Position Paper:

P2P Overlay, Internet Underlay and Their Mutual Impact

Amir Hassan Rasti
University of Oregon
amir@cs.uoregon.edu

Contents

1 Introduction

The Internet has evolved greatly since its first days in different aspects. The network infrastructure has grown from a few academic and research institutions to a huge global network with nodes in almost every home. In the meantime, a new class of applications have been designed and widely used over the Internet for a wide variety of functions; peer-to-peer (P2P) applications. In P2P applications, participating peers form overlays through which they exchange data. The load imposed by the P2P applications on the network has raised concerns in ISPs because of its higher level and different pattern compared to traditional client-server applications. These issues have motivated three areas of research : *(i)* Internet topology, *(ii)* design and characterization of P2P applications, and *(iii)* studying the mutual impacts between the P2P applications and the underlying network. In this position paper, we survey the research works published in these areas in order to locate any open issues and problems. Below, we present an overview of these areas.

Internet topology characterization: In this area, the researchers study the Internet connectivity graphs in order to learn about the structure of the Internet and how it is evolving. Such information is critical for Internet researchers as it provides knowledge about potential features and shortcomings that may result from certain connectivity structure. For instance, some studies (*e.g.*, [3]) have claimed that the Internet has a *scale-free* structure and therefore its connectivity is dependent on a small number of very high degree nodes (hubs) and concluded that the Internet is vulnerable to targeted attacks on these hubs.

The Internet topology is often studied at two different abstraction levels: *(i) Router-level topology* describes the connectivity graph of the routers that interconnect the Internet, while in *(ii) AS-level topology* the connectivity of autonomous systems (*i.e.*, networks with an independent management) is the subject of study. Since the expansion of the Internet to a global network, no complete topology of the Internet has been presented to date and such a task still remains infeasible to do due to the distributed nature of the Internet. Despite this fact and other challenges, a significant number of research studies have been working on capturing and characterizing the Internet topology at both AS- and router-level using innovative techniques that we will discuss in Section 2.

P2P application design and characterization: The attractive features of the P2P network application model has encouraged application developers to employ the P2P model in a variety of applications. Specifically, in the area of content delivery and sharing, P2P applications have been mostly successful and popular. Nevertheless, designing an efficient, reliable and high performance P2P application can be very challenging. In such systems dynamics of peer participation, heterogeneity of the peers in terms of available resources and bandwidth along with some other issues are the main challenges that an application designer has to overcome. We will discuss some of the research works in this area in Section 3.1.

Once a P2P application is widely adopted by the Internet users, there are still many questions that need to be answered about it. Due to the distributed and nature of these applications, there is often no central monitoring or controlling entity and therefore one cannot answer questions on issues such as the performance and efficiency of the working system without a thorough network measurement. Also, the researchers are often interested in studying large P2P overlay networks as samples of complex networks in order to discover their features and shortcomings. In Section 3.2, we discuss some of the mostly cited works in the area of P2P measurement and characterization.

Overlay-Underlay Interaction: We mentioned before that the participating peers in a P2P application form an overlay. An overlay is a virtual data communication network that is built over a real network (Internet) actually responsible to carry the data packets. In recent years, the traffic imposed by the P2P overlays has raised concerns in many ISPs urging them to limit or control this traffic. In the area of overlay-underlay interaction, the researchers discuss the following issues: *(i)* The impact (load) of the P2P overlays on the network, *(ii)* ISP efforts to limit the impact and the reaction of P2P applications, *(iii)* ISP-friendly P2P applications, and *(iv)* ISP-P2P cooperation.

The increasing popularity of the peer-to-peer applications has caused the traffic of such systems to become an issue for the ISPs. On one hand, the P2P model is attractive to the content providers because it empowers them to feed more users with little investment. For instance, NBC has re-branded a P2P streaming and file sharing platform, called Pando, for high definition rebroadcasting of their shows over the Internet. On the other hand, many ISPs have raised concerns about both level and pattern of the traffic caused by P2P applications. Furthermore, some ISPs have incorporated mechanisms to detect and limit the amount of traffic associated with certain P2P applications. In the summer of 2008, the Federal Communications Commission (FCC) issued a ruling against Comcast on “discrimination among applications” and ordering them to stop such practices. The ruling was based on a complaint accusing Comcast of blocking P2P traffic. This was after other attempts by P2P applications to make P2P traffic harder to detect and control by the ISPs (*e.g.*, encryption).

Why are the ISPs concerned about P2P traffic ? There are two important differences between the P2P applications and common client-server applications; *(i)* In most of the traditional client-server applications (*e.g.*, WWW), the uplink traffic of the users is relatively small. However, in P2P applications, participating peers may generate as much upload traffic as they download. This results in a significant increase in the amount of upload traffic that the ISPs have to transmit. *(ii)* In most traditional applications, the traffic flow has a temporal dependence with the human interaction. For example, when the user clicks on a WWW link, the client starts to download the targeted page or file and the flow stops as soon as the download is complete which normally takes between a few seconds to a few minutes for larger files that are requested less frequently. In contrast, in P2P applications, although the traffic flow starts with user interaction, it will often continue much longer without any user action. For instance, in BitTorrent (a popular peer-to-peer file distribution application) the network link is often utilized in both outbound and inbound directions during the downloading time. Even once the download is complete, the client automatically continues providing content to other participants until stopped by the user effectively keeping the uplink busy even after the download is complete. The uplink traffic is often sent to peers in other ISPs and therefore increases the load on the inter-ISP links. Therefore, the advent of P2P applications increases ISPs' costs by *(i)* increasing the ISPs' uplink traffic for the same volume of download *(ii)* changing the user traffic pattern from bursty (short flows of traffic that are originated by user interactions) to steady (continuous flow even without user interaction). Assuming fixed amount of download per user, increased uplink traffic often means that the ISPs should purchase more bandwidth for the same number of users. Also, bursty traffic pattern which was the dominant user utilization pattern, allowed a much larger provisioning ratio for the ISPs (the short flows by different users occur in different times and therefore the momentary load of the ISP is small) compared to a steady pattern, effectively forcing ISPs to purchase larger bandwidth for a fixed number of users.

The problem of P2P traffic for ISPs has motivated several sets of research projects. Some have proposed methods to make P2P applications "ISP-friendly" mainly by localizing their traffic within ISPs. Although localization can reduce ISP load for certain scenarios without degrading the application's performance, in many cases localization may limit the performance of P2P applications, mainly by making the groups of peers that can help each other smaller. Such limitations suggest that localization is not enough and some other mechanisms need to be used to differentiate between external peers.

Recently, there has been multiple research works suggesting cooperation between peer-to-peer applications and the ISPs. In summary, within such cooperative methods, ISPs help peer-to-peer applications select neighbors in order to minimize the load on the ISP's costly links. The ISP uses its information about the topology, link costs and utilization in order to adjust the amount of P2P traffic on its own external links.

1.1 Roadmap

This position paper, surveys and categorizes the research works in the three areas mentioned above.

We aim to understand the mutual effects of the P2P overlays and the Internet underlay. However, in order to characterize such effects, we first need to understand the characteristics of the P2P overlays and the underlying network.

Toward this end, in Section 2 we survey some outstanding research studies on the Internet topology. In Section 2.1, we discuss research studies characterizing the AS-level topology of the Internet and categorize their data gathering and characterization methods. Section 2.2 surveys the studies on router-level topology of the Internet and categorizes their data sources and characterizations.

In Section 3, we survey research studies on the design and characterization of P2P applications. Section 3.1 categorizes P2P overlays according to the function, structure and shape of the overlay. Section 3.2 surveys and groups research works in the area of characterizing P2P overlays through measurement, modeling and simulation.

Section 4, we focus on the mutual impacts of the P2P overlays and the underlying network. Section 4.1 surveys research works on the impact of P2P overlays on the network. In Section 4.2 discusses the actions made by the ISPs to manage the P2P traffic and why they are not acceptable by the network community. In Section 4.3, we summarize research works that try to form ISP-friendly overlays and in Section 4.4, we survey the recent works based on the cooperation between the ISP and the P2P applications.

Finally, Section 5 concludes the paper by reviewing the main challenges and shortcomings.

2 Studying the Internet Topology (Underlay)

Although the Internet is a man made phenomenon, because of its true distributed nature, no entity can claim to have a full map of its topology. Since the rapid evolution of the Internet in the 90s, capturing its topology has become an interesting challenge for the researchers. In addition to the network researchers who study Internet architecture in order to learn the associated features and shortcomings, some scientists have also shown interest in the Internet topology as a large scale complex network.

The Internet topology may be studied in two different levels. In AS-level topology, the connectivity graph is composed of nodes that each represent an Autonomous System (AS) and edges that represent a physical link between the two corresponding ASes. Roughly speaking, each AS represents an independent company's network and therefore AS-level topology depicts connectivity between companies. Since packet routing over the inter-AS links is handled by the BGP routing protocol and the main deciding factors in BGP routing are often predefined *policies*, having a simple connectivity graph of ASes is of little use when data paths are of any interest. Therefore, the edges of the AS-level connectivity graph are often annotated with the *peering relationships* among the corresponding ASes that also reflect the BGP policies applied on the link.

AS-level topology can provide a high-level view of the Internet and is very useful in describing the structure of the Internet, however, it may not provide enough details about the network technology. In router-level topology, the nodes of the connectivity graph represent routers and each edge of the graph represents a physical link between two routers. The common method to gather router-level connectivity practiced by the researchers is using *traceroute* to capture a massive number of router-level paths.

The main challenges in studying Internet topology are data gathering and characterization. Capturing connectivity data is each level has its own limitations and hurdles which need to be addressed. Once the data is available, a researcher will have to use right methods to look at the data in order to extract new and interesting features.

In this section we survey and categorize the most important research works in Internet topology characterization. In Section 2.1 we discuss the studies on AS-level topology and categorize them based on the data sources they have used, characterization method they have employed and also the type of model they provide for the Internet connectivity.

In Section 2.2 we survey important research conducted on the router-level topology of the Internet and categorize them according to their data source, characterization method and modeling class.

2.1 AS-level Topology of the Internet

The Internet is a network of networks. Each network operated and controlled by a separate and independent administrative entity is called an Autonomous System (AS). Since connectivity structure and packet routing within and among

ASes are each based on different goals and principles, AS-level and router-level topology need to be studied separately. Connectivity among ASes is often based on business decisions rather than technical ones and for this reason, packet routing also follows business policies. For instance, a small ISP often chooses a provider offering a lower price for their desired service level.

In the AS-level connectivity graph, each node represents an AS and each edge shows a physical connection between two ASes. Note that if two ASes cover a large geographical area, they may have multiple physical links connecting their networks in different locations. However, in AS-level topology the number of links between two ASes is usually not considered.

