Measuring what might be regulated; Regulating what might be measured in the internet

Dr. Jonathan Liebenau & Dr. Silvia Elaluf-Calderwood

Department of Management
LSE

WIE 2013 at CAIDA,
UC San Diego
Measuring what might be regulated;
Regulating what might be measured in the internet

Dr. Jonathan Liebenau &
Dr. Silvia Elaluf-Calderwood

Department of Management
LSE

WIE 2013 at CAIDA,
UC San Diego
Core Questions

• Is internet governance sustainable? Which metrics are needed for governance? Who has relevant data?
• Significance of response to NSA? People all over the world woke up to the fact that the system has not got the governance control that they had assumed How feasible? What effect on WWW?
From network of networks to multiple internets

Move from 1990s “network of networks” (Noam, 2002) to the idea of multiple internets (Yoo 2012; Claffy & Clark, 2013)

Underlying concept of containing the inter-relationship of the networks undermined by:
- Market shaping (regulatory boundaries & powers)
- Limited capabilities of incumbents (half-hearted efforts to establish CDNs, cloud services, content production)
- Strategic vertical integration of internet firms (Google, Apple, Amazon, etc.)
- National policies restricting data trade (ex. non-export of financial data)
- Lack of trust (NSA; identity theft; privacy infringements, security breaches)

NSA is not the whole explanation:
- NSA reveals forces that makes it apparent
- Starts 2007 w/ iPhone
- Business models, privacy, security, identity, architecture rapidly shift
- Special pleading arguments from telecom incumbents against OTTs; CDNs
Move from 1990s “network of networks” (Noam, 2002) to the idea of multiple internets (Yoo 2012; Claffy & Clark, 2013)
Underlying concept of containing the inter-relationship of the networks undermined by:

- Market shaping (regulatory boundaries & powers)
- Limited capabilities of incumbents (half-hearted efforts to establish CDNs, cloud services, content production)
- Strategic vertical integration of internet firms (Google, Apple, Amazon, etc.)
- National policies restricting data trade (ex. non-export of financial data)
- Lack of trust (NSA; identity theft; privacy infringements, security breaches)
NSA is not the whole explanation:
NSA reveals forces that makes it apparent
Starts 2007 w/ iPhone
Business models, privacy, security, identity, architecture rapidly shift
Special pleading arguments from telecom incumbents against OTTs; CDNs
Internet modularity

- Changing architecture, dynamics, usage & business models
- Value added dominated by US firms
- Some attributable to inefficiency, inadequate innovation & incompetence of European business vs. counterparts
- Also: structural & governance inhibitors
Platforms

- Standard approach
  - Intel legacy
  - Two-sided markets
- The internet is not a production line factory
- Model does not explain well:
  - Digital service delivery
  - Multi-sided markets
  - Dynamic, evolving architectures
  - Scale
Metrics

Legacy focus on:
- End users
- Access
- Speed

Role of standards bodies
- ITU
- Regulators
- Academics
Role of commercial entities
- Consultancies
- Akamai
- Sandvine
- Cisco

New needs:
- Traffic
- Volume
- Value added
- Economics of
  - routing
  - control points
  - load balance

Comparison among networks
- How we distinguish what is:
  - comprehensive
  - segmented
  - or partial views
- We don’t know what isn’t counted
Legacy focus on:

- End users
- Access
- Speed
New needs:
• Traffic
• Volume
• Value added
• Economics of
  • routing
  • control points
  • load balance
Comparison among networks
  • How we distinguish what is:
    • comprehensive
    • segmented
    • or partial views
  • We don't know what isn't counted
Role of standards bodies
- ITU
- Regulators
- Academics

Role of commercial entities
- Consultancies
- Akamai
- Sandvine
- Cisco
Metrics & Economics

- Economic variables of data
- Cost of data vs. cost of transport (Lehr et al. 2011)
- Asymmetric behavior
- Differences in perspective of traffic management
  - IP transit
  - private peering
  - public peering
  - CDN
- Market shaping and power issues
Economic variables of data
Cost of data vs. cost of transport (Lehr et al. 2011)
Asymmetric behavior
Market shaping and power issues
Differences in perspective of traffic management

