CenturyLink™

Evolution in Characterizing Internet Usage

Amie Elcan, CenturyLink
Principal Architect, Data Network Strategies and Development
amie.elcan@centurylink.com
Strata Conference + Hadoop World NYC 2013
October 30, 2013

Outline

Use of the Internet is evolving. Measurements that characterize its use must also evolve.

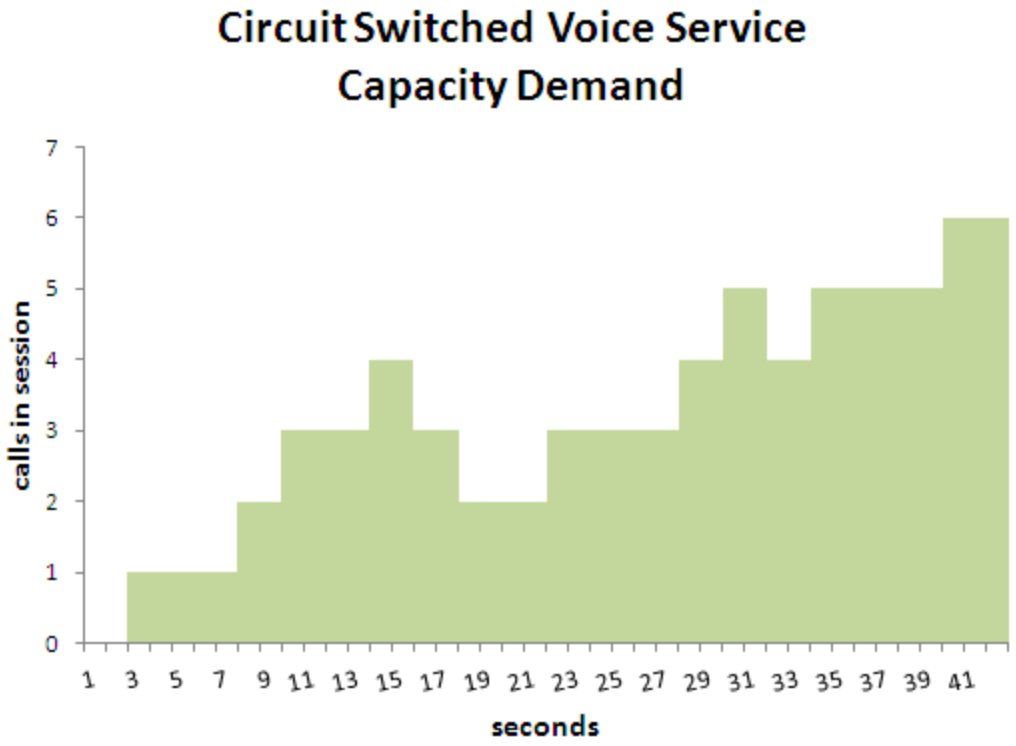
- The strategic need for large volumes of data and analytics
- Short review of some traditional telephony and data traffic measurements
- Dramatic increase in internet traffic over past 5 years
- Bandwidth demand accelerators
- Transformation in measurements
- Summary

Why Use Analytics?

- To make efficient data driven decisions
- To understand how customers are using their Internet service
  - Excellent service delivery over a converged network requires an understanding of the characteristics of the traffic types traversing the network
- To improve productivity. Revolution in measurement capabilities is occurring. Monitoring of more than fundamentals is essential. Information used for capacity planning, network design, marketing , and predicted network performance impact from changes to network
- To explore and discover meaningful patterns and relationships
- To validate and supplement external measurements (FCC Measuring Broadband America ,Netflix tech blog, Renesys IP Transit Rankings, …)

Circuit Switch Measurements

- Erlang traffic measurements derived from call arrival rates and hold times (durations)
- Erlangs computed for capacity planning purposes – Erlangs are an average number discrete number of calls in session/circuits
- Call arrivals and durations derived from machine generated measurements
- Traffic carried over phone network was homogeneous until fax, voice recording machines, dial up, etc. Dial up could be identified by unusually long hold times that are not supported by traditional sizing guidelines
- Single functionality of telephony until the 1970s

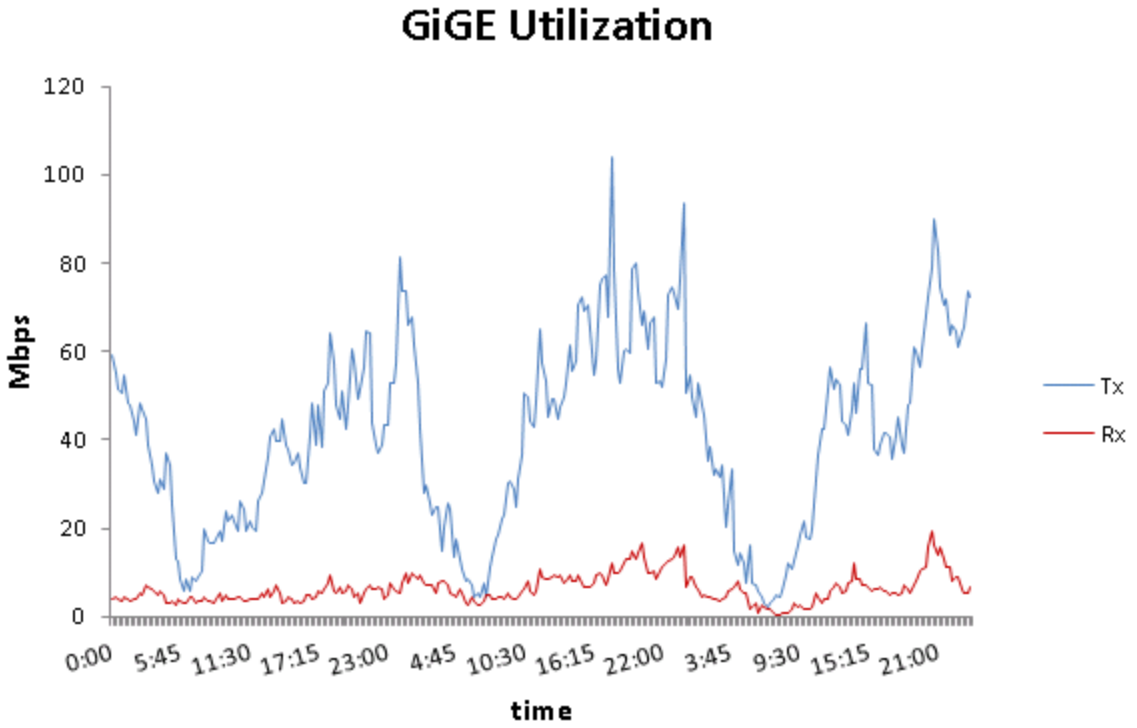


Calls in session are discrete values. Number in session changes over continuous time.