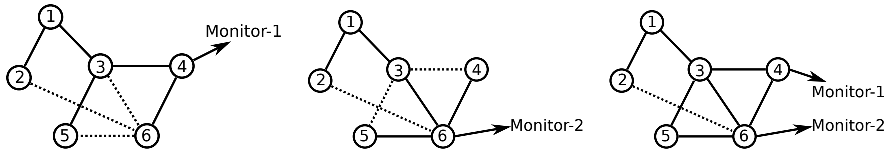
One of the main challenges of the studies on the AS-level topology is obtaining a reliable data source. In most of the studies, the AS-level connectivity data is obtained from BGP monitoring and archiving servers such as University of Oregon's RouteViews. Although the data from such sources is known to be incomplete, it is still used as the best source of information on AS-level connectivity. One of the main reasons for studying Internet topology is learning about the paths that the data packets traverse from source to destination. Since the inter-AS routing is policy-based and handled by BGP routing protocol, the BGP policies also need to be included in AS-level topology, otherwise, the connectivity information will not be useful. In Section 2.1.1, we discuss different data sources used in AS-level topology and categorize published works from this aspect.

AS peering policies are usually simplified in AS relationships. In this categorization the relationship between each pair of connected ASes belongs to one of the following groups: (i) customer-provider, (ii) peer-peer and (iii) sibling-sibling. The basic BGP policy that is commonly used is usually referred to as "valley-free" routing. This model associates a hierarchical model to the Internet in which each customer is located below its provider(s). In this hierarchy, the top level ASes have no providers, instead they are connected to each other over peer-peer relationships. In this hierarchy, the *tier* of each AS is simply its level in the hierarchy, where top level ASes are tier-1, their customers are tier-2 and so on. This hierarchical structure is an insightful characterization of the AS-level topology. In Section 2.1.2, we discuss the characterization techniques and methods used in AS-level topology and categorize the research works from this point of view.

In order to better understand the AS-level topology, some studies have taken the modeling approach. In some of these works, connectivity of the ASes and its pattern and evolution are the subject of mathematical and stochastic models. For example, the node degree distribution of the AS connectivity graph has been modeled with different stochastic models. We will discuss the modeling alternatives and survey research on modeling AS-level topology in Section 2.1.3.

2.1.1 Data Sources for AS-level Topology

One of the main challenges of studying AS-level topology of the Internet is obtaining accurate and complete data. Using inaccurate, outdated or biased



(a) 3 links are hidden from monitor-1.
(b) 3 links are hidden from monitor-2.
(c) One link is hidden from both monitors.

Figure 1: BGP monitors cannot observe all links.

data can mislead a researcher towards incorrect conclusions. For instance, Chen *et al.* [20] show that missing a large number of peering links interconnecting medium-sized ISPs in the BGP traces used by some earlier works has been the main cause of observing power law degree distributions and consequent incorrect results.

Research studies have used three sources of data in studying AS-level topology. Below, we discuss these sources and survey the studies using each.

- **Using BGP archives:** One group of common sources of information for capturing AS-level topology are public BGP monitoring and archiving servers. One of the mostly cited such projects is University of Oregon's RouteViews that has been actively monitoring and archiving BGP routing tables and updates since late 1997. In BGP, each routing update includes the complete *AS-path* from the update origin up to the router receiving the update, therefore, each BGP router maintains all the AS-level paths connecting it to all other reachable networks, which is essentially one view of the Internet's AS-level topology. We should note that this view, as shown in Figure 1 only includes the links appearing in the paths starting from our BGP router's AS and all other links of the AS connectivity graph are hidden from it.

  In RouteViews, a large number of BGP peerings are established to many volunteer ASes all over the Internet which will act as RouteViews' vantage points. Over these peerings, each vantage point relays all the updates visible from their points of view to RouteViews. Effectively, the set of all paths received by RouteViews includes a large portion of all the links between ASes.

  Using BGP archives one can produce an AS-level topology snapshot that includes all the active ASes. Also, using saved archives from different points in time, one can study the dynamics of the AS-level topology over a certain time period. On the down side, BGP snapshots often do not include backup links since they are not actively used and advertised by the corresponding ASes unless their main links stop working. Also, as mentioned earlier, by including more vantage points (BGP peers), a BGP monitoring service can extend their sight to observe a larger number of links, however, there are always links that remain hidden. Roughan *et* *al.* [70] try to identify and enumerate these missing links.

Several studies on AS-level topology such as papers by Govindan *et al.* [37], Faloutsos *et al.* [33], Medina *et al.* [60], Gao [34] and Mahadevan *et al.* [57], the authors have used BGP data to produce AS-level topology of the Internet. Chen *et al.* [20] expose incompleteness of BGP data as the main cause of the observed power-laws in earlier works such as the paper by Faloutsos *et al.* [33]. They claim that in BGP snapshots 20%-50% of the links are missing.

Chang *et al.* [17] compare RouteViews data sets with the BGP data sets they have gathered from a set of looking glasses and routing registries and find 25-50% more AS relationships and 2% more ASes. A looking glass is a web interface allowing public viewing access to an ISP's BGP routers, while in an Internet Routing Registry (IRR), the routing policies of each AS is maintained in a public database.

Zhang and Liu [84] compare the AS-level topology obtained from RouteViews snapshots with those they have produced by gathering data from looking glasses, routing registries and multiple route servers including RouteViews. In order to observe backup links that do not usually appear in the RouteViews paths, they obtain all BGP updates over a one year period from RouteViews and include the links observed in these updates to their data set as well. The final AS-level topology they produce includes 44% mode links and 3% more ASes than the average graph obtained from RouteViews data alone.

Roughan *et al.* [70] aim to estimate the number of missing links in the AS connectivity graph obtained from RouteViews using stochastic and information theory models. Their estimates approve 3 earlier works ([42, 61, 84]) that tried to produce a complete AS-level topology. They estimate the number of missing links to be about 37% of the observed links at a certain time. They also estimate that using 700 route monitors, we can observe 99.9% of the links in the AS connectivity graph.

- **Converting router-level paths to AS-level:** Some researchers including Chang *et al.* [18] have suggested gathering router-level path information such as traceroute logs and converting them to AS-level paths. In this method, AS-level connectivity can be obtained in finer granularity (*e.g.*, multiple links between ASes). Another advantage in comparison with BGP data is capturing ASes whose routes are aggregated in BGP with other ASes. However, using this method involves some serious data gathering challenges, *i.e.*, accessing a sufficient number of vantage points to run traceroute experiments. Also, traceroute data is known to have certain issues resulting in incomplete or in some cases erroneous data that we will discuss in Section 2.2. Besides these issues, mapping routers to ASes may also add some error due to using foreign IP addresses in border routers. Chang *et al.* [18] use traceroute logs from the Internet Mapping Project [21]. They present some techniques to avoid the effect of the issues mentioned above. The authors claim that the method addresses some shortcomings of the BGP-based method, however, due to the increasing security concerns, networks are blocking traceroute access to their networks and therefore capturing a nearly complete picture of the Internet using traceroute based methods is infeasible.

- **Extracting AS relationships from routing registries:** One of the services usually provided by the Regional Internet Registries (RIRs), including ARIN [7], RIPE [69], APNIC [6], LACNIC [51] and AfriNIC [1], is maintaining routing registries for their own geographical zones. Additionally, some third party organizations such as RADB [64] also run routing registries. A routing registry is a public database for keeping and publishing the BGP routing policies used by individual ASes over each of their peering relationships with their neighboring ASes.

  Using routing registries, a researcher not only can obtain AS connectivity information, he can also infer the relationship types using the policies listed. Routing registry data can be quite useful for an Internet topology researcher, since it provides all peering information including backup links as well as details of the policy without any measurement and these information are hard to obtain from sources discussed earlier. However, in practice routing registry data is of limited use in Internet topology studies because the entries are often out of date and incomplete due to the fact that the ISPs have little motivation to keep them up to date. Gao [34] uses ARIN's routing registry information to compare and evaluate her algorithm for inferring AS relationships while Chang *et al.* [17] use RIPE's routing registry to complement data obtained from RouteViews and looking glass websites, as described earlier in this section.

  In summary, although routing registry data is often considered incomplete and therefore it is not relied on as a sole source of information, some researchers have used it in order to complete the topology obtained from other sources or as a reference to compare and evaluate their methods, such as inferred sibling relationships.

2.1.2 Characterization Methods for AS-level Topology

Characterizing the AS-level topology of the Internet is a common goal that has been pursued using different techniques and methods. The common goal is discovering interesting characteristics and features of the AS connectivity graph that can provide an insight on better understanding the way Internet works and evolves. The most important subjects of AS-level topology characterization, according to the volume of research work, are : (i) Degree distribution in AS connectivity graph, (ii) Hierarchy of the AS-level topology, and (iii) Inferring inter-AS relationships. In this section we survey the research work addressing each of these subjects and mention the benefits and the challenges involved in each case.

- **Node degrees in AS connectivity graph:** In AS connectivity graph, each node, representing an AS, is connected to a number of other nodes. The number of ASes that each AS is connected to, determines the degree of the corresponding node. Node degree distribution provides the most basic view of a graph's connectivity and therefore it has been used in numerous research works to capture and present the structure of the AS connectivity graph.

  Faloutsos *et al.* [33] in one of the first works in AS-level topology characterization claim that the degree distribution of the AS connectivity graph follows a power-law distribution. They also discover other power law relationships in the Internet topology, including the number of nodes within h hops as a function of h. Based on these finding, they show that random graphs do not represent the Internet topology. Later, Medina *et al.* [60] analyze possible root causes for the observed power laws in the previous work. They identify *preferential connectivity* together with *incremental growth* as the key contributing factors to the power law relationships. Fabrikant *et al.* [32] propose an explanation for the power laws based on a toy model of Internet growth in which two objectives are optimized simultaneously: last mile connection costs and transmission delays measured in hops. Power laws tend to arise as a result of complex, multi-objective optimization.

  Chen *et al.* [20] identify incompleteness of BGP data as the main cause of the observed power-laws. The paper shows that by compensating for the missing links, the resulting degree distribution becomes heavy-tailed but not power-law. It also claims that the connectivity dynamics and growth processes assumed in [60] do not apply to the Internet. Later, Li *et al.* [53] show that degree distribution alone cannot capture the specifications of a graph completely by showing examples of different graphs with very different characteristics showing the same degree distributions. Although the heavy-tail degree distribution of the AS-level topology shows that there are a few ASes with very large degrees while the vast majority have very small degrees, such pattern should not be used to conclude a certain structure in the Internet.

  Joint degree distribution is proposed by Mahadevan *et al.* [57] as a definitive metric in order to capture the connectivity preferences with regards to node degrees. This paper shows that the Internet topology is disassortative, *i.e.*, nodes have a tendency for connecting to nodes with dissimilar degrees.

  Caveats: Although node degree is an important factor and it can reveal several features of the graph, care must be taken in identifying graphs based on their degree distributions alone. Also, considering the incompleteness of the available AS connectivity graphs, any findings regarding node degree may be an artifact of the missing links.