- IP transit
- private peering
- public peering
- CDN
Approaches

- Qualitative
- Quantitative
- Hybrids
- Problems of data validation
<table>
<thead>
<tr>
<th>Report</th>
<th>Comments</th>
<th>Economic Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collectors of traffic counters connected to routers (e.g. CiscoVNI data)</td>
<td>Information based on Cisco’s own collection. Considered industry benchmark. Lots of exceptions. No QoS. Traffic is analysed in categories that are not matched by other industry reports</td>
<td>No correlation between traffic and pricing. but regular reports comment on economic context.</td>
</tr>
<tr>
<td>Reports from firms generating traffic</td>
<td>Google, YouTube, Facebook <a href="http://www.google.com/transparencyreport/traffic/#expand=TJ%20or">http://www.google.com/transparencyreport/traffic/#expand=TJ%20or</a></td>
<td>No indicators for comparison</td>
</tr>
<tr>
<td>Classification of traffic: e.g. Sandvine</td>
<td>Focus sectors e.g. entertainment, home roaming, comm services. Sandvine Global Internet Phenomena Report</td>
<td>Some based primarily on billing, by type of traffic</td>
</tr>
<tr>
<td>Academic studies on the internet e.g. Economides</td>
<td>Theoretical and practical academic analysis of current status of the internet</td>
<td>Yes. Some of the pricing is indirectly estimated</td>
</tr>
<tr>
<td>Consultancy reports commissioned by interested parties</td>
<td>(incumbents, consumer groups, industry assoc., etc.) e.g. BCG, ATKearney (2013) Report on traffic and demand based on ETNO members data</td>
<td>Yes, based on own commissioned calculations</td>
</tr>
</tbody>
</table>
### Explanation of Traffic Categories

The table below describes each of the traffic categories used in the Global Internet Phenomena Report: 2H 2013.

<table>
<thead>
<tr>
<th>Traffic Category</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage</td>
<td>Large data transfers using the File Transfer Protocol or its derivatives. Services that provide file-hosting, network back-up, and one-click downloads</td>
<td>FTP, Rapidshare, Mozy, zShare, Carbonite, Dropbox</td>
</tr>
<tr>
<td>Gaming</td>
<td>Console and PC gaming, console download traffic, game updates</td>
<td>Nintendo Wii, Xbox Live, Playstation 2, Playstation 3, PC games</td>
</tr>
<tr>
<td>Marketplaces</td>
<td>Marketplaces where subscribers can purchase and download media including applications, music, movies, books, and software updates</td>
<td>Google Android Marketplace, Apple iTunes, Windows Update</td>
</tr>
<tr>
<td>Administration</td>
<td>Applications and services used to administer the network</td>
<td>DNS, ICMP, HTTP, SNMP</td>
</tr>
<tr>
<td>Filesharing</td>
<td>Filesharing applications that use a peer-to-peer or Newsgroups as a distribution models</td>
<td>BitTorrent, eDonkey, Gnutella, Ares, Newsgroups</td>
</tr>
<tr>
<td>Communications</td>
<td>Applications, services and protocols that allow email, chat, voice, and video communications; information sharing (photos, status, etc) between users</td>
<td>Skype, WhatsApp, iMessage, FaceTime</td>
</tr>
<tr>
<td>Real-Time Entertainment</td>
<td>Applications and protocols that allow “on-demand” entertainment that is consumed (viewed or heard) as it arrives</td>
<td>Streamed or buffered audio and video (RTSP, RTP, RTMP, Flash, MPEG), peercasting (PPStream, Octoshape), specific streaming sites and services (Netflix, Hulu, YouTube, Spotify,)</td>
</tr>
<tr>
<td>Social Networking</td>
<td>Websites and services focused on enabling interaction (chat, communication) and information sharing (photos, status, etc) between users</td>
<td>Facebook, Twitter, LinkedIn, Instagram</td>
</tr>
<tr>
<td>Tunneling</td>
<td>Protocols and services that allow remote access to network resources or mask application identity.</td>
<td>Remote Desktop, VNC, PC Anywhere, SSL, SSH,</td>
</tr>
<tr>
<td>Web Browsing</td>
<td>Web protocols and specific websites</td>
<td>HTTP, WAP browsing</td>
</tr>
</tbody>
</table>