Erlangs = Offered Traffic = Average Concurrent Number in Session = A * H

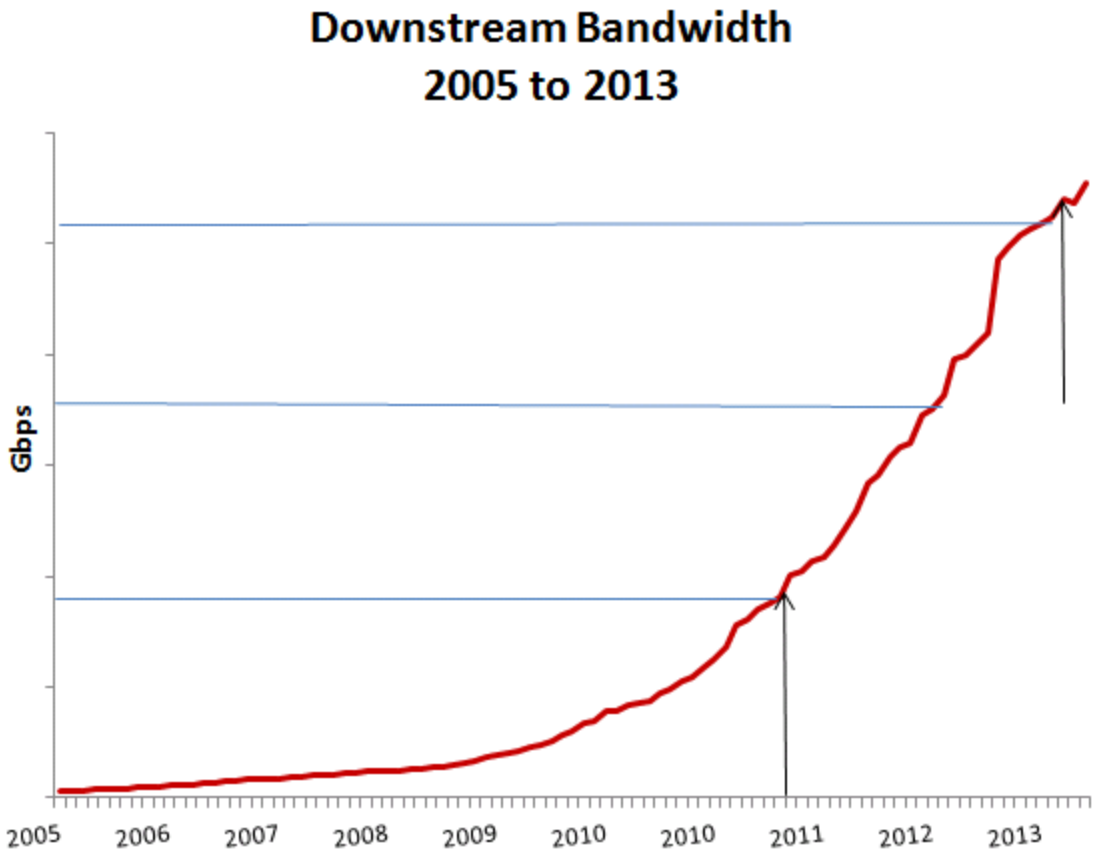
A = arrivals per time period H= duration in same time unit, 1/H is departure rate

Typical Data Network Capacity Planning Measurements



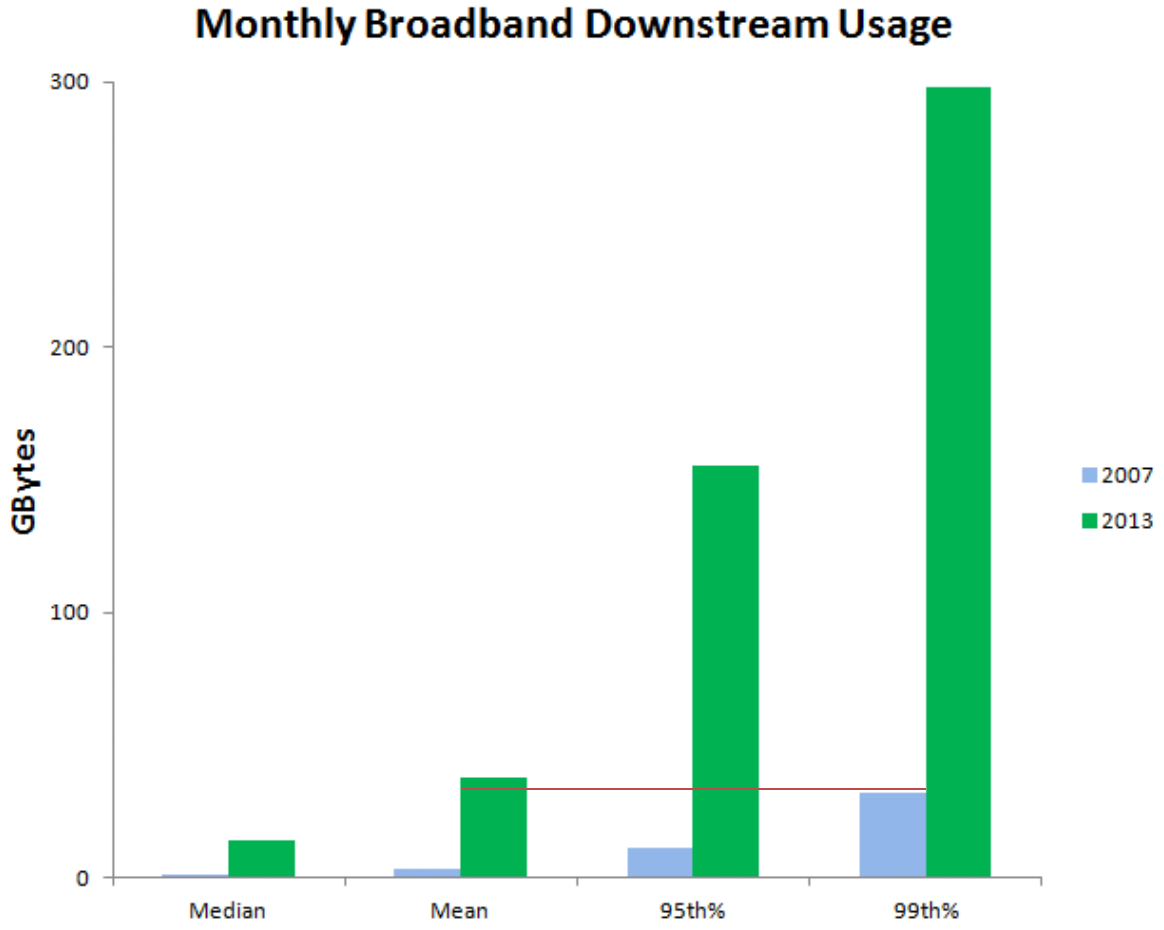
Data rate measured at regular time intervals. Pattern is bursty. New applications can create havoc very quickly. Much variability and uncertainty. Functionality always changing.

