

Internet Inter-Domain Traffic

Craig Labovitz, Scott Iekel-Johnson,
Danny McPherson
Arbor Networks
Ann Arbor, MI
{labovit, scottij, danny}@arbor.net

Jon Oberheide, Farnam Jahanian
University of Michigan
Ann Arbor, MI
{jonojono, farnam}@umich.edu

ABSTRACT

In this paper, we examine changes in Internet inter-domain traffic demands and interconnection policies. We analyze more than 200 Exabytes of commercial Internet traffic over a two year period through the instrumentation of 110 large and geographically diverse cable operators, international transit backbones, regional networks and content providers. Our analysis shows significant changes in inter-AS traffic patterns and an evolution of provider peering strategies. Specifically, we find the majority of inter-domain traffic by volume now flows directly between large content providers, data center / CDNs and consumer networks. We also show significant changes in Internet application usage, including a global decline of P2P and a significant rise in video traffic. We conclude with estimates of the current size of the Internet by inter-domain traffic volume and rate of annualized inter-domain traffic growth.

Categories and Subject Descriptors: C.2 [Computer Communication Networks]: Miscellaneous

General Terms: Measurement.

1. INTRODUCTION

Saying the Internet has changed dramatically over the last five years is cliché – the Internet is **always** changing dramatically: fifteen years ago, new applications (*e.g.*, the web) drove widespread consumer interest and Internet adoption. Ten years ago, new backbone and subscriber access technologies (*e.g.*, DSL/Cable broadband) significantly expanded end-user connections speeds. And more recently, applications like social networking and video (*e.g.*, Facebook and YouTube) again reshaped consumer Internet usage.

But beyond the continued evolution of Internet protocols and technologies, we argue the last five years saw the start of an equally significant shift in Internet inter-domain traffic demands and peering policies. For most of the past fifteen years of the commercial Internet, ten to twelve large transit providers comprised the Internet “core” interconnecting thousands of tier-2, regional providers, consumer networks

and content / hosting companies. Textbook diagrams of the Internet and research publications based on active probing and BGP routing table analysis generally produce logical Internet maps similar to Figure 1a [1]. This diagram shows a strict hierarchy of global transit providers at the core interconnecting smaller tier-2 and regional / tier-3 providers.

Over the past several years industry economic forces, including the continued decline of the price of IP wholesale transit and the growth of advertisement-supported content, significantly altered the interconnection strategies of many providers [2]. In the emerging new Internet economy, content providers build their own global backbones, cable Internet service providers offer wholesale national transit, and transit ISPs offer CDN and cloud / content hosting services [3, 4, 5, 6]. For example, we found that over the last two years Google migrated the majority of its video and search traffic (which we later show constitutes more than 5% of all inter-domain traffic) away from transit providers to its own fiber backbone infrastructure and direct interconnects with consumer networks.

The substantial changes in provider inter-connection strategies have significant ongoing implications for backbone engineering, design of Internet-scale applications, and research. However, most providers treat their Internet traffic statistics with great commercial secrecy as these values reveal insights into market penetration and competitive strategies. As a result, the significant shift in Internet inter-domain traffic patterns has gone largely undocumented in the commercial and research literature.

Most Internet traffic research has typically focused on secondary indicators of Internet traffic such as BGP route advertisements [7, 8, 9], DNS probing [10], broad industry surveys [11], private CDN statistics [12], or traffic measured on an individual provider or enterprise network [13].

A few more closely related efforts have studied global Internet traffic using publicly available exchange point statistics [14] or a small set of residential networks [15, 16, 17, 18]. Still other work used industry surveys and targeted discussions with providers [19, 20, 21]. Finally, traceroute analysis in [22] also identified a topological trend towards a more densely interconnected Internet especially with respect to large content providers.

In this paper, we provide one of the first large scale longitudinal studies of Internet inter-domain traffic using direct instrumentation of peering routers across multiple providers. We address significant experimental data collection and commercial privacy challenges to instrument 3,095 peering routers across 18 global carriers, 38 regional / tier-

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

SIGCOMM’10, August 30–September 3, 2010, New Delhi, India.
Copyright 2010 ACM 978-1-4503-0201-2/10/08 ...\$10.00.

2, and 42 consumer and content providers in the Americas, Asia, and Europe. At its peak, the study monitored more than 12 terabits per second of offered load and a total of more than 200 exabytes of Internet traffic over the two-year life of the study (July 2007 to July 2009). Based on independent estimates of total Internet traffic volume in [14, 23], we believe the probes directly monitor more than 25% of all Internet inter-domain traffic.

Our major findings include:

- **Evolution of the Internet “Core”:** Over the last two years, the majority of Internet inter-domain traffic growth has occurred outside the traditional ten to twelve global transit carriers. Today, most Internet inter-domain traffic by volume flows directly between large content providers, hosting / CDNs and consumer networks.
- **Consolidation of Content:** Most content by inter-domain traffic volume has migrated to a relatively small number of large hosting, cloud and content providers. Out of the approximately thirty-thousand ASNs in the default-free BGP routing tables [24], 30 ASNs contribute a disproportionate average of 30% of all Internet inter-domain traffic in July 2009.
- **Estimation of Google’s Traffic Contribution:** At an average of more than 5% of all inter-domain traffic in July 2009, Google represents both the largest and fastest growing contributor of inter-domain traffic. Google’s share of all inter-domain traffic grew by more than 4% between July 2007 and July 2009.
- **Consolidation of Application Transport:** The majority of inter-domain traffic has migrated to a relatively small number of protocols and TCP / UDP ports, including video over HTTP and Adobe Flash. Other mechanisms for video and application distribution like P2P have declined significantly in the last two years.
- **Estimation of Internet Size:** Using data from independent known inter-domain provider traffic volumes, we estimate both the volume and annualized growth rate of all inter-domain traffic. As of July 2009, we estimate inter-domain traffic peaks exceed 39 Tbps and grew an annualized average of 44.5% between July 2007 and 2009.

The rest of this report is organized as follows: §2 provides an overview of our data collection infrastructure and analysis methodology. §3 discusses significant changes in Internet topology and commercial interconnection relationships between providers. §4 analyzes changes in Internet protocols and applications. Finally, we conclude with validation of our data and estimates of both the volume of all inter-domain traffic and annualized rate of growth.

2. METHODOLOGY

Our analysis in this paper is based on traffic statistics exported by operational routers from a large and, we argue later, representative sample of Internet providers. Specifically, we leverage a widely deployed commercial security and traffic monitoring platform to instrument the BGP peering

edge routers of 110 participating Internet providers. Based on private commercial sales data, we believe the majority of the probe deployments enjoy complete coverage of the provider’s BGP peering edge. However, we lack specific visibility into the network probe coverage of any individual anonymous study participant.

The instrumented routers export both traffic flow samples (*e.g.*, NetFlow, cFlowd, IPFIX, or sFlow) and participate in routing protocol exchange (*i.e.*, iBGP) with one or more probe devices. A smaller number of providers have deployed inline or “port span” versions of the appliances to monitor traffic payloads and enact security policies. Per our anonymity agreement with participating providers, we did not collect more specific details on deployment configuration (*e.g.*, flow sample rates, router model number, etc.).

While sampled flow introduces potential data artifacts particularly around short-lived flows [25], we believe the accuracy of flow is sufficient for the granularity of our inter-domain traffic analysis. Further, we argue flow provides the only scalable and cost-effective monitoring approach given the scale of our study.

Each probe independently calculates traffic statistics based on user configured information and BGP learned topology. Calculated statistics include breakdowns of traffic per BGP autonomous system (AS), ASPath, network and transport layer protocols, ports, nexthops, and countries. A more detailed description of the probe capabilities is available in commercial datasheets and white papers at [26].

The probe configuration includes user supplied classification of the probe’s primary geographic coverage area (*i.e.*, North America, Europe, etc.) as well as market segment (*i.e.*, tier-1, tier-2, content, consumer or educational). We use the provider supplied self-categorizations in our aggregate data analysis discussed in later Sections.

