Characterizing the reliability of broadband Internet access

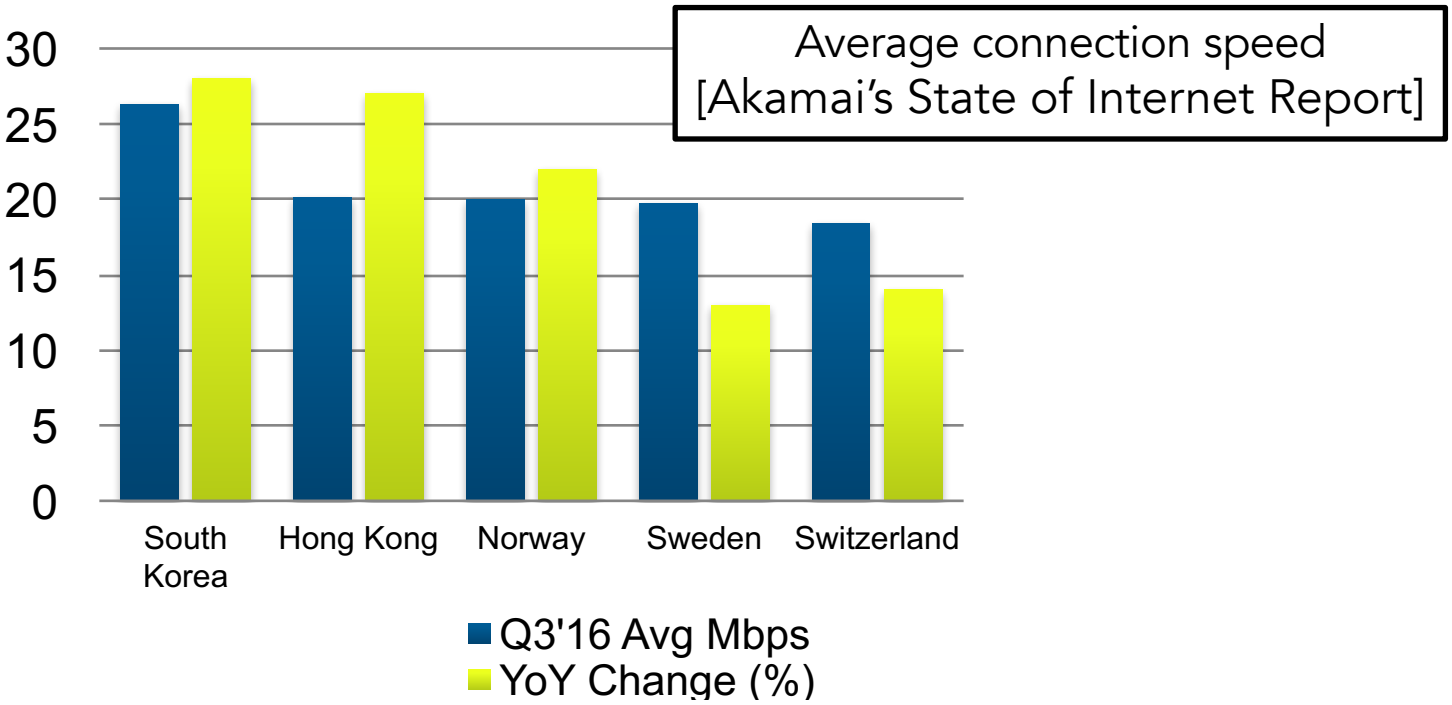
Zachary Bischof
Fabian Bustamante
Nick Feamster

AquaLab

Northwestern

The growth of broadband

- Nearly 1 billion fixed-line broadband subscriptions worldwide
  - Consistent share of total Internet usage, despite increase in mobile subscriptions [ITU State of Broadband report 2016]
- Speeds are increasing rapidly



The importance of being connected

- With higher capacities, a migration to "over-the-top" services

- And higher expectations of reliability
  - The main reason for complaints (71%)*

*Ofcom, UK broadband speed, 2014

theguardian

November 10, 2017

Broadband and landline users to get automatic compensation for poor service

Ofcom says new scheme could result in customers receiving a total of £142m a year in payouts

Broadband reliability – Key questions

- Does reliability matter to end users?
- How reliable are broadband services?
- If not sufficiently reliable, how can we improve them?

Impact of reliability – method

- Measure users' reactions to spontaneous network conditions

- Use FCC/SamKnows dataset
  - ~11k gateways in the US
  - Use ping, DNS and network usage data
  - Ping and network usage data aggregated by hour

- Use network usage as a *proxy* for QoE
  - *Assumption – If unhappy, you use the service less*

Frequent high loss & usage

- Hypothesis – *Frequent periods of high packet loss rates result in lower network demand during periods of normal operation*
- Natural experiment
  - Group users based on fraction of hours with loss ≥ 5%
  - Compare across groups, matching confounding factors

