Internet Inter-Domain Traffic

Craig Labovitz, Scott Iekel-Johnson, Danny McPherson
Arbor Networks

Jon Oberheide, Farnam Jahanian
University of Michigan

Motivation

- **Measuring the Internet is hard**
- **Significant previous work on**
  - Router and AS-level topologies
  - Individual link / ISP traffic studies
  - Synthetic traffic demands
- **But limited "ground-truth" on inter-domain traffic**
  - Most commercial arrangements under NDA
  - Significant lack of uniform instrumentation
- **Goal: longitudinal observations of Internet traffic**
  - Can we instrument representative distribution of ISPs?
  - Estimate of traffic volume / growth
  - Analysis of changes in Internet traffic demands

Conventional Wisdom

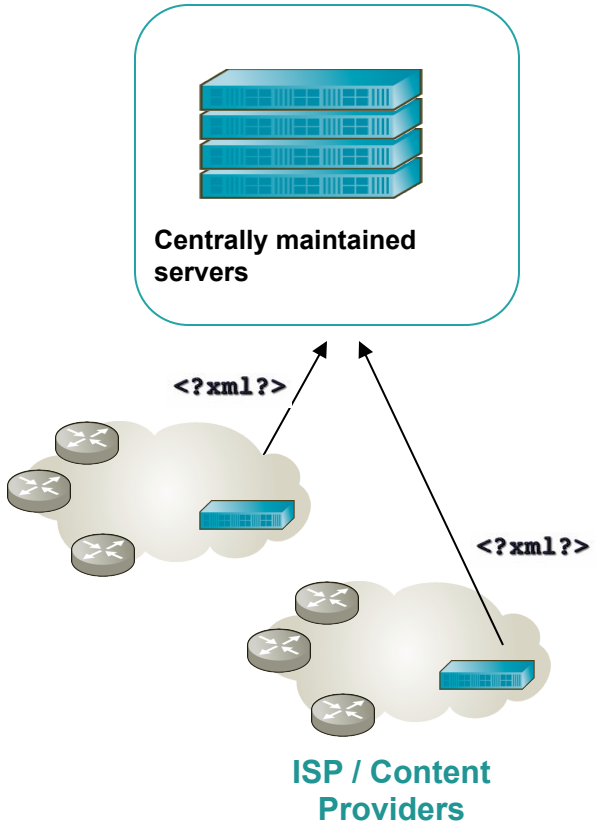
- **Internet is a global scale end-to-end network**
  - Packets transit (mostly) unmolested
  - Value of network is global addressability / reachability (metcalfe effect)
- **Broad distribution of traffic sources / sinks**
- **An Internet "core" exists**
  - Dominated by a dozen global transit providers
  - Interconnecting content, consumer and regional providers

Methodology

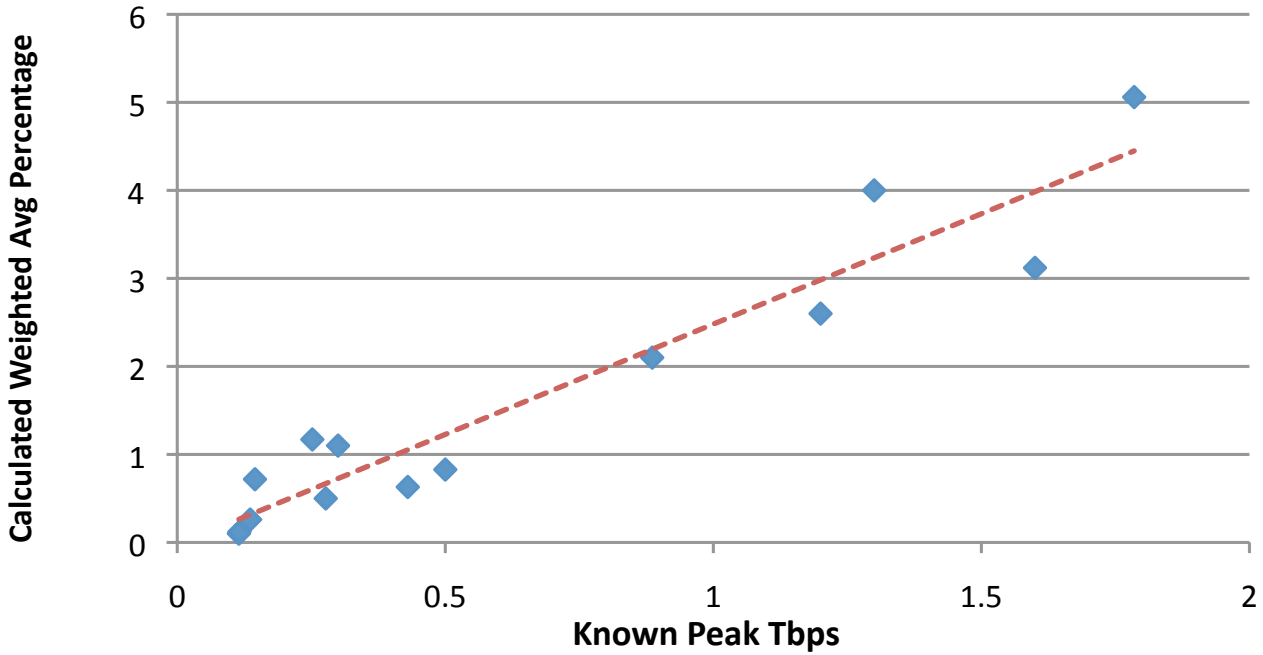
- **Focus on inter-domain traffic**
  - i.e. distinguish from web hits, tweets, VPN, etc.
- **Leverage widely deployed commercial Internet monitoring infrastructure**
  - Add export of coarse grain traffic statistics (ASNs, ASPaths, protocols, ports, etc.)
  - Via anonymous XML forwarded to central servers
- **Cajole carriers into participation**
  - 110+ ISPs / content providers
  - Including 3,000 edge routers and 100,000 interfaces
  - And an estimated ~25% all inter-domain traffic
- **Wait two years…**

