Measuring Network Quality for End-Users, 2021

An Internet Architecture Board virtual workshop

Day 1

Agenda

Introduction 1

14:00 Chairs' Intro

14:10 Stuart Cheshire. The Internet is a Shared Network

14:17 Jana Iyengar. The Internet Exists In Its Use

14:24 Yaakov (J) Stein. The Futility of QoS

14:31 Discussion

15:00 Keynote by Vint Cerf

15:30 Pedro Casas. 10 Years of Internet-QoE Measurements. Video, Cloud, Conferencing, Web and Apps. What do we need from the Network Side?

15:37 Lucas Pardue, Sreeni Tellakula. Lower layer performance not indicative of upper layer success

15:42 Discussion

16:00 Break

Introduction 2

16:10 Ahmed Aldabbagh. Regulatory perspective on measuring network quality for end users

16:17 Michael Welzl. A Case for Long-Term Statistics

16:24 Joachim Fabini. Objective and subjective network quality

16:31 Discussion

Metrics 1

17:00 Matt Mathis. Preliminary Longitudinal Study of Internet Responsiveness

17:07 Brandon Schlinker. Internet's performance from Facebook's edge

17:14 Discussion

18:00 End of Day 1

The Internet is a Shared Network

Stuart Cheshire

IAB Workshop

Measuring Network Quality for End-Users

14th - 16th September 2021

Stuart Cheshire, Apple

Measuring Network Quality for End-Users

Since the dawn of the web, three decades of improving Internet ***quantity***

- 9600 baud dial-up to 1Mb/s home connections to 1Gb/s home connections

Time to give more attention to improving Internet ***quality***

- Responsiveness: Consistently lower end-to-end round-trip delays
- Democratizing connectivity: An IPv6 address for every device
- Enabling innovation
  - Don’t block new protocols
  - Stay flexible — avoid ossification

A More Responsive Internet

Consistently lower end-to-end round-trip delays

Key points:

- Lower round-trip delays
- End-to-end (glass-to-glass or speaker-to-speaker)
- Consistent

The Internet is a Shared Network

draft-cheshire-internet-is-shared

All flows (or almost all) should be capacity-seeking and adaptive

All flows deserve both high throughput *and* low delay at the same time

- AQM (Active Queue Management)
- ECN (Explicit Congestion Notification)
- Preferably fine-grained, like L4S
- FQ (Flow Queueing) is not the answer
- Traffic management is not a zero-sum game

All current speed tests are wrong

The Internet is a Shared Network

draft-cheshire-internet-is-shared

The brains of the Internet are in the end systems

We need the end systems to behave responsibly

Non-Queue-Building Considered Harmful

Queue Protection needs harsh penalties

UDP is Not A Protocol

- The thing you are running over UDP is the transport protocol
- If you don’t understand that… we should all be very afraid

The Internet Exists In Its Use

Jana Iyengar

The Internet Exists In Its Use

Jana Iyengar
Fastly

Measuring Network Quality for End-Users

The workshop will discuss the following questions:

1. What are the fundamental properties of a network that contribute to good user experience?
2. What metrics quantify these properties, and how to collect such metrics in a practical way?
3. What are the best practices for interpreting those metrics, and incorporating those in a decision making process?
4. What are the best ways to communicate these properties to service providers and network operators?
5. How can these metrics be displayed to users in a meaningful way?

A program to find abstract metrics for general Internet quality

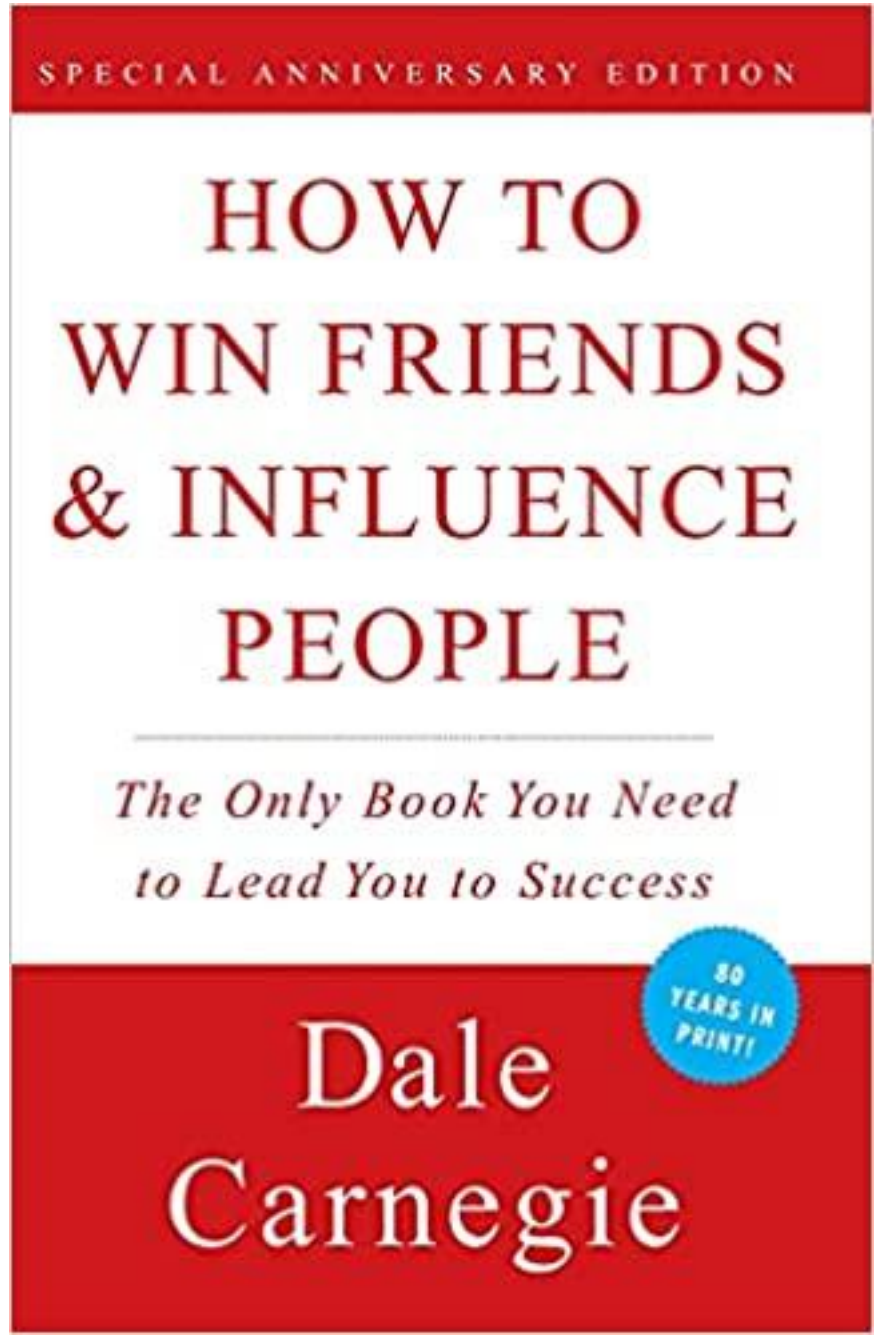
A program to find
abstract metrics
for general Internet quality
is doomed to be eventually useless.

There will not ever be
Maxwell’s equations for “Internet” quality!

Talking about metrics in the abstract is meaningless. Application metrics are what matter. Specific applications and their metrics ultimately drive adoption of network “improvements”. See ECN. See the multitude of congestion control papers over the past 4 decades. See AQM. Our industry is littered with the corpses of dead ideas that never saw deployment because they were demonstrably great in some abstract metric, but didn’t move the needle for real applications.

Talking about metrics in the abstract is meaningless. Application metrics are what matter. Specific applications and their metrics ultimately drive adoption of network “improvements”. See ECN. See the multitude of congestion control papers over the past 4 decades. See AQM. Our industry is littered with the corpses of dead ideas that never saw deployment because they were demonstrably great in some abstract metric, but didn’t move the needle for real applications. The best example is congestion control. This community went through many phases of trying to figure out how to evaluate congestion controllers using the exact network conditions that were considered reflective of “the Internet”: first, by distilling “the Internet” to enough representative topologies (remember the barbell topology? the parking lot?), then distilling it to enough representative background traffic effects (first random loss, then background traffic with elastic and inelastic sources curated to generate enough self-similarity, then with just TCP and other apps, because it was finally just easier to do the real thing anyway), and now finally we have some new thinking that seems to consider what a congestion controller does to real applications (see Ranysha Ware and Justine Sherry’s work at CMU).

Talking about metrics in the abstract is meaningless. Application metrics are what matter. Specific applications and their metrics ultimately drive adoption of network “improvements”. See ECN. See the multitude of congestion control papers over the past 4 decades. See AQM. Our industry is littered with the corpses of dead ideas that never saw deployment because they were demonstrably great in some abstract metric, but didn’t move the needle for real applications. The best example is congestion control. This community went through many phases of trying to figure out how to evaluate congestion controllers using the exact network conditions that were considered reflective of “the Internet”: first, by distilling “the Internet” to enough representative topologies (remember the barbell topology? the parking lot?), then distilling it to enough representative background traffic effects (first random loss, then background traffic with elastic and inelastic sources curated to generate enough self-similarity, then with just TCP and other apps, because it was finally just easier to do the real thing anyway), and now finally we have some new thinking that seems to consider what a congestion controller does to real applications (see Ranysha Ware and Justine Sherry’s work at CMU). So, I’ll propose short-cutting the entire thing, don’t go looking for abstract metrics, look at real applications, and think about network quality as what a user can and cannot do well on the network. And beware of abstract applications as well: video-conferencing is not skype is not FaceTime is not zoom is not webex. Video-conferencing is abstract. If you don’t think of the concrete, you are doomed! This program is doomed!


SPECIAL ANNIVERSARY EDITION
HOW TO WIN FRIENDS & INFLUENCE PEOPLE
The Only Book You Need to Lead You to Success
80 YEARS IN PRINT!
Dale Carnegie

Proposition 1:

The Internet exists in slices.

Every user sees only the slice of the Internet that their usage exercises.