Users with 1-10% hours of ≥ 5% loss

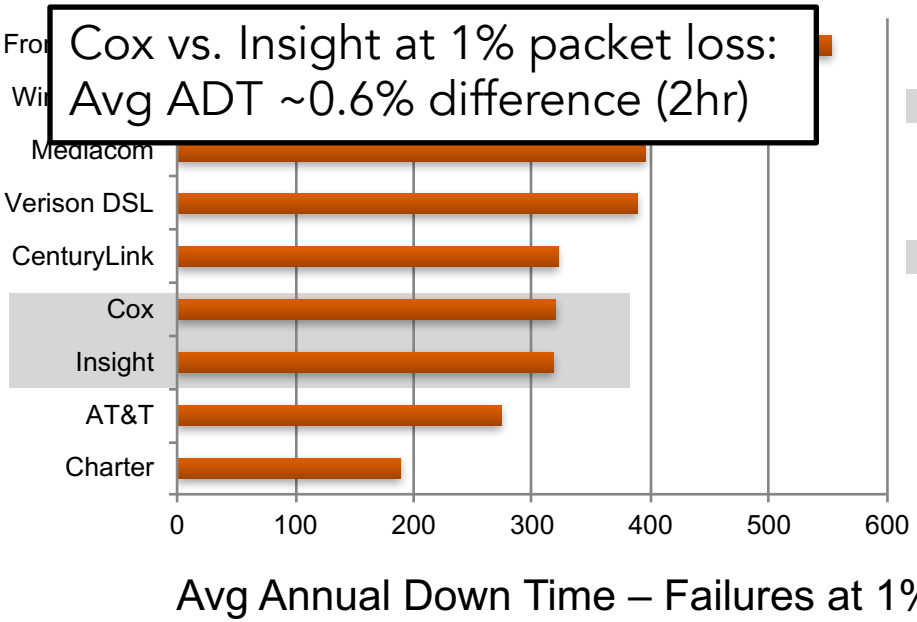
Control group	Treatment group	% H holds	P-value
(1%, 10%)	>10%	68.3	3.65×10^{-5}
(0,5%, 1%)	>10%	70.0	6.95×10^{-6}
(0.1%, 0.5%)	>10%	70.8	2.87×10^{-6}
(0%, 0.1%)	>10%	72.5	4.34×10^{-7}

Increasing difference between control and treatment group's services

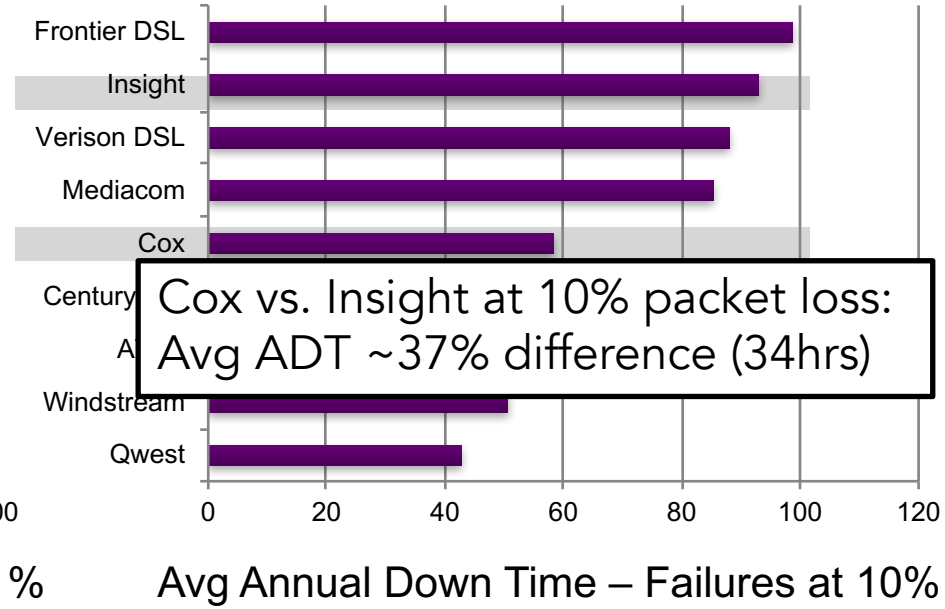
Greater impact

Characterizing reliability

- Metrics of reliability: Mean Time Between Failure (MTBF), Down Time, Availability
- Defining a *failure* for a best-effort service