- **Inferring relationships between ASes:** As mentioned earlier, AS relationship information is an important part of the AS-level topology since such information is necessary in order to understand the BGP policies that are used in routing among ASes. However, the relationships are business information and can be private. Therefore, the researchers have tried to infer the relationships from other information such as AS connectivity graph as well as the routing registry information. These inference techniques are often based on common conditions that one expects to observe in these relationships. For instance, it is expected that the degree of a provider be larger than that of its customer. However, since there are always exceptions and special cases, such assumptions lead to a certain amount of error. Although inferring customer-provider relationships might be less challenging, inferring peer-peer and sibling-sibling relationships often requires additional information.

Gao [34] proposes a method for inferring relationships from the AS-paths obtained from RouteViews. Her proposed algorithm is based on the *valley-free* routing principle according to which no customer lies between two providers of its own in an AS-path since a customer does not provide transit service to its providers. It is also assumed that in each customer-provider relationship, the degree of the provider is larger than that of the customer. In this algorithm, in each AS-path the AS with the highest degree is chosen as the top AS and the other relationships are inferred based on the valley-free principle. By processing each path, one vote is cast towards the inferred relationships along that path and the final decision is based on the total votes resulting from processing all the paths after certain adjustment and refinement. Subramanian *et al.* [78] present the AS relationship assignment as an optimization problem and propose a heuristic algorithm to solve this problem by combining AS-paths from multiple vantage points in the Internet. Other works by Xia and Gao [81] and Dimitropoulos *et al.* [29] evaluate the proposed algorithms and suggest incremental improvements over those algorithms by accounting for missing relationships and including routing registry information in inferring sibling relationships, respectively. The Cooperative Association for Internet Data Analysis (CAIDA) [13] generates AS relationship snapshots of the Internet using algorithms from [29] applied to RouteViews data on a regular basis and archives and publishes the results for the public use.

The main challenges involved in this problem are inferring sibling-sibling relationships as well as accounting for the missing connectivity information. The inferred relationships are widely used in a variety of research works involving the Internet topology and traffic.

- **Hierarchy of the AS-level topology** The relationships between ASes are commonly used in the area of AS-level topology to depict the hierarchy of the Internet. According to this hierarchy, each AS is assigned a tier number reflecting a level of the hierarchy. Generally, tier-1 ASes are those who have no providers and a tier-n AS has at least one tier-$(n-1)$

provider. Understanding the hierarchical structure of the Internet in insightful and can be used to explain many characteristics of the Internet and the way traffic flows over it. However, there are a few challenges that makes this work nontrivial. First, there is some controversy on defining tier-1 ASes and the instances. Since the business contracts among top-level ASes are confidential, accurate inference of the type of relationships becomes challenging. Also, some peer-peer and sibling-sibling relationships make shortcuts linking different tiers to each other that makes some of the assumptions invalid.

Ge *et al.* [35] provide an algorithm to classify ASes in their respective tiers according to the inferred customer-provider relationships using the above definition. They also make available a tool called *TierClassify*) implementing their algorithm for public use.

Dimitropoulos *et al.* [28] provides an alternative classification of the ASes. They define 6 classes of ASes, namely, Large ISPs, Small ISPs, Customer ASes, Universities, Internet exchange points and Network information centers. They use AdaBoost machine learning tool and manually classify more than 1000 ASes in order for the machine learning algorithm to learn the characteristics of each class. The AS attributes include IP space size and type and number of AS relationships along with boolean attributes that reflect the results of searching certain words in the AS description field from the registry.

2.1.3 Modeling the AS-level Topology

Modeling is often used in characterizing complex systems. This method can be very helpful in simplifying and understanding the basic rules governing the system behavior and it can possibly enable the researchers to predict the system's behavior in response to anomalies or unexpected events. Modeling the AS-level topology has been pursued in a number of research works. The main challenge of a useful modeling work would be finding the right model that not only fits measured data but also can provide an insight into the limitations and tradeoffs governing the AS-level topology.

In this section we categorize some of the research works on modeling the AS-level topology according to the type of the model they use.

- **Descriptive Models:** In this class of modeling works, certain characteristics of the AS-level topology are captured by measurement and then mathematical models are provided trying to fit the captured data. Faloutsos *et al.* [33] fit the degree distribution of the AS connectivity graph with a power-law model and Medina *et al.* [60] find *preferential connectivity* and *incremental growth* as the main causes of the power-laws. However, Chen *et al.* [20] questions the Barabasi-Albert model for AS-level topology based on the fact that the observed degree distributions were artifacts of incompleteness of the AS connectivity graph. They suggest that the actual degree distribution of the AS connectivity graph does not fit a BA model although it is heavy-tailed and suggest adapting a HOT-based model for AS-level topology.

In another example of modeling, Roughan *et al.* [70] in an effort to discover the missing links of the AS-level topology, employs the capture-recapture idea from biology, to derive a Binomial Mixture Model(BMM) for the number of observations of each link across all view points. They estimate the model parameters using an Expectation Maximization (EM) algorithm.

Since descriptive models are only based on measured data, they can be vulnerable to measurement errors.

- **Generative Models:** In multiple areas of networking, the researchers need to set up simulations. These simulation often need a topology graph that has similar characteristics as the Internet. Generative models are algorithms designed to generate graphs with similar characteristics as the modeled graph. BRITE [59]is a topology generator tool that is able to generate topology graphs using a variety of models. In particular *ASWaxman* generates AS-level topologies with the properties of a random graph in which nodes with shorter distance are more likely to get connected to each other while *ASBarabasiAlbert* results in topologies that have power-law degree distribution and try to represent the hierarchical structure of the Internet. Some older examples of the generative models of the Internet are *GT-ITM and Transit-stud* [15] and *Tiers* [30]. A representative generative model can be a quite useful tool for evaluating a design using simulation or verifying a hypothesis about the Internet however since each model focuses on representing the Internet from a certain aspect or a number of aspects they always miss some other characteristics of the real Internet.

- **HOT-based Models:** Any complex system can be thought of as a solution to an optimization problem with certain constraints and tradeoffs. Highly Optimized Tolerance (HOT) denotes a class of models that are based this very principle. In HOT-based modeling, researchers try to find use these optimizations and tradeoffs in order to build a model that describes behavior of the system. Fabrikant *et al.* [32] are the first to provide a HOT-based model of the AS-level topology. They propose a toy model of the incremental access network design optimizing a tradeoff between connectivity distance and node centrality. They also show that the relative importance of these factors can significantly change the resulting topology. Alderson *et al.* [4] make a proposal of identifying the economic and technical tradeoffs involved in network access design for building a HOT-based model of the Internet topology. They suggest that the "Buy-at-Bulk" scheme is an optimization to a tradeoff on the bandwidth provisioning problem according to which "larger capacity cables have higher overhead costs , but lower per-bandwidth usage costs."

  Chang *et al.* [19] also apply HOT concept to the AS connectivity problem. They extend earlier works by presenting a multivariate optimization problem that determines AS decisions in choosing an upstream provider: (i) AS-geography *i.e.*, location and number of ASes within each AS, (ii) AS-specific business models and (iii) AS evolution *i.e.*, a historic account of each AS in the dynamic market.

Although HOT-based models are much more challenging to develop compared to the descriptive models, they are quite more robust against measurement errors. On the other hand, since the Internet evolution is a distributed process driven by many independent entities with potentially different goals and limitations, assuming that the same set of tradeoffs are controlling this process in different places seems questionable.

2.2 Router-level Topology of the Internet

As mentioned earlier, the Internet topology studied in two different abstraction levels. While in AS-level topology the connectivity among ISPs is the focal point and the most important factor forming the topology is business relationships, in router-level topology, the network infrastructure is the primary subject of study and the network technology is the major factor.

The most important challenge in studying router-level topology of the Internet is data gathering. Although a simple tool such as *traceroute* is potentially able to capture the router-level paths between any two points in the Internet, practical limitations significantly reduce the usability of the results. Commonly in router-level Internet topology, the main source of information is the data resulting from of the Internet is captured via a large-scale series of *traceroute* operations.

The router-level topology, if captured with acceptable accuracy, provides a higher resolution over AS-level topology. In spite of the AS-level topology, multiple paths may be captured as well. However, the main problem remains accurate data gathering due to limitations that the traceroute and other tools have.

2.2.1 Data Sources for Router-level Topology

In this class of studies, *traceroute* has been the basic tool used when a global scope is desired while in studies with local scope, topology information is usually provided by the ISPs. Traceroute [47] can provide the router-level path from a source over which the researcher has control to any arbitrary destination host over the Internet. Traceroute, originally developed by Van Jacobson, sends a series of packets with controlled TTL values. TTL (time to live) of an IP packet determines the maximum number of routers it can pass before reaching destination, a mechanism designed to dispose of the packets that get stuck in routing loops as a result of routing problems. In order to capture the router-level path from host A to B, traceroute must be run on host A. Upon execution, it starts sending packets (ICMP or UDP packets depending on version and parameters used) with TTL value of zero. As a result, the first router on the path will dispose of the packet and send an ICMP error message back to the sender. In each round, traceroute increments the TTL value by one until either the packet reaches the destination or the TTL reaches a predefined maximum value (usually 30). The error messages returned by the transit routers as a result of TTL expiration are used by traceroute to identify routers on the path and thereby produce a list of IP addresses of the routers on the path.

Although traceroute has been very useful for determining routing problems, its ability to capture the global router-level topology is limited for the following reasons. First, in order to produce the Internet topology a researcher needs to capture a large number of paths. The usefulness and representativeness of the resulting topology highly depends on the number and distribution of the endpoints of the paths. In order to capture the Internet topology with an acceptable coverage, a researcher would need access to a large number of hosts worldwide which is very hard to obtain. Second, increasingly many networks are using firewalls that block traceroute packets into their networks, specially at the edge of the Internet. This will limit the coverage and accuracy of the captured paths and the resulting topology. Third, there are known limitations in traceroute technique that result in erroneous results in presence of dynamic routing. Remember that each hop is identified by a separate packet and due to dynamics of routing, different packets may take different paths between the same pair of end-points. Using such erroneous paths can mislead the researcher to including false links in the topology.

There are a number of projects and tools built on top of the basic traceroute technology with the goal of achieving higher accuracy and wider coverage. Since data gathering is a major challenge in studying the Internet topology, below we compare these data gathering tools and projects and the research works using each.