Additional Methodology Details

- **Within a given ISP, commercial probes**
  - Monitors NetFlow / Jflow / etc and routing across multiple edge routers
  - Probes are topology aware of ISP, backbone and customer boundaries
  - Some deployments include payload / DPI observations
- **Post-process data**
  - Focus on distributions / share
  - Calculate percentages per category
  - Calculate weighted averages using number of routers in each deployment
- **Augment analysis with**
  - Provider interviews / surveys
  - Known traffic volumes



Methodology Validation



- **Validate predictions based on "ground-truth"**
  - Linear fit of 12 known ISP traffic demands
  - Significant variety in measurement technology and definitions
  - Linear R squared (coefficient of determination) value of 0.91
- **Further validate with extensive discussions / surveys of providers**
- **Also provides estimate of inter-domain size / growth (45 Tbs and 45%)**

Change in Carrier Traffic Demands

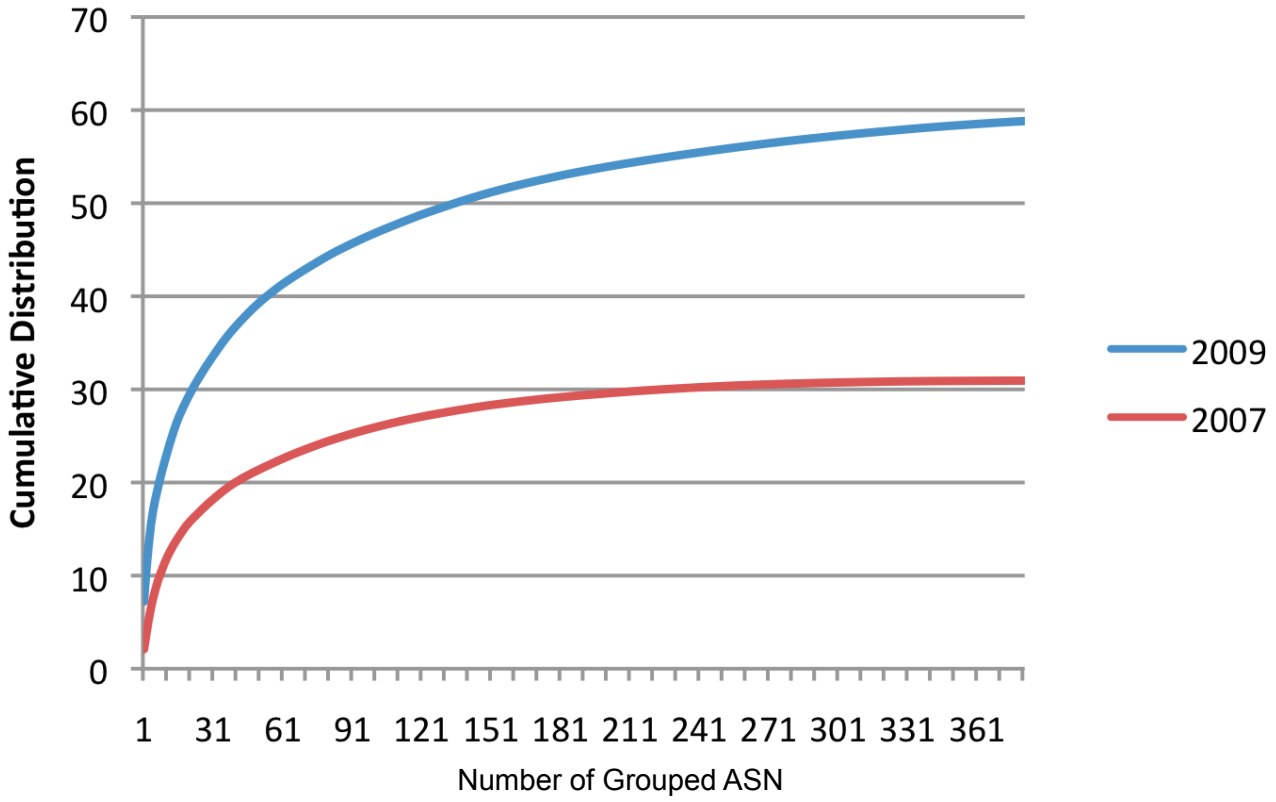
Rank	2007 Top Ten	%
1	*ISP A*	5.77
2	*ISP B*	4.55
3	*ISP C*	3.35
4	*ISP D*	3.2
5	*ISP E*	2.77
6	*ISP F*	2.6
7	*ISP G*	2.24
8	*ISP H*	1.82
9	*ISP I*	1.35
10	*ISP J*	1.23