We worked extensively with the provider community to address commercial privacy concerns. For example, every participating probe strips all provider identifying information from the calculated statistics before forwarding an encrypted and authenticated snapshot of the data to central servers. We also agreed to not publish any per provider traffic rates nor customer data derived from ASPath traffic analysis.¹

We pursued several approaches to mitigate sources of possible error in the data. We began by excluding three ISPs (out of 113) from the dataset that exhibited signs of obvious misconfiguration via manual inspection (*i.e.*, wild daily fluctuations, unrealistic traffic statistics, internally inconsistent data, etc.).

Unfortunately, our measurement infrastructure suffered from the real-world operational exigencies of providers. Throughout the course of the study, providers expanded deployments with new probes, decommissioned older appliances and otherwise modified the configuration of their probes and backbone infrastructure. As a result, the absolute traffic volumes reported by probes exhibited occasional discontinuities. For example, one probe consistently

¹While we discuss several Internet providers by name in this paper, we base all provider-specific analysis on anonymized ASN and ASPath datasets aggregated across all study participants. Any overlap or correlation with providers who may (or may not) be sharing data or have research or commercial affiliations with the institutions or authors of this paper is unintended and coincidental.

| Segment | Percentage |
|--------------------------|------------|
| Regional / Tier2 | 34 |
| Global Transit / Tier1 | 16 |
| Unclassified | 16 |
| Consumer (Cable and DSL) | 11 |
| Content / Hosting | 11 |
| Research/ Educational | 9 |
| CDN | 3 |

(a) Market Segment

| Region | Percentage |
|---------------|------------|
| North America | 48 |
| Europe | 18 |
| Unclassified | 15 |
| Asia | 9 |
| South America | 8 |
| Middle East | 1 |
| Africa | 1 |

(b) Geographic Region

Table 1: Distribution of anonymous Internet provider participants in our study by market segment and geographic region.

reported hundreds of gigabits of traffic until dropping to zero abruptly in early 2009 as the provider migrated traffic to other routers and newer probe appliances.

The probe data exhibited less variance with respect to traffic ratios (*i.e.*, the ratio of ASN, port, protocol, etc. to all inter-domain traffic in each deployment). Specifically ratios such as TCP port 80 or Google ASN origin traffic remained relatively consistent even as the number of monitored routers, probe appliances and absolute volume of reported traffic fluctuated in a deployment. Given the relative consistency of ratios and our inability to distinguish changes in absolute traffic volumes from artifacts due to provider measurement infrastructure changes, most of the analysis in this paper focuses on traffic percentages (*i.e.* share of traffic) rather than absolute traffic values. The focus on ratios also simplifies our aggregate analysis across a large set of heterogeneous providers.

Throughout every 24 hour period, the probes independently calculated the average traffic volume every five minutes for all members of all datasets (*i.e.*, traffic contributed by every nexthop, AS Path, ASN, etc.) as well as the average volume of total inter-domain network traffic. The probes then calculated a 24 hour average for each of these items using the five minute averages. Finally, the probes used the daily traffic volume per item and network total to calculate a daily percentage for each item.

The first chart in Table 1 provides a market segment breakdown of anonymous provider participants by percentage of all deployments in our study. The second table shows a breakdown of percentage of deployments by geographic region. Regional and tier-2 providers comprise the largest component at 34% of anonymous statistics followed by unclassified and tier-1 at 16% each.

We observe that the relative high cost of the commercial probes used in our study may introduce a selection bias towards larger providers. We further note that both analyst data and our study participant set reflect a continued weighting towards North America and Europe both in traffic volume and number of providers [27, 11, 6, 28].

While our study included a large and diverse set of Internet providers, evaluation of sample bias is a challenge given the anonymity of the study participants and the lack

of “ground-truth” quantitative market data (*i.e.*, most available data on provider Internet traffic volumes is based on qualitative surveys [27, 11]).

We evaluated several mechanisms for weighting the traffic ratio samples from the 110 deployments to reduce selection bias. However, the anonymity of the study participants and the narrow scope of our data collection provided a limited number of weighting options. Ultimately, we found a weighted average based on the number of routers in each deployment provided the best results during data validation in §5 and represents a compromise between the relative size of an ISP while not obscuring data from smaller networks.

Specifically, for each day d we calculate the weighted average percent share of Internet traffic $P_d(A)$ for a specific traffic attribute A , where A is an ASN, TCP port, country of origin, etc. The weights are calculated based on the total number of routers reporting traffic on that day at each of the N study participants reporting data for that day. Thus, on day d for participant i with router count $R_{d,i}$ we calculate the weight:

$$W_{d,i} = \frac{R_{d,i}}{\sum_{x=1}^N R_{d,x}}$$

We then calculate day d ’s weighted average percent share $P_d(A)$ based on each provider’s measured average traffic volume for A on day d , $M_{d,i}(A)$, and total average inter-domain traffic for day d , $T_{d,i}$. This gives a weighted average percent share of traffic for A as

$$P_d(A) = \sum_{x=1}^N W_{d,x} * \frac{M_{d,x}(A)}{T_{d,x}} * 100$$

We excluded any provider more than 1.5 standard deviations from the true mean in order to focus on values that were less likely to have measurement errors due to transient provider issues (misconfiguration, network problems, or probe failures). With the exception of Comcast’s peering ratios discussed in §3, we used the sum of traffic both in and out of the provider networks for $M_{d,i}(A)$ and $T_{d,i}$.

In some cases, our analysis may underestimate categories of inter-domain traffic. Specifically, the probes lack visibility into traffic exchanged over direct peering adjacencies between enterprise business partners or between smaller tier-2 and tier-3 Internet edge providers. Similarly, the study may underestimate inter-domain traffic associated with large content providers such as Google who are increasingly pursuing edge peering strategies. We also emphasize that our study is limited to inter-domain traffic and excludes all internal provider traffic, such as intra-domain cache traffic, VPNs, IPTV and VoIP services.

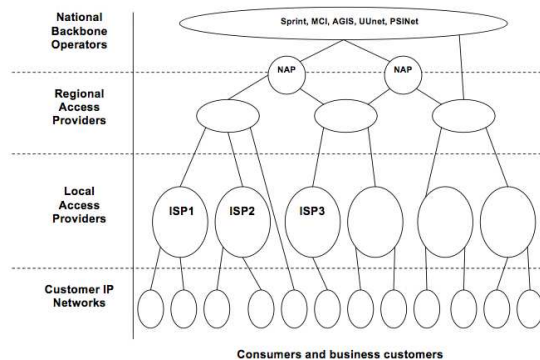
Finally, we validated our findings with private discussions with more than twenty large content providers, transit ISPs and regional networks. These discussions provided “ground-truth” and additional color to better understand the market forces underlying our observed inter-domain traffic trends. We note that our derived data matched provider expectations both in relative ordering and magnitude of ASN traffic volumes. In addition, twelve providers supplied independent inter-domain traffic measurements for validation of our analysis. We use these twelve known provider traffic values in §5 to add confidence to our calculated inter-domain ASN traffic distributions as well as to estimate the overall volume of global inter-domain traffic.

3. ASN TRAFFIC ANALYSIS

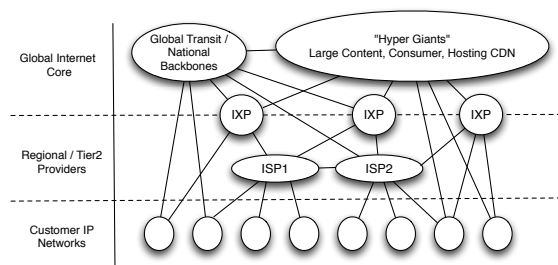
In this section, we present a coarse grained analysis of changes in inter-domain traffic patterns. We begin with a look at the ten largest contributors (based on our analysis) of inter-domain traffic in the months of July 2007 and July 2009. With the exception of content providers (*i.e.*, Google, Microsoft) and Comcast, we anonymize provider names in sensitivity to the potential commercial impact of this data.