Avg Annual Down Time – Failures at 1%



Avg Annual Down Time – Failures at 10%

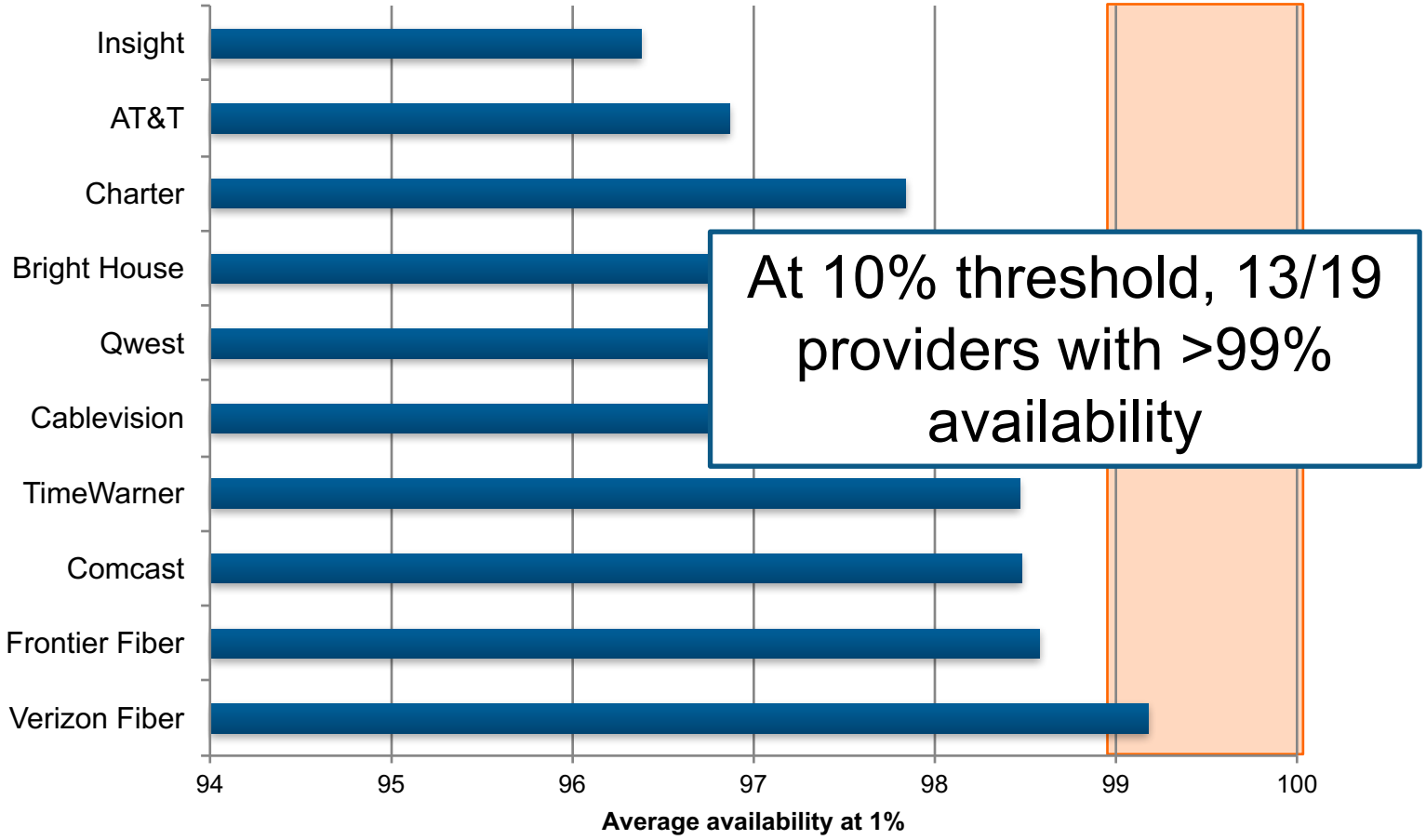
Use three thresholds: 1%, 5% and 10%

Broadband reliability in the US

- Effect of service provider
- Effect of access technology
- Effect of service tier
- Effect of demographics
- ISP and DNS reliability

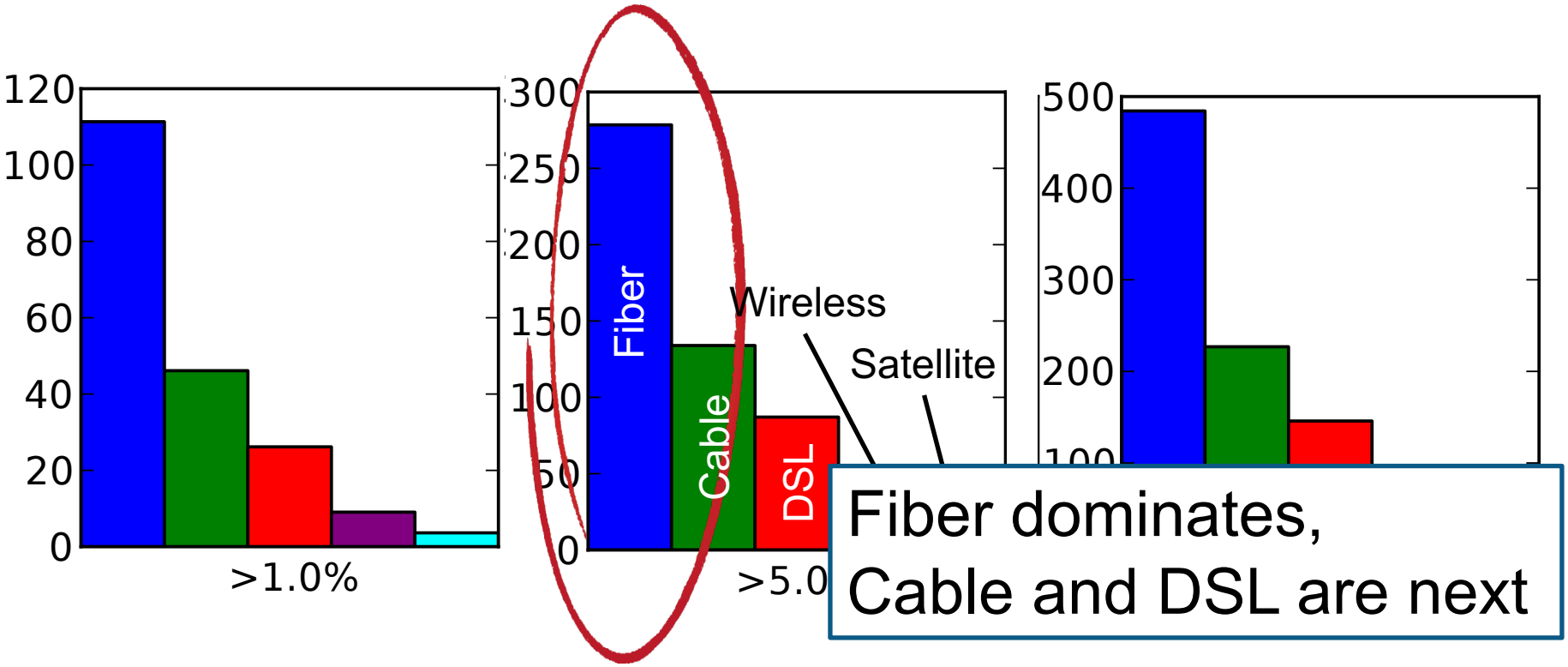
ISP and reliability

- At 1% threshold, one provider with >99% avail.