Demand for Bandwidth on the Internet Growing Dramatically



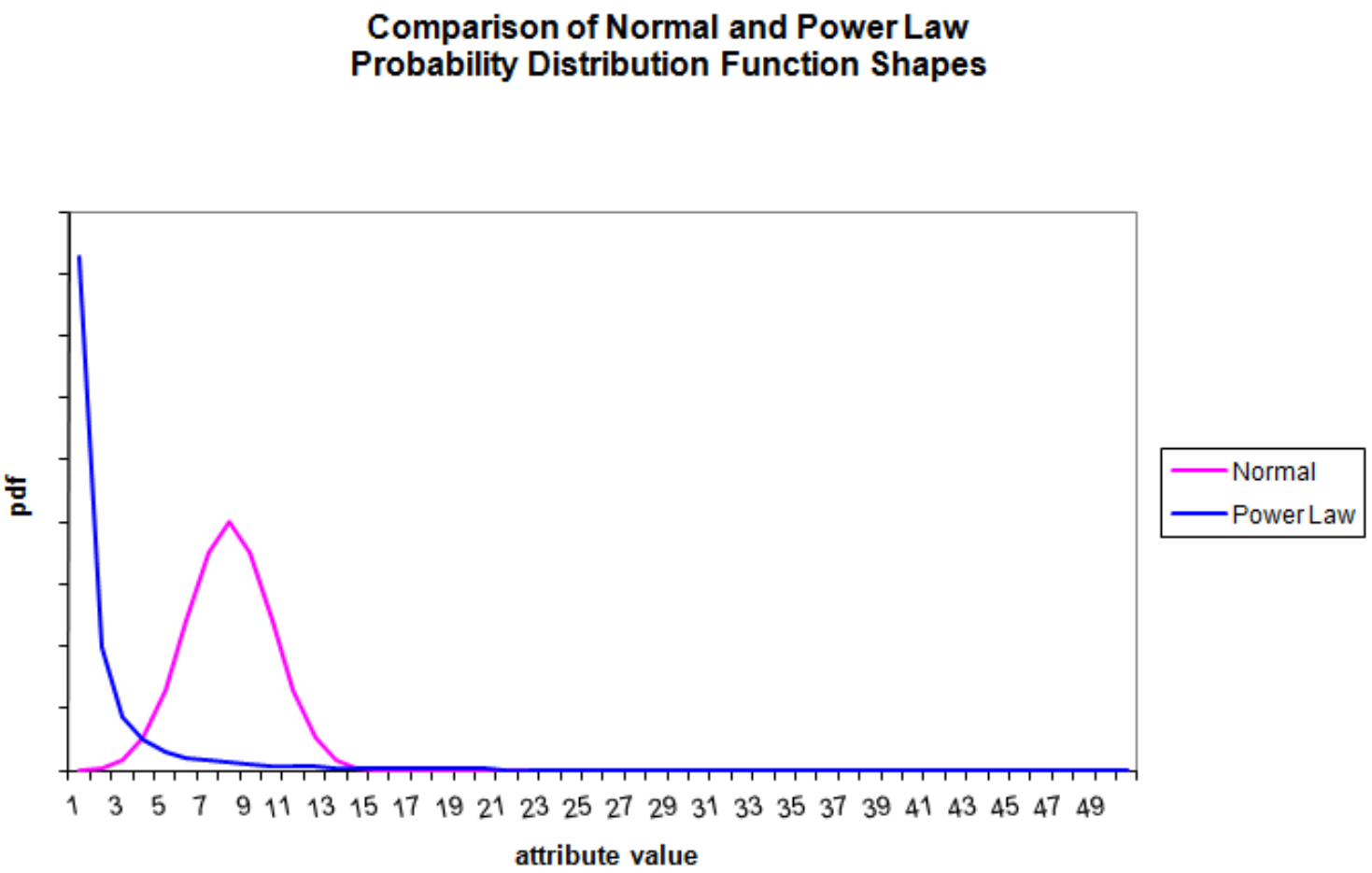
Growth in total bandwidth demand grew in one year by the total amount up to 2011.