3.1 Provider Inter-domain Traffic Share

We calculate the ten largest contributors of inter-domain traffic in the first two charts of Table 2 using the weighted average percentage of inter-domain traffic (*i.e.*, $P(A)$) reported by each Internet provider in our study either originating or transiting each ASN A . We then aggregate all ASNs which are managed by the same Internet commercial entity (*e.g.*, Verizon’s AS701, AS702, etc.). This last step is required since many large transit providers manage dozens of ASNs reflecting geographic backbone segmentation and merger or acquisition lineage. Finally, we exclude stub ASNs from the aggregation step which we only observed downstream from other corporate ASN (*e.g.*, DoubleClick (AS 6432) traffic transits Google (AS 15169) in all our observed ASPaths).



(a) Traditional Internet logical topology



(b) Emerging new Internet logical topology

Figure 1: The hierarchical old and more densely interconnected emerging Internet. Figure A generally reflects historical BGP topology while Figure B illustrates emerging dominant Internet traffic patterns.

As a category, the ten largest providers by inter-domain traffic volume in Table 2a account for 28.8% of all inter-domain traffic. ISP A represents the largest provider traffic share in 2007 with an average of 5.77% of all inter-domain traffic, followed by ISP B (4.55%) and ISP C (3.35%).

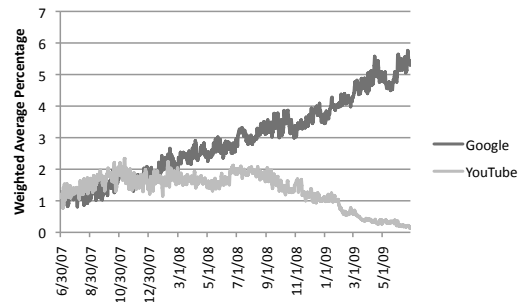


Figure 2: Growth in Google inter-domain traffic contribution. Graph shows weighted average percent of all inter-domain traffic contributed by YouTube and Google ASNs. Over time, Google migrated YouTube traffic and back-end infrastructure into Google peering / transit and data centers.

Our analysis of traffic data from July 2007 suggests traffic patterns consistent with that of logical topological textbook diagrams in Figure 1a. Specifically, we find the largest Internet providers by inter-domain traffic volume correlate with the twelve largest transit networks popularly regarded as the global transit core [29].

In the second chart of Table 2b, we show the impact of subsequent commercial policy and traffic engineering changes on the ten largest Internet providers by inter-domain traffic contribution as of July 2009. We note that the 2009 list includes significant variance from 2007, including the addition of non-transit companies to the list. Specifically, both a content provider (Google) and a consumer network (Comcast) now rival several global transit networks in inter-domain traffic contribution. Provider A and B continue to hold the top two spots at 9.4 and 5.7 percent of all inter-domain traffic, respectively. We discuss both Google and Comcast in more detail later in this Section.

Table 2c provides another view of the data showing the gain in providers’ average percentage of all inter-domain traffic between July 2007 and July 2009. We note that growth in this table requires a provider gain “market share”, *i.e.*, the provider exceed the overall growth of inter-domain traffic (currently growing at 35-45% annualized).

Google inter-domain traffic enjoyed the largest growth in our two year study period by gaining 4% of all inter-domain traffic. Figure 2 provides the weighted average percent of inter-domain traffic due to Google ASNs (including properties) and YouTube (AS36561) between July 2007 and July 2009.

Discussions with providers and analysis of the data in Figure 2 suggests much of Google’s traffic share increase came through the post-acquisition migration of YouTube inter-domain traffic to Google’s ASNs (from both LimeLight and YouTube ASN) [30]. At the start of the study period, both Google and YouTube represent slightly more than 1% of all inter-domain traffic. Figure 2 shows YouTube ASN inter-domain traffic decreasing as Google traffic continues to grow through the summer of 2009.

ISP A and ISP B also showed significant growth in Table 2c. Private discussion with analysts and providers sug-

| Rank | Provider | Percentage | Rank | Provider | Percentage | Rank | Provider | Increase in Traffic Share |
|------|----------|------------|------|----------|------------|------|-----------|---------------------------|
| 1 | ISP A | 5.77 | 1 | ISP A | 9.41 | 1 | Google | 4.04 |
| 2 | ISP B | 4.55 | 2 | ISP B | 5.7 | 2 | ISP A | 3.74 |
| 3 | ISP C | 3.35 | 3 | Google | 5.2 | 3 | ISP F | 2.86 |
| 4 | ISP D | 3.2 | 4 | ISP F | 5.0 | 4 | Comcast | 1.94 |
| 5 | ISP E | 2.6 | 5 | ISP H | 3.22 | 5 | ISP K | 1.60 |
| 6 | ISP F | 2.77 | 6 | Comcast | 3.12 | 6 | ISP B | 1.36 |
| 7 | ISP G | 2.24 | 7 | ISP D | 3.08 | 7 | ISP H | 1.21 |
| 8 | ISP H | 1.82 | 8 | ISP E | 2.32 | 8 | ISP L | 0.66 |
| 9 | ISP I | 1.35 | 9 | ISP C | 2.05 | 9 | Microsoft | 0.62 |
| 10 | ISP J | 1.23 | 10 | ISP G | 1.89 | 10 | Akamai | 0.06 |

(a) Top Ten 2007

(b) Top Ten 2009

(c) Top 10 Growth

Table 2: The ten largest contributors of inter-domain traffic by weighted average percentage of all Internet inter-domain traffic. Includes average percentage of all traffic from study participants originating, terminating, or transiting the ASNs managed by each provider in July 2007 and July 2009. The third table includes providers with the most significant inter-domain traffic share growth over the two-year study period.

gest these providers enjoyed growth both due to their CDN business (ISP A) and role providing transit to large content providers (both ISP A and ISP B). Comcast also showed significant growth with a gain of close to 2% of all inter-domain traffic.

We briefly focus on changes in Comcast’s inter-domain traffic contribution as an illustration of possible commercial policy and traffic engineering changes belying some of the results in Table 2c. In 2007, we found Comcast inter-domain traffic share (distributed across a dozen regional ASN) represented less than 1% of all inter-domain traffic. Also in 2007, Comcast inter-domain traffic patterns resembled that of most traditional consumer providers with traffic ratios of 7:3, or the majority (70%) of traffic coming into Comcast. In the language of the industry, Comcast represented a typical “eyeball” consumer network [19, 6].

Figure 3a shows the weighted average percent of all inter-domain traffic both a) originating or terminating in Comcast managed ASNs (*i.e.*, origin) and b) transiting Comcast to reach other ASNs (*i.e.*, transit). In the summer of 2007, Comcast origin traffic contributed an average of 0.13% of all inter-domain traffic – a percentage in line with other large North American cable operators. During the same time period, Comcast transit traffic represented 0.78% of all inter-domain traffic. While Comcast origin traffic saw modest growth over the two year study period, the majority of Comcast’s traffic increase stemmed from transit – nearly a 4x growth.

Figure 3b shows another view of the data. We calculate the weighted average percentage of inter-domain traffic into all Comcast ASNs versus outbound. We use this In / Out peering ratio as an approximation of the Comcast’s content contribution versus consumption. The graph shows that over the two year study period Comcast’s traffic ratios inverted with the cable operator becoming a net Internet inter-domain contributor by July of 2009.

Discussions with analysts and ISPs provide some insight into Comcast’s transformation. Over the last five years, Comcast executed on a number of technology and business strategies, including consolidation of several disparate regional backbones into a single nationwide network and roll-out of a “triple play” (voice, video, data) consumer product. Most significantly, Comcast began offering wholesale transit (GigE and 10GigE IP), cellular backhaul and IP video distribution (though Comcast Media Center subsidiary) [6].

3.2 Inter-domain Traffic Consolidation

In this subsection, we explore consolidation in inter-domain traffic demands. We argue the growth of Google, Comcast, Microsoft and Akamai traffic in Table 2c provides a bellwether of broader traffic engineering, commercial expansion and content consolidation trends.