- **Skitter** [14] is a project of CAIDA with the goals of (i) determining forward IP paths, (ii) measuring RTTs, (iii) tracking persistent routing changes and (iv) visualizing network connectivity.Skitter uses the traceroute technique in addition to some kernel hacks in order to increase the accuracy of RTT measurements. Barford *et al.* [10]employ Skitter traces between 8 sources and more than 1000 destinations spread all over the world to build up a partial picture of the Internet backbone in the year 2000. While the sources are all hosts owned by the project placed in volunteer networks, the destinations are a web servers distributed over the Internet. They argue that towards the goal of characterizing the Internet backbone, the utility of adding more vantage points in a traceroute study is marginal. Specifically, they claim that a careful selection of two or three vantage points will result in nearly same coverage as all the 8 sources used by skitter. *Archipelago (Ark)* is the evolution of the skitter project including the skitter monitors, measurement tool, several other data processing tools. Later the skitter measurement tool was replaced with *scamper.* Scamper [39]is an extended version of the skitter tool that also supports IPV6 and is able to flexibly use TCP or UDP probing packets. Luckie *et al.* [55] use scamper from 8 vantage points distributed across the globe and 3 different sets of destinations including random routable addresses, top 500 websites according to Alexa [80], and a list of known routers from an earlier study. They show that although ICMP traceroute probing is able to reach more destinations and discover more AS links, UDP probes infer the greatest number of IP links.

- **Mercator** proposed by Govindan *et al.* [38] focuses on the problem of finding useful destination addresses in a traceroute-based technique. They use *informed random address probing* to make guesses about which prefixes might contain addressable nodes by heuristics from common patterns of IP space allocation. They also employ source routing (supported by a only 8% of the Internet routers) to include cross links considering that they only employ one vantage point. Mercator addresses the problem of IP address aliasing by sending probes to the discovered address of the router and comparing the discovered address with the responding address to verify whether or not the two addresses belong to the same router.

- **Rocketfuel** proposed by Spring *et al.* [75], is a tool for mapping the router-level topology of an ISP using traceroute, RouteViews data, and reverse DNS. They perform traceroute experiments sourced from 800 vantage points hosted by nearly 300 traceroute web servers (servers that provide traceroute service from their location to any desirable host). The authors focus on improving the efficiency of probing. Using their path reduction techniques, they manage to reduce the number of probes needed by three orders of magnitude compared to a brute-force all-to-all probing without any significant accuracy loss. They capture a much more complete graph with roughly seven times as many links.

- **Paris Traceroute**was proposed by Augustin *et al.* [8]. The authors focus on the traceroute errors in presence of dynamic routing. They list possible traceroute anomalies such as loops, cycles and diamonds and show how they can happen as a result of different forms of dynamic routing such as load balancing.

Also, in a number of studies, such as the work by Li *et al.* [53], Abilene is used as a source of data. Abilene Network (now known as Internet2 Network) is a high performance backbone network in the U.S. mainly connecting academic and research centers throughout the country using high speed links (10 Gbps). Abilene Network provides a useful research case for Internet researchers because it makes all the topology and traffic information publicly available. Although such data does not represent the Internet, it still provides useful insights particularly as a real testbed to evaluate methods and tools for measurement and characterization of the Internet topology and traffic.

Despite all the efforts, finding a reliable source of date that provides highly accurate and representative data on the router-level topology of Internet is still a problem.

2.2.2 Characterization Methods for Router-level Topology

In characterizing router-level topology, the researchers usually search for interesting characteristics and features of the routers' connectivity. In some research works with pure science theme, the router connectivity graph is studied as a complex network with little attention to the context and the root causes. Another group of works study the router-level connectivity in order to understand the Internet and possibly discover its features and shortcomings.

Below, we survey some of the subjects discussed in the router-level topology characterization and discuss their advantages and shortcomings.

- **Node degree distribution** is commonly used as a characterization metric for router-level topology of the Internet. Similar to many other graphs, degree distribution is often considered the most basic piece of information that can capture and present some characteristics of the router-level connectivity graph mainly by showing the heterogeneity level across the nodes. Faloutsos *et al.* in their SIGCOMM paper [33] in addition to the AS-level topology that we discussed in Section 2.1, use a router-level topology map from an earlier work from 1995 and show that the degree distribution follows power law similar to the AS-level topology. This result has been rejected from different aspects in the works published later. Yook *et al.* [83] suggest a fractal model for the Internet topology and show that the power laws do not represent the Internet and the degree distributions are in fact exponential. Lakhina *et al.* [52] show that the power laws are an artifact of sampling the router-level topology using traceroute. They perform simulations showing that traceroute-like sampling will result in power-law degree distributions even if the original graph is a random ER graph.

  Nonetheless, all the studies agree that the router-level topology of the Internet has a heavy-tailed distribution. Some papers such as the work by Albert *et al.* in the Nature journal [3] have warned that in this heavy-tailed distribution, there are extremely high degree nodes that act as the central hubs of the Internet and failure of each can disconnect a large portion of the network. This idea was rejected by Li *et al.* [53] who showed that several graphs with very different characteristics may have similar power-law degree distributions. Although degree distribution provides a first level of understanding about the router connectivity graph, care must be taken not to read too much from it.

  Tomography of the Internet Since the researchers do not have direct access to the core of the Internet, they use information gathered from several endpoints in order to provide an image of the core. This practice is commonly referred to as *tomography* in many disciplines. According to this definition, we can categorize any traceroute-based studies of the Internet router-level topology as tomography. Coates *et al.* [24] provide a survey of the techniques for making inferences about the Internet based on the observed behavior. They include two classes of network tomography:

(i) estimating link-level characteristics from path-level data and (ii) estimating path-level characteristics from link-level data. The inferred data may be loss rate, packet delay or the connectivity. Although the common perception is that having more vantage points, the tomography results will be more accurate and reliable, Barford *et al.* [10] question the "more is better" approach and show that increasing the size of the network measurement infrastructure only leads to marginal improvement in Internet tomography.

2.2.3 Modeling the Router-level Topology

The modeling approaches for router-level topology of the Internet are similar to those we discussed for the AS-level topology in Section 2.1.3. They aim to find mathematical models that describe and explain the connectivity patterns. A good model not only should match the measured and confirmed data from the router-level topology, it should also provide an insight for understanding how the network grows and evolves. Using a reliable model, a researcher can detect vulnerabilities or predict potential malfunctioning threatening the Internet. Developing models that bear the mentioned capabilities has been a challenging task. Below we provide a survey of the modeling studies on the router-level topology of the Internet.

- **Descriptive Models:** In this group of studies, certain measured data on the router-level topology is examined for similarities against known mathematical models. Faloutsos *et al.* [33] use a router-level topology data set captured in 1995 and find similarities with the power-law model which was later rejected. This work is explained in more detail in Section 2.1.3.

- **Generative Models:** Similar to the description given for AS-level topology modeling, generative models of the router-level topology are algorithms or programs designed to generate synthetic topologies resembling the real router-level topology of the Internet. These models are widely used in simulation-based evaluation of network applications. Furthermore, they often provide the flexibility of generating a range of topologies by one or more controlling parameters.

  BRITE [59] (also an AS-level topology generators) is a tool that is able to generate topology graphs using a variety of models. In the class of *flat router-level* models, it places the nodes on a plane based on random or heavy-tailed model and after establishing the links using either *RouterWaxman* or *RouterBarabasiAlbert*, assigns link bandwidth according to either constant, uniform, exponential or heavy-tailed models with controlled parameters. *RouterWaxman* generates AS-level topologies with the properties of a random graph in which nodes with shorter distance are more likely to get connected to each other while *RouterBarabasiAlbert* results in topologies with power-law degree distribution by using *incremental growth* technique with *preferential attachment*. Medina *et al.* [59] also categorize earlier generative models into two groups of (i) *ad-hoc models* that are mostly built based on educated guesses such as the hierarchical structure of the Internet (*e.g.*, GT-ITM [15]) and (ii) *measurement-based models* that try to reproduce the measurement results such as Barabasi-Albert models that reproduce power-law degree distributions using preferential attachment.

Li *et al.* [53]use a first-principles approach in developing a generative model for the router-level topology of the Internet. They apply technological limitations and economical considerations into a performance optimizing design process yielding a generative model of the Internet's router-level topology.

While a generative model can be useful in evaluating a new design using simulation or verifying a hypothesis about the Internet structure, one may not assume that a generated topology resembles the network in every aspect. Limitations of each generative model should be recognized before employing them.

- **HOT-based Models:** General description of HOT-based models is provided in Section 2.1.3. Li *et al.* [53]pursue a first-principles approach aligned with the idea of HOT-based modeling in which the technological constraints and economical considerations are identified as the primary factors determining a network's decisions at the time of topology construction. According to this paper, the router building technology limits the bandwidth-degree product due to the limited bandwidth of the router's data bus. They use a number of state of the art Cisco routers in 2004 in order to identify the technological limits at the time and argue that the market mostly demands for relatively low bandwidth ports while the core of the network requires very high bandwidth ports. Therefore the solution to the optimization problem would be configuring routers with maximum number of ports at the edge (low bandwidth) and maximum bandwidth ports (small number) at the core of the network. They compare graphs generated by different generative models and show that the HOT graph has the highest performance (throughput) and lowest likelihood. The authors publish another paper [5] in which they extend the previous work by evaluating their HOT graph with Abilene and Rocketfuel data. HOT-based models are still a hot topic in studying the Internet topology and due to not relying on measurements, they are not subject to measurement errors. However, it seems that the idea is not yet developed enough to produce useful models representing the Internet topology from multiple aspects.

  Yook *et al.* [83] suggest a fractal model (scale-free) for the Internet topology in which the links are placed by competition between preferential attachment and linear distance dependence. According to their scale-free model, the Internet connectivity depends on a small number of very high degree nodes that representing the Internet hubs. They conclude that although the Internet is robust to random node failures, it is quite fragile to targeted attacks on these hubs. Doyle *et al.* [31] extend their earlier works on modeling the router-level topology by suggesting a "robust-yet-fragile" model for the Internet. They show that the characteristics of the scale-free model does not match those of the Internet while the HOT model they had suggested earlier [5] shows similar features and characteristics as the Internet using the two metrics of *performance* and *likelihood*. In their view, the Internet's fragility does not lie directly within its topological aspect. By focusing on the protocol stack, they mention that the lowest layers of the Internet are highly constrained by technological and economical limitations while the higher layers have more flexibility and freedom. The flexibility on the higher levels of the protocol stack such as the application layer is what makes the Internet robust and yet the same flexibility makes the network fragile to malicious exploitation.

3 Overlay Networks

P2P applications are used to provide a variety of network services in a decentralized fashion. Such systems are: *(i) robust*, since they do not have a single point of failure; *(ii) scalable*, as each user adds resources to the system, and capable of functioning at *(iii) very low cost*. The collection of participating peers in a P2P network form a *P2P overlay* which is a virtual network over which the peers exchange data. In most of today's P2P systems, the overlay networks are formed without considering the underlying network. For instance, in a random overlay network two peers that are in the same physical network have only a small chance of getting connected to each other while each may have neighbors from across the globe. Besides random overlays, in another group of P2P applications the overlay construction may have particular goals. For instance, in a gaming overlay, it is fair to assume that minimizing delay (between interacting peers) should be the goal of overlay construction while in streaming, maximum bandwidth from the source might be as important. In this class of P2P overlays, the goals of minimizing delay and bandwidth may indirectly cause overlay connections to become more localized.