Proposition 1:
The Internet exists in slices.
Every user sees only the slice of the Internet that their usage exercises.

Proposition 2:
To be relevant to a user,
the quality of the Internet must be measured and defined in terms of the various applications that the user might use over it.

Proposition 1:
The Internet exists in slices.
Every user sees only the slice of the Internet that their usage exercises.

Proposition 2:
To be relevant to a user,
the quality of the Internet must be measured and defined in terms of
the various applications that the user might use over it.

Proposition 3:
A user's perceived quality of the Internet
is inextricably tied to this evolving human artifact,
the properties of which are defined locally in time and space
by physical, social, political, and economic forces.

takeaways

To be successful,
any quality metric or measurement framework
must include (or show correlation to)
real applications on real networks.

takeaways

To be successful,
any quality metric or measurement framework
must include (or show correlation to)
real applications on real networks.

Applications change.
Any quality metric or measurement framework must
therefore be continually updated to reflect reality.

The Futility of QoS

Yaakov (J) Stein

The Futility of QoS
for Rich Communications Services

IAB virtual workshop on Measuring Network Quality for End-Users

2021

Yaakov (J) Stein
RAD
yaakovjstein@gmail.com

QoS parameters are proxies

Consumers are willing to pay for higher **QoE**
but conventional SLAs charge by **QoS**

QoS parameters are used as proxies because they

- are objective
- are easy to measure
- correlate well with subjective QoE

We know explicit estimates for subjective QoE given QoS parameters

QoE = f (application; QoS_n)

for many applications, such as

- conversational voice – MOS = f(codec, PLR, delay)
- streaming video – MOS = f(codec, PLR, PDV)
- general web browsing – ApDex = f(delay)

and many more …

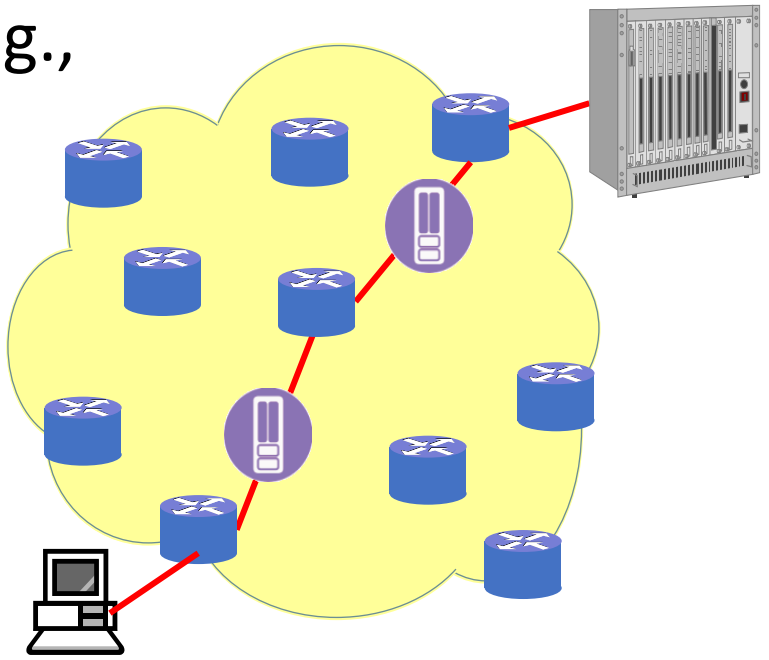
Rich communications services

A major paradigm major shift has recently occurred in networking

Traditional communications services were *pure transport services*

Today, services are much *richer,* and may perform
on-path, nontrivial, (virtual) network functions, e.g.,

- firewall
- WAN optimization (e.g., TCP proxy, compression, transcoding)
- CDN (hashing + cache server)



One can **prove** that for such rich services
there is, in general, no QoS→QoE relationship
QoE ≠ f (application; QoS_n)

Hence **QoS** and **QoS-based SLAs** are **meaningless**

The proof ?

The proof is based on
a sequence of *thought experiments* (*gedanken experiments*)

In each such thought experiment
- we pick a **QoS parameter**

and
- find a **network function**
that makes that parameter irrelevant to QoE for *some* **application**

see specific thought experiments in the paper …

NFV makes things even worse
since it enables dynamically inserting/reconfiguring functionalities

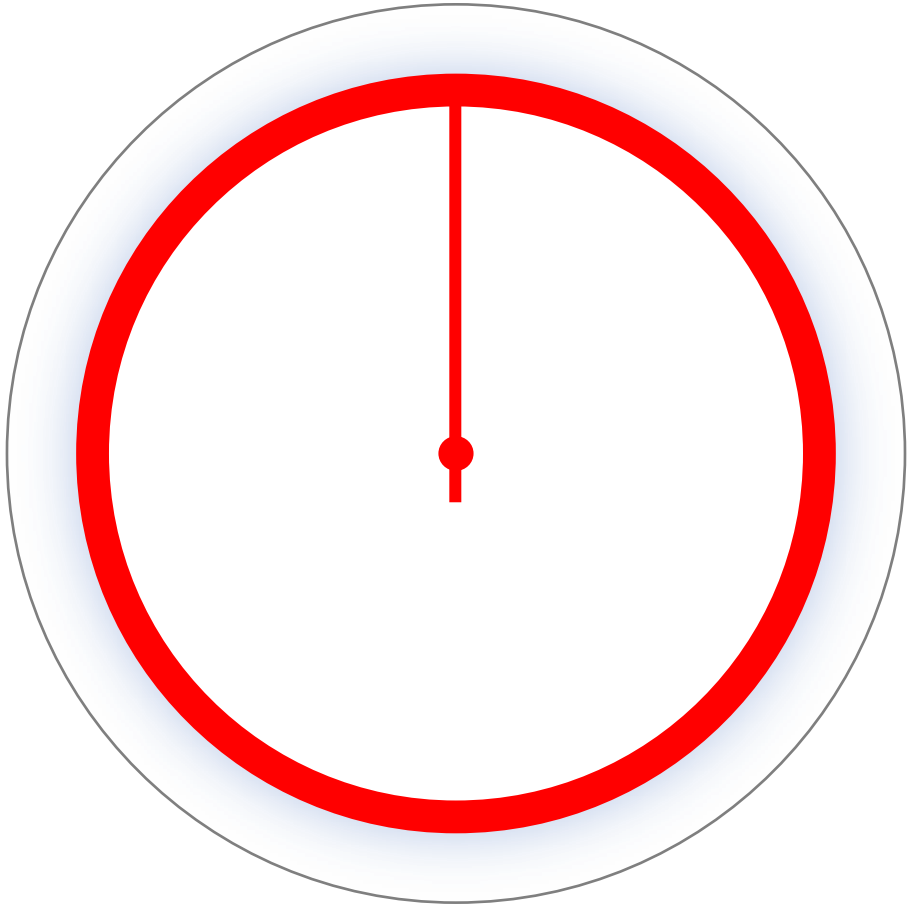
So, with NFV, we must pessimistically assume
that QoE ≠ f(QoS) problems may occur *anywhere* in the network