We first aggregate the 200 fastest growing ASNs described earlier in this Section into four broad categories using classifications from CAIDA [31] and manual inspection. As a category, ASNs in the content / hosting group grew by 58%, and consumer networks by 38%, while tier-1/2 both grew under 28% (*i.e.*, less than the average rate of aggregate inter-domain growth).

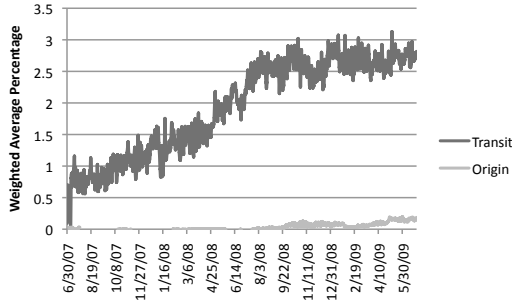
While tier-1 providers still carry significant volumes of traffic, observed Internet inter-domain traffic patterns in July 2009 suggest Figure 1b. In this emerging new Internet, the majority of traffic by volume flows directly between large content providers, datacenter / CDNs and consumer networks. In many cases, CDNs (*e.g.*, Akamai, LimeLight) and content providers (*e.g.*, Google, Microsoft, Facebook) are directly interconnected with both consumer networks and tier-1 / tier-2 providers.

Figure 4 shows a graph of the cumulative distribution of the weighted average percentage of all inter-domain traffic per origin ASN. The vertical axis shows the cumulative percentage and the horizontal axis provides the number of unique ASNs in both July 2007 and 2009.

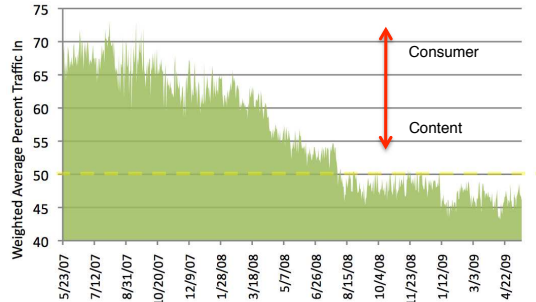
The main interpretation of the graph in Figure 4 is that as of July 2009, 150 ASNs originate more than 50% of all inter-domain traffic. The remainder of inter-domain traffic originates across a heavy-tailed distribution of the other 30,000 BGP ASNs. If traffic were evenly distributed across all ASNs, we would expect the top 150 ASNs to contribute only 0.15% of inter-domain traffic. By way of comparison, the top 150 ASNs contributed only 30% of all inter-domain traffic in July of 2007.

We observe that the Internet ASN traffic distribution in Figure 4 approximates a power law distribution. While discussion of power law properties and processes is beyond the scope of this paper, we note power laws have been observed (and debated) in Internet AS-level topology [32].

Table 3 shows the top ten origin ASNs as a weighted average percentage of all inter-domain traffic during July 2009. As of July 2009, Google’s origin ASNs contribute a weighted



(a) Origin and transit growth



(b) Ratio change

Figure 3: Changes in Comcast inter-domain traffic patterns between July 2007 and July 2009. The first graph shows weighted average percentage of inter-domain traffic originating / terminating and transiting Comcast ASNs. The second graphs shows the change in Comcast In / Out peering ratio over the two year period.

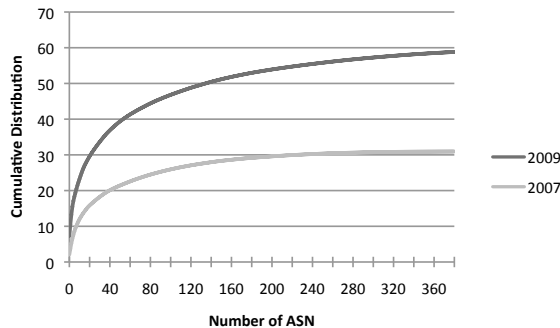


Figure 4: Graph shows the cumulative distribution of inter-domain traffic contributed by origin ASNs by weighted average percentage of all inter-domain traffic throughout the months of July 2007 and July 2009.

average 5% of all inter-domain traffic followed by ISP A’s enterprise / CDN business at 1.7% and LimeLight at 1.52%.

CDNs comprise one of the largest categories of consolidating traffic sources in Figure 4. As a grouping, we estimate CDNs contribute a weighted average percentage of approximately 10% of all Internet inter-domain traffic as of July 2009. We further observe that our estimates may significantly underestimate CDN contribution as we cannot easily distinguish CDN traffic from other sources of data within providers, *e.g.*, between CDN, transit, hosting, etc. We also note that our inter-domain analysis excludes Akamai CDN traffic since most Akamai content is served from caches co-located within provider infrastructure and IP address space.

Finally, we observe these large sources of inter-domain traffic are also increasingly highly interconnected with other providers and each other. We analyze the percentage of anonymous providers in our study utilizing a direct peering adjacency with each large content ASN in Table 3. We focus only on adjacencies representing the majority of traffic between the anonymous study provider and content ASN (*i.e.*, we exclude backup and secondary BGP paths). We

| Rank | Provider | Percentage |
|------|-------------------|------------|
| 1 | Google | 5.03 |
| 2 | ISP A | 1.78 |
| 3 | LimeLight | 1.52 |
| 4 | Akamai | 1.16 |
| 5 | Microsoft | 0.94 |
| 6 | Carpathia Hosting | 0.82 |
| 7 | ISP G | 0.77 |
| 8 | LeaseWeb | 0.74 |
| 9 | ISP C | 0.73 |
| 10 | ISP B | 0.70 |

Table 3: Top ten origin ASNs as an average weighted percentage of all inter-domain traffic in July 2009.

find that as of July 2009, the majority (65%) of study participants use a direct adjacency with Google. Similarly, 52% maintained a direct peering relationship with Microsoft, 49% with Limelight and 49% with Yahoo.

4. APPLICATION TRAFFIC ANALYSIS

In this section we explore changes in Internet inter-domain application traffic patterns between July 2007 and 2009. We examine both protocols and grouped TCP / UDP per port breakdowns of inter-domain traffic across different provider groupings.

We first provide some additional methodology. The commercial appliances used in this study classify applications by protocol and TCP/UDP port in the flow record. Since each flow record may contain multiple port numbers, the appliances follow heuristics (such as preferring a well-known port over an unassigned port and preferring a port less than 1024 to a higher port) to select a single probable application.

Unfortunately, port numbers alone provide a severely limited mechanism for classifying applications [33]. In particular, port-based heuristics could not identify a probable application in more than 25% of all observed inter-domain traffic in our study. The unclassified traffic includes applications using either non-standard ports, ephemeral port numbers, or otherwise unrecognized protocols. For example, port-based classification only classifies the control traffic associated with protocols like FTP and not the semi-random

ports used by subsequent data transfer (the bulk of the traffic). Port heuristics also do not identify tunneled applications such as video or other protocols running over HTTP, nor applications like P2P using encryption or random port numbers.

Given the limitations of port-based application classification, we augment our study dataset with a smaller set of application statistics based on payload classification (*i.e.*, DPI). Specifically, we leverage data from inline appliances deployed across the consumer edge of five cooperating provider deployments in Asia, Europe and North America [34]. These five deployments include traffic representing several million cable and DSL subscribers.

The inline probe appliances use a combination of proprietary rule-based payload signatures and behavioral heuristics to classify applications. Based on third-party testing and provider commercial evaluation, these inline probes achieve a high rate of classification accuracy and represent the best available “ground-truth” with respect to the classification of inter-domain application traffic within these deployments.

While the payload dataset is not large enough to provide meaningful extrapolation to all inter-domain traffic, the dataset does provide additional validation and insight into our port / protocol application analysis in this Section. However, given the nature of the inline deployments we believe the data likely includes a bias towards consumer applications and P2P since many of the providers purchased the payload inspection appliances in part based on a perceived need to manage P2P traffic.

