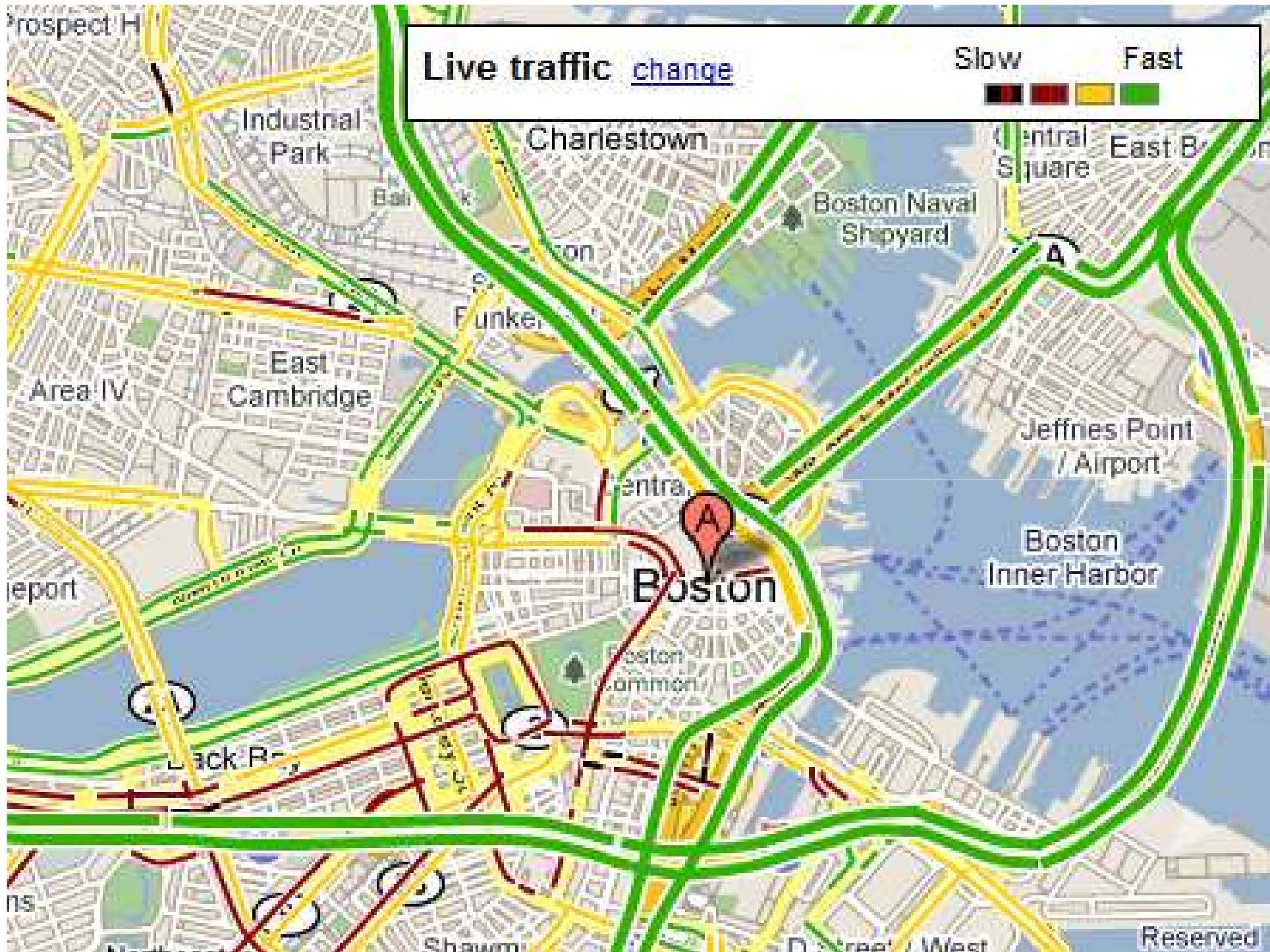


Broadband microfoundations: the need for traffic data

Steven Bauer

Advanced Network Architecture Group

CSAIL, MIT



The problem

- Major of stakeholders have little visibility into what is happening with traffic on a network
- Results in widely diverging opinions about the true state of networks
 - what are the congestion and utilization levels now and in the predicable future?
 - what are the underlying cost structures for carrying traffic and expanding capacity?
 - what are the effects of different traffic management policies?