Discussion

To be enqueued, please write ‘+q’ in the chat

To cancel being enqueued, please write ‘-q’ in the chat

The duration of each comment is limited by 60 seconds



Keynote

Vint Cerf

10 Years of Internet-QoE Measurements. Video, Cloud, Conferencing, Web and Apps. What do we need from the Network Side?

Pedro Casas

10 Years of Internet-QoE Measurements
What do we Need from the Network Side?

Dr. Pedro CASAS

AIT Austrian Institute of Technology @Vienna

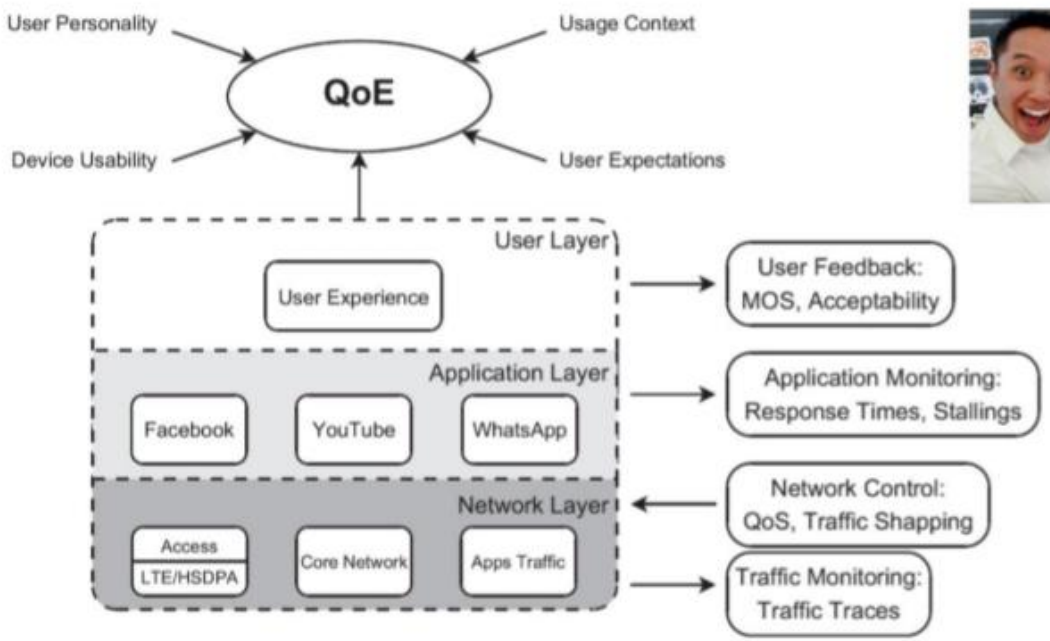
Data Science & AI

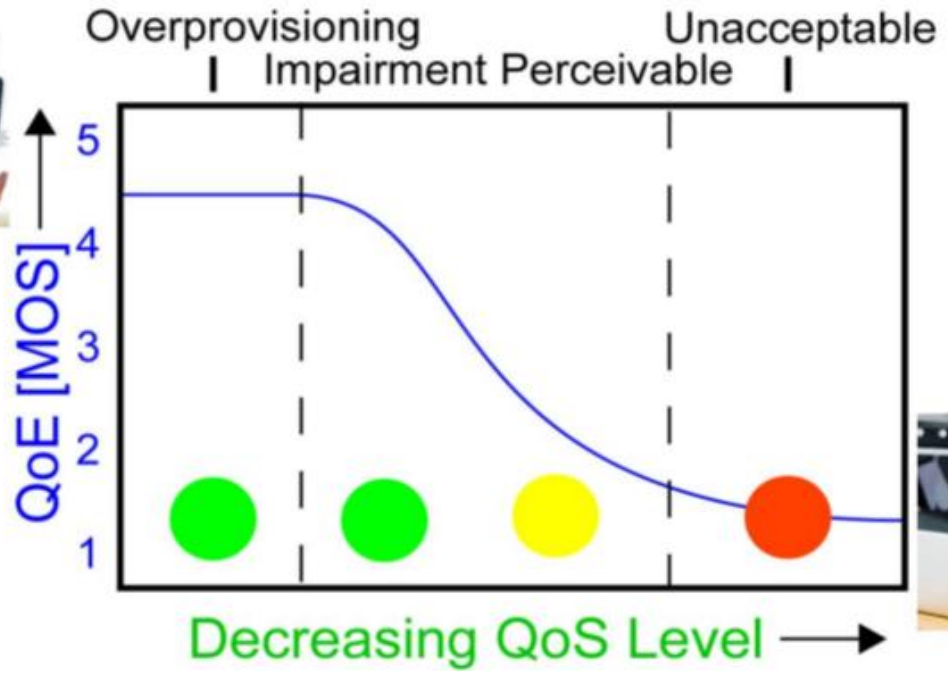
IAB Workshop on Measuring Network Quality for End-Users

September 14, 2021

Internet QoE: User eXperience in Real World Networks & Services at Scale

- Internet **QoE** – enables a more ***user-centric understanding*** and ***operation*** of the quality of data communication ***networks and services***
- ***QoE-driven/QoE-aware algorithms*** can help the ***network function*** in a ***more efficient and effective way*** (from **Traffic Engineering** to **on-line monitoring** & diagnosis)
- ***QoE*** modeling, measurement and assessment is ***complex*** and ***time-consuming***



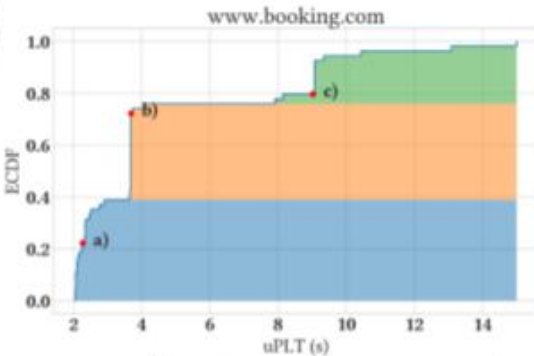


10 Years of Internet-QoE Measurements

- ***Poor QoE*** significantly ***reduces user engagement***:
  - **Video-streaming** (re-buffering, video quality, quality switching)
  - **Teleconferencing** (video quality, audio/video sync, interactivity)
  - **Web services** (page load times, responsiveness)

- ***Poor QoE*** impacts usability and ***reduces productivity***:
  - **Cloud services** (responsiveness, interactivity)

- ***Internet-QoE*** is significantly ***impacted*** by the ***end-device***:
  - ***Capabilities*** (smartphone vs desktop PC) and ***expectations***



- ***Perception@scale*** is ***partially subjective***
  - **Multi-modality in u-PLT** (perceived page load time)
  - *We have **fundamental laws** on **QoE** perception…*
  - *…but these **do not capture the whole picture***

What do we Need from the Network?

- Video Streaming
- QoS – ***downlink bandwidth***
- User – ***re-buffers, video quality***

- Web browsing
- QoS – ***latency***
- User – ***AFT, TTI, FID***

- Web/Cloud services
- QoS – ***downlink/uplink bandwidth, latency***
- User – ***responsiveness***

- Depends on ***what the user perceives*** from the ***specific service***
- ***Fast connections***, ***low-latency***, ***responsive*** communication protocols, ***time-stable performance***
- ***Performance stability*** is highly ***relevant*** for ***user experience*** and ***QoE*** (bandwidth/latency fluctuations, outages)
- Quick answer: we ***need all of these*** to realize ***seamless user experience***
- Holistic-view: ***QoE is session*** (***temporal***-integration) and ***multi-service*** dependent
- Data-driven (***AI/ML***) approaches to ***capture session QoE***

Lower layer performance not indicative of upper layer success

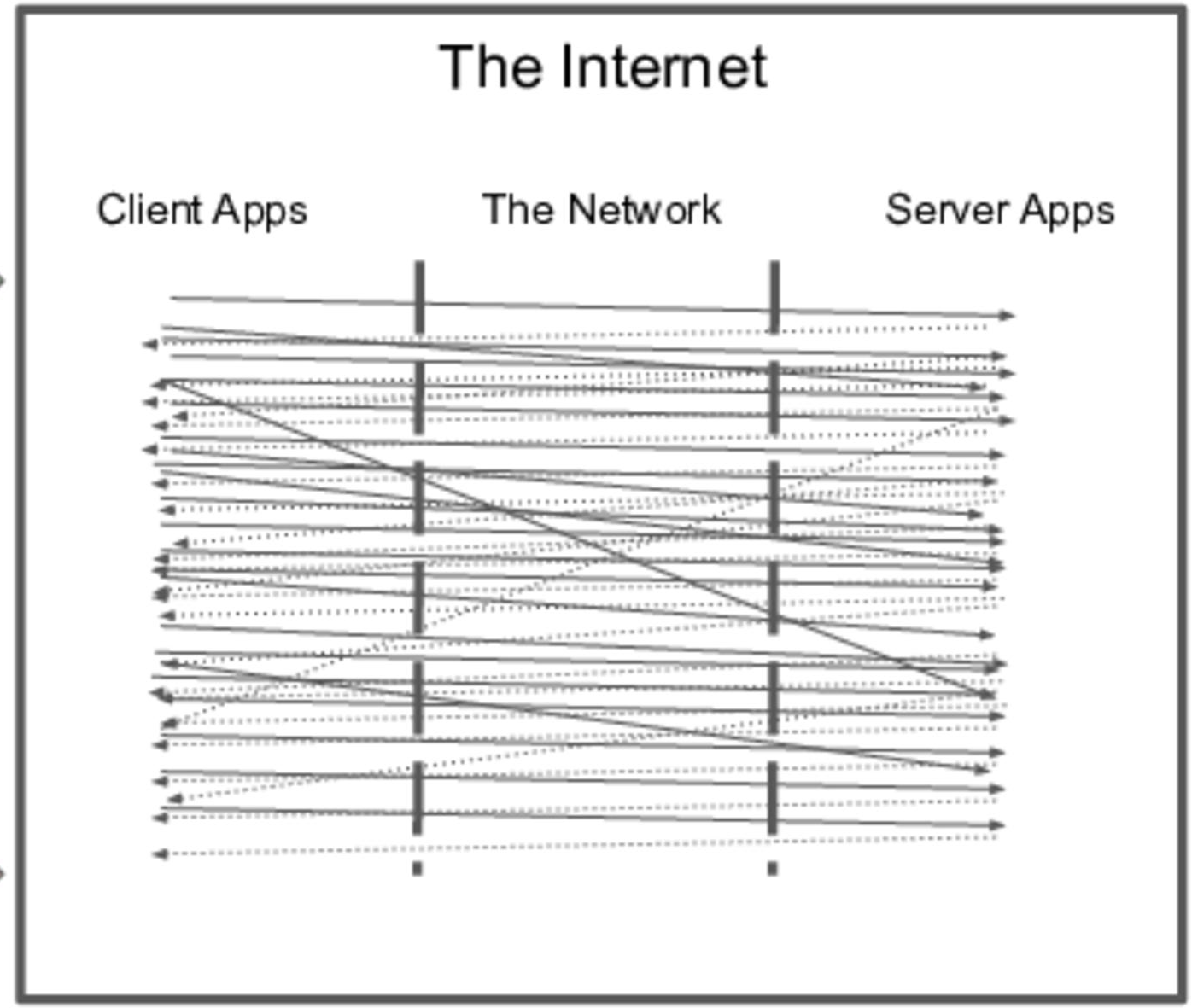
Lucas Pardue, Sreeni Tellakula

Lower-layer performance is not indicative of upper-layer success

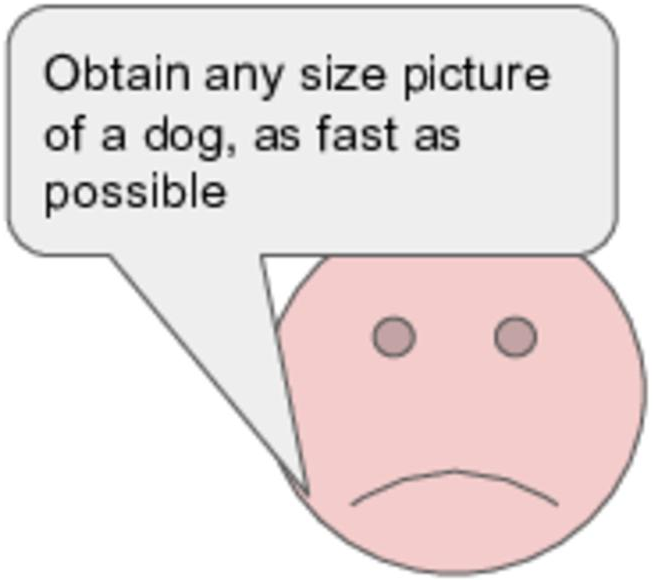
Presentation to IAB Measuring Network Quality for End-Users Workshop 2021

Lucas Pardue

Retrieve a medical scan for diagnosis in a timely manner



Obtain any size picture of a dog, as fast as possible



The Internet is slow

My client device is old

My ISP is slow

The server is overloaded

Corporate security software

Bufferbloat

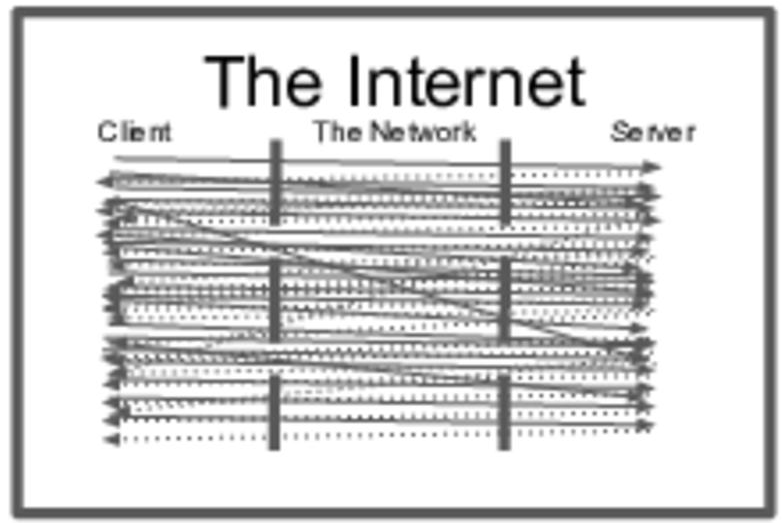
My mum was microwaving a potato

BGP routing problem

Tuesday

Newsflash

Don't care why!