Finally, a few recently proposed P2P overlays explicitly follow the goal of locality-awareness or network-awareness either with or without the support from the underlying network [1]. Considering the popularity of the P2P applications and the load they impose on the underlying network, it is important to study different types of P2P applications as well as the research works aiming at characterizing P2P overlays.

The impact of a P2P overlay on the underlying network depends on: *(i)* overlay connectivity structure, *(ii) traffic generation pattern* and *(iii)* packet forwarding and routing mechanism in the overlay. In order to study this imapct, we need to learn about the structure, packet generation and data paths in the overlays. The P2P applications are used for a variety of functions and their respective overlay networks have different shapes, structures and characteristics according to the functionality they are designed for. In Section 3.1, we categorize most well-known P2P overlay networks in research and user communities according to the overlay's functionality, structure and shapes and compare the subgroups, accordingly. In Section 3.2 we overview a number of research works on characterizing P2P overlays while categorizing them according to their approaches.

3.1 Categorizing P2P Overlays

P2P applications can be categorized from numerous aspects. In this report, we focus on the overlays and therefore we categorize P2P applications with this focus. Although overlays may be used for a variety of purposes, generally one overlay is constructed and used for a single functionality. In some P2P applications, multiple overlays are formed and used for multiple functionalities.

[1] We will discuss this issue in detail in Section 4

This is because the structure, shape, and characteristics of the overlay depends on its functionality.

The two main classes of overlay functionalities are *(i) Signaling and control* and *(ii) Content delivery*. In content delivery overlays, large amounts of data are transferred through the overlay to reach interested peers while in signaling overlays only queries and responses that are often short are transmitted through the overlay. We discuss and further subgroup signaling and content delivery overlays in Sections 3.1.1 and 3.1.2, respectively.

3.1.1 Signaling and Control Overlays

In a variety of P2P applications, overlays are used for maintaining membership and exchanging queries and responses. In this class of overlays, the main goal of the overlay construction are *reachability* and *resiliency* and therefore the overlays are often richly connected. The following are some examples of the signaling overlays categorized by the functions.

Categorizing signaling overlays based on functionality

- **Searching** is an important problem in file sharing applications. In these applications each user shares a number of files with other users and is interested in finding and downloading other files shared by other users. In order to avoid single point of failure issue associated with a central indexing server (*e.g.*, Napster), a *decentralized search* mechanism is used by some file sharing applications such as Gnutella. In Gnutella, participating peers form an overlay to handle the decentralized search functionality. Peers send their search queries to their neighbors and each peer checks the query against their own shared files. If they have a matching file, they will send back a positive response, otherwise they relay the query to their neighbors. Although simple, searching over a large scale flat overlay may become quite inefficient. In Semantic Small World [54], peers form a highly clustered overlay. The clusters are based on the semantics of the content shared by peers. Taking advantage of similar interests by groups of people, the semantic based clustering makes searching much more efficient in SSW.

- **Store and lookup services** is handled by a group of popular P2P applications called *Distributed Hash Tables* (DHT). A DHT is responsible for distributed storage of key-value pairs, similar to a local hash table. In DHTs each peer is assigned with an ID and is responsible for a part of the hash space according to the assigned ID. Each peer maintains a *routing table* consisting of a set of links to other peers that are its neighbors. Together these links form the overlay network. A node picks its neighbors according to a certain structure that is the main difference between different DHTs and is often referred to as the DHT's topology. Commonly, the routing table size and the routing algorithm complexity in DHTs are $O(log(n))$ where n denotes the number of participants. In

CAN [67], peers form an overlay over a virtual multi-dimensional Cartesian coordinate space. This d-dimensional coordinate space is a virtual logical address, completely independent of the physical location and physical connectivity of the nodes. In Chord [76], node keys are arranged on a circle. Each peer's routing table includes its successor and predecessor which are the next and the previous node on the circle, respectively. Each peer is responsible for the ID space contained between that peer and its successor. In addition to the successor and predecessor the routing tables also include a few shortcuts to other locations in the circle for the sake of faster routing.

Other well-known proposed DHTs are Pastry [71], Tapestry [86] both using circular ID spaces and Kademlia [58] uses the XOR metric to calculate the binary distance between two peers, in order to determine neighboring and routing information. Due to the efficient decentralized store/lookup service they offer, DHTs are widely used in different applications for indexing and state keeping. For instance, in Vuze, a popular BitTorrent client, a DHT is formed to act instead of a BitTorrent tracker, in case it becomes unavailable. In Freenet [23] a DHT-like overlay is formed for anonymized distribution of data to protect freedom of speech. The protocol design ensures anonymity of the publisher and downloaders of the data.

Categorizing signaling overlays based on structure:

Signaling overlays are generally divided into two groups based on the their structure. Below we compare and contrast the two groups with examples.

- **Unstructured overlays:** In this group of overlays, peers connect to each other in an arbitrary fashion. Each peer can individually select its own neighbors after a *peer discovery* phase in which peers acquire information about other participating peers. The resulting overlay topology is often close to a random graph, and thereby, highly *resilient* to *churn* (*i.e.*, dynamics of peer participation).

  In Gnutella, peers upon joining follow a peer discovery mechanism and learn about a number of other participating peers. Among those peers, they randomly select a subset and try connecting to them and continue until a predefined number of neighbors is reached. However, our earlier study on Gnutella [65] shows that there is a certain level of connectivity preference towards geographically close peers that may allow one to argue that Gnutella overlay is not purely random.

  Although unstructured overlays are easy to build and maintain, their *performance* and *efficiency* are often points of concern. Searching for popular content in an unstructured overlay is often easy and fast, while the search performance for unpopular content is lower. This is because the query should eventually reach all the participating peers to ensure that a rare content can be found. There is also a trade-off between efficiency and performance of searching that a P2P application can control by adjusting the forwarding range of each query.

Although signaling overlays usually do not carry heavy traffic, high packet rate may become an issue for large flat overlays. To alleviate this problem multiple techniques have been used including the two-tier topology in modern Gnutella.

- **Structured overlays:** In structured overlays, also known as distributed hash tables (DHT), globally consistent protocols are used for neighbor selection and query routing in order to ensure efficient routing and resolution of queries. CAN [67], Chord [76], Pastry [71], Tapestry [86] and Kademlia [58] are the most well-known DHTs and we briefly discussed them earlier.

  Structured overlays can offer high levels of performance and efficiency. Most operations, such as joining the overlay and looking up a key value are performed in $O(log(n))$ where n denotes the overlay size. However, maintaining the overlay in presence of churn is often quite costly. When a peer leaves the DHT, its responsibility should be transferred to other peers. Also when a new peer joins the system, it should find its place and often load the keys previously stored in its responsibility zone from other peers. Additionally, most DHTs require periodic maintenance to keep the space allocation balanced and their routing tables up-to-date.

  Although most structured overlays are used for store-lookup services, there are exceptions such as Freenet [23] in which published files by the users are stored in the overlay in order to provide an anonymous and non-traceable file sharing environment.

3.1.2 Content Delivery Overlays

In this class of overlays, participating peers assist each other in downloading the content by contributing their upload *bandwidth.* The content is either a file or a stream which all peers are interested in receiving. The content is often broken into chunks and transmitted through the overlay and relayed by each peer to reach all other peers. In the traditional client-server content distribution, the server needs to have a large bandwidth as well as other resources in order to serve all the clients. However, in P2P content delivery, the source will only need to upload the content a small number of times (ideally once) and then the peers will download the content from each other. Thereby, content delivery overlays provide a *scalable*, *resilient* and *low cost* method for distributing large files and streams and therefore have become very popular.

In this section we divide the content delivery overlays from 3 aspects: *(i) content-type*, *(ii) overlay shape*, and *(iii) content delivery mechanism.*
Categorizing content delivery overlays based on the content type: The content distributed through the overlay may be a *file* or a *stream.*

- **File distribution overlays:** In these P2P applications, a file or a set of files are shared by a source among the participating peers. Although the downloading peer's goal is to complete the download as fast as possible, there is no hard timing constraints and therefore this class of content is also referred to as *elastic* content. BitTorrent [25] is the most popular P2P file distribution application. In BitTorrent, users interested in downloading the same file or set of files form a dynamic content delivery overlay. The files are divided into small blocks. Each peer receives the list of blocks available in its neighboring peers and subsequently sends requests for the blocks that it needs. While the *tit-for-tat* mechanism ensures bandwidth contribution by all peers, the *rarest-first* policy used by each peer for selecting which block to download, facilitates diffusion of all the blocks across the overlay. The content delivery method used by BitTorrent in which the data is broken down to small blocks which are undeterministically distributed in an overlay is also called *swarming*.

- **P2P streaming overlays:** In this class of P2P applications, multimedia streams are shared among interested users. In comparison to file distribution, streams are more challenging to distribute through an overlay due to strict timing requirements. In particular, each block of the stream will be useful at each peer only if it arrives before its playout time (non-elastic). Also, a sustained average delivery rate, equal to the stream bandwidth is necessary to each peer in order to ensure uninterrupted playback of the live stream.

  These overlays are used in delivering two types of streams. While some of the streaming overlays such as PRIME [56] and Coolstreaming [85] focus on delivering *live* audio-video streams, another group of P2P streaming applications such as Pando provide streaming of pre-recorded media often with VCR functionality. In pre-recorded streaming, a longer portion of the stream can be buffered to prevent interruptions during the playback, making the timing requirements looser. In live streaming, the amount of acceptable buffering is usually shorter. On the other hand, with pre-recorded streaming, participating peers play different parts of the stream at the same time and therefore the possibility of mutual uploading between two peers is very limited.

Categorizing content delivery overlays based on the overlay shape: Content delivery overlays are designed in one of the following shapes: (i) tree, (ii) multiple-tree, and (iii) mesh.

- **Tree:** In a *tree-based* overlay, peers form a single source-rooted tree and the content is distributed to all peers along the tree. In tree-based overlays such as Narada [43], each peer has only one parent from which all the content is downloaded. Tree-based overlays are simple to build, however they often suffer from multiple shortcomings including limited robustness and stability in presence of churn and limited scalability in terms of control overhead and latency.

- **Multiple-trees:**In more recent proposed works such as CoopNet [62] and Splitstream [16], *multiple trees* are built for content delivery overlays. In multiple-tree based overlays, the content (usually a stream) is divided to multiple parts and each part is delivered through one tree. This mechanism has three main advantages over a single tree approach: (i) In a single-tree overlay, the leaves do not contribute any bandwidth to the system while in multiple tree, each leave in one tree may have children in other trees. (ii) In a multiple-tree overlay, each peer's departure will disrupt receiving the content for all its descendents while in a multiple-tree overlay, each peer concurrently receives content from multiple parents and a temporary disconnection from one tree will only limit the rate or quality of the content. (iii) Peer heterogeneity can be supported in multiple-tree approach by joining a number of trees proportional to the peer bandwidth.