Rank	2009 Top Ten	%
1	*ISP A*	9.41
2	*ISP B*	5.7
3	Google	5.2
4	-	
5	-	
6	Comcast	3.12
7	-	
8	-	
9	-	
10	-	

Based on analysis of anonymous ASN (origin/transit) data (as a weighted average % of all Internet Traffic). Top ten has NO direct relationship to study participation.

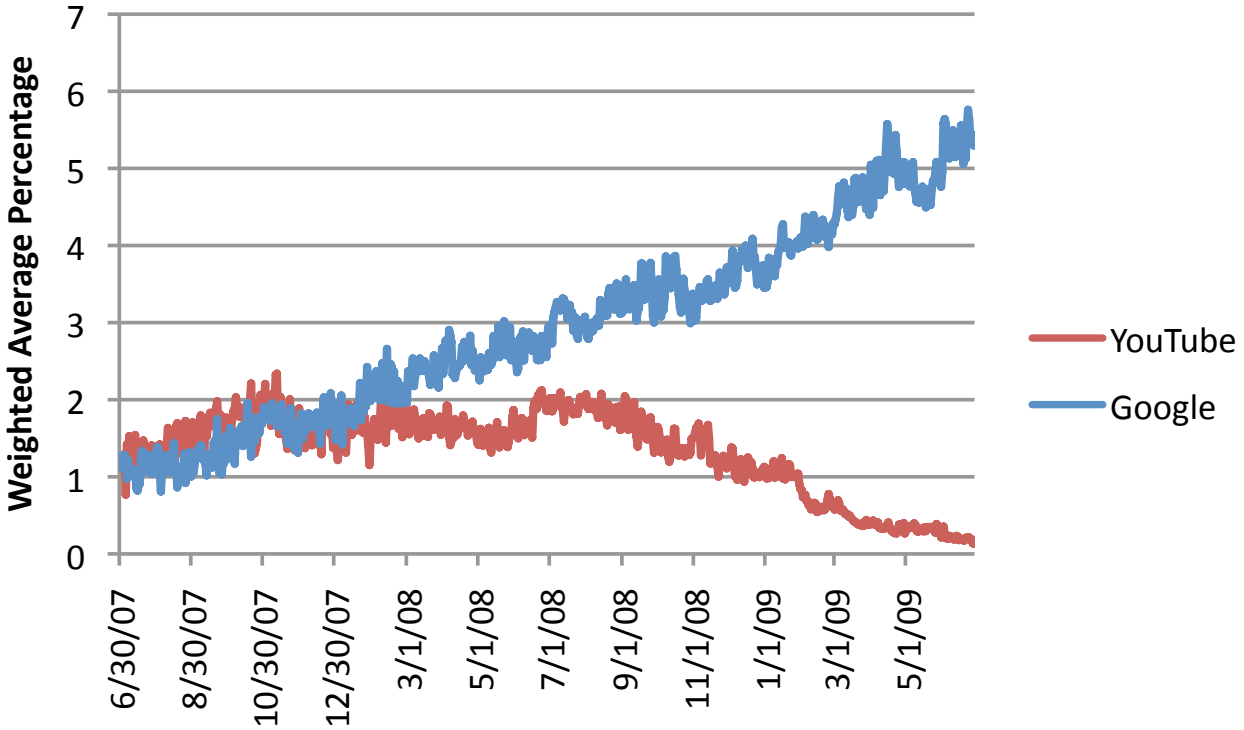
- **In 2007, top ten match "tier-1" ISPs (e.g., Wikipedia)**
- **In 2009, global transit carry significant traffic volumes**
  - But Google and Comcast join the list
  - And a significant percentage of ISP A traffic is Google transit

Consolidation of Content *(Grouped Origin ASN)*



- In 2007, thousands of ASNs contributed 50% of content
- In 2009, 150 ASNs contribute 50% of all Internet traffic