Access technology and reliability

Mean Time Between Failures in hours

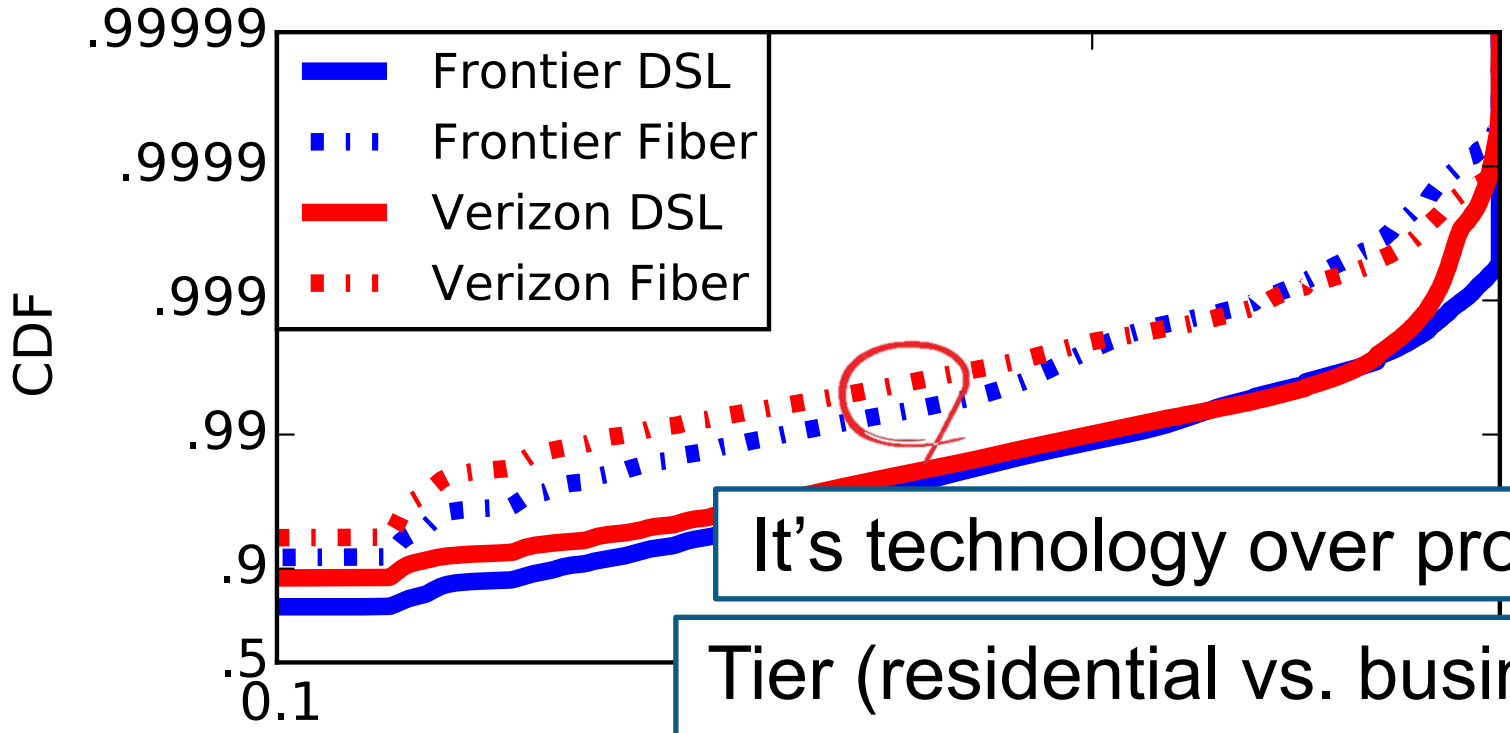


Fiber dominates,
Cable and DSL are next

Technology, service tier and reliability

- Two providers offering services over two different access technologies

CDF service availability

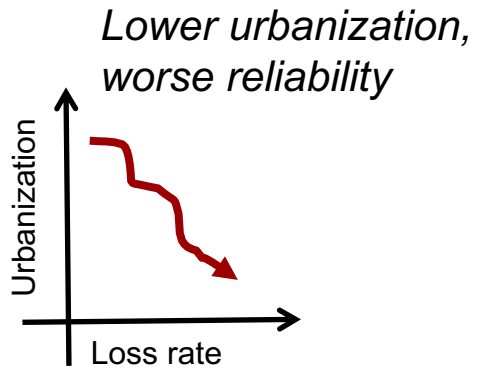


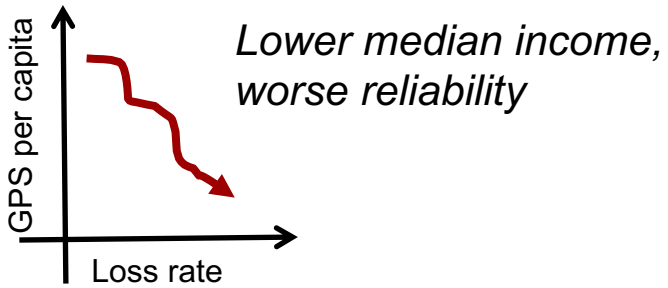
It's technology over provider

Tier (residential vs. business) has very little effect

Broader context – demographics

- Combine FCC MBA dataset with US Census Bureau, explore:
  - Urbanization level per state - urbanized areas, urban clusters and rural areas
  - State median income
- Found weak/moderate correlations
  - With urbanization levels – *r = - 0.397*
  - With median income – *r = - 0.569*





Broader context – DNS reliability

- To users, DNS and network failures are indistinguishable
  - But their reliability is not always correlated

Top 6 ISPs by connection and DNS availability

ISP	Availability @ 5%
Verizon Fiber	99.67
Cablevision	99.53
Frontier Fiber	99.47
Comcast	99.45
Charter	99.29
Bright House	99.28

Only one provider in common

ISP	DNS
Insight	99.97
Windstream	99.90
Qwest	99.90
Hughes	99.90
Frontier Fiber	99.90