4.1 Largest Applications by Traffic Volume

In Table 4, we show the ten largest applications by a weighted average percent of all inter-domain traffic as of July 2009. Table 4a shows data from port / protocol classification of applications and Table 4b displays statistics from the five inline / port span payload deployments.

We calculate the rankings in Table 4a using weighted average percentage of all inter-domain traffic using each well-known port and protocol. For purposes of highlighting Internet traffic trends, Table 4 groups multiple well-known ports and protocols into high level application categories. We observe that many application groupings include dozens of associated ports / protocols (*e.g.*, P2P).

From Table 4a, we see the majority of inter-domain traffic in 2009 consists of web as a category at 52% (SSL and other ports besides TCP port 80 account for less than 5% of this number). Video as a category represents the second largest application group at 2.64% and VPN protocols rank third at 1.41% followed by email at 1.38%. Overall, our findings are consistent with other recent consumer traffic studies [18].

All other protocols including games, ftp, and news account for fractions of one percent of inter-domain traffic. As noted earlier, ports and protocols alone provide a limited view of Internet application usage and Table 4a includes a sizable 46% and 37% percentage of unclassified traffic in 2007 and 2009, respectively. Since port-based classification only discovers the control traffic for many file transfer and multimedia protocols, we believe Table 4a significantly under-represents traffic for video, P2P and file transfers.

We next look at payload based traffic breakdowns from the five consumer deployments in Table 4b. All values represent the average percentage of subscriber traffic attributed

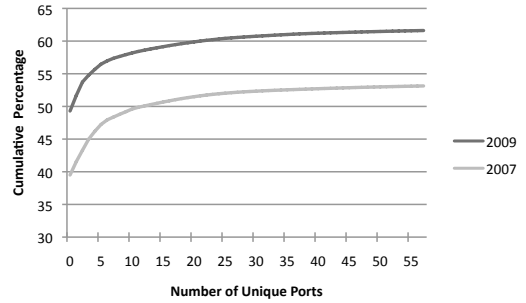


Figure 5: Cumulative distribution of the weighted average percentage of inter-domain traffic contributed by well-known ports and protocols for July 2007 and July 2009.

to each application group. We note that the configured application classifications used by the inline commercial appliances differ from the categories in Table 4a, including the lack of an explicit matching category for SSH and FTP. The “Other” category in Table 4b includes dozens of less common enterprise, database and consumer applications.

Overall, the application breakdowns correspond closely between the two tables with the notable exception of P2P. Both Table 4a and Table 4b show Web contributing 52% of Internet traffic and similarly close percentages for games and email. VPN and News shows a slightly larger variance between the two tables likely due to the consumer bias of the five inline deployments.

Data from the inline deployments also suggest that HTTP video may account for 25-40% of all HTTP traffic. In particular, one of the largest video sites, YouTube, uses progressive HTTP download. Payload analysis also suggests encrypted P2P / other ports represent another 10-15% of uncategorized traffic in Table 4a and other video / audio streaming protocols make up 3-5% of uncategorized traffic. Finally, the payload statistics show the remaining traffic consists of a heavy-tailed distribution across hundreds of less common applications.

4.2 Application Traffic Changes

In the remainder of this Section, we explore longitudinal changes in inter-domain application traffic patterns. We examine both changes in the relative traffic contribution of application categories as well as specific ports and protocols. As with earlier analysis, we use the weighted average percentage of inter-domain traffic across all providers. We again observe that growth in this dataset equates to “market share,” where a growing application gains traffic at the expense of other applications.

Not unexpectedly, our analysis finds TCP and UDP combined account for more than 95% of all inter-domain traffic. VPN protocols including IPSEC’s AH and ESP contribute another 3% and tunneled IPv6 (protocol 41) adds a fraction of one percent. The remaining percentage of protocol traffic populates a heavy-tailed distribution across the entire possible protocol number range and likely represents configurations errors and denial of service attacks.

Figure 5 shows the cumulative distribution of the average

| Rank | Application | 2007 | 2009 | Change |
|------|--------------|-------|-------|--------|
| 1 | Web | 41.68 | 52.00 | +10.31 |
| 2 | Video | 1.58 | 2.64 | +1.05 |
| 3 | VPN | 1.04 | 1.41 | +0.38 |
| 4 | Email | 1.41 | 1.38 | -0.03 |
| 5 | News | 1.75 | 0.97 | -0.78 |
| 6 | P2P | 2.96 | 0.85 | -2.11 |
| 7 | Games | 0.38 | 0.49 | +0.12 |
| 8 | SSH | 0.19 | 0.28 | -0.08 |
| 9 | DNS | 0.20 | 0.17 | -0.04 |
| 10 | FTP | 0.21 | 0.14 | -0.07 |
| | Other | 2.56 | 2.67 | +0.11 |
| | Unclassified | 46.03 | 37.00 | -9.03 |

(a) Port / Protocol

| Average Percentage | |
|--------------------|-------|
| Web | 52.12 |
| Video | 0.98 |
| Email | 1.54 |
| VPN | 0.24 |
| News | 0.07 |
| P2P | 18.32 |
| Games | 0.52 |
| SSH | N/A |
| DNS | N/A |
| FTP | 0.16 |
| Other | 20.54 |
| Unclassified | 5.51 |

(b) Payload Matching

Table 4: Top application categories. The first table shows top applications by weighted average percent of inter-domain traffic in July 2007 and 2009 based on port / protocol classification. The second table shows average application breakdowns in July 2009 across five consumer providers using proprietary payload and application behavioral classification heuristics.

weighted percentage of inter-domain traffic per each TCP / UDP port and other protocols over the two year study period. In July 2007, 52 ports contributed 60% of the traffic. By 2009, only 25 ports / protocols contributed 60% of inter-domain traffic. Overall the CDF data suggests a migration of Internet application traffic to a smaller set of application ports and protocols.

We show a specific example of video migration later in this Section and Figure 6. The popular Xbox Live service provides another example. On June 16, 2009, we found Microsoft migrated all Xbox Live (originally TCP / UDP port 3074) traffic to use port 80 in a minor system update [35, 6].

Discussion with network operators suggests the consolidation of application port and protocols in Figure 5 is due both to the growing dominance of the browser as an application front end and the efforts of content owners / developers to redress the deployment burdens introduced by near ubiquitous network layer security policies. Specifically, the majority of deployed firewalls will pass HTTP by default but less commonly used applications may require configuration changes such as port forwarding or explicit pass rules.

4.2.1 Applications Exhibiting Growth

We next look at the application categories with growth in inter-domain traffic share. We begin with the fastest growing application category, the web. As discussed earlier in this Section, web protocols account for a weighted average 52% of all inter-domain traffic as of July 2009.

Table 4a shows that well-known Web ports (*i.e.*, TCP 80, 443 and 8080) gained 10 percentage points between July 2007 and 2009. Discussions with providers and analysis of the data from payload based classification of applications suggests much of the Web (and particularly HTTP) growth is due to video.

As a category, video represents both the second largest and second fastest growing application class. Table 4a shows a 1.05% growth in video protocol (*i.e.*, Flash, RTSP, RTP, and RTCP) percentage points between July 2007 and July 2009. At the end of the study period, these video protocols represented a 2.64% weighted average of all inter-domain

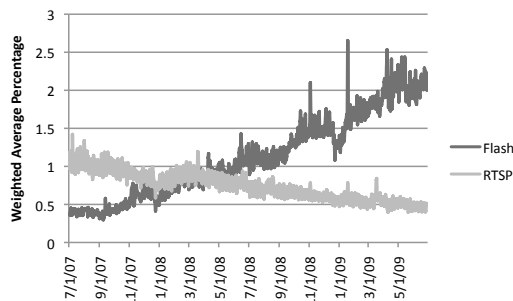


Figure 6: Change in weighted average percent of video protocols inter-domain traffic contribution between July 2007 and July 2009.

traffic. This growth in video corresponds to a widely documented increase in the popularity of Internet-based movie and television-based applications, including Hulu, YouTube, Veoh, and the BBC’s iPlayer [20]. Further, data from payload based classification suggests up to 10% of HTTP traffic in Table 4a may be due to progressive HTTP download (*e.g.*, YouTube).