- **Mesh:** In a *mesh-based* overlay, such as BitTorrent [25] and PRIME [56], a random directed or undirected overlay among participating peers is formed. Each peer may download some part of the content from any of its neighbors. In contrast to tree-based approach, the mesh-based approach does not need to construct and maintain an explicit overlay structure for delivery of content to all peers. This further simplifies the overlay maintenance in presence of churn.

Categorizing content delivery overlays based on content delivery mechanism: According to their content delivery mechanism, content delivery overlays belong to either of the following groups:

- **Push:** In push-based content delivery, often used over tree-based overlays, each parent is responsible for forwarding the content to its children. The content flow to all peers is predetermined with the overlay shape. For instance, SplitStream [16] is a high-bandwidth content distribution system based on application-level multicast. In this application, multiple trees are formed and the shared stream is divided into multiple sub-streams, each pushed down through one of the trees. All proposed end-system multicast projects such as Narada [43], NICE [9] and Overcast [48] also follow the push mechanism.

- **Pull:** In pull-based content-delivery, usually used over mesh-based overlays, peers exchange their block availability status and then each peer requests or *pulls* the blocks it needs from neighbors who have them. With this mechanism, no peer is responsible for providing certain content to another and the data exchanges are based on availability and request. For instance, in BitTorrent [25] peers receive a *bitmap* depicting content availability at each neighbor and then use *rarest-first* policy to decide which blocks to request from their neighbors ensuring maximum block diversity in each neighborhood. Peers will only provide a sustained upload if the receiving party also provides them with a "high" upload rate on the blocks that they request. Non-contributing peers will get *choked* by other peers and may not receive a sustained download.

3.2 Characterizing P2P Overlays

Due to the increasing popularity of P2P applications, several research studies are published that try to characterize P2P applications through *(i) network measurement*, *(ii) modeling* and *(iii) simulation*. In this section, we review some outstanding examples of these research studies.

- **Network Measurement:** In this class of studies, Internet measurement is performed over an active P2P overlay in order to assess performance, show possible shortcomings or provide an analytical model.

  Saroiu *et al.* [72] perform a measurement study on Gnutella and Napster P2P overlays. They measure peer properties including session times and number of shared files, as well as network properties such as end-to-end latency, reported and available bandwidth. Their measurements show that there is significant heterogeneity and lack of cooperation across peers participating in Gnutella and Napster.

  Stutzbach *et al.* [77] introduce *cruiser* a high performance crawler for the Gnutella overlay. Using cruiser they capture full snapshots of the Gnutella overlay taken in a few minutes. They show that snapshots taken with slow crawlers lead to erroneous results biased towards short-lived peers. The authors observe an *Onion-like structure* according to which peer connectivity is related to uptime. Moreover, they show the existence of a *stable core* in Gnutella overlay that ensures reachability despite peer participation dynamics.

  Izal *et al.* [46] provide a measurement study of the BitTorrent using a 5-month long BitTorrent tracker log file. Using this source of information the authors capture several metrics related to a popular swarm including population, each peer's upload and download volumes and rates and downloading times. They show that the seeds (the peers who stay in the system after download completion) significantly contribute to the system and they show that BitTorrent can successfully sustain handle flash crowds.

  In our research work [65], we capture a large number of snapshots from the Gnutella overlay during a 15-month time-span. We characterize the evolution of Gnutella during this time period and show how the revisions of the popular Gnutella clients have effectively managed to keep the overlay balanced and efficient despite the population becoming quadrupled.

- **Modeling P2P applications:** Analytical and stochastic modeling is used in a number of research studies, in order to capture and explain some of their characteristics. Qiu and Srikant [63] propose a fluid model of BitTorrent using game theory and validate the model by simulation and experiments. They assign exponential distributions to peer arrival rate, abort rate and departure times and model peer evolution from the joining time until it leaves the system using a fluid model. They provide formulas for the number of seeds and downloaders and downloading time accordingly assuming a Nash equilibrium.

Some other research works also target modeling of different aspects of P2P applications. Ge *et al.* [36] model a generic P2P file sharing system as a multiple-class closed queuing network. Zou and Ammar [87] provide a “file-centric model” for P2P file sharing systems. In their model, they focus on a file’s movement through the system and its interaction with the peers.

None of the observed modeling studies focus on the overlay structure and characteristics.

- **Simulation studies on P2P applications:** Many research studies on P2P applications use some kind of simulation. Simulation is often used as a low cost method for evaluating a proposed system or a modification to an existing system. Simulations may be performed in different levels. A *session-level* simulation of a P2P system provides a simple environment for testing basic functionalities of a P2P system without getting involved in packet-level details and dynamics.

  Bharambe *et al.* [11] develop a session-level simulator of the BitTorrent system that models peer activity (joins, leaves, block exchanges) as well as many of the associated BitTorrent mechanisms (local rarest first, tit-for-tat). Using their simulator, they study effectiveness of BitTorrent’s mechanisms and show that their proposed technique can improve fairness in BitTorrent. As another example, in our paper [66], we perform session-level simulation of an unstructured P2P overlay and our proposed sampling technique, in order to study the effect of churn on the accuracy of sampling.

  A *packet-level* simulation is closer to a real experiment. Such simulations are often used in evaluating lower layer protocols such as congestion control, routing and data link layer protocols. However, in the cases that the packet dynamics are important to the applications functionality, they can also be used. The network simulator (NS-2) [45] is widely used by the researchers as a reliable and flexible packet-level simulator. For instance, Magharei *et al.* [56] implement their proposed P2P streaming application over NS-2 and use it to evaluate its functionality and performance.

  Although packet-level simulations are more realistic, they may not be used for simulating very large networks. In this case, session-level simulation may be used if the packet-level details are not very important.

4 Interactions between Overlay and Underlay

In this section we focus on the mutual effects, interactions and possible cooperation between the P2P overlay and the Internet underlay. A number of research studies have focused on the impact of the P2P overlays on the underlying network using measurement and simulation. We discuss this group of studies in Section 4.1. In Section 4.2, we discuss the unilateral efforts by the ISPs in limiting the impact of the P2P overlays and the network neutrality concept. Next, in Section 4.3, we overview a number of research studies proposing P2P overlays that try to minimize their impact on the underlying network, called ISP-friendly or network-aware overlays. Finally, Section 4.4, introduces a number of research projects and engineering efforts proposing cooperation between the overlay and the underlay in order to build overlays that are desirable for both underlay and the P2P application.

4.1 Overlay Impact on the Underlay

The direct effect of the overlay network on the underlay is the *traffic* associated with the P2P overlay that can lay a costly and unexpected load on the ISPs. As we discussed in detail in Section 1, the P2P traffic in costly for the ISPs because of its temporal pattern and symmetric load. In this section, we survey two research studies that try to characterize the impact of the P2P overlay on the ISPs. They both rely on packet traces captured at vantage points connecting an ISP or campus network to the Internet. They both show that the P2P traffic consumes a large portion of the gateway links and thereby they motivate modifications in the P2P overlays by localizing or caching in order to save a considerable amount of traffic on the Internet gateways of the ISPs.

- Karagiannis *et al.* [49] compare the load on the ISP for the cases of traditional client-server, P2P, local caching and their proposed mechanism. They propose a locality-aware overlay for peer-assisted content-delivery and show that its performance and the external load (impact on the ISP) is the best among the compared cases. They show that current P2P content distribution overlays (*e.g.*, BitTorrent) are not ISP-friendly because they generate a large amount of external traffic that can be avoided.

- Gummadi *et al.* [40] capture a 200-day trace of KazaA traffic at their campus gateway. They observe that most requests are for small files while most of the traffic volume is formed by large files. Although they do not capture the internal traffic, they show that there is a considerable amount of requests going outside the network while they can be resolved locally and therefore they suggest that a locality-aware scheme can help in reducing external traffic of KazaA.

4.2 Underlay Limiting Overlay

The ISPs have tried to control the P2P traffic in different ways. Toward this end, the P2P overlay traffic needs to be identified first. The simple methods of using TCP and UDP port numbers is now of little use because most P2P applications are flexible in the port number that they use and in some cases NAT traversal techniques - which are now very common - require using non-standard ports. There are a number of commercial protocol analyzers that combine a variety of techniques in order to identify the application responsible for each flow of traffic. These technique include deep packet inspection and traffic pattern analysis. Some researchers including Suh *et al.* [79] and Branch *et al.* [12] propose techniques based on temporal patterns of packets and packet sizes to identify Skype traffic.

The next step after identifying a P2P flow would be applying some type of restriction. The following methods have been reportedly used to contain or block the P2P traffic:

- *Packet Filtering:* This method requires implementation on routers, can be costly and limiting the router performance. In this method all packets identified as the target class will be dropped by the router. It will quickly alarm the users because their P2P applications will often stop working and therefore it is rarely used by ISPs who have to compete customer satisfaction. This method has been used in some campus residential networks where the users have limited options.

- *Traffic Policing:* This method (also known as rate limiting) also requires implementation on the routers however it is more flexible and less likely to be noticed by user. In this method, the network administrator defines an access-list that identifies target traffic flows and then assigns a maximum data rate to each class of traffic or to any flow belonging to that class. All packets exceeding the predefined maximum rates, will be dropped and as a result users will experience slow P2P transactions. Since the low speed may be associated with many factors including the P2P application itself, this method does not alert most of the P2P users against the ISP. Class based rate limiting can be technologically costly for the ISPs.

- *Connection Resetting:* This method has been reportedly used by some ISPs and its advantage is that the intervening device does not need to be on the path of the traffic therefore it can be implemented on a regular computer (not necessarily a router) with monitoring access to the traffic. In this technique after detecting a P2P flow, in order to terminate the connection, TCP reset packets are sent to both ends of the connection on behalf of the other end. In order to avoid alarming the users, this method can be applied on a random subset of the matching flows.

- *Transparent traffic redirection:* In this method, designed for localizing BitTorrent-like traffic, the ISP runs a transparent tracker proxy. When BitTorrent clients try to access a tracker to join a swarm, the connection will be redirected to the transparent proxy. The proxy server then controls the external traffic related to the swarm by connecting the local peers to each other and preventing local peers to connect to external peers. This method aims at smoothly limiting of the external traffic with minimum service degradation for the P2P application. However, in BitTorrent-like P2P overlays, certain level of random connectivity is needed to ensure that the blocks can diffuse all neighborhoods. Excessive localization will result in a heavily clustered overlay and thus may degrade the performance of the overlay by limiting the opportunity for peers to help each other.