The reason the Internet was slow was because of everyone else on "the Internet".

The reason the Internet was slow was because my mum was microwaving a potato.

Measuring the world is hard, there is no universal frame of reference, there are no grand unified measures of success. > 99% of the world population don't know what LUL means.*

Does knowing there is *any kind of problem* help identify where it might be?

Does knowing the theoretical parameters vs operational parameters help?

Everybody wants data but who is empowered to make change?

Changes are expensive, hard to test in advance of expenditure, hard to prove in the long term.

There's an illusion of choice.

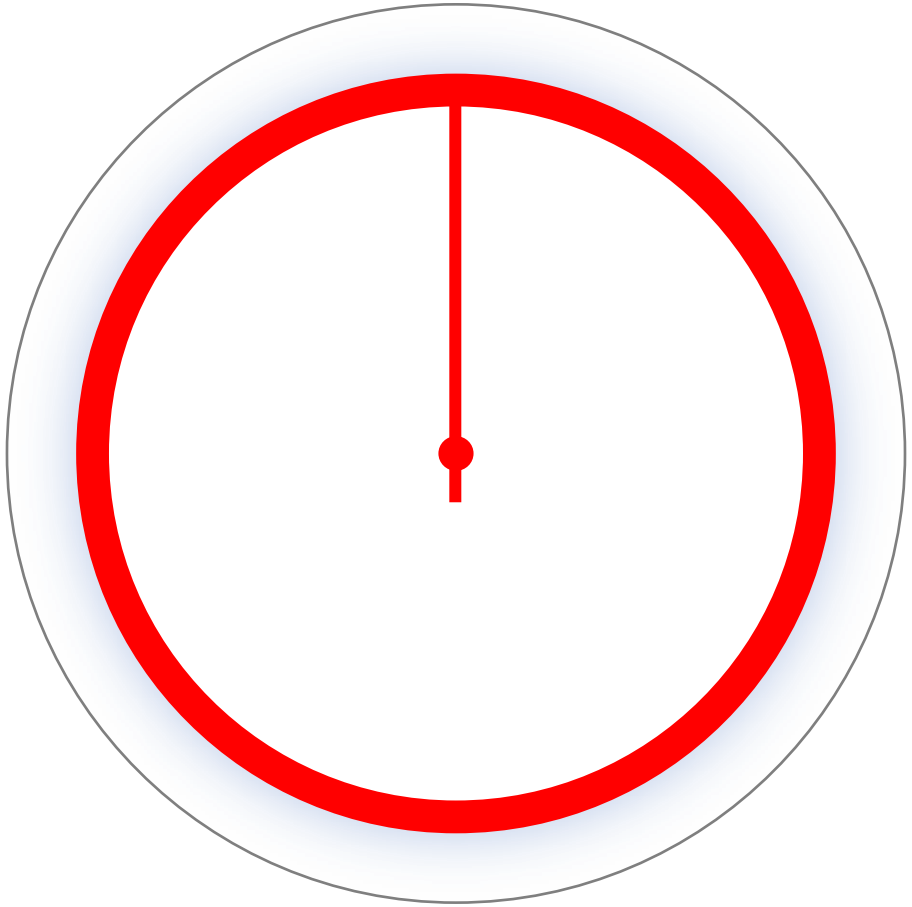
* 50% of statistics are made up on the spot

Discussion

To be enqueued, please write ‘+q’ in the chat

To cancel being enqueued, please write ‘-q’ in the chat

The duration of each comment is limited by 60 seconds



Break

Introduction 2

MAKING COMMUNICATIONS WORK FOR **EVERYONE**

Regulatory perspective on measuring network quality for end-users

Ahmed Aldabbagh, Ofcom (UK)
14-19 Sept 2021

Disclaimer: The analyses, opinions and findings in this presentation represent the views of the author, should not be interpreted as an official position of Ofcom and are not intended to be an official statement of Ofcom's policy.

Operator requirements

- Understand the performance of their networks.
- Understand the performance of the equipment.
- The impact that such performance has on the experience of their customers.
- Generally within the context of network management system.
  - Troubleshooting and
  - Pre-fault diagnostics

Regulatory perspective

- Promoting the ability of consumers to access and distribute information or run applications and services of their choice.
- Universal service directive aims at promoting a satisfying quality-of-service and ensuring better information for consumers on its actual quality of service which will be received.

Consumers access the Internet

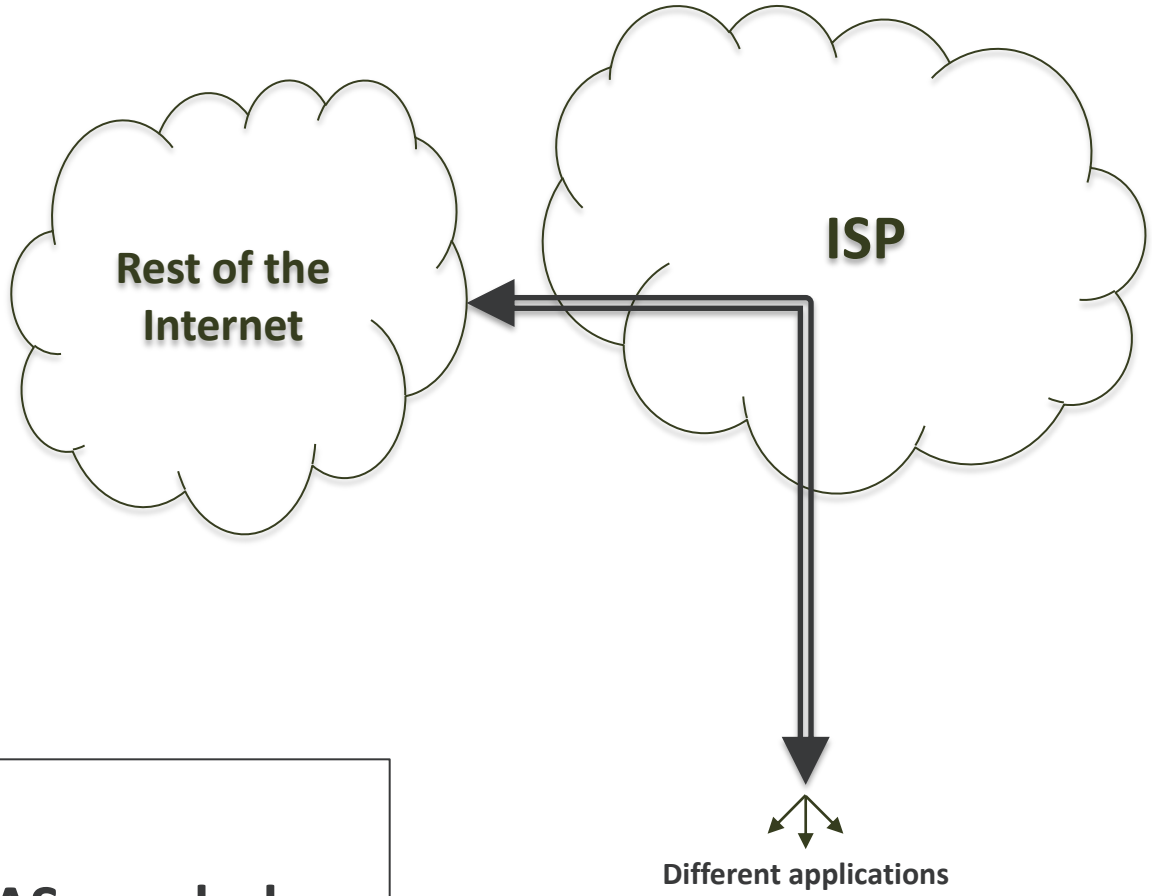
Transparency: measurement for consumers

- What am I getting?
- etc

Measurements for policy makers

- Evolution of national infrastructure
- Etc

- Innovation



Risk

- Degradation of IAS as whole
- App-specific degradation

Internet access quality monitoring system requirements and features

- **Accuracy**
- **Comparability**
  - Different IAS providers and their offerings
  - Different countries when possible
- **Trustworthiness**
  - Independence (from operators)
  - Accountability and legal value
- **Privacy**
- **Security**
- **Openness**
- **Future proof-ness**

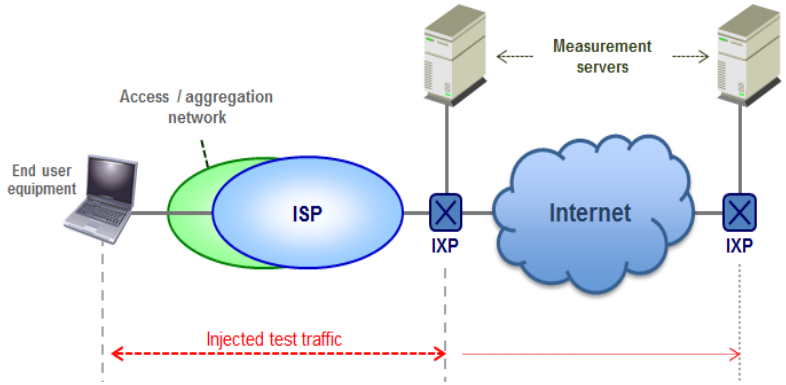
- **Active measurements using test traffic**
- **Passive measurements using ordinary user traffic**