Sandvine traffic categories (2013)
<table>
<thead>
<tr>
<th>Metric</th>
<th>Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Penetration</strong></td>
<td>• Business surveys: total number of subscriptions&lt;br&gt;• Household and consumer surveys: proportion of houses connected to Internet/broadband</td>
</tr>
<tr>
<td><strong>Speed</strong></td>
<td>• Business surveys and market research: advertised speeds (e.g., OECD, FCC)&lt;br&gt;• Content delivery networks and web services: download speeds (e.g., Akamai, Netflix)&lt;br&gt;• Distributed client-side hardware: download and upload speeds (e.g., government partnerships with SamKnows)&lt;br&gt;• Crowdsourcing: download and upload speeds (e.g., Ookla’s Speedtest, M-Lab)</td>
</tr>
<tr>
<td><strong>Price</strong></td>
<td>• Market research: comparison of offers across different ISPs and countries (e.g., OECD, FCC)&lt;br&gt;• Crowdsourcing: user-submitted information on prices (e.g., Ookla’s Net Index)</td>
</tr>
<tr>
<td><strong>Infrastructure:</strong>&lt;br&gt;• location, size, and routing</td>
<td>• IP address distribution&lt;br&gt;• Allocation of domains&lt;br&gt;• Number of Internet hosts&lt;br&gt;• Number, size, and relationships of autonomous systems (AS)&lt;br&gt;• Network bandwidth estimates&lt;br&gt;• Internet exchange (IX) location and traffic&lt;br&gt;• Route identification and analysis&lt;br&gt;• National network status (e.g., Renesys, Arbor Networks)&lt;br&gt;• International pipe location, traffic, and dependencies</td>
</tr>
</tbody>
</table>

*Traditional metrics - Source Berkman Institute (2013)*
<table>
<thead>
<tr>
<th>General Data Type(s)/Origin(s)</th>
<th>Specific Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reporting on individual behavior</td>
<td>• Client-side behavioral monitoring software (e.g., ComScore, Alexa)</td>
</tr>
<tr>
<td></td>
<td>• Cookies and browsing history</td>
</tr>
<tr>
<td></td>
<td>• Consumer surveys</td>
</tr>
<tr>
<td>Network monitoring: location, type, and quantity of traffic</td>
<td>• Monitoring by ISPs</td>
</tr>
<tr>
<td></td>
<td>• Monitoring by network services (e.g., content distribution networks, Internet security companies)</td>
</tr>
<tr>
<td>Data collection by websites and services: visitors, contributors, content, links, comments,</td>
<td>• Websites, including social media platforms (e.g., user-generated content sites, social network sites, blogging and micro-blogging sites)</td>
</tr>
<tr>
<td>languages, locations</td>
<td>• Search data</td>
</tr>
<tr>
<td>Social media mapping: link- and/or content-based</td>
<td>• Landscape mapping: platform/service-based mapping (e.g., Twitter, Facebook, blogosphere)</td>
</tr>
<tr>
<td></td>
<td>• Topical or issue-based mapping</td>
</tr>
<tr>
<td>Qualitative assessments</td>
<td>• Expert opinion surveys</td>
</tr>
<tr>
<td>General Data Type(s)/Origin(s)</td>
<td>Specific Source(s)</td>
</tr>
<tr>
<td>-------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Reporting on individual behavior</td>
<td>• Client-side behavioral monitoring software (e.g., ComScore, Alexa)</td>
</tr>
<tr>
<td></td>
<td>• Cookies and browsing history</td>
</tr>
<tr>
<td></td>
<td>• Consumer surveys</td>
</tr>
<tr>
<td>Network monitoring: location, type, and quantity of traffic</td>
<td>• Monitoring by ISPs</td>
</tr>
<tr>
<td></td>
<td>• Monitoring by network services (e.g., content distribution networks, Internet security companies)</td>
</tr>
<tr>
<td>Data collection by websites and services: visitors, contributors, content, links, comments, languages, locations</td>
<td>• Websites, including social media platforms (e.g., user-generated content sites, social network sites, blogging and micro-blogging sites)</td>
</tr>
<tr>
<td></td>
<td>• Search data</td>
</tr>
<tr>
<td>Social media mapping: link- and/or content-based</td>
<td>• Landscape mapping: platform/service-based mapping (e.g., Twitter, Facebook, blogosphere)</td>
</tr>
<tr>
<td></td>
<td>• Topical or issue-based mapping</td>
</tr>
<tr>
<td>Qualitative assessments</td>
<td>• Expert opinion surveys</td>
</tr>
</tbody>
</table>

*Behaviour metrics - Berkman (2013)*
### Table 4: Methods for Measuring Internet Activity