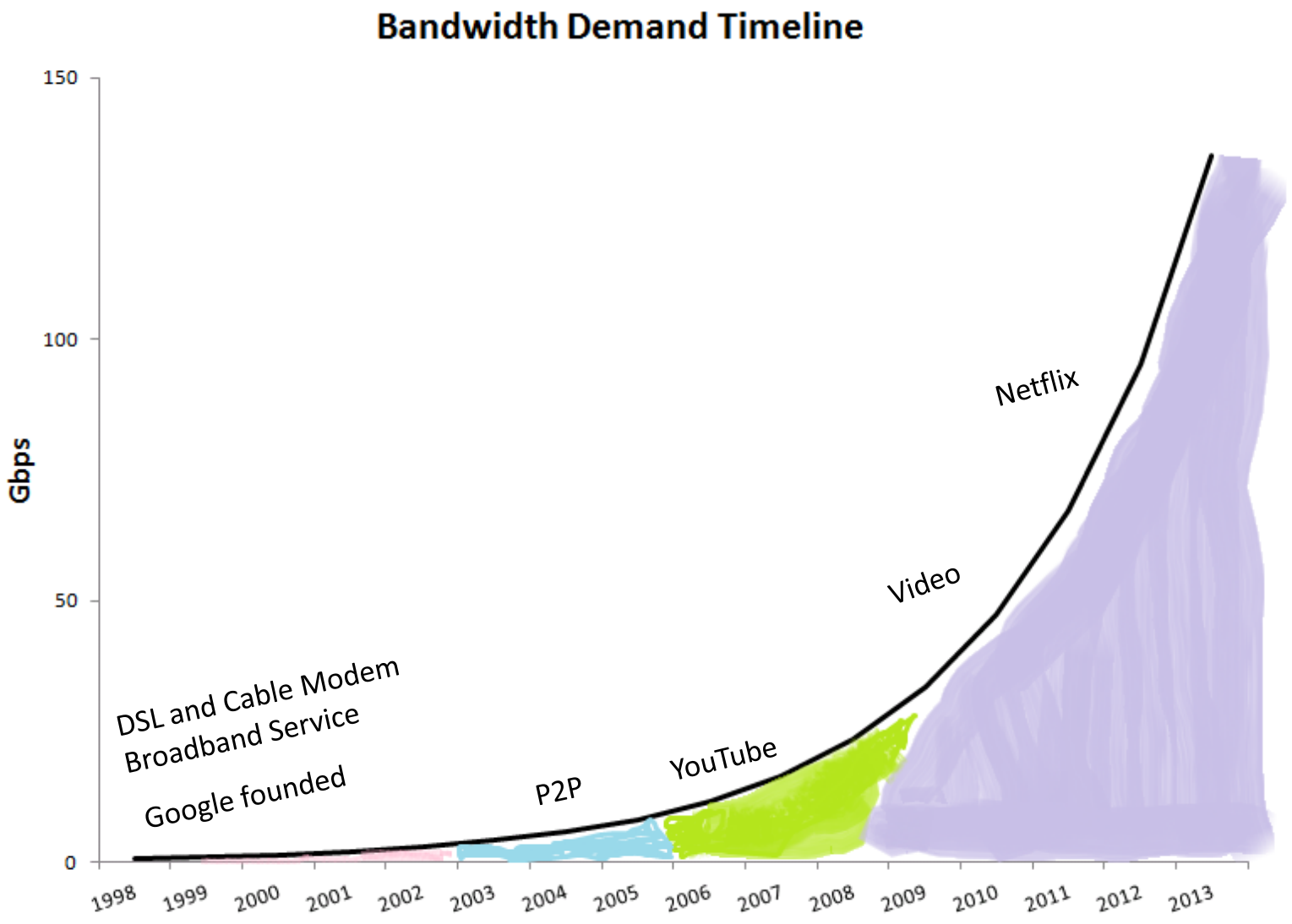
What factors are causing this dramatic increase in demand for bandwidth?



The 99th % of usage has become the mean …

Monthly Download Use - Power Law Type Distribution


Comparison of Normal and Power Law
Probability Distribution Function Shapes
pdf
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
attribute value
Normal
Power Law


Bandwidth Demand Timeline
Gbps
150
100
50
0
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
DSL and Cable Modem
Broadband Service
Google founded
P2P
YouTube
Video
Netflix

Transformation of the Internet

- The Internet was originally a distributed system with user generated content (e-mail, web sites, file transfers)
- Arbor 2009 Rise of the Hyper Giants report – it was observed that a small numberof dominant content providers were generating most of the bandwidth consumed by users
- Evolution to concentration of commercial sources with asymmetrical network traffic patterns
- Commercial transformation and an evolution in how the Internet is being used is driving what needs to be measured on a data network
- The Internet was originally connected only computers. Now it connects a wide range of devices including tablets, TV monitors, gaming, consoles, wireless devices, streaming devices, smart phones, etc.
- Internet use has exploded. Data traffic is growing at nearly 50% each year. What is causing this?