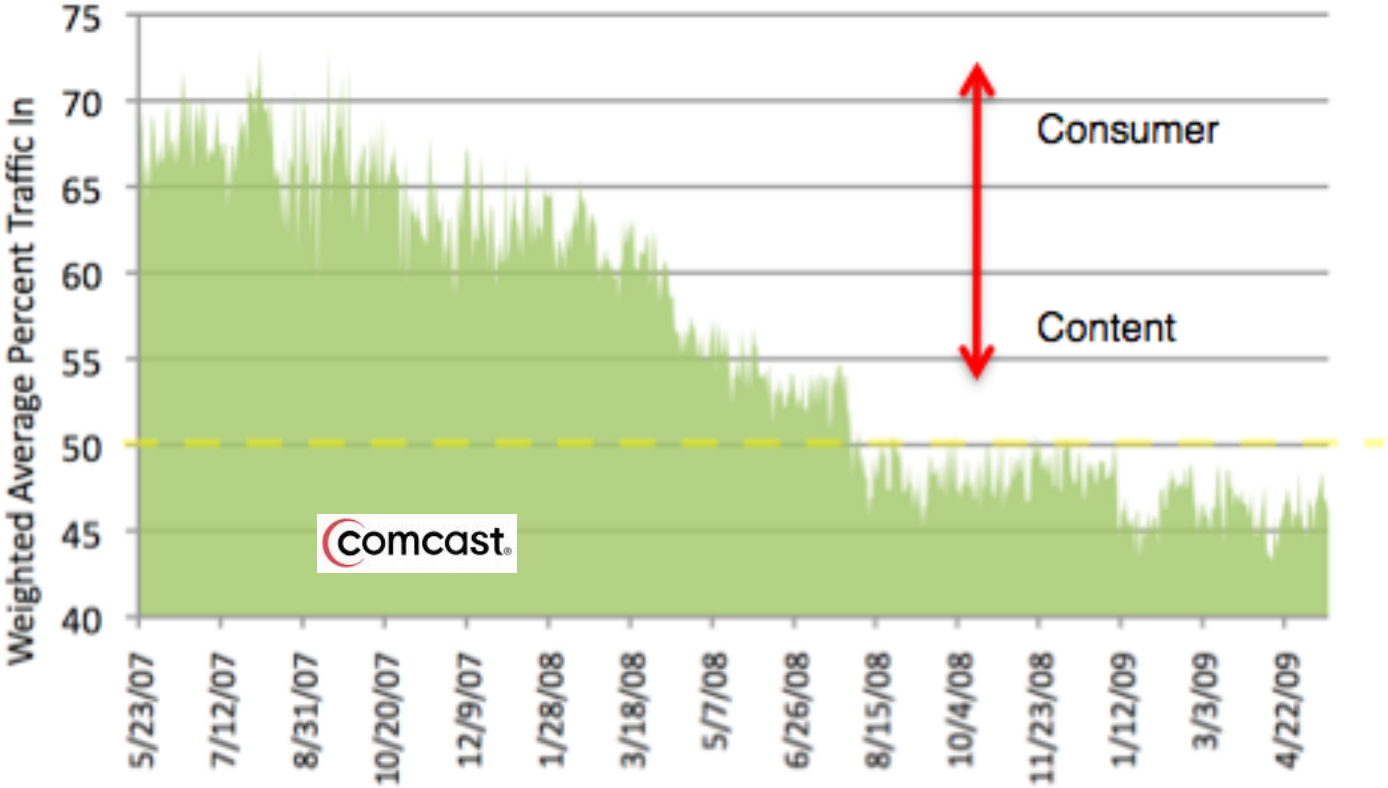
A Google Case Study



Graph of weighted averaged grouped ASNs

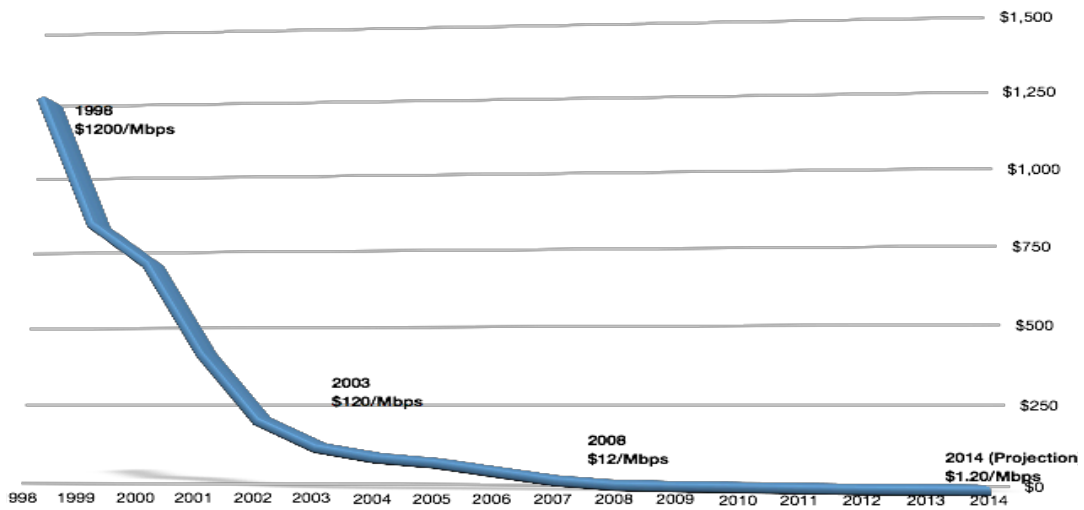
- Over time Google absorbs YouTube traffic
- As of July 2009, Google accounts for 6% of all Internet inter-domain traffic
- Google the fastest growing ASN group