We show a graph of the growth in video protocols in Figure 6. The graph shows the weighted average percentage of inter-domain traffic contributed by Flash and RTSP over the two year period of our study. Flash grew from .5% to 3.5% in two years, or more than 600% growth. Conversely, RTSP declined by .05% during the same period.

Discussions with network operators suggests most of the RTSP traffic migrated to Flash and HTTP. These two protocols offer both more widely supported and simpler alternatives to RTSP. We note that many Internet IPTV offerings still use RTSP internally.

We also note the spike of Flash traffic in Figure 6 corresponding to the Obama inauguration on January 20, 2009 [36]. Over the day of the inauguration, Flash traffic climbed to a weighted average of more than 4% of all inter-domain traffic. While the Tiger Woods US Open playoff generated

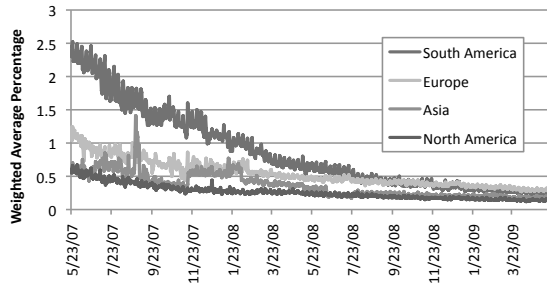


Figure 7: Weighted average percentage of Internet traffic due to P2P over well-known ports by geographic region.

a spike in North American traffic in June 2008 [37], this spike does not appear in the global analysis as it was largely localized to the US.

As categories, VPN and game well-known ports / protocols also exhibit small percentage point growth during our study period, growing at 0.38 and 0.12 percentage points, respectively. We observe that the top three game protocols contribute more than a half percent of all inter-domain traffic as of June 2009.

4.2.2 Applications Exhibiting Decline

Excluding Web, Video, VPN and Games, all other application groups in Table 4 saw a decline in weighted average percentage of all inter-domain traffic during our study period. We focus on the most prominent application category exhibiting decline, P2P.

As a category, P2P saw the largest decline with a drop of 2.8% percentage points between July 2007 and July 2009. Given the provider and regulatory concern over P2P traffic in 2007 [38], any change in relative P2P volumes has significant provider traffic management, regulatory and research implications.

Figure 7 breaks down average percentage of inter-domain traffic using P2P well-known ports by geographic region. All four regions (South America, North America, Asia and Europe) show significant declines in P2P over the two year study period. South America exhibits the largest decrease dropping from an average of 2.5% of inter-domain traffic to under a half percent.

Analysis of P2P traffic using payload analysis from inline / portspan commercial ISP deployments shows a similar trend. Specifically, in July 2007 application payload analysis of the five consumer deployments shows P2P percentages at 40% of all traffic. At the end of the study period, application payload analysis of these deployments found P2P traffic percentages at less than 20% of all traffic. Our results mirror related research findings and press observations of P2P decline [18, 39, 40, 41].

Discussions with Internet providers and a survey of research literature and press articles suggests several possible explanations for the decline in P2P, including: improvements in P2P client and algorithm efficiency [42], stealthier P2P clients and algorithms (*i.e.*, evasion of payload application classification), migration to tunneled overlays

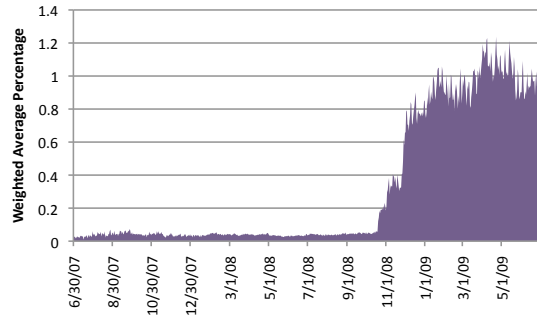


Figure 8: Weighted average percentage of inter-domain traffic due to Carpathia Hosting, home to several of the largest direct download file sharing sites on the Internet.

(*i.e.*, IPv6), provider traffic management policies and the increased use of P2P encryption. We note, however, that our payload inline / port span dataset does not show any significant growth in encrypted traffic.

Private discussions with network operators suggest significant volumes of P2P traffic may have migrated to other distribution alternatives, including direct download and streaming video [40]. These distribution alternatives may avoid many of the problems associated with P2P such as ISP traffic management, poor seeding of torrents and the threat of litigation over the exchange of copyrighted materials [43].

Direct download sites examples include MegaUpload, RapidShare, and Mediafire [44, 43]. Similarly, video commercial sites like Hulu, YouTube, Veoh, and MegaVideo provide streaming access to thousands of popular movies and television shows².

As an illustration of the possible migration of P2P towards other distribution mechanisms, we graph inter-domain traffic to a large direct download site in Figure 8. Normally, our ASN based analysis cannot identify traffic contributed by any individual co-located hosting customer. Figure 8 provides an exception.

Carpathia Hosting hosts several large customer direct download and video streaming sites including MegaVideo and MegaUpload (Alexa rank of 72) [45, 44]. We graph the weighted percentage of all inter-domain traffic originating or terminating in Carpathia’s ASN (AS29748, AS46742, and AS35974) in Figure 8.

Private discussions with providers indicate the abrupt and significant jump in Carpathia inter-domain traffic percentages after January 2009 is due to the migration and consolidation of MegaUpload and associated sites on Carpathia servers. As of July 2009, Carpathia represents a weighted average of more than 0.8% of all inter-domain traffic.

5. INTERNET SIZE ESTIMATES

In this final analysis section, we use independent measurements of provider inter-domain traffic to both validate our

²While most video and direct download sites police copyright infringement, some sites demonstrate a more flexible policy towards intellectual property [40, 44].

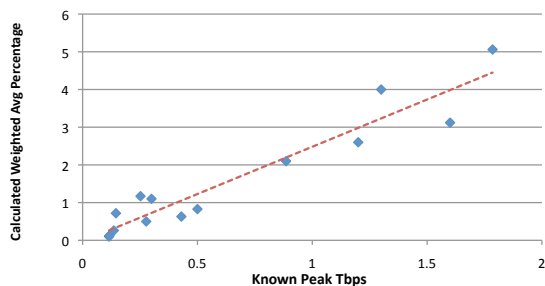


Figure 9: A graph of independent inter-domain traffic volumes from twelve reference providers plotted against the calculated aggregate ASN share of all inter-domain Internet traffic for each provider from our data. Graph includes a linear fit of these values.

results as well as estimate the current volume of all inter-domain traffic and the annualized rate of inter-domain traffic growth. We compare our findings with recent research and commercial estimates of global Internet traffic.

5.1 Traffic Volume

To provide independent verification of our study measurements, we solicited inter-domain traffic statistics from twelve large topologically and geographically diverse providers and content sites. We focused our solicitation on datasets disjoint from the 110 anonymous providers in our study.³

Each provider supplied peak inter-domain traffic volumes for July 2009. These providers use a combination of in-house Flow tools or SNMP interface polling to determine their inter-domain traffic volumes. We use these twelve known inter-domain traffic values as “ground-truth.”

Figure 9 shows a plot of each of the provided traffic volumes against that provider’s weighted average percentage of all inter-domain traffic (based on ASN) from our data. We also shows a linear fit of these measurements. The resulting line has a slope of 2.51, meaning that a 2.51% share of all inter-domain traffic represents approximately 1 Tbps of inter-domain traffic. This provides an extrapolation to the overall size of the Internet at $1 / 2.51 = 39.8$ Tbps as of July 2009.

While exhaustive validation of our results is difficult given the commercial secrecy surrounding provider traffic statistics, the plot in Figure 9 lends confidence to our findings. The linear fit has an R^2 value of 0.91, indicating that our data and statistical analysis is consistent with the independent “ground-truth” measurements supplied by the twelve providers.