Connection reliability alone is not enough

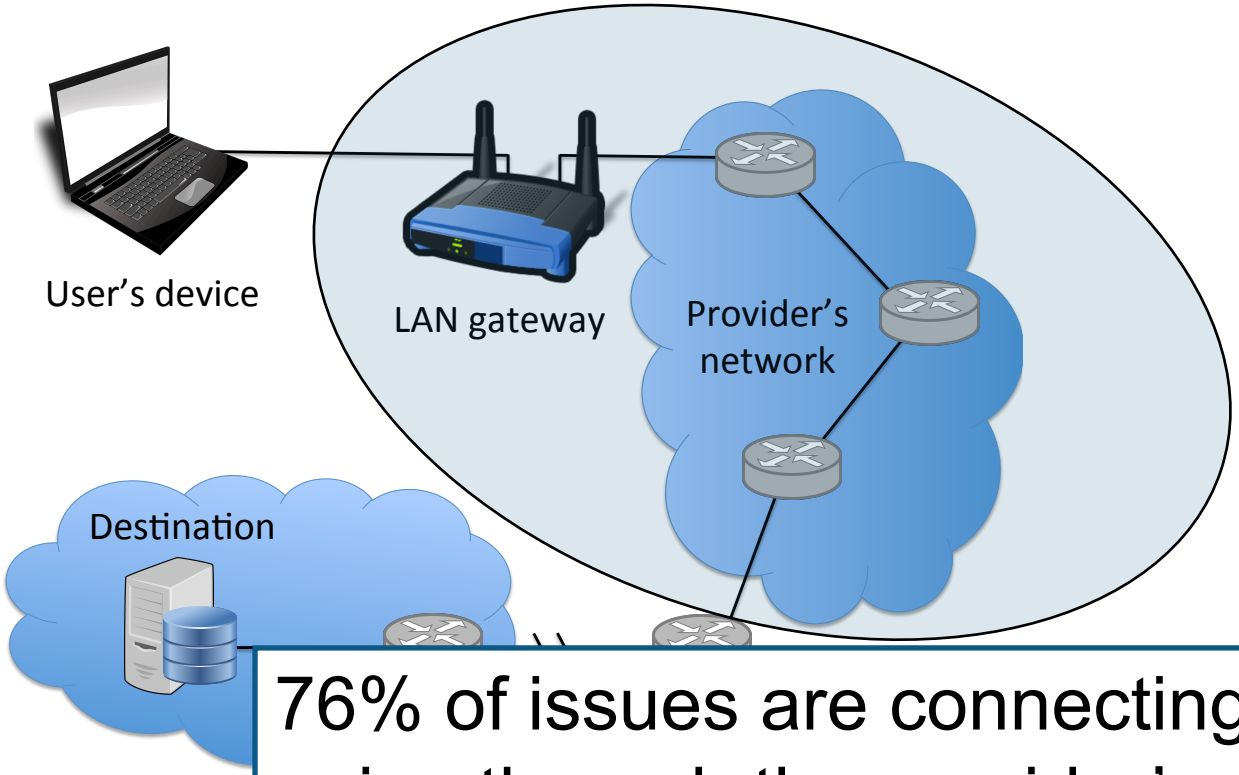
Improving reliability

- Two ways to improve reliability
  - Reduce the probability of a component failure
  - Bypass failures by adding redundancy

- Improving the technology itself is a long, expensive process
  - E.g., upgrading DSL to fiber means laying new cable

Where do reliability issues occur?

- *What is the cause of broadband reliability issues?*
  - *End host, ISP, or destination?*