tv

Roku

vudu





Sprint



SAMSUNG









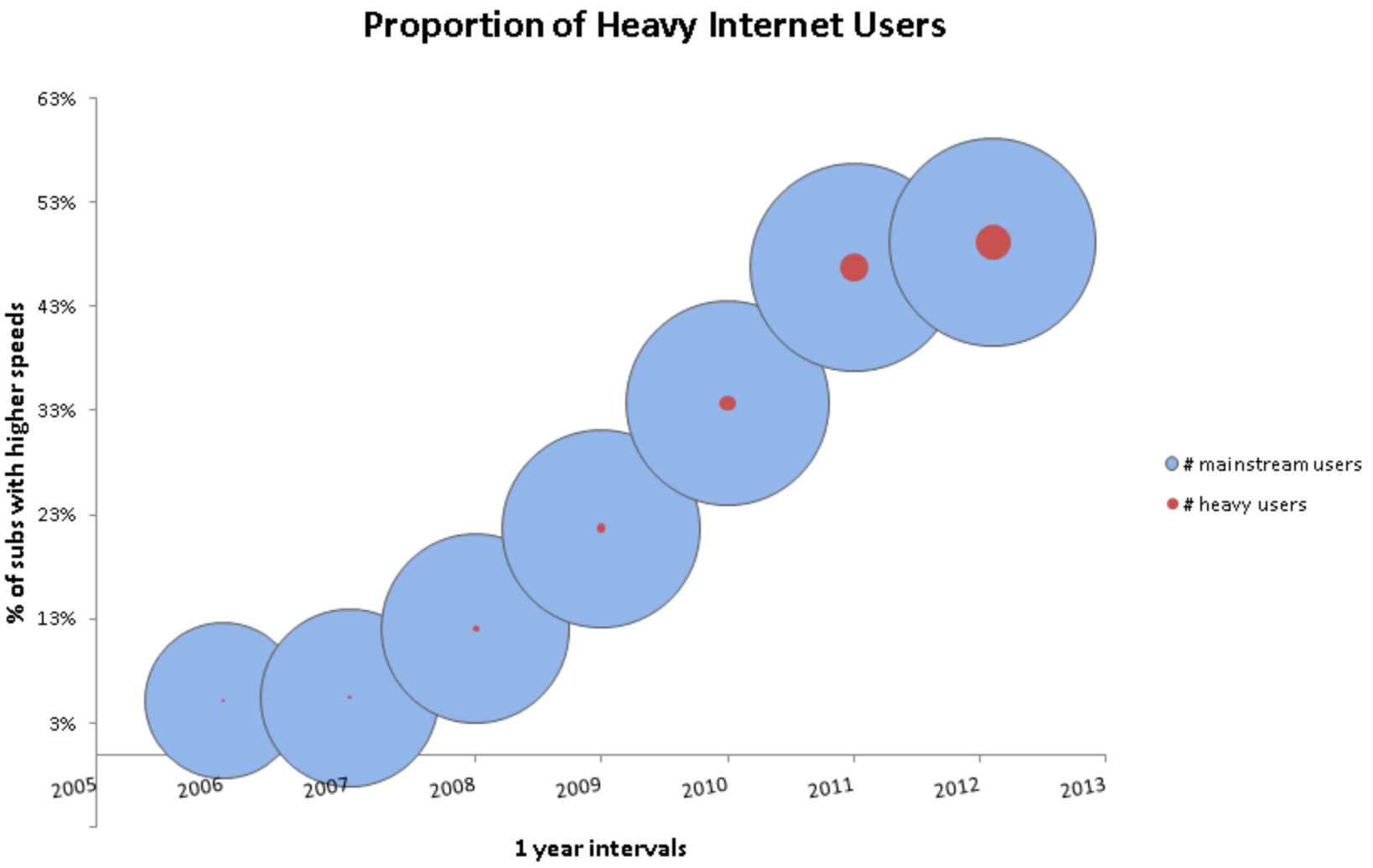
PLAYSTATION 3

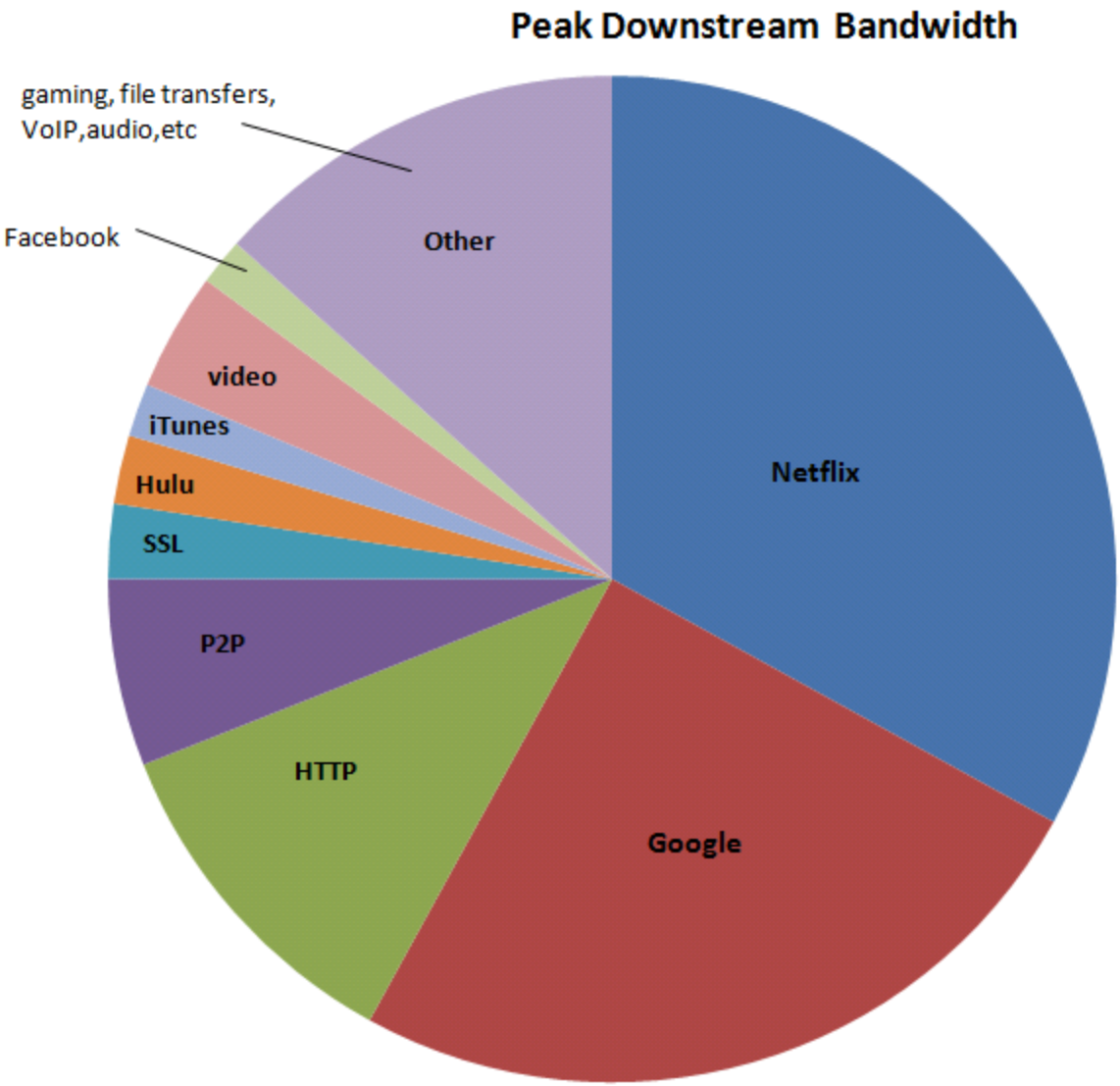

From the physical layer to higher layers …

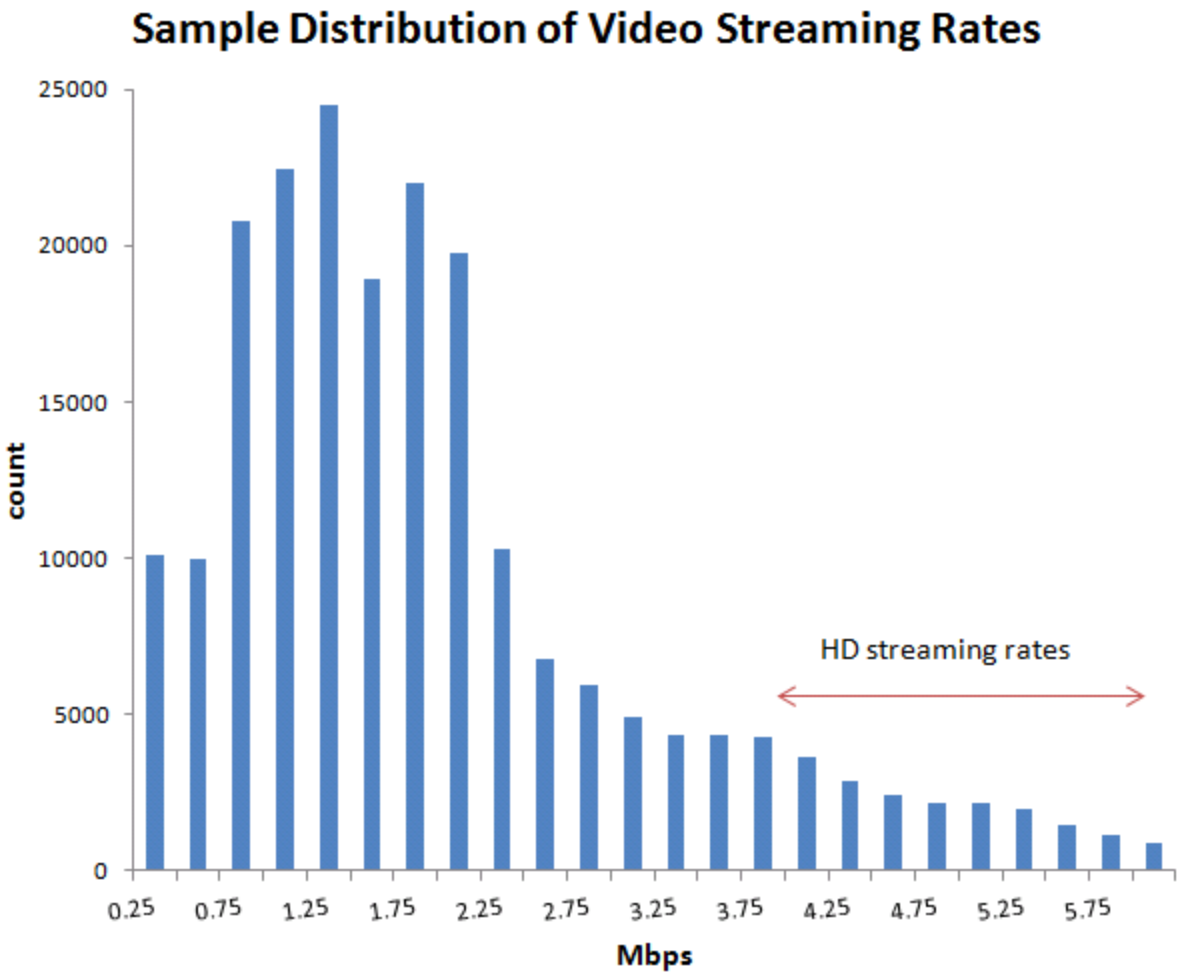
- The commodity of a data network is information, not just the physical characteristics of generic bits or bytes. (Tim Wu, *The Master Switch*)
- Measurements need to be more granular to understand services are being provided. Can only know what is being carried by monitoring at higher layers.