<table>
<thead>
<tr>
<th>Method</th>
<th>Strengths</th>
<th>Limitations</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual surveys</td>
<td>• Robust sampling can offer data that is representative of general population.</td>
<td>• Expensive to employ</td>
<td>• Peer Internet, Surveys</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requires baseline data on individuals to accurately record and report information.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Opt-in survey less representative.</td>
<td></td>
</tr>
<tr>
<td>Business surveys</td>
<td>• Can provide near-real-time coverage of users.</td>
<td>• Reporting based on user intention to enter or under report.</td>
<td>• Workers and market subscription data</td>
</tr>
<tr>
<td></td>
<td>• Reflect actual transaction data.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Report surveys</td>
<td>• Able to obtain complete report on users.</td>
<td>• Other highly subjective</td>
<td>• World Economic Forum, Web Index, Facebook, Twitter, YouTube, Reddit, Oktos</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Difficult to ensure non-response consistency</td>
<td></td>
</tr>
<tr>
<td>Cardcaching</td>
<td>• Often able to achieve broad coverage.</td>
<td>• Climate and environmental factors.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Can be highly responsive to changing contexts and events.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web analytics</td>
<td>• Can provide comprehensive view of platforms/websites.</td>
<td>• Tend to be limited to specific platforms and laws.</td>
<td>• YouTube, Wikipedia, Facebook, Twitter, Reddit, YouTube, Reddit, Oktos</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Limited to transactional behavior.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Often proprietary</td>
<td></td>
</tr>
<tr>
<td>Data-scape monitoring</td>
<td>• Detailed individual online behavior.</td>
<td>• Sampling/representativeness issues.</td>
<td>• Social media, Facebook, Twitter, Reddit, YouTube, Reddit, Oktos</td>
</tr>
<tr>
<td>Network monitoring</td>
<td>• Can offer broad view of users’ Internet transactions and traffic.</td>
<td>• Proprietary</td>
<td>• Social media, Facebook, Twitter, Reddit, YouTube, Reddit, Oktos</td>
</tr>
<tr>
<td>Botnet and botneting infrastructure</td>
<td>• View of infrastructure: monitoring Internet breaches.</td>
<td>• Useable dataset of real access and behavior.</td>
<td>• Internet bots, IP addresses, International bandwidth</td>
</tr>
</tbody>
</table>
Catching up with Business Models

- Microsoft (SLAs, Azure & cloud ownership)
- Amazon (value to non-US based customers; NSA links)
- Network operators (esp. mobile) & collaboration (e.g. Blackberry in India, Indonesia, Saudi Arabia)
- Facebook/Google/Amazon & data center businesses
- Emerging national or regional models
- Cloud challenges (national segmentation; transportability; accessibility)
- Traffic monitoring and control in real time
Regulatory practices & the changing worldwide digital economy

- Regulators’ inadequate accommodation of architecture.
  - Some incumbent networks held to be distinct
  - Some networks “governed” by “market forces”
- Legacy boundaries of regulation
- Incompatible international differentiation

For terminology
Requirements
Metrics
Mgt. practices, strategic choices & relations w/ financial services differentiate between network operators & internet businesses

Networks as legacies of state-owned enterprises (Europe) / monopolies and duopolies (in USA)

Practices common among leading internet companies are rare among network operators:
- detailed data mining
- capacity monitoring
- rapid shifting of routing
- use of outsourced services

And...
Mgt. practices, strategic choices & relations w/ financial services differentiate between network operators & internet businesses

Networks as legacies of state-owned enterprises (Europe) / monopolies and duopolies (in USA)
And...

Mgt. practices, strategic choices & relations w/ financial services differentiate between network operators & internet businesses

Networks as legacies of state-owned enterprises (Europe) / monopolies and duopolies (in USA)

Practices common among leading internet companies are rare among network operators:
• detailed data mining
• capacity monitoring
• rapid shifting of routing
• use of outsourced services
Future Impact

- New layer of segmentation based on privacy and data requirements
- Severe undermining of the value proposition of USA based companies in regards to data assurance for cloud.
- Cloud services will require a revision on their unique selling point and determination of SLAs with customers.
- New regulatory practices will be put in place to protect data that is geographically sensitive.
Future Impact (2)

- Revision of scope of big data & open data
- Lobbying from privacy groups & citizens rights advocates re.:
- Scope and limitations of internet data collection
- Sharing & storage practices
- New business models for international customers perhaps to offer some customers contracts that specify their data is NSA-monitored and some not monitored contracts
- New internet metrics can help for commercial purposes value how these new business models will be priced and negotiated
What is certain is that...

- The internet from 1969 is now mature and going into middle age
- Business models innovation is driving change
- Regulators are still focused on earlier forms and problems...and new demands for privacy are emerging
- Metrics need to be redefined to full econometric analysis

Easy to get lost in the debate of net neutrality vs. sustainable business models