The consequences

- harder to have confidence in the regulatory and investment decisions that affect networks
- risk of making decisions that have undesirable or unexpected results
 - traffic management policies that are efficient and ‘fair’ could be disrupted
 - private investments in expanding capacity could be deterred
- symmetric risk of failing to make decisions that would have been beneficial
 - network operators could be exploiting their control to thwart or discourage disruptive new innovations and competitors (either intentionally or accidentally)

Rest of this talk...

- data we are gathering
- how this data may prove important in answering questions that are relevant to the entire community of stakeholders
- why collecting the data and making it more generally available is challenging

MIT Internet Traffic Analysis Study (MITAS)

- Per subscriber usage data (up and down byte counts over different measurement periods ranging from 15 minutes, 1 hour, to 1 month) from sources such as IPDR data records
- Link utilization records that represent the aggregate data flow for subscribers (up and down byte counts over different measurement periods again ranging from 15 minutes, 1 hours, to 1 month) collected from sources such as SNMP byte counters
- Percentages of traffic in different classifications such as http, email, video, peer-to-peer etc as determined by the rules of 3rd party traffic analysis boxes.
- Historical utilization data at both for both subscriber level and aggregate data flows
- Reports detailing interesting traffic management incidents and challenges
- Results of individual experiments that for instance explore the correlation between link utilization and packet loss, latency, and jitter

Traffic data

- hundreds of different statistics, values, and events from each network element
 - routers,
 - switches,
 - servers,
 - caches
 - subscriber modems
 - etc

Traffic data is input into strategic and operational decision-making across virtually all ISP functions

- informs decisions about the capacity of internal links
- routing policies
- security policies
- interconnection contracting
- high availability and disaster recovery planning
- financial projections
- employee evaluations
- technical strategy discussions
- sales and marketing

Broadband traffic characterization

- develop representative aggregate and subscriber traffic models
- analyze and forecast market trends
- plan network provisioning and management

Traffic diagnosis

- enable the analysis of significant traffic events
 - de-peering incidents
 - significant cable cuts
 - routing incidents,
 - major media events
 - security incidents
- lessons learned in terms of understanding the actual and potential effects of the incidents
- Understand the effects of such incidents on communications, business, and end users

Traffic management

- non-discrimination
 - cannot discriminate against particular Internet content or applications. This means they cannot block or degrade lawful traffic over their networks, or pick winners by favoring some content or applications over others in the connection to subscribers' homes
- Transparency
 - stating that providers of broadband Internet access must be transparent about their network management practices

Promoting awareness of the challenges, successes, and opportunities in broadband

- How broadband is being used to support novel applications improving
 - education
 - health care
 - business processes
 - entertainment
 - communication
- Which regions are leading and lagging in the adoption of these innovations

Technical challenges

- missing data, spurious data, missing metadata, and ambiguous fields
- varying network measurement methodologies are common over time, across ISPs and measurement equipment providers, and even within a single provider's network
- location of traffic probes in a network determines what traffic is measured
- Megabytes to terabytes, comma separated files to specialized databases