We also calculate the absolute number of bytes for the month of May 2008 for comparison with Cisco’s last published data [23]. This monthly calculated value matches Cisco’s Internet traffic estimate of 9 exabytes per month in 2008. Table 5 shows the result of our calculations combined with data from a private survey of providers and published reports from Cisco [23] and MINTS [14]. We also corrob-

³Though study participants are anonymous, we obtained a complete customer list of providers that had purchased the commercial probes used in our study. We then solicited inter-domain traffic data from providers without deployed probes.

orate the data with a private survey of ISPs and content providers [46].

5.2 Traffic Growth

To estimate the rate of inter-domain traffic growth, we compute an annual growth rate (AGR), which represents the estimated increase in inter-domain traffic volume over a year period. This annual growth rate is based on daily traffic samples collected over a one year period at each router associated with a participant deployment. To calculate the AGR for a particular router, we employ a methodology similar to MINTS [14]. Specifically, we determine an exponential fit of the form $y = A * 10^{Bx}$, where x is the day ($[1, 365]$) and y is the traffic sample in bps for day x for the router. An example curve fit over the daily sample points between May 2008 and May 2009 for a for an anonymous provider can be seen in Figure 10a. From the results of our linear least squares fit, we calculate the annual growth rate as $AGR = 10^{365 * B}$. For example, an AGR of 0.5 represents a 50% decrease in traffic, 1.0 represents no change, 2.0 represents a 100% increase, 3.0 represents a 200% increase, and so on.

However, as discussed in Section 2, changes in the commercial probe infrastructure can complicate our growth estimation. For example, a provider may add, remove, or reconfigure routers associated with a particular probe deployment over time. Frequent changes in measurement infrastructure combined with misconfiguration and other anomalies can lead to noise within the dataset. Such noise may be present at three levels of granularity in the dataset: (1) datapoint-level: datapoints for a single router may be lacking in terms of the number of valid or non-zero data points; (2) router-level: fitting a growth curve to an inadequate set of traffic samples for a router may result in an inaccurate fit, and (3) deployment-level: deployments may have misconfigured or anomalous routers or a small number of total routers, resulting in the unstable routers having a large effect on the overall deployment.

To deal with these sources of noise within the dataset, we apply a pass at each level of granularity to smooth out noise and exclude anomalous and misconfigured routers. For sample-level noise, we exclude sample sets that do not have at least 2/3 valid data points throughout the year period. For router-level noise, we exclude AGR calculations that exhibit a high standard error when fitting a curve to noisy sample points. Lastly, we smooth out per-deployment noise by only considering routers with AGRs between the 1st and 3rd quartiles of the routers within that deployment.

We calculate the overall AGR for a deployment as the mean of the AGRs of the eligible routers within that deployment. The computed annual growth rates between May 2008 and 2009 for a number of provider deployments (Tier-1, Tier-2 and cable / DSL providers) in our data are shown in Figure 10b. In addition to per-deployment AGRs, we calculate AGRs by market segment by taking the mean of the per-deployment AGRs of the providers within that market segment. Table 6 breaks down the growth of each market segment and includes the number of deployments and eligible routers used to compute each AGR.

In Table 5, we compare our results of inter-domain traffic growth to similar measurement studies from Cisco and MINTS [23, 14], as well the average growth rates reported in survey of 25 ISPs. We note that both Cisco and MINTS report a slightly higher rate of 50%, but the difference may

| Estimate | 110 ISPs | ISP Survey | Cisco | MINTS |
|----------------------------|------------|------------|------------|--------------|
| Traffic Volume Per Month | 9 exabytes | N/A | 9 exabytes | 5-8 exabytes |
| Traffic Annual Growth Rate | 44.5% | 35-45% | 50% | 50-60% |

Table 5: Estimates of inter-domain traffic volume and annualized growth.

| Market Segment | Annual Growth Rate | Deployments | Routers |
|----------------|--------------------|-------------|---------|
| Tier 1 | 1.363 | 6 | 82 |
| Tier 2 | 1.416 | 21 | 152 |
| Cable / DSL | 1.583 | 8 | 79 |
| EDU | 2.630 | 4 | 13 |
| Content | 1.521 | 3 | 6 |

Table 6: Annual growth rate (AGR) and number of eligible deployments and routers by market segment.

be due to inclusion of internal / backbone traffic while our study focused on inter-domain traffic.

6. CONCLUSION

In this paper, we provide one of the first large-scale longitudinal studies of Internet inter-domain traffic. Specifically, over a two year period we analyzed more than 200 Exabytes of commercial inter-domain traffic through the direct instrumentation of more than 3,000 peering routers across 110 geographically and topologically diverse Internet providers.

Our main contribution is the identification of a significant ongoing evolution of provider interconnection strategies and resultant inter-domain traffic demands, including the rapid transition to a more densely interconnected and less hierarchical inter-domain Internet topology. In particular, we find the majority of inter-domain traffic now flows directly between large content providers, data center / CDNs, and consumer networks. We show that as of July 2009, 150 ASNs (out of 30,000 ASN in default-free BGP tables) originate more than 50% of all Internet inter-domain traffic by weighted average percentage. We also identify changes in Internet inter-domain application traffic patterns, including a significant rise in video traffic and a corresponding decline in P2P.

While analysts and the press have provided anecdotal discussion of these macro level Internet trends (*e.g.*, [3]), we believe this paper represents the first quantitative characterization. We again observe that the emerging new provider inter-connection strategies have significant ongoing implications for backbone engineering, design of Internet-scale applications and research. Given the significant obstacles intrinsic to commercial inter-domain traffic measurement, we hope to make our data available to other researchers on an ongoing basis pending anonymization and privacy discussions with provider study participants.

Overall, we argue the Internet industry is in the midst of an inflection point out of which new network engineering design, business models and economies are emerging. Economic changes including the decline of wholesale IP transit prices [2] and the dramatic growth in advertisement-supported services reversed decade-old business dynamics between transit providers, consumer networks and content providers. For example, providers that used to charge content networks for transit now offer settlement-free interconnection or, in some cases, may even pay the content networks for access [5, 6].

As measured in this paper, provider inter-domain traffic

demands provide a key measure of emergent network engineering and commercial strategies. As Google, Microsoft, Facebook, Baidu and other large content owners and consumer networks compete for virtual real estate and Internet market share, we expect the trend towards Internet inter-domain traffic consolidation to continue and even accelerate.

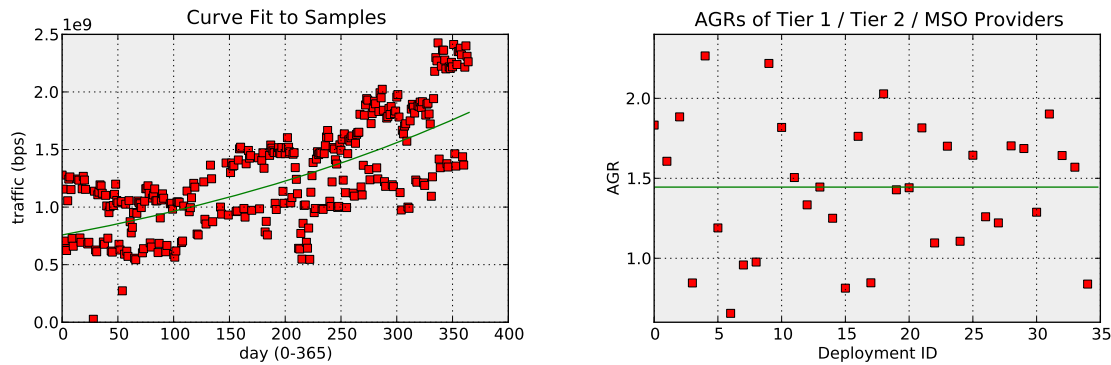
7. ACKNOWLEDGMENTS

The authors wish to thank Haakon Ringberg for contributions to an earlier incarnation of this research. We also thank Jennifer Rexford, Randy Bush, Vijay Gill, Bill Norton, Nasser El-Aawar, Shane Amante, Andrew Odlyzko, Kim Claffy, Darren Anstee, Emile Aben, Bradley Huffaker, and Mike Hollyman for their constructive comments. The authors also acknowledge the anonymous SIGCOMM 2010 referees for their feedback.