- **User initiated tests**
- **Background tests**

- **Hardware probes**
- **Software probes**

- **Mobile**
- **Fixed**
- **Fixed-wireless**

- **Cost**



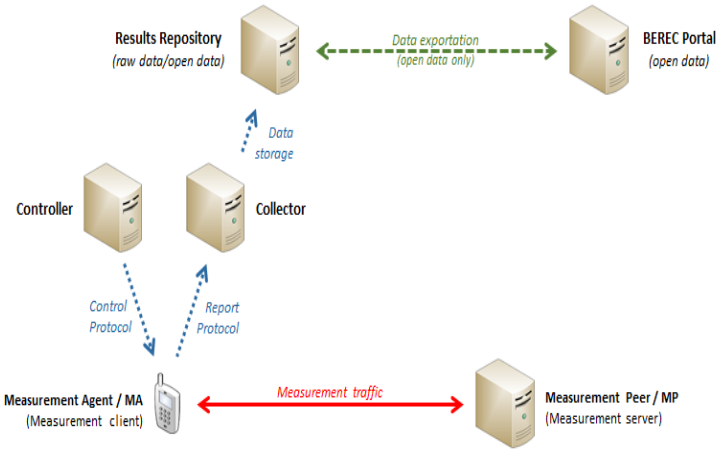
Measurement systems

Considered

- Large-Scale Measurement of Broadband Performance (IETF/LMAP)
- IP Performance Metrics (IETF/ IPPM)
- RIPE Atlas
- M-Lab
- etc

BEREC

- Federated system
- Drew a great deal from LMAP



Crowdsource

- Rising in importance as it is beginning to open new insights
- Lack of standards
- Privacy issues
- Collection is not future proof

Approach: metrics

QoS

- Speed (upload and download)
- Delay and its variations
- Packet loss
- DNS look up time
- etc

- How to measure in view of earlier requirements ?

QoE

- Becoming more and more important
- But is it difficult?
- Lack of common theoretical framework?
- Lack of standardisation

Final remark: where should the future be?

Federation of ‘Interoperable’ systems

- Standardisation in measurements methods
- Common software implementations
- Standardisation in common data format and APIs

QoS to QoE??

- Difficult to solve?
- What about Delta Q as the basis for future?
- Need for standardisation

A Case for Long-Term Statistics

Michael Welzl

A Case for Long-Term Statistics (1)

- Quality metrics are inputs to improve the network
  - How can “normal” users do that?
    Change ISP; change equipment; upgrade firmware; tell people in the household “no TikTok while daddy talks”

- Sitting in front of a console and obtaining “aha, quality is reduced now” does not help much
  - This is a case against such interactive use

Michael Welzl
IAB Workshop on Measuring Network Quality for End-Users, 2021

A Case for Long-Term Statistics (2)

Useful metrics:

- Is my connection the bottleneck? (how often in general / was it the bottleneck 5 minutes ago, during our telco?)
- Do applications influence each other?
  (Does “stay away from TikTok while I work” even help? Is it bad to plan conversations in my household at the same time?)
  - Or: did they, during the last 10 minutes?
    At which time of day do applications influence each other, in general?

A Case for Long-Term Statistics (3)

- How to obtain these metrics: passive listening
  - and associate traffic with applications

- How often does packet loss / RTT growth happen?
  - Did it, during the last X minutes?

- Which application(s) were involved? Did they influence each other (share a bottleneck)?
  - Passive shared bottleneck detection: RFC 8382
    This uses OWD; various variants exist (research)

Objective and subjective network quality

Joachim Fabini

Network Quality: Definition and Context?

- **User's expectation on network quality**?
  - Internet (or service?) connectivity **available**?
  - Experienced quality **underperforms**, **matches** or **exceeds** (user) **expectations**?
  - **Timeframe:** evaluate **past** network quality, **current** quality, or **anticipate quality**?
    - Repeatability, continuity: revised/questioned by RFC7312 [1] - "Identical" parameters?
  - **Easy to evaluate and compare**: map results to a value or scale?
  - **Automated post-processing**: request "increase network quality to value X"?
- **Theory:** map user perception metrics to technical ones (IPPM)
  - Connectivity, one-way delay, one-way loss, transfer capacity (**existing RFCs**)
  - Focus on bottleneck link(s)
  - "Fair": **no bias** for specific operators and/or technologies deployed in the networks
  - **Simplicity**: network quality metric comprehensible and usable by any Internet user.
- **Challenges?**

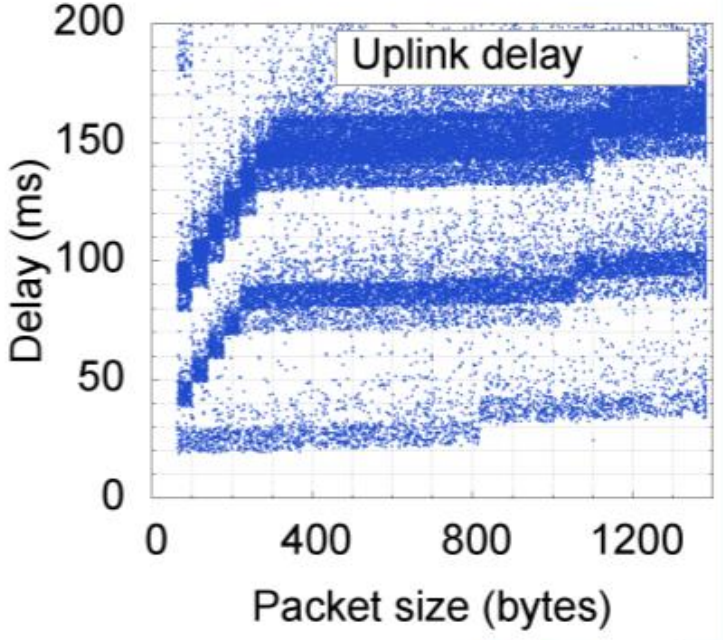
[1] J. Fabini and A. Morton, "Advanced Stream and Sampling Framework for IPPM," RFC 7312, 2014.

Network Quality: Technology and Bias

- **Users experience service quality and NOT network quality!**
  - How to differentiate **network** vs. service/**application** quality?
  - **Services define the network traffic pattern**
    - User's configuration options on how services use the network: limited
  - **User behavior, service use and traffic patterns differ**
    - Depending on societies, cultures, geographies, age groups, tariffs, etc.

Source: http://commons.wikimedia.org/wiki/File%3ALautsprecherkabel_Makro_nah.jpg

- **Network is no stateless copper wire** ([2], [3])
  - Networks react to user traffic statefully (at various layers)
  - Schedulers allocate and deallocate capacities
  - **HUGE number of uncertainty factors**
    - Can include data stream: **packet size, content and timing!**
  - Capture? Privacy concerns - welcome to the GDPR…



[2]: J.Fabini: "Access network measurements", Lightning talk, IRTF Workshop: Research and Applications of Internet Measurements (RAIM) 2015, Yokohama, Japan. https://irtf.org/raim-2015-slides/fman/fabini.pdf
[3]: J.Fabini: "Delay measurement tools: RDM", Lightning talk, IRTF Workshop: Research and Applications of Internet Measurements (RAIM) 2015, Yokohama, Japan. https://irtf.org/raim-2015-slides/mpt/fabini.pdf

Quality Measurement Framework

- Define **requirements for network quality rating**
  - **Audience: end-user, operator, automation? SLA verification? Actionable item?**
  - **Past, presence or future?** Requirements wrt **accuracy?**
- Can (and should) we **ignore services** when evaluating network quality?
  - (Service + user input + OS + network response + X) = network traffic pattern(s)
  - Networks react to these patterns: Users access networks only through services!
- **One option: services** offer an **abstract network profile description**
  - NPD includes protocols, network traffic patterns, requirements. User input?
  - Networks instrumented to monitor existing traffic. Anticipate response to NPD?
  - LMAP + service profile? Simplicity and use by non-technical users? Complexity?
- Mapping network quality to one single value? Orthogonal metrics…
  - Visual (multi-dimensional, graphical) representation?
- **Contact: mailto:Joachim.Fabini@tuwien.ac.at**

Network Quality: Definition and Context?

- **User's expectation on network quality**?
  - Internet (or service?) connectivity **available**?
  - Experienced quality **underperforms**, **matches** or **exceeds** (user) **expectations**?
  - **Timeframe:** evaluate **past** network quality, **current** quality, or **anticipate quality**?
    - Repeatability, continuity: revised/questioned by RFC7312 [1] - "Identical" parameters?
  - **Easy to evaluate and compare**: map results to a value or scale?
  - **Automated post-processing**: request "increase network quality to value X"?
- **Theory:** map user perception metrics to technical ones (IPPM)
  - Connectivity, one-way delay, one-way loss, transfer capacity (**existing RFCs**)
  - Focus on bottleneck link(s)
  - "Fair": **no bias** for specific operators and/or technologies deployed in the networks
  - **Simplicity**: network quality metric comprehensible and usable by any Internet user.
- **Challenges?**