- It is inadequate to monitor only basic physical network characteristics without knowledge of their relationship to applications in a converged network
- The boundaries in service delivery between applications, end user devices, OS and connectivity are blurring
- A loose coordination/structure exists between media creators, distributors and devices

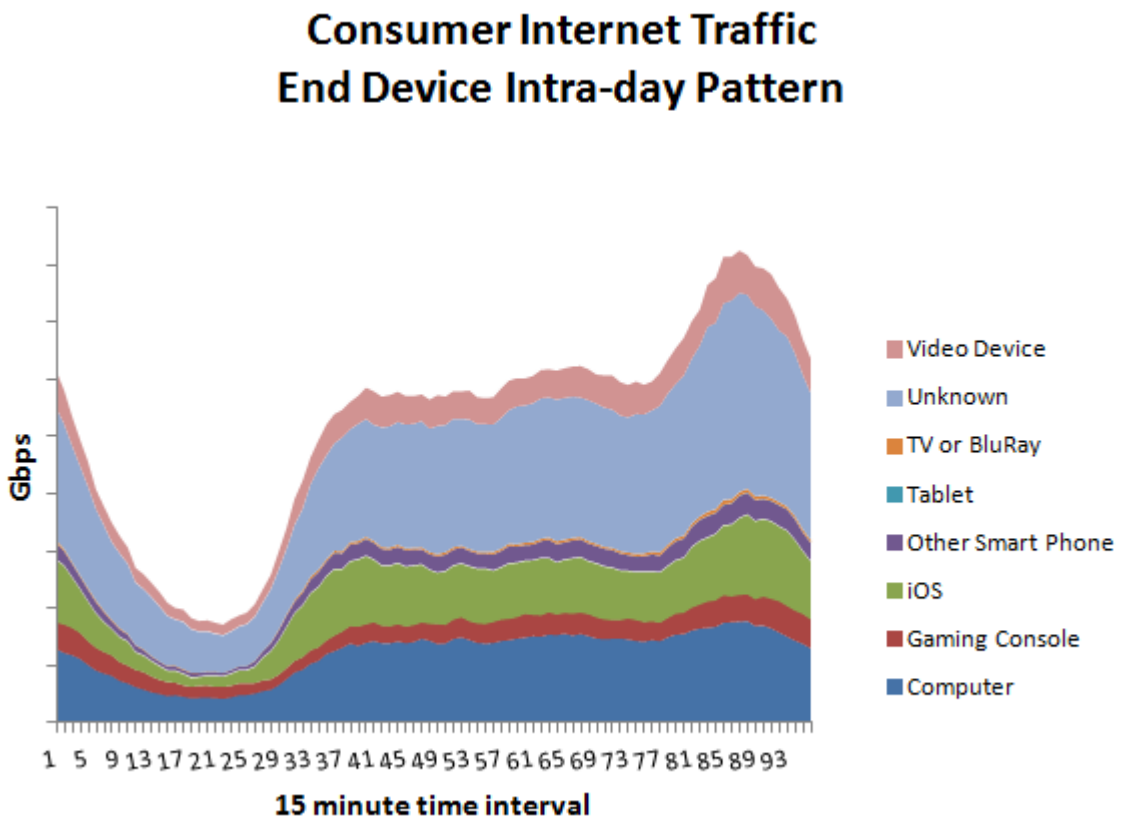
Bandwidth Demand Accelerators

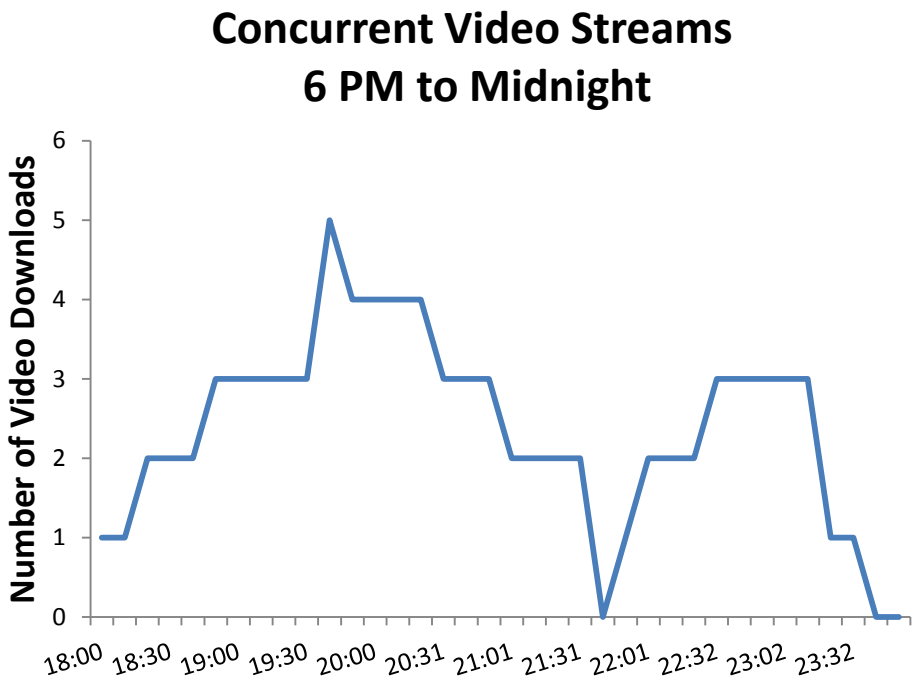
- Faster speed connectivity
- Video
- Higher quality video
- More end devices
- Wi-fi offload
- Multiple devices in use per connection
- More people using the Internet during peak hours
- Economic conditions