Finally, we thank multiple unnamed reviewers at Internet providers and content networks for their generous insights. We also thank the 110 Internet provider participants for their extraordinary willingness to contribute data and make this research possible.

8. REFERENCES

- [1] C. Huitema, *Routing in the Internet (2nd ed.)*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2000.
- [2] O. Malik, "Wholesale Internet Bandwidth Prices Keep Falling." GigOM Blog, <http://gigaom.com>, October 2008.
- [3] NetCompetition.org, "A First-Ever Research Study: Estimating Google's U.S. Consumer Internet Usage and Cost." Unpublished white paper, 2008.
- [4] L. Dignan, "Comcast Feeling the Heat from Competition." ZDNet, <http://blogs.zdnet.com>, October 2007.
- [5] G. Goth, "New Internet Economics Might Not Make it to the Edge," in *IEEE Internet Computing*, vol. 14,1, ACM, January 2010.
- [6] Private communication with network operators., July 2009.
- [7] A. Dhamdhere and C. Dovrolis, "Ten Years in the Evolution of the Internet Ecosystem," in *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, pp. 183–196, ACM New York, NY, USA, 2008.
- [8] J. Wu, Z. M. Mao, J. Rexford, and J. Wang, "Finding a needle in a haystack: Pinpointing significant BGP routing changes in an IP network," in *Proc. Symposium on Networked Systems Design and Implementation*, 2005.
- [9] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, "The (in)Completeness of the Observed Internet AS-level Structure," *IEEE/ACM Transactions on Networking (ToN)*, 2010.
- [10] M. Rajab, F. Monrose, A. Terzis, and N. Provos, "Peeking Through the Cloud: DNS-Based Estimation and Its Applications," in *Applied Cryptography and Network Security: 6th International Conference, ACNS 2008, New York, NY, USA, June 3-6, 2008, Proceedings*, p. 21, 2008.



(a) Example aggregate growth rate calculation. (b) Per-deployment aggregate growth rate calculations.

Figure 10: First graph shows example curve fit to sample points of an anonymous provider during annualized growth rate (AGR) calculation. The second graph providers per-deployment AGRs of Tier-1, Tier-2, and cable providers from May 2008 to May 2009.

[11] "Global Internet Geography." Telegeography Research <http://www.telegeography.com>, September 2009.

[12] Akamai, "State of the Internet," 2009.

[13] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and C. Diot, "Packet-level Traffic Measurements from the Sprint IP Bbackbone," in *IEEE Network*, November 2003.

[14] "Minnesota Internet Traffic Studies (MINTS)." MINTS <http://www.dtc.umn.edu/mints>, July 2009.

[15] D. Antoniadis, M. Polychronakis, N. Nikiforakis, E. Markatos, and Y. Mitsos, "Monitoring three National Research Networks for Eight Weeks: Observations and Implications," in *IEEE Network Operations and Management Symposium Workshops, 2008. NOMS Workshops 2008*, pp. 153–156, 2008.

[16] G. Maier, A. Feldmann, V. Paxson, and M. Allman, "On Dominant Characteristics of Residential Broadband Internet Traffic," in *Proc. ACM IMC*, 2009.

[17] K. Cho, "Trends in Japanese Residential Traffic," *ISOC Panel on Internet Bandwidth: Dealing with Reality*, 2009.

[18] G. Maier, A. Feldmann, V. Paxson, and M. Allman, "On Dominant Characteristics of Residential Broadband Internet Traffic," in *IMC '09: Proceedings of the 9th ACM SIGCOMM conference on Internet Measurement Conference*, (New York, NY, USA), ACM, 2009.

[19] B. Norton, "Internet Service Providers and Peering." Equinix White Paper, 2001.

[20] B. Norton, "Video Internet: The Next Wave of Massive Disruption to the US Peering Ecosystem." Equinix White Paper, September 2006.

[21] P. Faratin, D. Clark, P. Gilmore, S. Bauer, A. Berger, and W. Lehr, "Complexity of internet interconnections: Technology, incentives and implications for policy," in *The 35th Research Conference on Communication, Information and Internet Policy (TPRC)*, ACM, 2007.

[22] P. Gill, M. F. Arlitt, Z. Li, and A. Mahanti, "The Flattening Internet Topology: Natural Evolution, Unsightly Barnacles or Contrived Collapse?," in *Proceedings of PAM*, 2008.

[23] Cisco Systems, "Cisco Visual Networking Index – Forecast and Methodology." A Cisco White Paper, 2008.

[24] "University of Oregon RouteViews Project." <http://www.outerviews.org>.

[25] B. Choi and S. Bhattacharyya, "On the Accuracy and Overhead of Cisco Sampled Netflow," in *Sigmetrics Workshop on Large Scale Network Inference (LSNI): Methods, Validation, and Applications*, June 2005.

[26] Arbor Networks, "Peakflow." Product data sheet and whitepapers, www.arbornetworks.com/peakflowsp.

[27] P. Marshall, "Link data: Global network." Yankee Group Report <http://www.yankeegroup.com>, October 2009.

[28] J. Markoff, "Internet Traffic Begins to Bypass the U.S.," *New York Times*, August 2008.

[29] Wikipedia, "Tier1 Network." Wikipedea <http://en.wikipedia.org>.

[30] R. Miller, "Google-YouTube: Bad News for Limelight?." Datacenter Knowledge Blog, <http://www.datacenterknowledge.com>, October 2006.

[31] CAIDA, "Internet Topology." <http://www.caida.org/research/topology>.

[32] Q. Chen, H. Chang, S. J. Shenker, R. Govindan, and W. Willinger, "The Origin of Power Laws in Internet Topologies Revisited," in *Proc. of IEEE Infocom*, 2007.

[33] M. Roughan, S. Sen, O. Spatscheck, and N. Duffield, "Class-of-Service Mapping for QoS: a Statistical Signature-Based Approach to IP Traffic Classification," *ACM Sigcomm Internet Measurement Workshop*, 2004.

[34] Arbor Networks, "Arbor E100." Product data sheet, www.arbornetworks.com.

[35] T. Magrino, "Xbox Live Going Dark June 16." GameSpot, <http://www.gamespot.com>, June 2009.

[36] C. Labovitz, "The Great Obama Traffic Flood." Arbor Networks Blog, <http://asert.arbornetworks.com/2009/01/the-great-obama-traffic-flood/>, January 2009.

[37] C. Labovitz, "The Tiger Effect." Arbor Networks Blog, <http://asert.arbornetworks.com/2008/06/the-tiger-effect/>, June 2008.

[38] J. Pigg, "P2P: Damn This Traffic Jam." Yankee Group Report <http://www.yankeegroup.com>, July 2008.

[39] J. Erman, A. Gerber, M. Hajiaghayi, D. Pei, and O. Spatscheck, "Network-aware Forward Caching," in *Proceedings of the 18th International Conference on World wide web*, ACM New York, NY, USA, 2009.

[40] N. Anderson, "P2P Traffic Drops as Streaming Video Grows in Popularity." Ars Technica, <http://arstechnica.com>, September 2008.

[41] J. Cheng, "Report: UK File Sharing Drops, Even Among Teens." Ars Technica, <http://arstechnica.com>, June 2009.

[42] H. Xie, Y. Yang, A. Krishnamurthy, Y. Liu, and A. Silberschatz, "P4P: Provider Portal for Applications," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4, 2008.

[43] Wikipedia, "Direct Download." <http://en.wikipedia.org>.

[44] C. Adamsick, "Warez' the Copyright Violation? Digital Copyright Infringement: Legal Loopholes and Decentralization," *TechTrends*, vol. 52, no. 6, pp. 10–12, 2008.

[45] Alexa, "The top 500 Sites on the Web." <http://www.alexa.com>.

[46] D. McPherson and C. Labovitz, "2009 Survey of ISP Traffic Trends." Private Survey of 25 Large ISPs and Content Providers., July 2009.