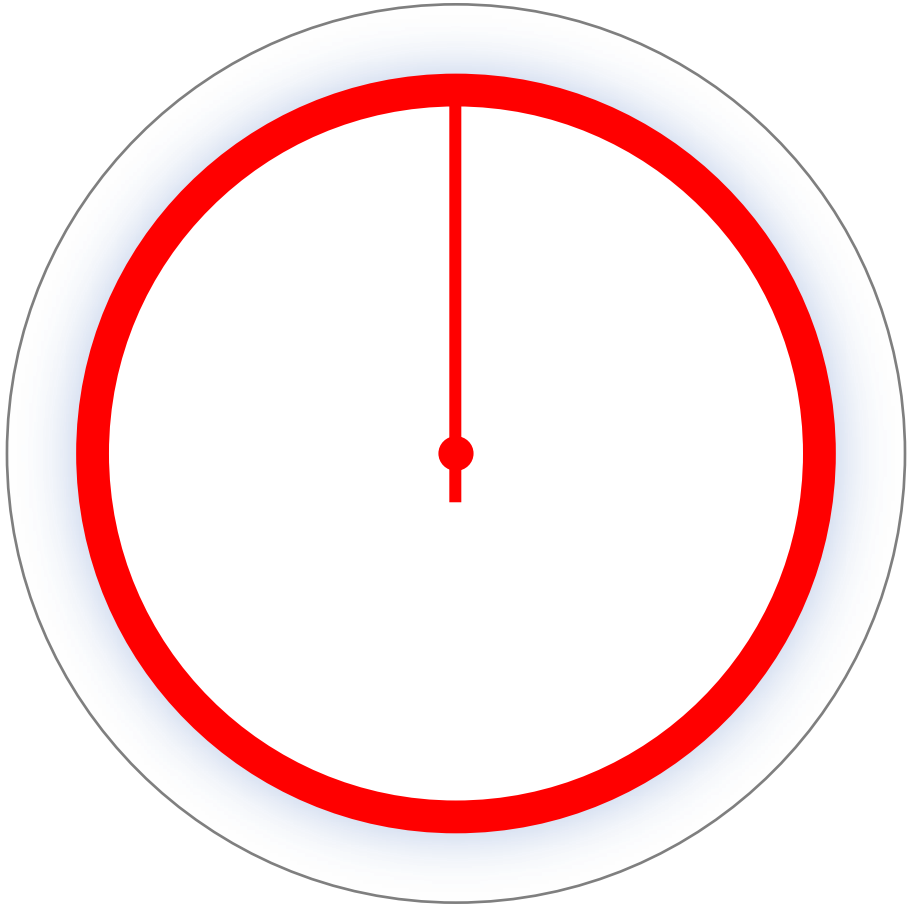
[1] J. Fabini and A. Morton, "Advanced Stream and Sampling Framework for IPPM," RFC 7312, 2014.

Discussion

To be enqueued, please write ‘+q’ in the chat

To cancel being enqueued, please write ‘-q’ in the chat

The duration of each comment is limited by 60 seconds



Metrics 1

Preliminary Longitudinal Study of Internet Responsiveness

Matt Mathis

Preliminary Longitudinal Study of Internet Responsiveness

Matt Mathis
September 2021
Measurement Lab & Google

Exploring an Internet Responsiveness Metric

- Patterned after Apple's "responsiveness" developer tool
- Rate of delivering network rounds
  - Network rounds per second or per minute (rpm)
  - TCP level rounds: data->ACK->data
    - (The Apple tool measures application level rounds with application pings)
    - Both are important
  - In principle transport Responsiveness can be estimated from:
    - Header traces and the reconstructed critical path (canonic definition?)
    - Polled TCP state (SRTT)
      - Via Web100 or TCP_info
- For this preliminary study we use the **Final** value of the SRTT
  - Measurement lab has 12+ years of archived SRTT data

Future Goals

- Calibrated responsiveness computed from polled SRTT

  Accumulated rounds = SUM(poll_interval / SRTT)
  - Can compute this metric retroactively on all archived TCP snapshots
  - Can compute this metric in real time during a test
    - And include it in the completion report
  - Cross calibrate various responsiveness like metrics
- Validate that responsiveness predicts page load times for most web pages
  - Most web data objects run out of data while in slowstart
    - Transport responsiveness would predict completion time
  - Object start times depend on application responsiveness
    - Perhaps plus DNS (responsiveness) and browser cpu times

Figure 1: Mean Responsiveness circa NYC

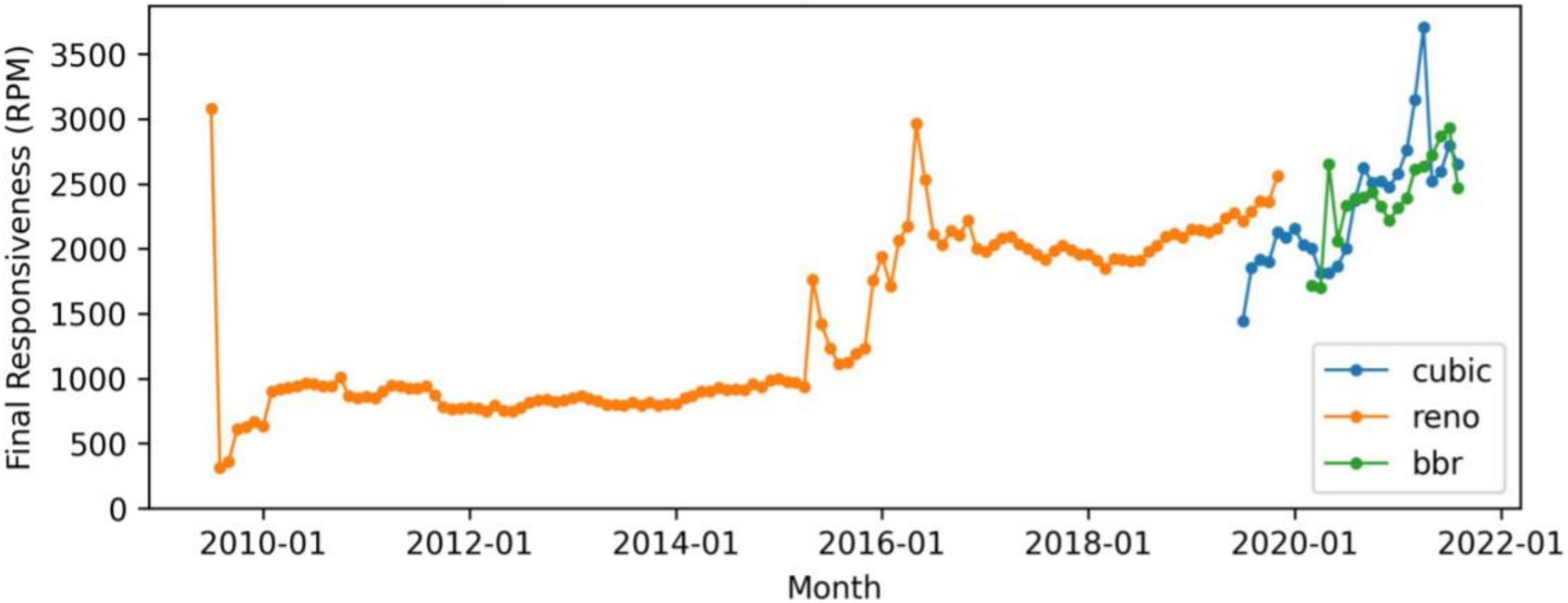
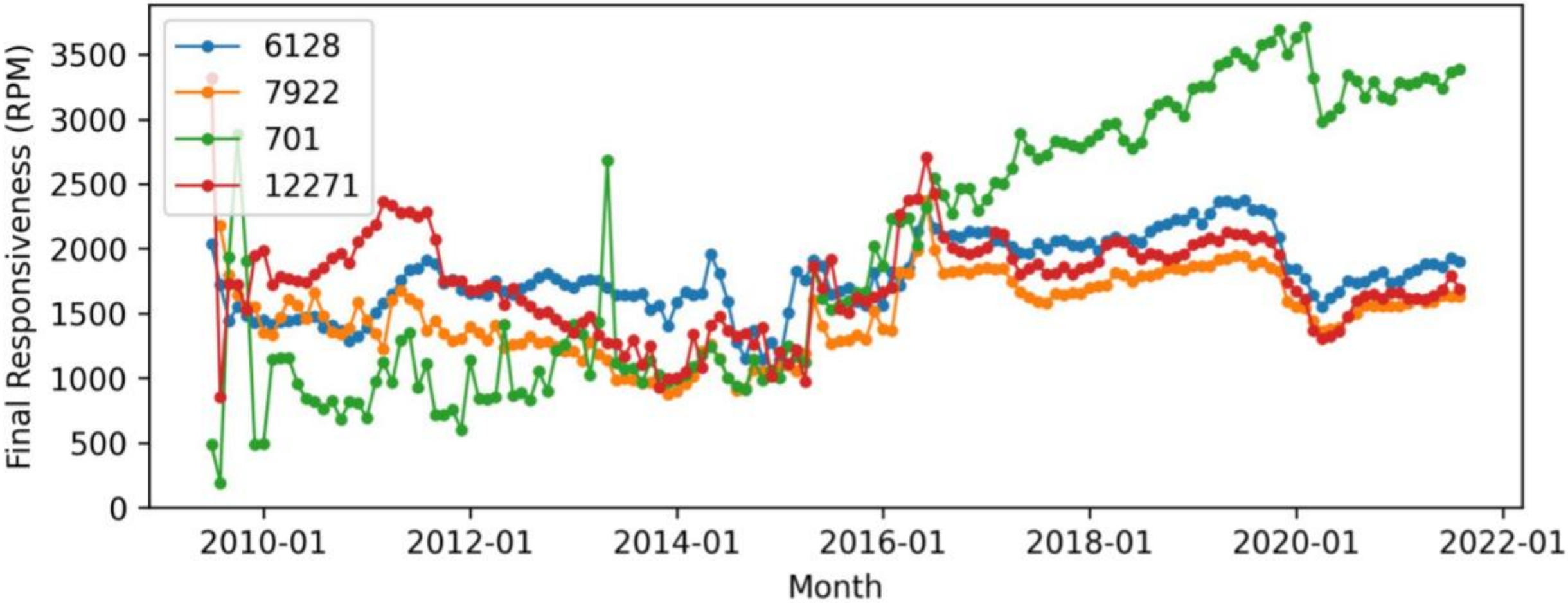


Figure 4: Responsiveness of the 4 ISPs with the largest test volumes circa LGA



Future Responsiveness Work

- Calibrate different methods for measuring responsiveness against each other
- Calibrate (various) responsiveness metrics against web page load times
- Reflect on the relationship between responsiveness and jitter (e.g real time)
- Look at some additional derived metrics
  - Harmonic mean of the SRTT
  - Network power = throughput * responsiveness
- How to treat "infinite lossless buffers" in aggregate stats
  - Observed responsiveness depends only on TCP and running time