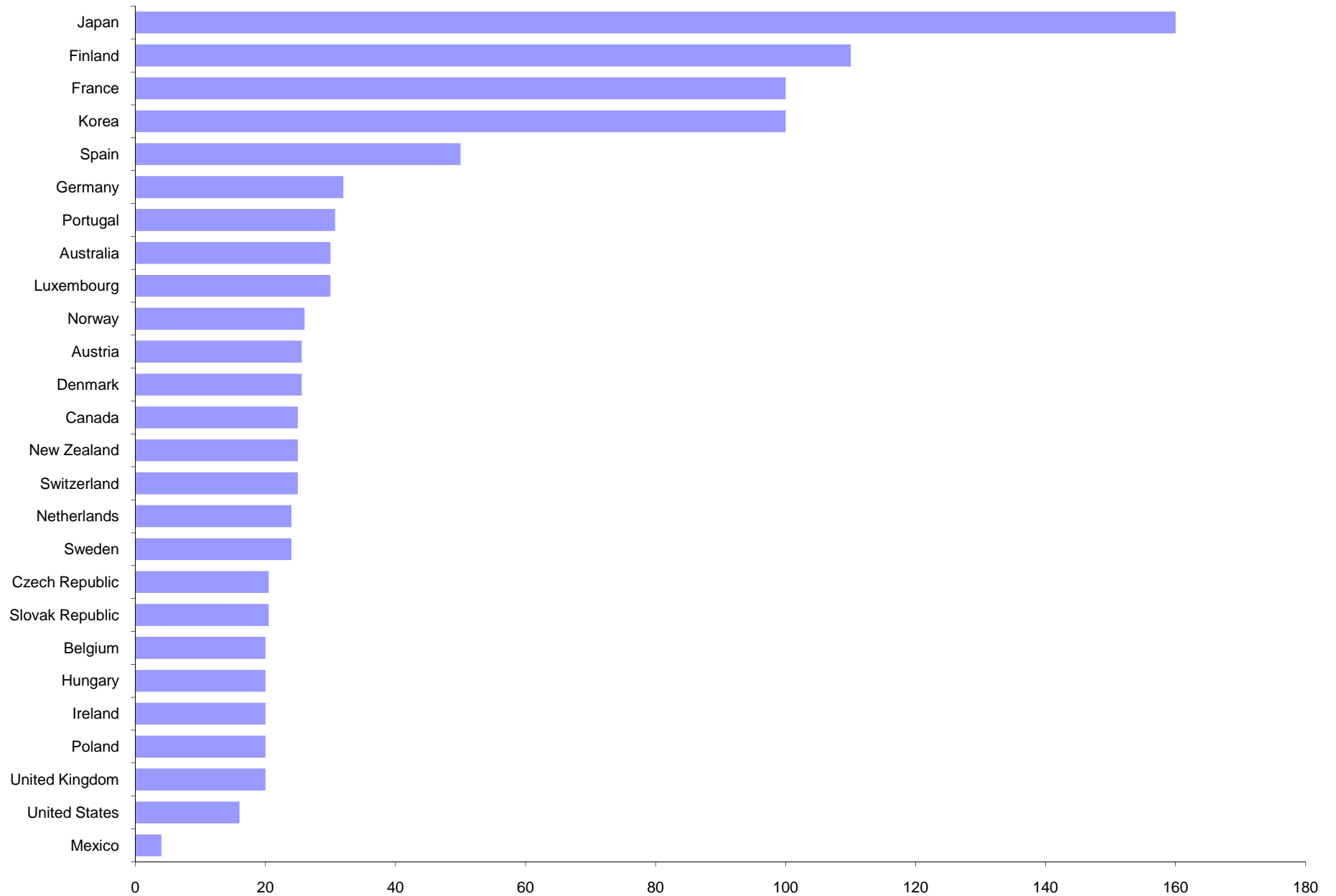
Analytic Challenges

- match traffic characterizations with other types of data in ways that allow analysts to better understand aggregate and per-user behavior
 - what other services (telephony, video, premium video, etc) a subscriber takes,
 - the advertised service characteristics (peak rate, service pricing, etc.) of each subscriber,
 - subscription timing (when was service first initiated, when changed, when terminated),
 - geographic location data about the subscribers, and other types of demographic data

Analytic Challenges

- Often no generally accepted "right" metrics for many of the questions that come up in discussion of traffic data
 - Definition of congestion
 - Measurement locations
 - Measurement intervals

Fastest advertised broadband speeds, using cable, Mbit/s, Sept 2008



Legal Challenges

- privacy implications of measuring individual subscriber behaviors must be carefully managed
 - no clear or universally accepted norms or rules for protecting user data on the Internet
- Institutional Review Board (IRB)

Business Challenges

- We believe efforts to collect data would be most successful if the data is voluntarily contributed
- Rich economics literature documenting the importance of private and asymmetric information, and its potential to effect the allocation of resources and profits

Conclusion

- Better visibility by outside stakeholders into the traffic data of networks is required to improve the
 - regulatory processes,
 - investment/market decision-making,
 - technical research
- Promote understanding and trust between what ultimately has to be a cooperative community of interconnecting and communicating parties

“good data outlives bad theory”

- Data can be useful to later generations of researchers in ways not yet understood
- historical data set of traffic data for the Internet might provide future networking research a baseline for evaluating the large-scale impact of both evolutionary and revolutionary changes in the Internet