A Comcast Case Study



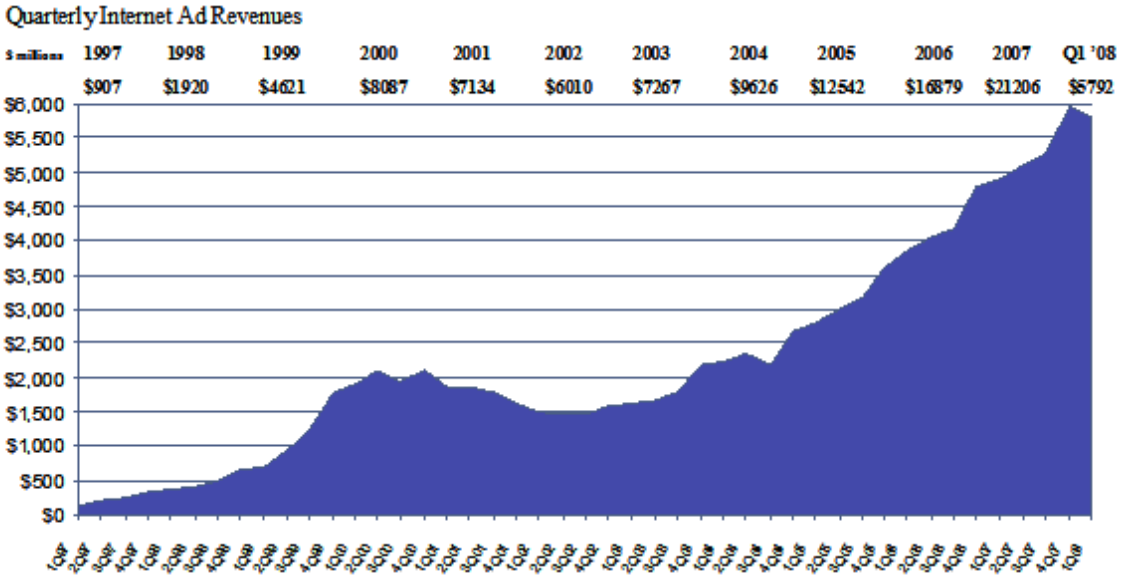
- In 2007, Comcast has typical “eyeball” peering ratios
- By 2009, Comcast resembles a transit / content provider
  - Wholesale transit, cell backhaul, video distribution, backbone consolidation

