experience. In *Proceedings of the SPIE International Conference and Exhibits on the Convergence of IT and Communications (ITCom'02)*, July 2002.
- [19] AboveNet IP and Fiber Maps
<http://www.above.net/products/maps2/index.html>
- [20] Rocketfuel: An ISP Topology Mapping Engine
<http://cs.washington.edu/research/networking/rocketfuel>
- [21] Spring, N., Mahajan, R., and Wetherall, D. Measuring ISP topologies with Rocketfuel. In *Proceedings of the ACM SIGCOMM 2002 Conference*, August 2002.
- [22] Padmanabhan, V. N., and Subramanian, L. An investigation of geographic mapping techniques for Internet hosts. In *Proceedings of the ACM SIGCOMM 2001 Conference*, August 2001.
- [23] Taft, N., Bhattacharyya, S., et al. Understanding traffic dynamics at a backbone PoP. In *Proceedings of the SPIE ITCOM Workshop on Scalability and Traffic Control in IP Networks*, 2001.
- [24] Spring, N., Mahajan, R., and Anderson, T. Quantifying the causes of path inflation. In *Proceedings of the ACM SIGCOMM 2003 Conference*, August 2003.
- [25] CAIDA's Inferred AS Relationships Dataset
<http://as-rank.caida.org/data>
- [26] Dimitropoulos, X., Krioukov, D., et al. AS relationships: inference and validation. In *Proceedings of the ACM SIGCOMM Computer Communication Review (CCR)*, Vol. 37(1), 2007.
- [27] Dimitropoulos, X., Krioukov, D., et al. Revealing the autonomous system taxonomy: the machine learning approach. In *Proceedings of the 7th Passive and Active Measurements Workshop (PAM)*, 2006.
- [28] Kern, W. F., and Bland, J. R. *Solid Mensuration with Proofs*. Second Edition, John Wiley & Sons, 1954.
- [29] Kee, Y., Casanova, H., and Chien, A. A. Realistic modeling and synthesis for computational grids. In *Proceedings of the SC2004 High Performance Computing, Networking and Storage Conference*, November 2004.
- [30] TeraGrid
<http://www.teragrid.org>
- [31] iVDGL: International Virtual Data Grid Laboratory.
<http://www.ivdgl.org>
- [32] Bilgin, A., and Marcellin, M. W. JPEG2000 for digital cinema. In *Proceedings of the 2006 IEEE International Symposium on Circuits and Systems (ISCAS'2006)*, May 2006.
- [33] Kangasharju, J., Roberts, J., and Ross, K. W. Object replication strategies in content distribution networks. *Computer Communications*, Vol. 25(4), April 2002.
- [34] Lowekamp, B., Tierney, B., et al. Enabling network measurement portability through a hierarchy of characteristics. In *Proceedings of the 4th International Workshop on Grid Computing (Grid'2003)*, November 2003.
- [35] Wong, B., Slivkins, A., and Sifer, E. G. Meridian: a lightweight network location service without virtual coordinates. In *Proceedings of the ACM SIGCOMM 2005 Conference*, August 2005.
- [36] den Burger, M., Kielmann, T., and Bal, H. E. TopoMon: a monitoring tool for grid network topology. In *Proceedings of the 2002 International Conference on Computational Science*, April 2002.
- [37] Madhyastha, H. V., Isdal, T., et al. iPlane: an information plane for distributed services. In *Proceedings of the 7th USENIX Symposium on Operating Systems Design and Implementation*, November 2006.
- [38] Govindan, R., and Tangmunarunkit, H. Heuristics for Internet map discovery. In *Proceedings of the INFOCOMM 2003 Conference*, June 2003.
- [39] Veeraraghavan, M., Lee, H., et al. A varying-bandwidth list scheduling heuristic for file transfers. In *Proceedings of the 2004 IEEE International Conference on Communications (ICC'04)*, June 2004.
- [40] Yang, X., Shen, L., et al. An efficient scheduling scheme for on-demand lightpath reservations in reconfigurable WDM optical networks. In *Proceedings of OFC/NFOEC'2006*, March 2006.
- [41] Taesombut, N. and Chien, A. A. Distributed Virtual Computer (DVC): simplifying the development of high performance grid applications. In *Proceedings of the 2004 Workshop on Grids and Advanced Networks (GAN'04)*, April 2004.

10. APPENDIX

In our simulations, we evaluated the proposed network information models across a range of real ISPs' metropolitan, national and international networks. Here, we give the details of these network topologies as summarized in Table A.1-3.

Table A.1. Details of AboveNet metro-area network topologies

City	Number of PoPs	Number of links	Average edge degree	Average link latency (msec)
Boston	15	19	2.533	0.016
Chicago	33	43	2.606	0.030
Houston	25	34	2.720	0.036
Los Angeles	24	24	2.000	0.022
Philadelphia	22	25	2.273	0.031
San Francisco	38	47	2.474	0.031
Seattle	23	25	2.174	0.030
Washington DC	35	44	2.514	0.028

Table A.2. Details of ISP backbone network topologies

ISP	Network Presence	Number of PoPs	Number of links	Average edge degree	Average link latency (msec)
AT&T	US	110	140	2.546	1.918
Ebone	US/Europe	27	46	3.407	2.110
Exodus	US/Europe	22	36	3.273	5.672
Level3	Global	48	400	16.667	6.678
Sprint	Global	44	86	3.909	6.792
Telstra	Australia	55	57	2.073	4.648
Tiscali	Europe	47	80	3.404	2.986
Verio	Global	119	229	3.849	3.633

Table A.3. Details of the studied multi-ISP, global network

ISP	Network Presence	Number of PoPs	Number of links	Average edge degree
AT&T	Global	115	170	2.957
British Telecom	Global	109	189	3.468
Cogent	North America/ Europe	87	100	2.299
Global Crossing	Global	329	386	2.347
Level3	US/Europe	59	97	3.288
NTT/Verio	Global	39	74	3.795
Qwest	US	53	99	3.736
Sprint	Global	282	379	2.688
Time Warner	US	48	73	3.042
Verizon	Global	98	187	3.816
Interdomain Links			597	
Total		1219	2351	3.857