Network Neutrality: All the methods described above, regardless of the technique used, are criticized by a large group of people in the networking community. They believe that the network should treat all packets equally regardless of the application they belong to. In other words the network should avoid discrimination among applications. This thesis, consistent with the end-to-end argument, is referred to as *network neutrality* and was the basis of the FCC's ruling against Comcast [26]. In this ruling, the Federal Communications Commission ordered Comcast, a large ISP with a national market in the U.S., to "end discriminatory network management practices".

4.3 Topologically Aware Overlays

In response to the ISP concerns, the P2P research community proposed ideas towards ISP-friendly P2P applications. The common goal across these research works is trying to decrease the inter-ISP traffic by increasing the relative number of local P2P connections and reducing number of external connections. Although these methods are often successful in limiting the ISP load, the effect on P2P performance is not evaluated from a neutral point of view. Additionally, since there is no authoritative topology or link cost information, such systems cannot use low cost external links or unpaid peering links between ISPs. In this section, we survey some outstanding research works on this topic.

- Ratnasamy *et al.* [68] suggest a binning scheme to find nearby nodes for peering and server selection. The bins are formed by sorted closeness to well-known landmarks(*e.g.*, 12 root DNS servers). They assign coordinates to each node in n-dimensional space where each dimension can take 3 values. The authors suggest a modification on CAN to selection node coordinates based on its network location.

- Harvey *et al.* [41] present *SkipNet*, a DHT-like overlay that allows for content locality and path locality. The locality is based on the node's DNS domain name.

- Kim and Chon [50] present a topologically-aware application-layer multicast overlay. In their scheme, close-by nodes are using network distance measurements to a few landmarks. Nodes are partitioned into topologically-aware clusters and local paths are determined between local nodes.

- Choffnes and Bustamante [22] propose *Ono*, . In Ono, nearby peers are identified according to their CDN server choice. In the CDNs they use, including Akamai and Limelight, a smart DNS server designates closest CDN server to each peer by a DNS lookup. Ono takes advantage of this system and tries to connect peers with the same CDN server together in order to *(i)* reduce the load on external ISP links, and *(ii)* improve system performance by avoiding bandwidth bottlenecks in the network. The authors claim average improvements of between 30% to 200% in download rates on the BitTorrent clients using Ono plugin. An advantage to previous works is that Ono does not need any costly network measurement or probing, instead, it only depends on periodic DNS lookups.

4.4 Cooperation between Overlay and Underlay

Considering the limitations of independent (unilateral) ISP-friendly P2P applications described earlier, it has become evident that the proper way to make the applications ISP-friendly is by using information provided by the ISP. *P4P* and *Oracle* were recently proposed based on the idea of an interface between the ISP and the P2P application over which the ISP shares information with the application regarding the ISP's relative preference among candidate peers. In addition to the mentioned research works, there have been ongoing efforts in IETF on the idea of Application Layer Traffic Optimization (ALTO). As a result, multiple Internet drafts have been published addressing different aspects of the problem and their proposed solutions. Below we provide an overview of the outstanding publications on this topic.

- Xie *et al.* [82] propose *P4P*, an interface that provides ISP preferences to the application layer in order to enable the application to redirect its traffic to satisfy the ISP preferences in its neighbor selection. In P4P, the ISP runs a server called iTracker which is aware of the ISP's topology, current link loads and costs associated to each link. The iTracker is then responsible for translating these factors into a single *cost* metric that can be looked up on a per-destination basis. The local application tracker (*e.g.*, BitTorrent tracker) should contact the iTracker to look up the cost values and include the ISP goals as well as the application goals in the neighbor selection process. The paper also demonstrates, using simulation and experiments that the method improves or at least maintains application performance while reducing the cost on the ISP.

- Aggarwal *et al.* [2] propose *Oracle*, an interface between ISP and P2P application that takes the list of prospective neighbors from each peer, sorts them according to the ISP preferences before they are returned to the peer. This method is simpler however it requires implementation in each application. Also, the scalability is questionable since the neighbor selection duty is on the shoulder of one server for each ISP.

- ALTO [44] is a working group in the Internet Engineering Task Force

(IETF) with the goal of "designing and specifying an Application-Layer Traffic Optimization (ALTO) service that will provide applications with information to perform better-than-random initial peer selection". Here we provide an overview of three Internet drafts published within this working group.

Seedorf and Burger [73] provide a problem statement of the application layer traffic optimization problem. According to their draft, in current P2P applications, peers choose neighbors without reliable information (*e.g.*, based on measurements or simply randomly) leading to suboptimal choices. This document describes problems related to optimizing traffic generated by peer-to-peer applications and associated issues. Such optimization problems arise in the use of network-layer information. Crowley [27] argues that the problem of P2P traffic optimization is not solved by standardization at this point due to lack of motivation in the user community. He suggests that ISPs should deploy pricing models based on the amount of each user's external traffic. Shalunov *et al.* [74] discuss the format and standardization of the ISP-P2P information export service. The suggested method is similar to P4P [82] and an ISP controlled agent sets priority values on each potential peering relationship. The peers will then select their neighbors according to their own preference, as well as the ISP's.

5 Conclusion

In this survey, we reviewed a number of important research studies on the P2P overlays, the underlying network, and their mutual impacts on each other. We cover a set of fundamental design and evaluation issues by surveying previous studies. We find and report an array of open problems and challenges in the covered area.

In Section 2, we studied research works on the AS-level and router-level topology on the Internet. We observed that one important challenge in studying Internet topology is gathering data that is reasonably complete. In studying AS-level topology, the hidden links between low-tiered ASes cause incomplete topology snapshots while in studying router-level topology, limitations of traceroute technique and blocking of probe packets cause incompleteness of the data.

In Section 3, research works on P2P overlays were surveyed. We categorized P2P overlays according to their function, structure, shape and content type. These differences have key importance when we study the mutual effects between the underlay and the overlay. One main challenge in P2P overlays is providing incentives for the users to contribute their resources. Towards this end, P2P applications should be designed with selfishness as a basis rather than depending on people's altruism. We observe that although several research works have been published on characterizing P2P applications, the attention on the overlay structure, specifically in modeling areas, has not been significant. Modeling of P2P overlays and their traffic is an important prerequisite for understanting the impact of the overlays on the underlay.

Finally, in Section 4, we provided a survey of the research and engineering efforts on the issues involving both the P2P overlay, and the underlying network. We observed that although there are methods proposed for network aware overlay construction with the cooperation of network layer, they are not widely deployed by the ISPs and P2P applications due to the lack of motivation on the user's side which depends on the P2P application performance. There is little unbiased study reporting significant benefits of such cooperation for the user and the P2P application. On the other hand, ISPs still have concerns about the possible abuses and vulnerabilities resulting from an ISP-P2P interface such as P4P.

References

[1] AFRINIC. African Network Information Center, 2009.

[2] Aggarwal, V., Feldmann, A., and Schneideler, C. Can ISPs and P2P systems cooperate for improved performance? *CCR 37*, 3 (July 2007), 29–40.

[3] Albert, R., Jeong, H., and Barabasi, A.-L. Error and attack tolerance of complex networks. *Nature 406*, 6794 (July 2000), 378–382.

[4] Alderson, D., Doyle, J., Willinger, W., and Govindan, R. Toward an Optimization-Driven Framework for Designing and Generating Realistic Internet Topologies. In *HotNets* (2002).

[5] Alderson, D., Li, L., Willinger, W., and Doyle, J. C. Understanding Internet topology: Principles, models, and validation. *IEEE/ACM Transactions on Networking 13*, 6 (2005), 1205–1218.

[6] APNIC. Asia Pacific Network Information Centre, 2009.

[7] ARIN. American Registry for Internet Numbers, 2009.

[8] Augustin, B., Cuvellier, X., Orgogozo, B., Viger, F., Friedman, T., Latapy, M., Magnien, C., and Teixeira, R. Avoiding traceroute anomalies with Paris traceroute. In *IMC* (2006).

[9] Banerjee, S., Bhattacharjee, B., and Kommareddy, C. Scalable Application Layer Multicast. In *SIGCOMM* (Aug. 2002).

[10] Barford, P., Bestavros, A., Byers, J., and Crovella, M. On the marginal utility of network topology measurements. In *IMW* (2001).

[11] Bharambe, A., Herley, C., and Padmanabhan, V. Analyzing and Improving a BitTorrent Network's Performance Mechanisms. In *INFOCOM* (Barcelona, Spain, Apr. 2006).

[12] Branch, P. A., Heyde, A., and Armitage, G. J. Rapid Identification of Skype Traffic. In *nossdav* (2009).

[13] Caida. Cooperative Association for Internet Data Analysis, 2008.

[14] CAIDA. skitter, 2008.

[15] Calvert, K., Doar, M., Nexion, A., and Zegura, E. Modeling Internet Topology. *IEEE Transactions on Communications* (Dec. 1997), 160–163.

[16] Castro, M., Druschel, P., Kermarrec, A.-M., Nandi, A., Rowstron, A., and Singh, A. SplitStream: High-bandwidth content distribution in a cooperative environment. In *International Workshop on Peer-to-Peer Systems* (2003).

[17] Chang, H., Govindan, R., Jamin, S., Shenker, S. J., and Willinger, W. Towards capturing representative AS-level Internet topologies. *Computer Networks Journal 44*, 6 (2004), 737–755.

[18] Chang, H., Jamin, S., and Willinger, W. Inferring AS-level internet topology from router-level path traces. In *SPIE ITCom* (2001).

[19] Chang, H., Jamin, S., and Willinger, W. Internet connectivity at the AS-level: An optimization-driven modeling approach. In *ACM SIGCOMM Workshop on MoMeTools* (Karlsruhe, Germany, 2003), pp. 33–46.

[20] Chen, Q., Chang, H., Govindan, R., Jamin, S., Shenker, S., and Willinger, W. The Origin of Power-Laws in Internet Topologies Revisited. In *INFOCOM* (New York, NY, June 2002).

[21] Cheswick. online, 2000.

[22] Choffnes, D. R., and Bustamante, F. E. Taming the torrent: A practical approach to reducing cross-ISP traffic in P2P systems. In *SIGCOMM* (Aug. 2008).

[23] Clarke, I., Sandberg, O., Wiley, B., and Hong, T. W. Freenet: A Distributed Anonymous Information Storage and Retrieval System. *Lecture Notes in Computer Science* (2000).

[24] Coates, M., Hero, A., Nowak, R., and Yu, B. Internet Tomography. *IEEE Signal Processing Magazine* (May 2002).

[25] Cohen, B. Incentives Build Robustness in BitTorrent. In *First Workshop on Economics of Peer-to-Peer Systems* (May 2003).

[26] Commission, F. C. Commission Orders COMCAST to End Discriminatory Network Management Practices, Aug. 2008.

[27] Crowley, P. On the Relative Importance of P2P Peer Selection. Internet Draft, July 2008.