Market Forces Intuition



Revenue from Internet Transit

Source: Dr. Peering, Bill Norton



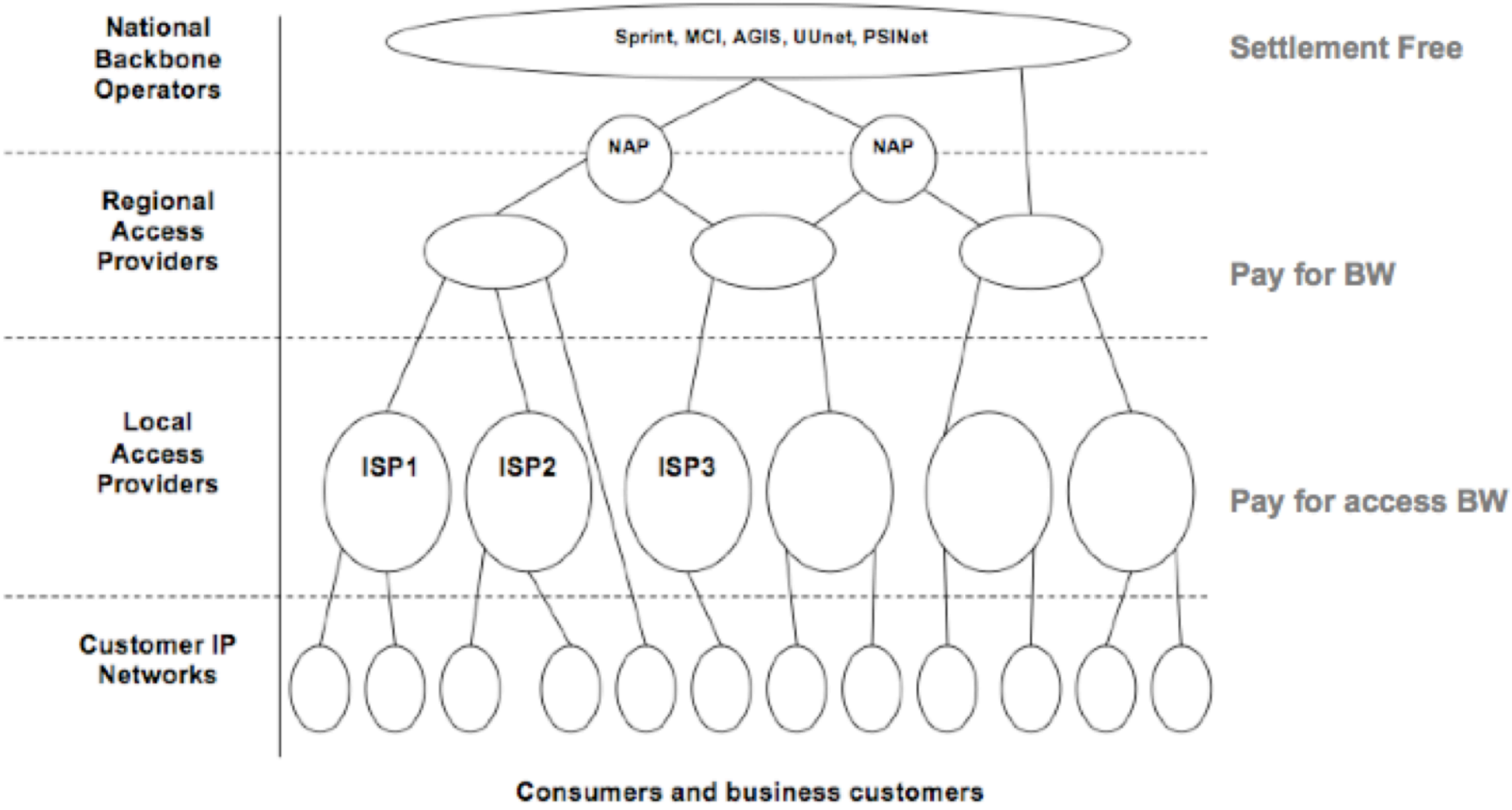
Revenue from Internet Advertisement

Source: Interactive Advertising Bureau

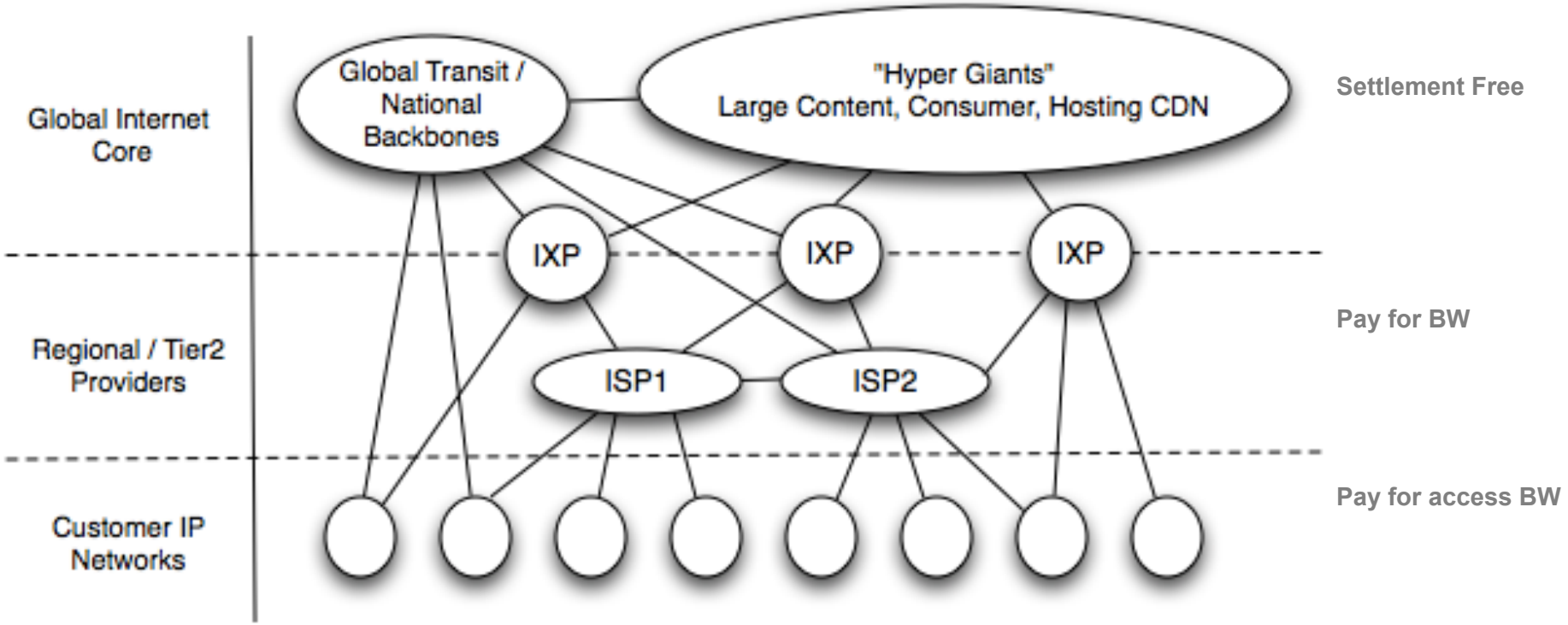
Market Intuition

- **Commoditization of IP and Hosting / CDN**
  - Drop of price of wholesale transit
  - Drop of price of video / CDN
  - Economics and scale drive enterprise to “cloud”
- **Consolidation**
  - Bigger get bigger (economies of scale)
  - e.g., Google, Yahoo, MSFT acquisitions
- **Success of bundling / Higher Value Services**
  - Triple and quad play, etc.
- **New economic models**
  - Paid content (ESPN 3), paid peering, etc.
  - Difficult to quantify due to NDA / commercial privacy
- **Disintermediation**
  - Direct interconnection of content and consumer
  - Driven by both cost and increasingly performance

Traditional Internet Model



A New Internet Model



- Flatter and much more densely interconnected Internet