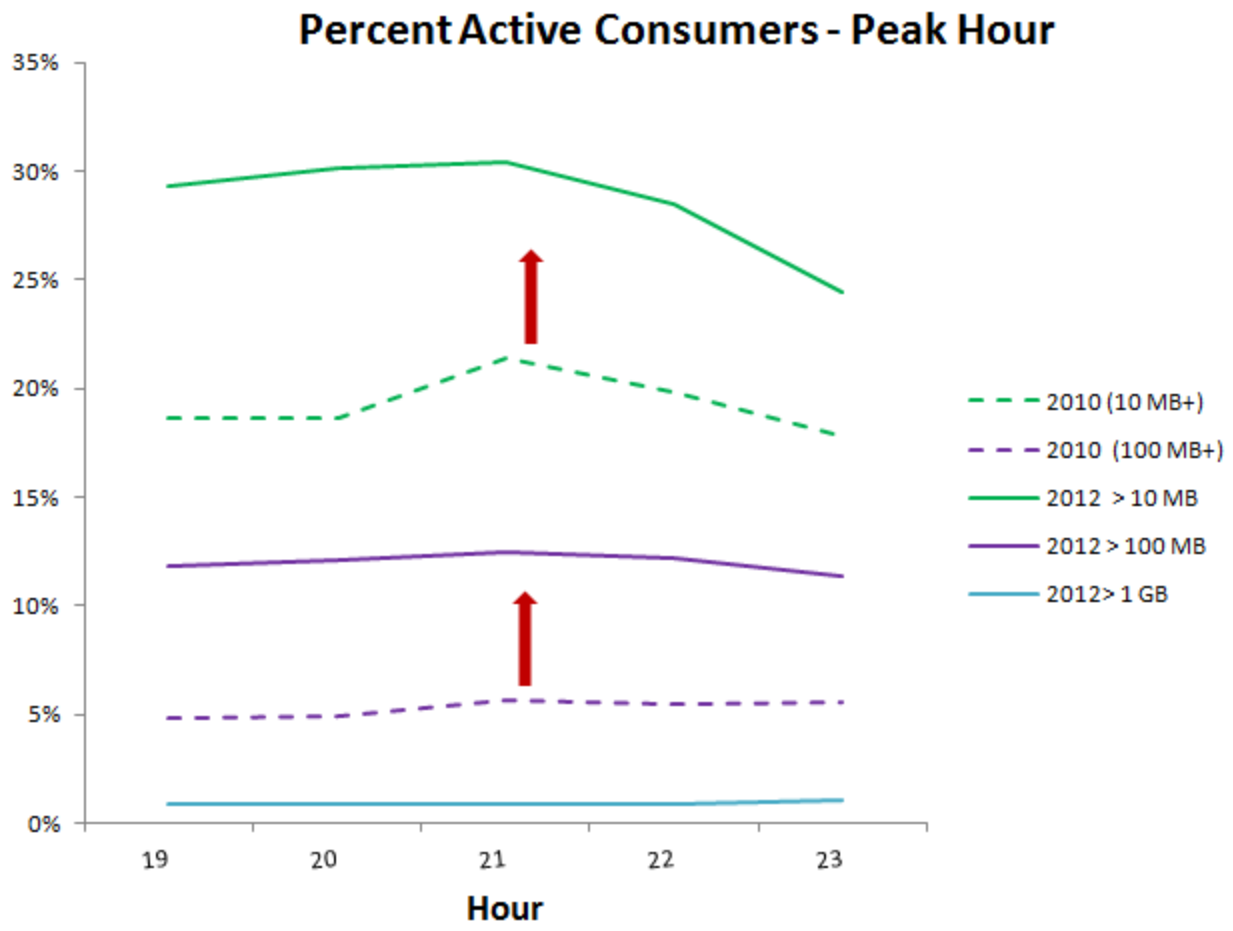

Proportion of Heavy Internet Users
63%
53%
43%
33%
23%
13%
3%
% of subs with higher speeds
2005
2006
2007
2008
2009
2010
2011
2012
2013
1 year intervals
mainstream users
heavy users


Peak Downstream Bandwidth
gaming, file transfers, VoIP,audio,etc
Facebook
Other
video
iTunes
Hulu
SSL
P2P
HTTP
Netflix
Google


Sample Distribution of Video Streaming Rates
25000
20000
15000
10000
5000
0
count
HD streaming rates
0.25
0.75
1.25
1.75
2.25
2.75
3.25
3.75
4.25
4.75
5.25
5.75
Mbps


Consumer Internet Traffic
End Device Intra-day Pattern
Gbps
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89 93
15 minute time interval
Video Device
Unknown
TV or BluRay
Tablet
Other Smart Phone
iOS
Gaming Console
Computer

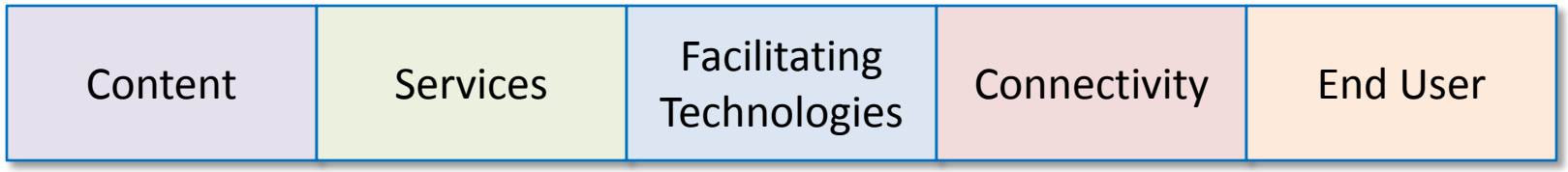

Concurrent Video Streams
6 PM to Midnight
Number of Video Downloads
6
5
4
3
2
1
0
18:00
18:30
19:00
19:30
20:00
20:31
21:01
21:31
22:01
22:32
23:02
23:32


Percent Active Consumers - Peak Hour
35%
30%
25%
20%
15%
10%
5%
0%
19
20
21
22
23
Hour
2010 (10 MB+)
2010 (100 MB+)
2012 > 10 MB
2012 > 100 MB
2012> 1 GB

What are the top ten Internet applications?