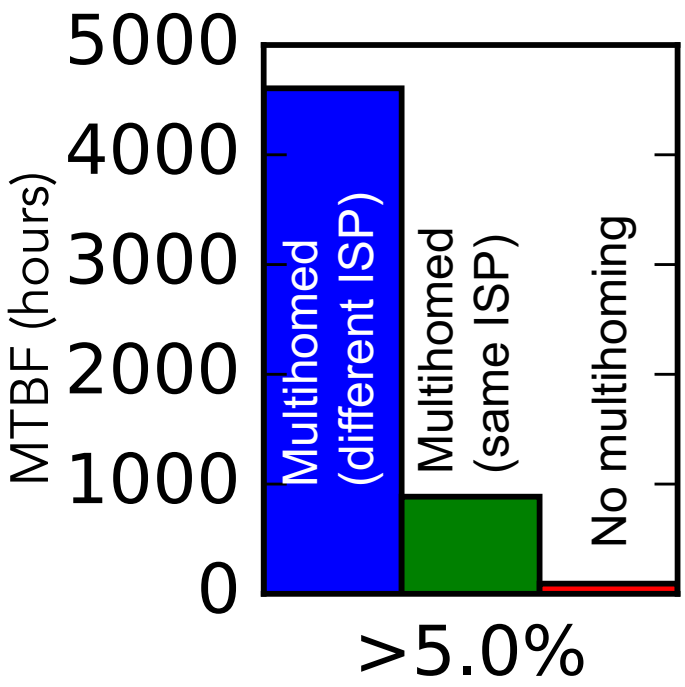


76% of issues are connecting to or going through the provider's network

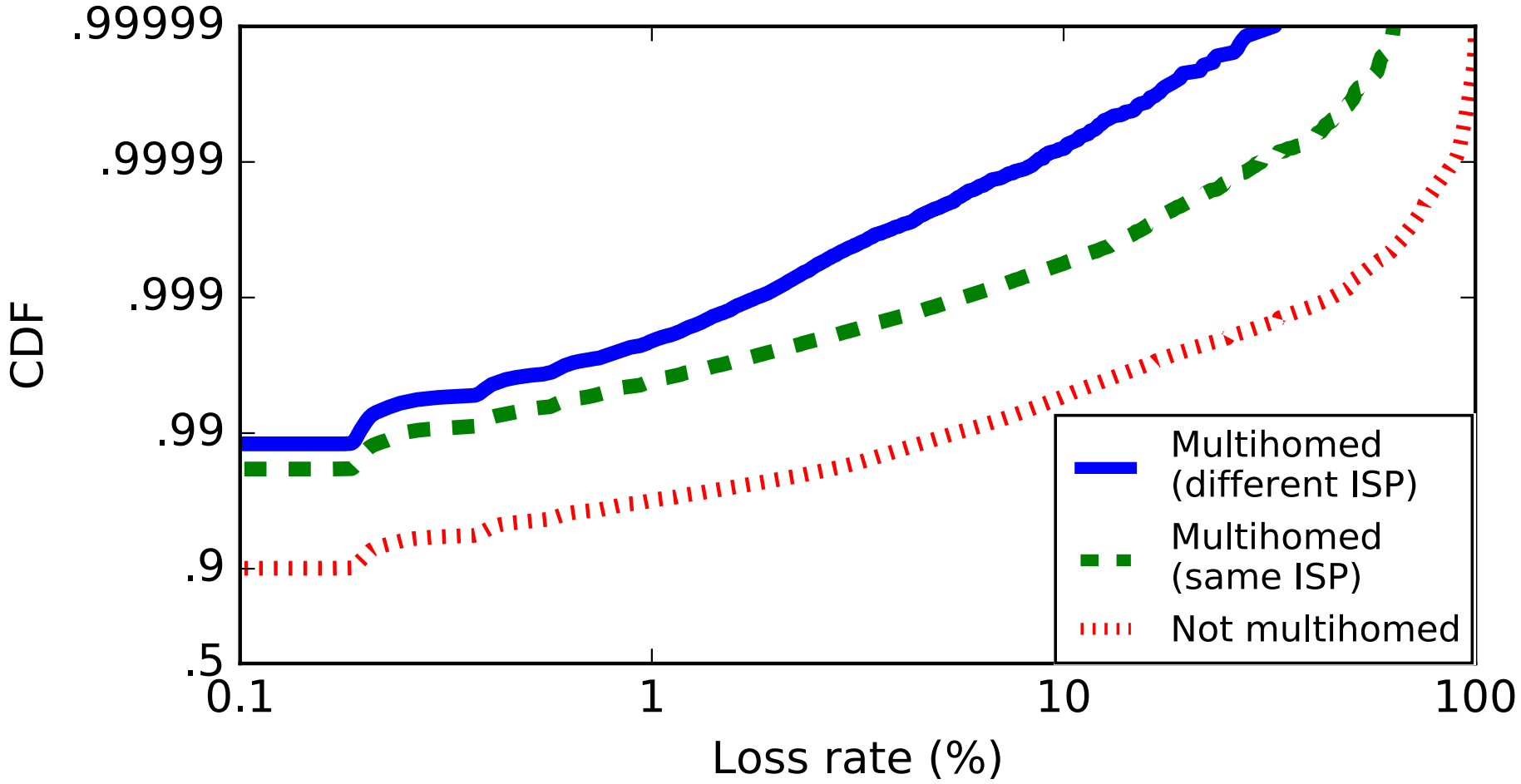
End-system multihoming

- End-system multihoming
  - Neighbors lending networks as a backup
  - ISP provided 3/4G backup connection

- To get a sense of its potential
  - Group users per census block
  - Online during the same period



End-system multihoming



By multihoming with different ISPs – four 9s availability

Summary and open issues

- An empirical demonstration of the impact of broadband reliability on user demand
- A characterization of today's broadband reliability
- And a practical proposal to improve on it

- How to capture QoE at scale, diagnose and localize its impairments?

Do users care?

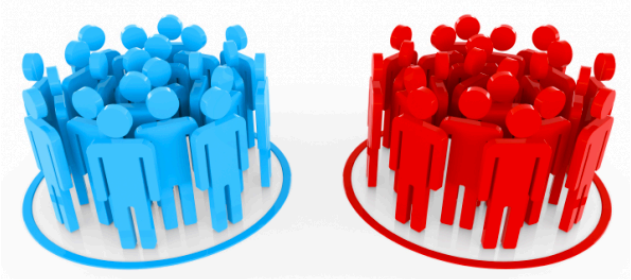
- Or, does reliability impact users' experience?
  - Standard challenges to capturing users' experience

- To evaluate this, we would like:
  - Scale – Different ISPs, different technologies, different regions, different contexts …
  - Natural settings
  - Reproducibility


Arnon Grunberg, Writing while wired NYT 2013

Reliability & QoE – Controlled experiments

- Classical controlled experiments
  - Control and treatment user groups, randomly selected
  - Treated with lower/higher reliability
  - Difference in outcome likely due to treatment



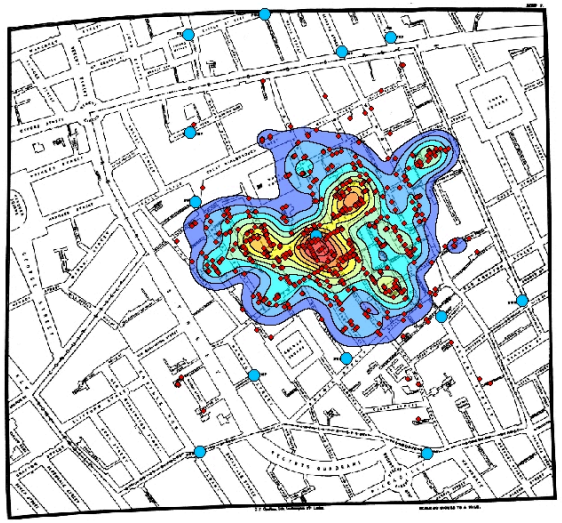
- Reproducibility, but
  - Poor scalability
  - No natural settings
  - Ethical and practical issues

- *Instead …*

Reliability & QoE – Natural experiments

- Common in epidemiology and economics

- Assignments to treatment is *as-if random*, controlling for co-founding factors
  - E.g., identifying Cholera's method of transmission