The real problem

- Bufferbloat is too complicated for users to complain about
- Lag is too subtitle for users to complain accurately
- We need a consumer grade tool that makes sense to people
  - A future version of the Apple tool is likely to do so
  - A future version on MLab NDT is likely to do so
  - So the users can get the attention of the marketing people
- And we need network engineer grade tools that show the same problem
  - To make the reports actionable

Backup slides

- The following plots take under half minute each, using colab and BigQuery

Figure 2: Mean US Responsiveness

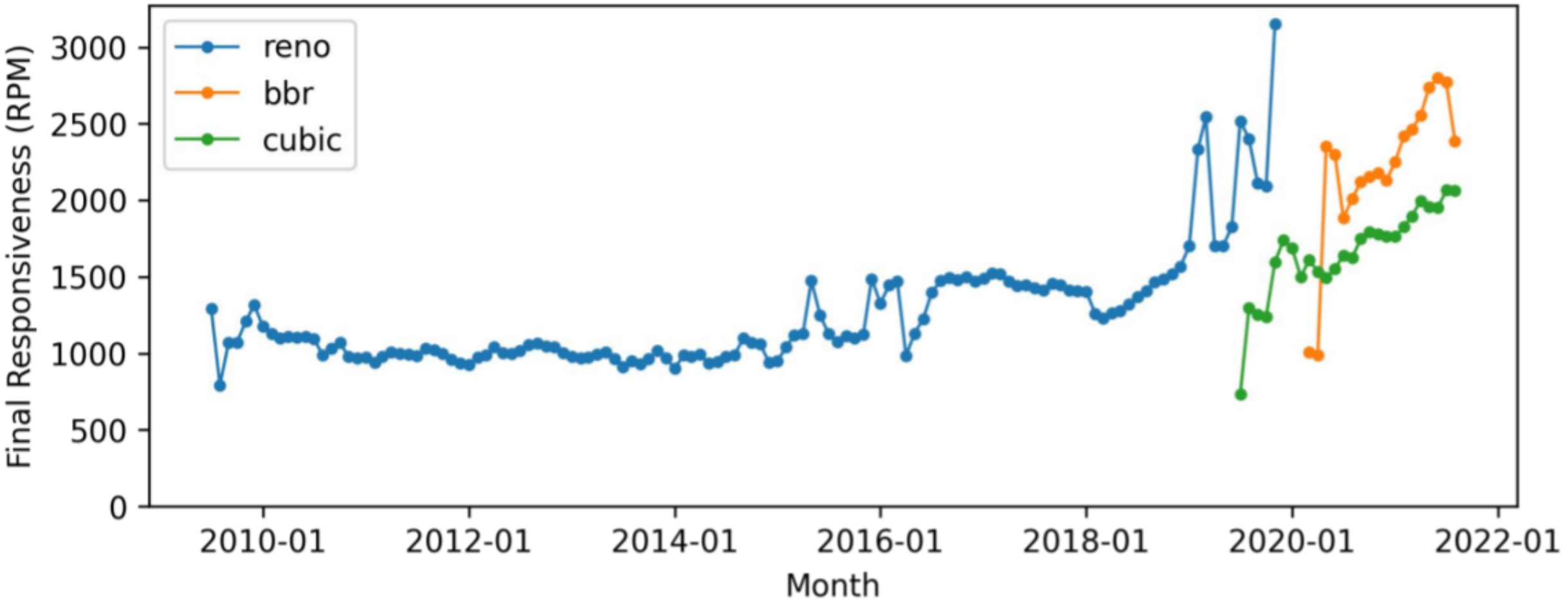
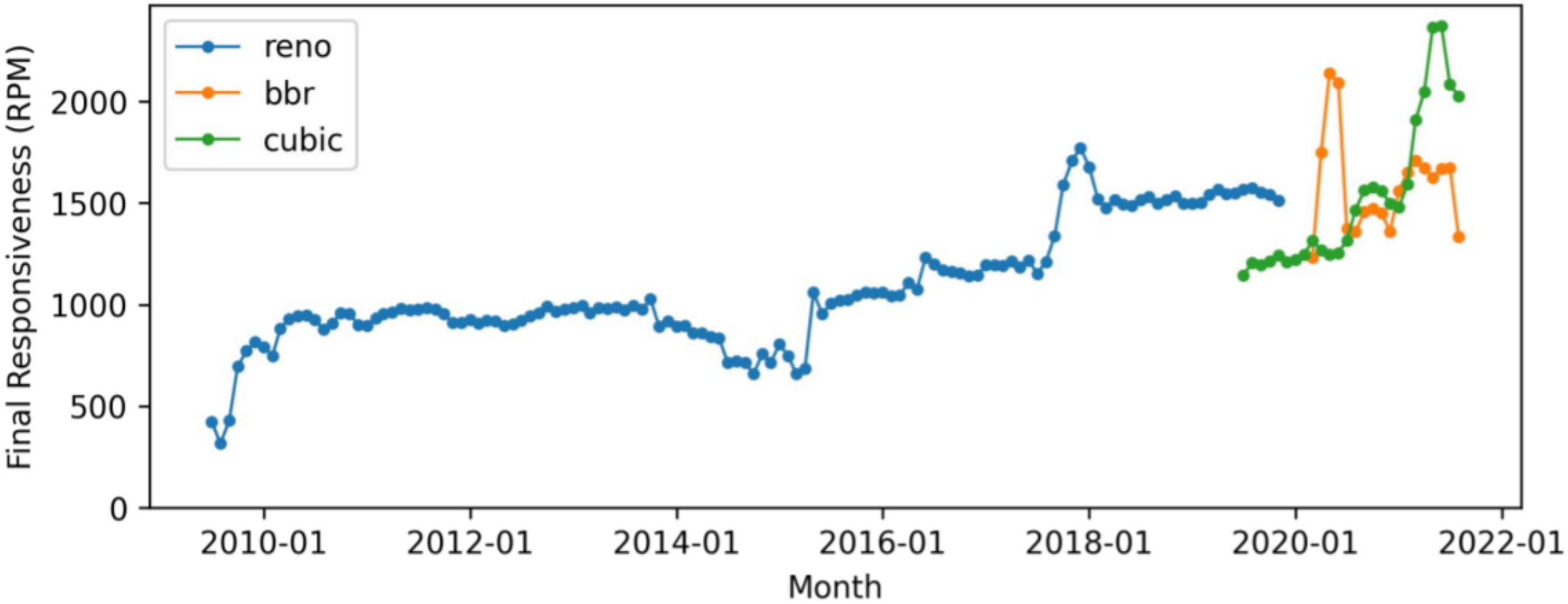
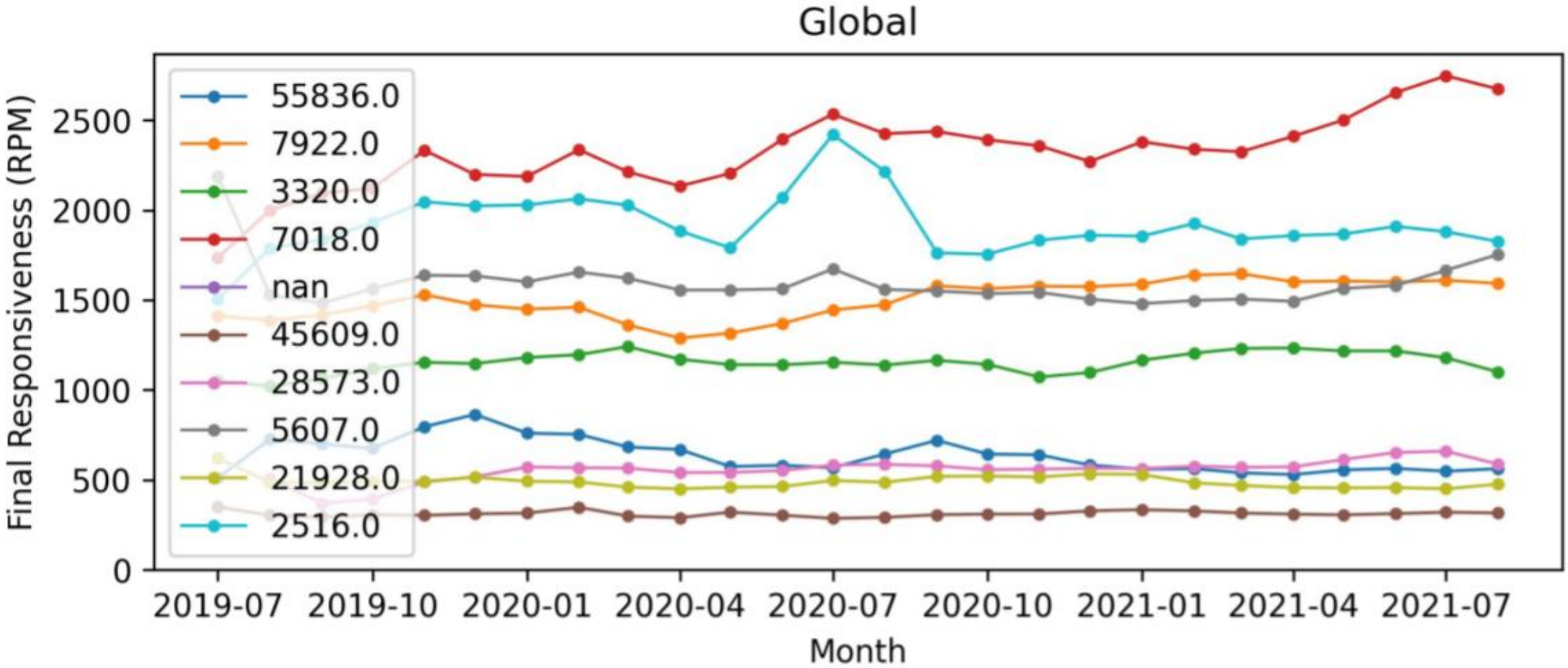
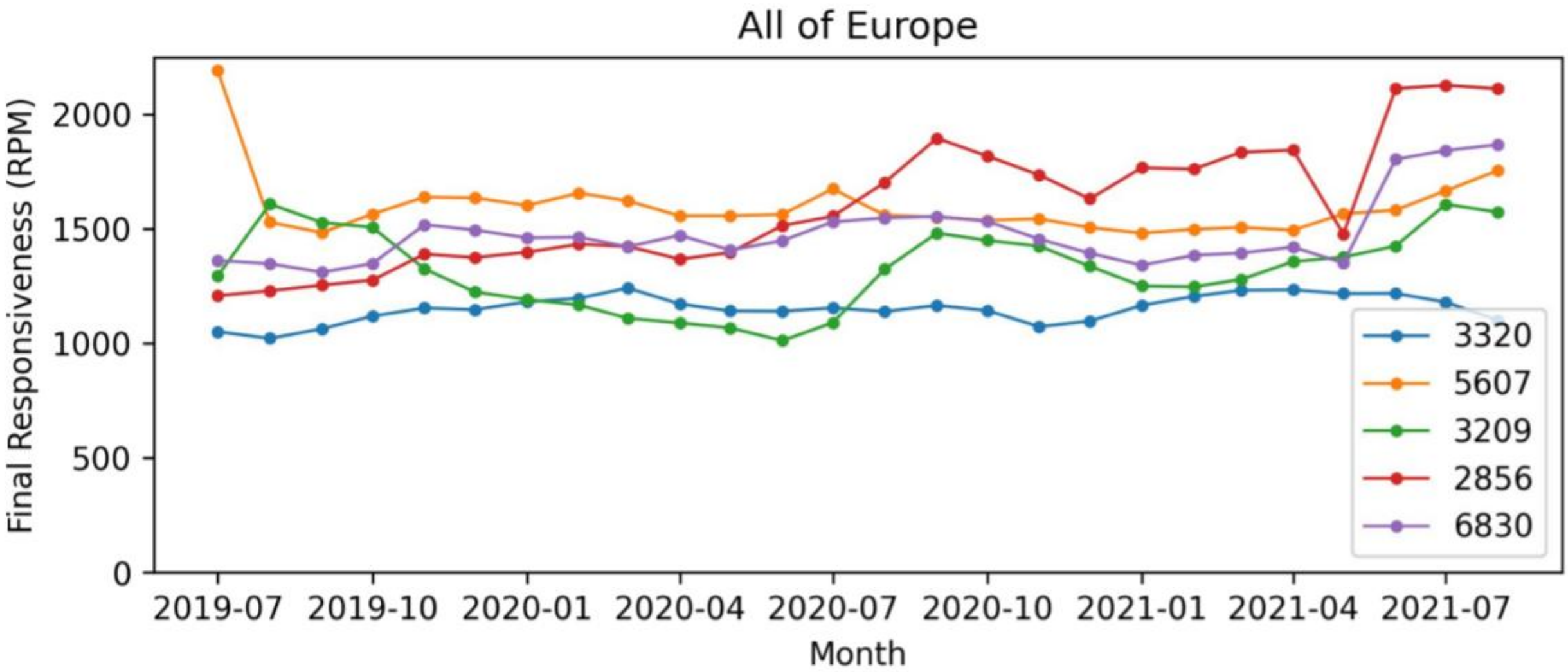
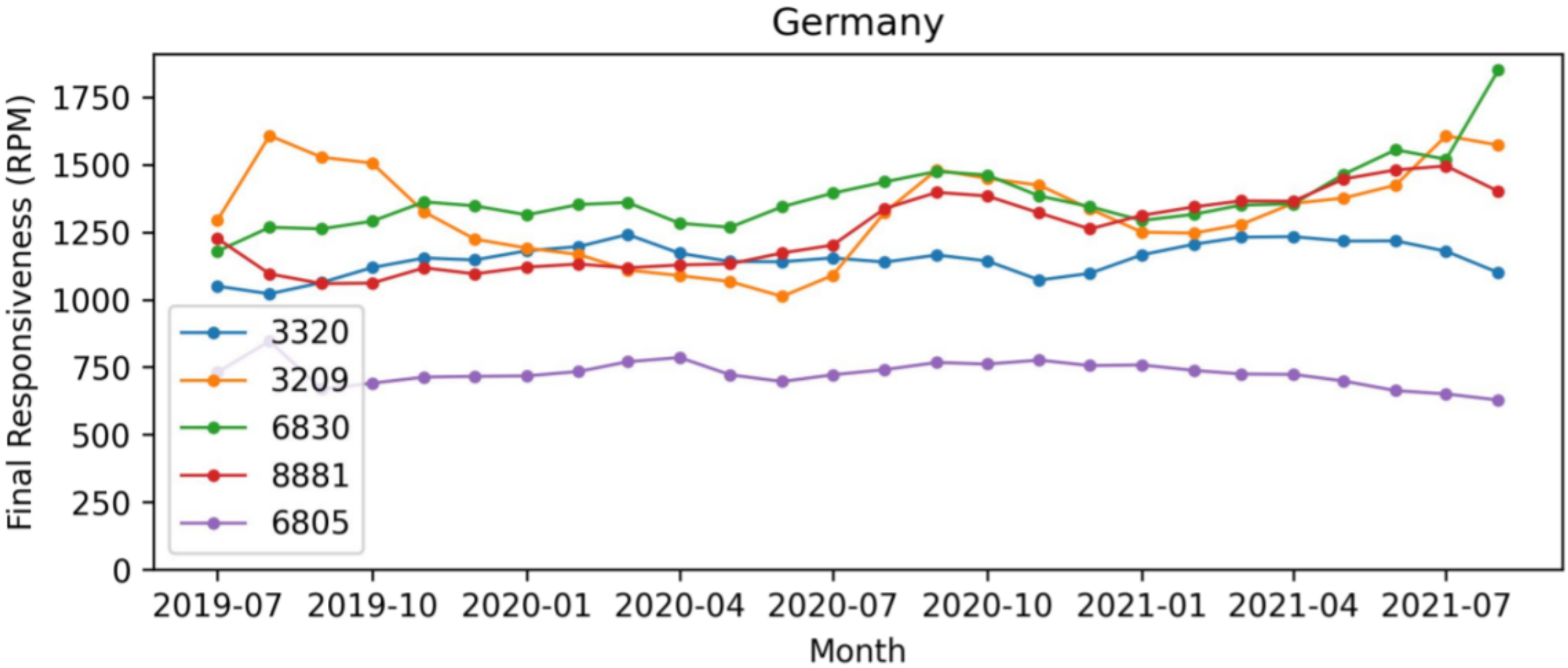