- Disintermediation between content and “eyeball” networks
- New commercial models between content, consumer and transit

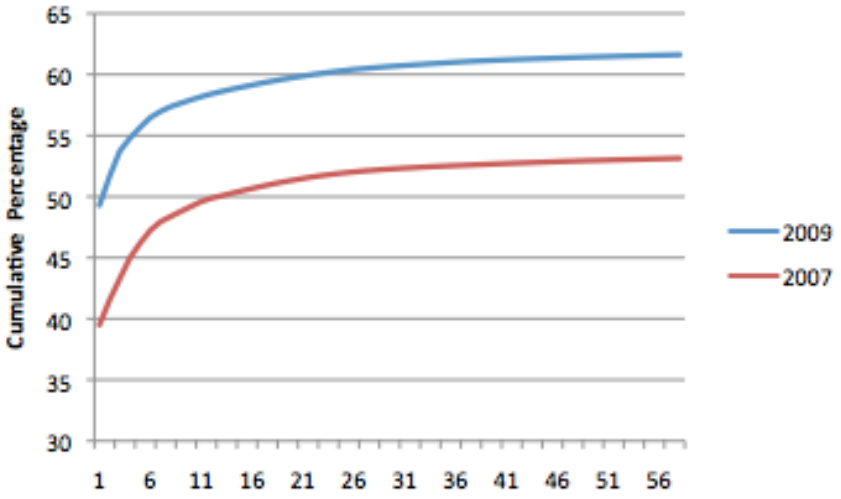
Applications

Rank	Application	2007	2009	Change
1	**Web**	41.68%	52.00%	**24.76%**
2	**Video**	1.58%	2.64%	**67.09%**
3	**VPN**	1.04%	1.41%	**35.58%**
4	**Email**	1.41%	1.38%	**-2.13%**
5	**News**	1.75%	0.97%	**-44.57%**
6	**P2P** (*)	2.96%*	0.85%	**-71.28%**
7	**Games**	0.38%	0.49%	**28.95%**
8	**SSH**	0.19%	0.28%	**47.37%**
9	**DNS**	0.20%	0.17%	**-15.00%**
10	**FTP**	0.21%	0.14%	**-33.33%**
	Other	2.56%	2.67%	**4.30%**
	Unclassified	46.03%	37.00%	**-19.62%**

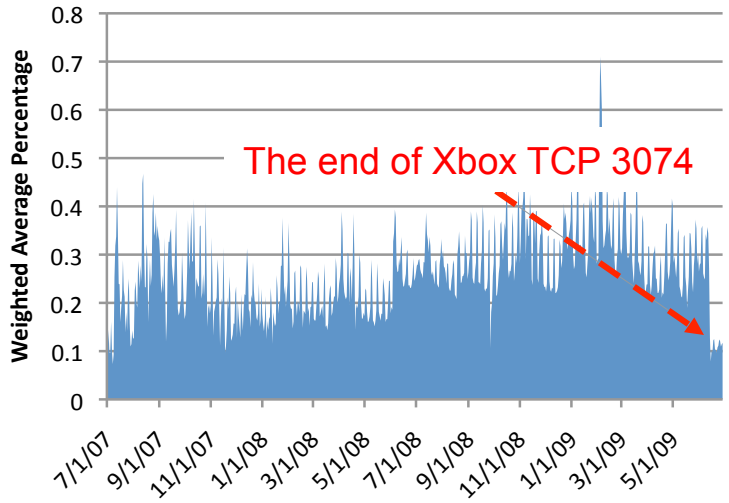
(*) 2009 P2P Value based on 18% Payload Inspection
Weighted average percentage of all Internet traffic using well-known ports

- **Growing volume of Internet traffic uses port 80 / 443**
  - Includes significant video component and source of most growth
- **Unclassified includes P2P and video**
  - Payload matching suggests P2P at 18%
  - P2P is fastest declining