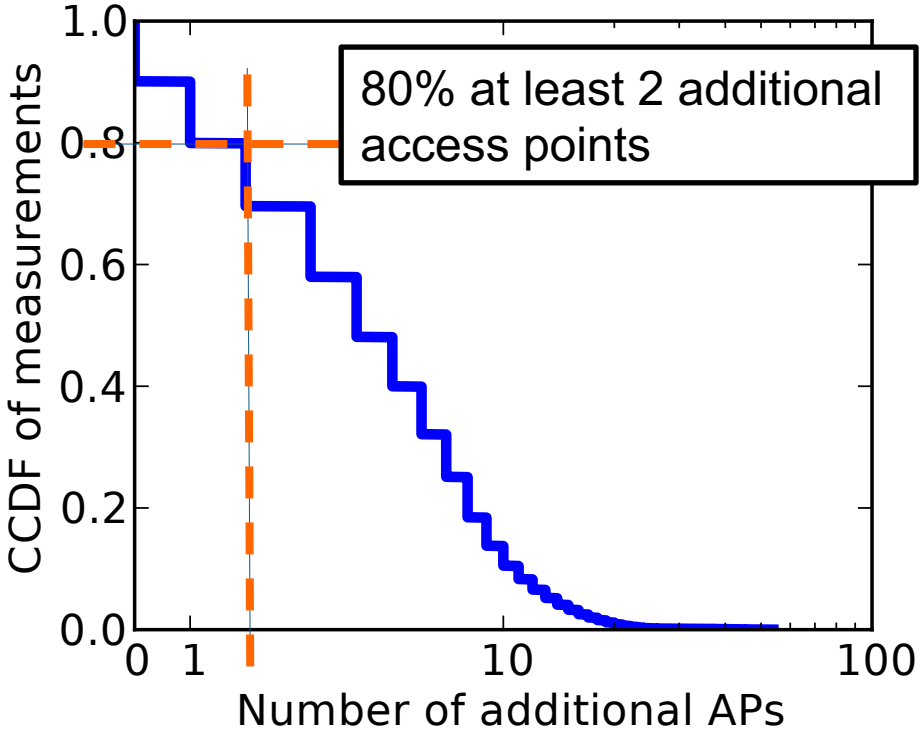
London's cholera epidemic, 1854

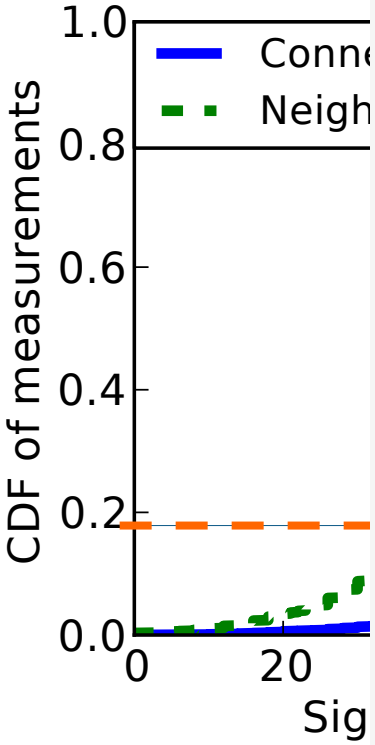
Reliability – Solution requirements

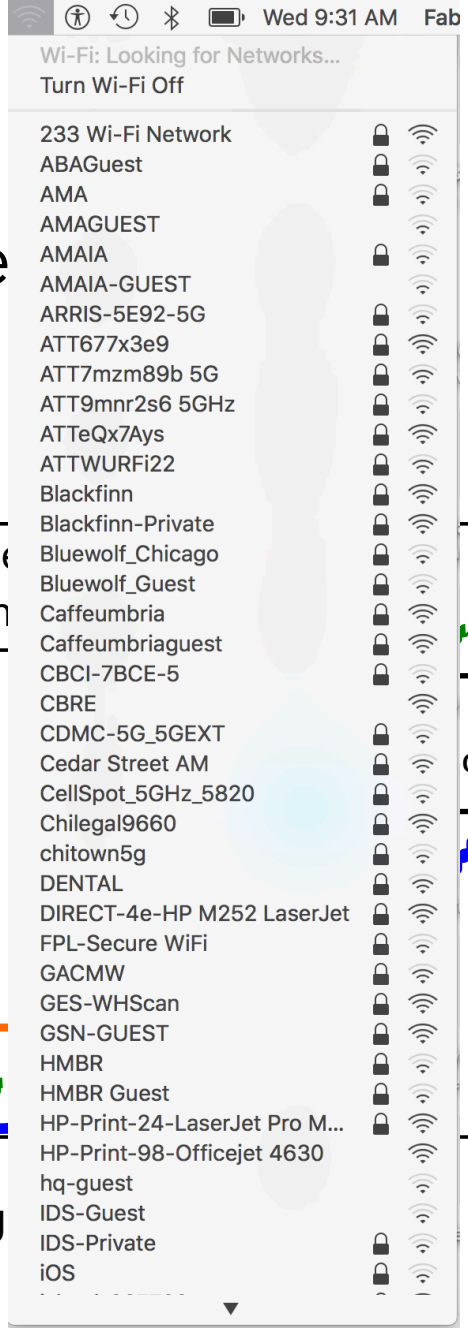
- Easy to deploy
  - Low-cost, useful despite diversity of home network configurations
- Transparent to end users
  - Step in when need, low/no overhead otherwise
- Improve resilience at the network level
  - Not just one application (e.g., no browser-based solutions)

Can we improve reliability?