[28] Dimitropoulos. Revealing the Autonomous System Taxonomy: The Machine Learning Approach. In *Passive and Active Measurements Workshop* (Mar. 2006).

[29] Dimitropoulos, X., Krioukov, D., Fomenkov, M., Huffaker, B., Hyun, Y., Claffy, K., and Riley, G. AS Relationships: Inference and Validation. *ACM SIGCOMM Computer Communication Review 37*, 1 (2007), 29–40.

[30] Doar, M. A Better Model for Generating Test Networks. In *Global Telecommunications Conference* (London, UK, Nov. 1996).

[31] Doyle, J. C., Alderson, D., Li, L., Low, S., Roughan, M., Shalunov, S., Tanaka, R., and Willinger, W. The "robust yet fragile" nature of the Internet. *Proceedings of the National Academy of Sciences 102*, 41 (2005), 14497–1452.

[32] Fabrikant, A., Koutsoupias, E., and Papadimitrio, C. H. Heuristically optimized trade-offs: A new paradigm for power laws in the Internet. In *ICALP* (2002).

[33] Faloutsos, M., Faloutsos, P., and Faloutsos, C. On Power-Law Relationships of the Internet Topology. In *SIGCOMM* (1999).

[34] Gao, L. On Inferring Autonomous System Relationships in the Internet. *IEEE/ACM Transactions on Networking 9* (2000), 733–745.

[35] Ge, Z., Figueiredo, D. R., Jaiswal, S., and Gao, L. On the Hierarchical Structure of the Logical Internet Graph. In *SPIE ITCom* (Denver, Colorado, USA, Nov. 2001).

[36] Ge, Z., Figueiredo, D. R., Jaiswal, S., Kurose, J., and Towsley, D. Modeling Peer-Peer File Sharing Systems. In *INFOCOM* (2003).

[37] Govindan, R., and Reddy, A. An Analysis of Internet Inter-Domain Topology and Route Stability. In *INFOCOM* (1997), pp. 850–857.

[38] Govindan, R., and Tangmunarunkit, H. Heuristics for Internet Map Discovery. In *INFOCOM* (Apr. 2000).

[39] Group, W. N. R. scamper, 2009.

[40] Gummadi, K. P., Dunn, R. J., Saroiu, S., Gribble, S. D., Levy, H. M., and Zahorjan, J. Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload. *ACM SIGOPS Operating Systems Review 37*, 5 (Dec. 2003), 314–329.

[41] Harvey, N. J., Jones, M. B., Saroiu, S., Theimer, M., and Wolman, A. SkipNet: A Scalable Overlay Network with Practical Locality Properties. In *USENIX Symposium on Internet Technologies and Systems* (2003).

[42] He, Y., Siganos, G., Faloutsos, M., and Krishnamurthy, S. V. A systematic framework for unearthing the missing links: Measurements and impact. In *NSDI* (Cambridge, MA, USA, Apr. 2007).

[43] hua Chu, Y., Rao, S. G., Seshan, S., and Zhang, H. A Case for End System Multicast. *IEEE Journal on Selected Areas in Communication (JSAC), Special Issue on Networking Support for Multicast 20*, 8 (2002).

[44] (IEFT), I. E. T. F. Application Layer Traffic Optimization (ALTO), 2008.

[45] ISI. The network simulator - ns-2, 2009.

[46] Izal, M., Urvoy-Keller, G., Biersack, E. W., Felber, P. A., Hamra, A. A., and Garces-Erice, L. Dissecting BitTorrent: Five Months in a Torrent's Lifetime. In *PAM* (Apr. 2004).

[47] Jacobson, V. traceroute, 1989.

[48] Jannotti, J., Gifford, D., Johnson, K. L., Kaashoek, M. F., and Jr., J. W. O. Overcast: Reliable Multicasting with an Overlay Network. In *OSDI* (Oct. 2000).

[49] Karagiannis, T., Rodriguez, P., and Papagiannaki, K. Should Internet Service Providers Fear Peer-Assisted Content Distribution? In *Internet Measurement Conference* (Berkeley, CA, Oct. 2005), pp. 63–76.

[50] Kim, Y., and Chon, K. Scalable and Topologically-aware Application-layer Multicast. In *Globecom* (Dallas, TX, Nov. 2004).

[51] LACNIC. Latin American and Caribbean Internet Addresses Registry, 2009.

[52] Lakhina, A., Byers, J. W., Crovella, M., and Xie, P. Sampling Biases in IP Topology Measurements. In *INFOCOM* (2003).

[53] Li, L., Alderson, D., Willinger, W., and Doyle, J. C. A first-principles approach to understanding the Internet's router-level topology. In *SIGCOMM* (2004), pp. 3–14.

[54] Li, M., Lee, W., and Sivasubramaniam, A. Semantic Small World: An Overlay Network for Peer-to-Peer Search. In *International Conference on Network Protocols* (Berlin, Germany, Oct. 2004).

[55] Luckie, M., Hyun, Y., and Huffaker, B. Traceroute Probe Method and Forward IP Path Inference. In *IMC* (2008).

[56] Magharei, N., and Rejaie, R. PRIME: Peer-to-Peer Receiver-drIven MEsh-based Streaming. In *INFOCOM* (Anchorage, Alaska, USA, May 2007), pp. 1415–1423.

[57] Mahadevan, P., Krioukov, D., Fomenkov, M., Huffaker, B., and Dimitropoulos, X. The Internet AS-level topology: Three data sources and one definitive metric. *Computer Communication Review 36*, 1 (2006), 17–26.

[58] Maymounkov, P., and Mazieres, D. Kademlia: A Peer-to-peer Information System Based on the XOR Metric. In *International Workshop on Peer-to-Peer Systems* (2002).

[59] Medina, A., Lakhina, A., Matta, I., and Byers, J. BRITE: An Approach to Universal Topology Generation,. In *MASCOTS* (Cincinatti, Aug. 2001), pp. 346–353.

[60] Medina, A., Matta, I., and Byers, J. On the Origin of Power Laws in Internet Topologies. *ACM Computer Communication Review* (2000).

[61] Muhlbauer, W., Feldmann, A., Maennel, O., Roughan, M., and Uhlig, S. Building an AS-topology model that captures route diversity. *ACM SIGCOMM Computer Communication Review 36*, 4 (Oct. 2006), 195–206.

[62] Padmanabhan, V. N., and Sripanidkulchai, K. The Case for Cooperative Networking. In *International Workshop on Peer-to-Peer Systems* (2002).

[63] Qiu, D., and Srikant, R. Modeling and Performance Analysis of Bit Torrent-Like Peer-to-Peer Networks. In *SIGCOMM* (2004).

[64] RADB. Routing Assets Database, 2009.

[65] Rasti, A., Stutzbach, D., and Rejaie, R. On the Long-term Evolution of the Two-Tier Gnutella Overlay. In *Global Internet* (Barcelona, Spain, Apr. 2006).

[66] Rasti, A., Torkjazi, M., Rejaie, R., Duffield, N., Willinger, W., and Stutzbach, D. Respondent-driven Sampling for Characterizing Unstructured Overlays. In *IEEE INFOCOM Mini-conference* (2009).

[67] Ratnasamy, S., Francis, P., Handley, M., Karp, R., and Shenker, S. A Scalable Content-Addressable Network. In *SIGCOMM* (2001).

[68] Ratnasamy, S., Handley, M., Karp, R., and Shenker, S. Topologically-Aware Overlay Construction and Server Selection. In *INFOCOM* (June 2002).

[69] RIPE. European IP Networks, 2009.

[70] Roughan, M., Tuke, S. J., and Maennel, O. Bigfoot, sasquatch, the yeti and other missing links: what we don't know about the as graph. In *IMC* (Vouliagmeni, Greece, Oct. 2008).

[71] Rowstron, A., and Druschel, P. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In *IFIP/ACM International Conference on Distributed Systems Platforms (Middleware)* (Heidelberg, Germany, Nov. 2001), pp. 329–350.

[72] Saroiu, S., Gummadi, P. K., and Gribble, S. D. Measuring and Analyzing the Characteristics of Napster and Gnutella Hosts. *Multimedia Systems Journal 9*, 2 (Aug. 2003), 170–184.

[73] Seedorf, J., and Burger, E. Application-Layer Traffic Optimization (ALTO) Problem Statement. Internet Draft, 2008.

[74] Shalunov, S., Penno, R., and Woundy, R. ALTO Information Export Service. Internet Draft, Oct. 2008.

[75] Spring, N., Mahajan, R., Wetherall, D., and Anderson, T. Measuring ISP Topologies with Rocketfuel. In *SIGCOMM* (2002).

[76] Stoica, I., Morris, R., Liben-Nowell, D., Karger, D. R., Kaashoek, M. F., Dabek, F., and Balakrishnan, H. Chord: A Scalable Peer-to-peer Lookup Protocol for Internet Applications. *IEEE/ACM Transactions on Networking* (2002).

[77] Stutzbach, D., Rejaie, R., and Sen, S. Characterizing Unstructured Overlay Topologies in Modern P2P File-Sharing Systems. *TON 16*, 2 (Apr. 2008).

[78] Subramanian, L., Agarwal, S., Rexford, J., and Katz, Y. H. Characterizing the Internet hierarchy from multiple vantage points. In *INFOCOM* (2002).

[79] Suh, K., Figueiredo, D., Kurose, J. F., and Towsley, D. Characterizing and detecting Skype-Relayed Traffic. In *INFOCOM* (Barcelona, Spain, Apr. 2006).

[80] the Web Information Company, A. Alexa, 2009.

[81] Xia, J., and Gao, L. On the Evaluation of AS Relationship Inferences. In *Globecomm* (2004).

[82] Xie, H., Yang, Y. R., Krishnamurthy, A., Liu, Y., and Silberschatz, A. P4P: Provider Portal for Applications. In *SIGCOMM* (2008).

[83] Yook, S.-H., Jeong, H., and Barabasi, A.-L. Modeling the Internet's large-scale topology. *pnas 99*, 21 (Oct. 2002), 13382–13386.

[84] Zhang, B., and Liu, R. Collecting the Internet AS-level Topology. *ACM CCR 35* (2005), 53–61.

[85] Zhang, X., Liu, J., and shing Peter Yum, T. Coolstreaming/donet: A data-driven overlay network for peer-to-peer live media streaming. In *INFOCOM* (Miami, FL, Mar. 2005).

[86] Zhao, B. Y., Huang, L., Stribling, J., Rhea, S. C., Joseph, A. D., and Kubiatowicz, J. D. Tapestry: A Resilient Global-Scale Overlay for Service Deployment. *IEEE Journal on Selected Areas in Communications 22*, 1 (Jan. 2004), 41–53.

[87] Zou, L., and Ammar, M. A File-Centric Model for Peer-to-Peer File Sharing Systems. In *International Conference on Network Protocols* (Atlanta, GA, Nov. 2003).