Evolution of End-to-End



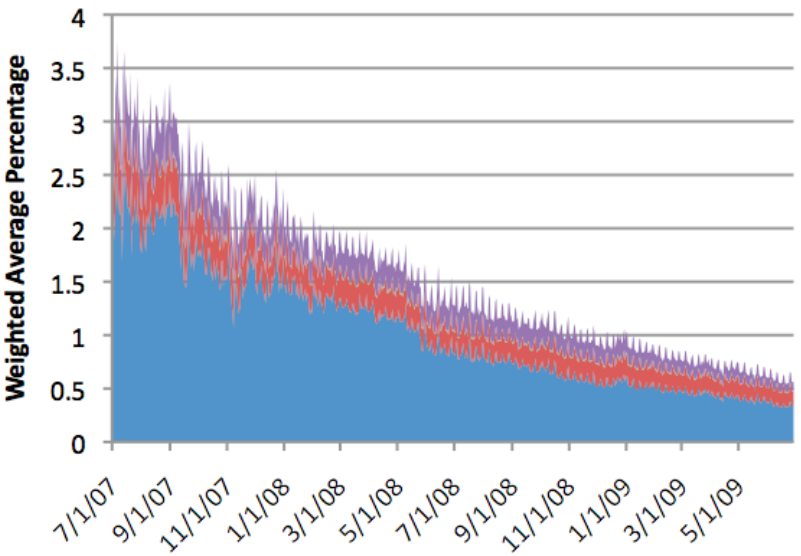
Cumulative Distribution of Traffic to TCP / UDP Ports

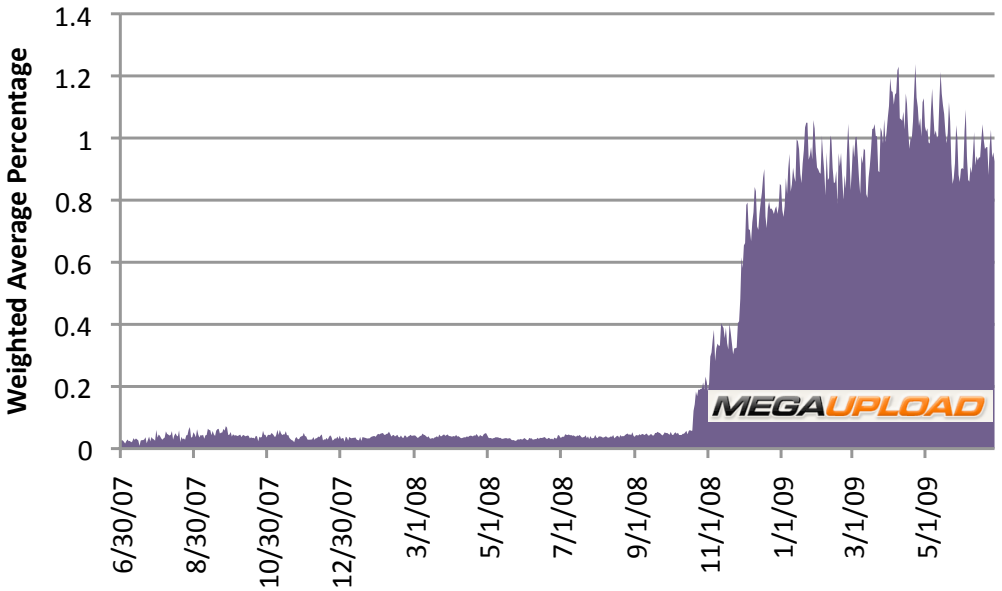


Weighted average percentage of Xbox Internet traffic

- **Growing dominance of web as application front-end**
- **Plus burden of ubiquitous network layer security policies**
- **Results in growing concentration of application traffic over a decreasing number of TCP / UDP ports**
  - Especially port 80
  - Especially video

Migration of File Sharing to the Web





- **In 2006, P2P one of largest threats facing carriers**
  - Significant protocol, engineering and regulatory effort / debate
- **In 2010, P2P fastest declining application group**
  - Trend in both well-known ports and payload based analysis
- **Significant corresponding growth in direct download and streaming video**
  - Carpathia small hosting company by traffic volume in Fall 2008
  - Mega becomes Carpathia customer in November 2008
  - Carpathia Hosting grows overnight to more than 0.8% of all traffic

Discussion

- **Significant changes in inter-domain traffic patterns**
- **Not quite Wired's "The Web is Dead"**
- **But significant shift from connectivity to content**
  - Aggregation of content / traffic sources
  - Shift from transit to direct interconnection
  - Most significant growth in ~150 large content ASN
- **And concurrent shift in applications to port 80**
  - i.e. the web may represent the new end-to-end
- **Implications on engineering and research**
  - ACL / port based security model
  - Fault tolerance
  - Routing, traffic engineering, network design
  - Rapid growth of non-interactive traffic demands (i.e. DC)

Questions

labovit@arbor.net

http://www.monkey.org/~labovit