- Observation: Most users in urban se
  connect to multiple WiFi networks

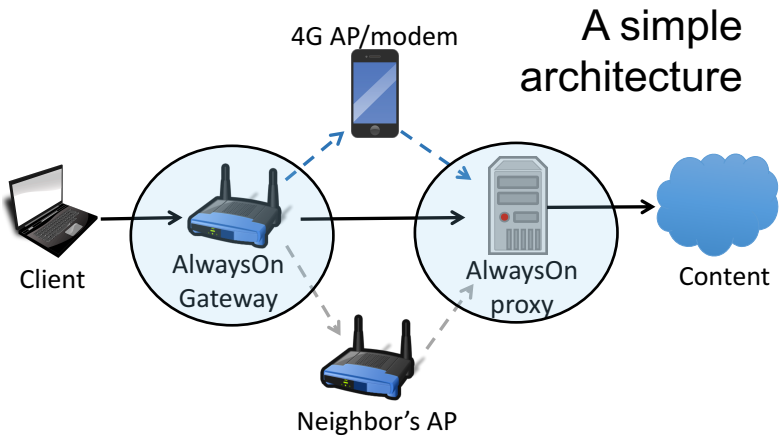






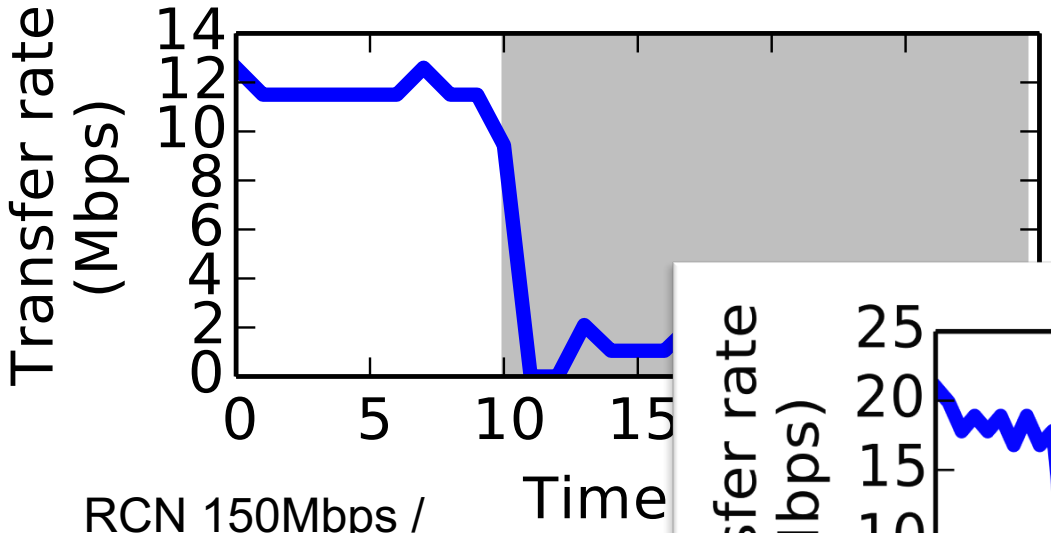
AlwaysOn – A prototype

- To components: Extended client and a server
- Multipath TCP to seamlessly switch between primary and backup
- Encrypted tunnel to the proxy and "guest" network for privacy
- Traffic policies implemented at gateway and proxy
  - e.g., inbound, outbound limits
  - Time restrictions
  - Website bans



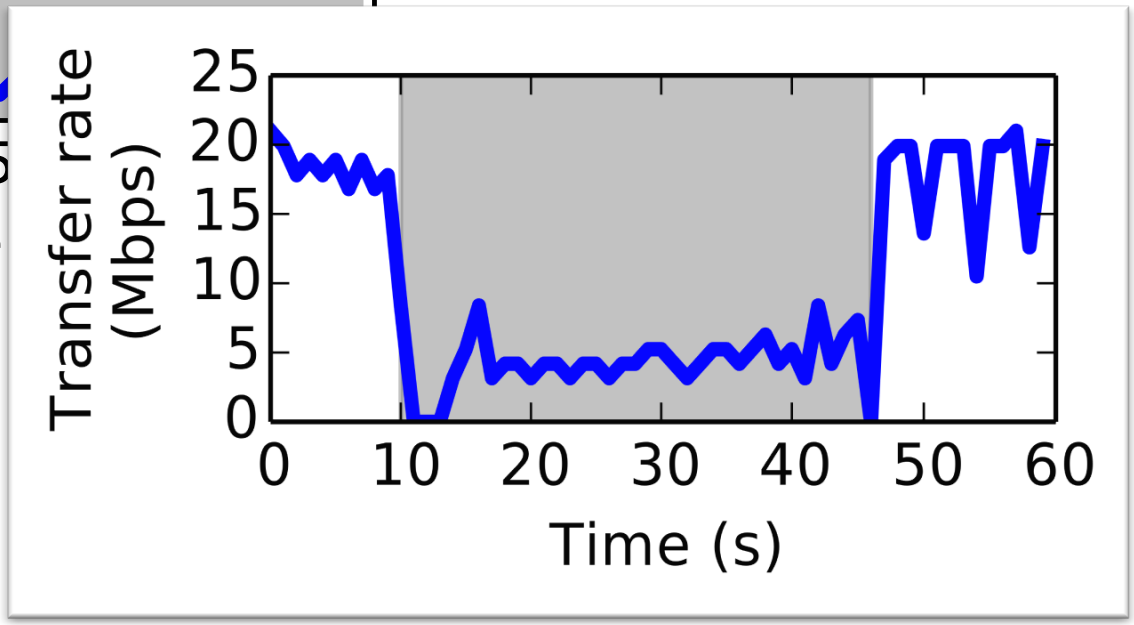
AlwaysOn's quick recovery

- Quick reaction to failure
  - Measured using `iperf` from a client, different settings and failure scenarios



RCN 150Mbps /
Verizon Wireless 4G LTE

Comcast 75Mbps /
AT&T 3Mbps



AlwaysOn's low overhead

- Downloading objects from Akamai's CDN with and without the AlwaysOn proxy
  - Distribution of download time for different objects