- From a bandwidth perspective?
- From a popularity perspective?
- From the commercial distributor perspective?
- From the generic application perspective?
- From a device perspective?
- From a subscriber profile perspective?
- From a protocol perspective?
- May be simplest to start with the function of the application

Linear Perspective on Internet Ecosystem (source to destination)



Content - commercial media or user generated (video, audio, images, voice, texts, books, games, Information files)

Services – communications (e-mail, IM, social networking, VoIP), search, entertainment (video, gaming), commerce transactions (banking, Amazon, E-bay), information/news/opinion, online billing and payment

Facilitating/Intermediate technologies – CDN caching, web hosting, advertising

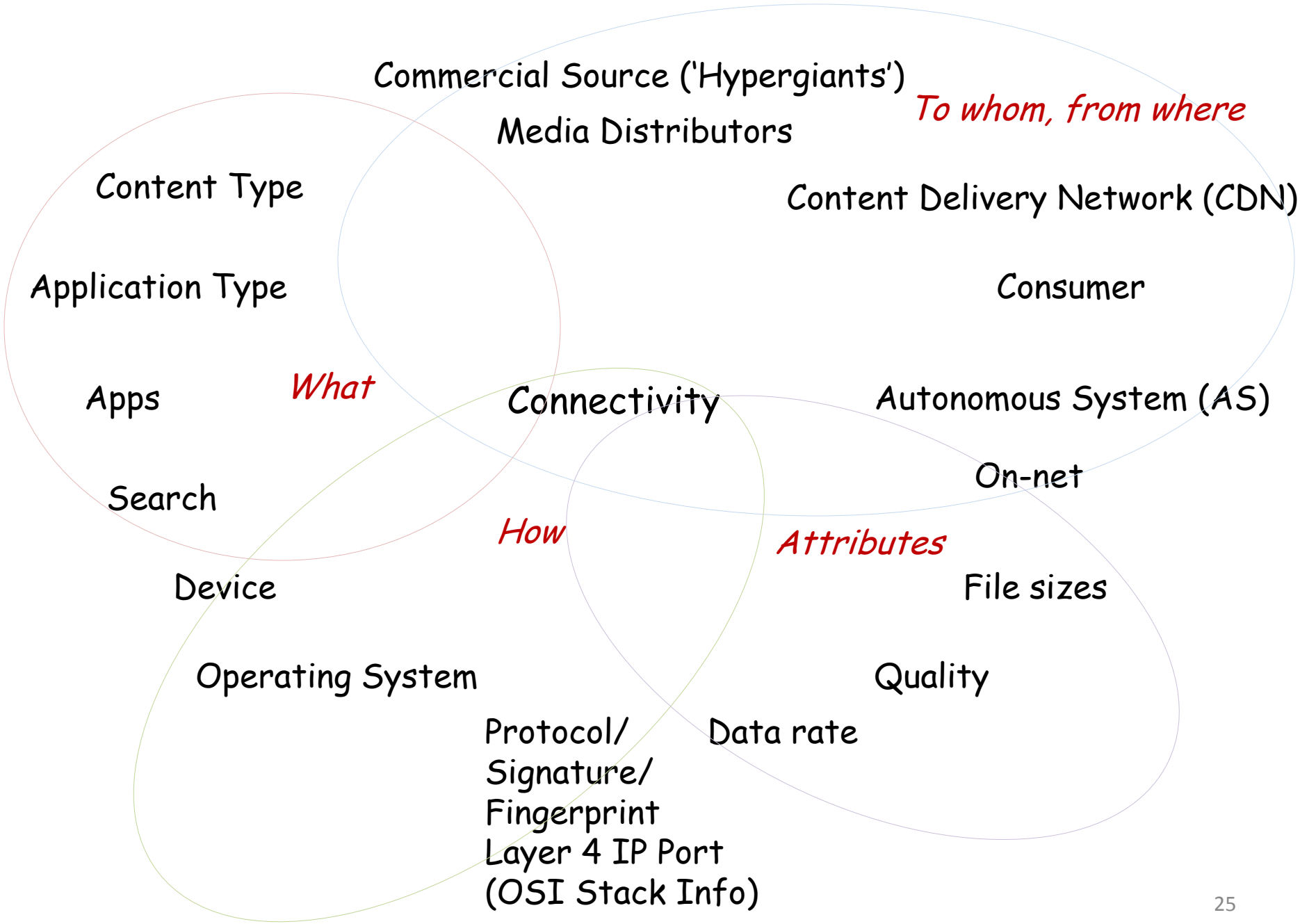
Connectivity – core network, interconnection between network operators, local access (technology, protocols, IP addressing)

End User - operating systems, browsers, security software, devices (gaming consoles, Apple TV, Roku, PC, Mac, Chromecast, smart phones, tablets, femtocell, sling box), wireline or wi-fi

What is a service?

- A service comprises combinations of connectivity, applications, protocol and devices.
- How may a service be categorized?
  - Who is providing the content?
  - Who is distributing the content?
  - What is the content?
  - How is content being delivered?
- Dramatic transformation from a network delivering one type of service

Elements Combined to Create an Internet 'Service'



Privacy

- Privacy on the Internet is fundamental to protect individual control over what personal information one selects to reveal about themselves publically (health, political beliefs, finances, etc.)
- Dignity, liberty, personal security and social welfare are at stake
- Granular level data may be collected and immediately aggregated to summary statistics on large groups to protect individual privacy
- Privacy is protected through careful design of data collection and processing procedures

Understanding a Complex System Requires Analysis of Large Data Sets

- As traditional communications networks collapse to a common one, an understanding of the services carried by it is essential for the Internet Service Provider
- Understanding the traffic requires gathering and analyzing large volumes of data that was not fathomable or needed a decade ago
- Dramatic improvements in monitoring equipment performance and reduction in data storage costs have enabled the collection of measurements needed to optimally design networks
- Data driven decisions are required to design cost effective/ high performance networks and to optimize coordination between elements enabling services in the Internet ecosystem