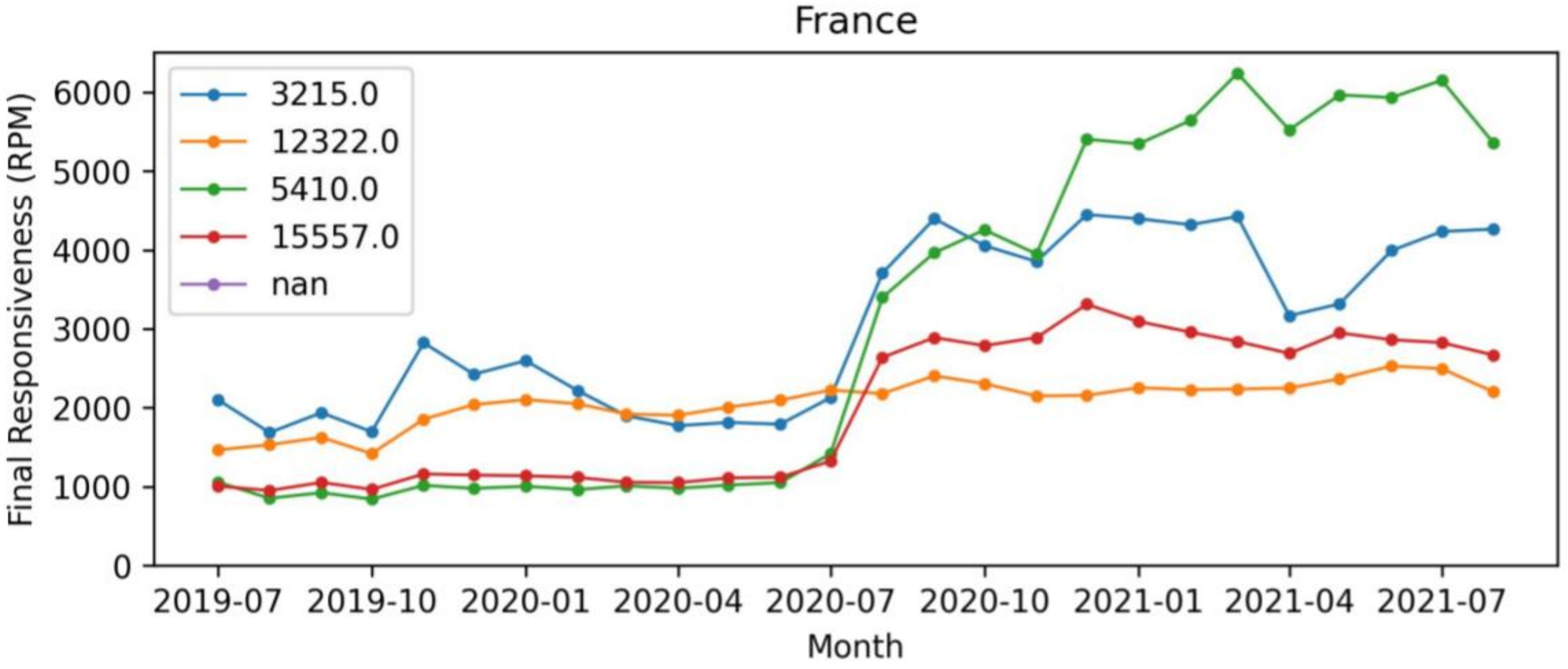
Figure 3: Mean German Responsiveness




Global
55836.0
7922.0
3320.0
7018.0
nan
45609.0
28573.0
5607.0
21928.0
2516.0
Final Responsiveness (RPM)
0
500
1000
1500
2000
2500
2019-07
2019-10
2020-01
2020-04
2020-07
2020-10
2021-01
2021-04
2021-07
Month


All of Europe
Final Responsiveness (RPM)
2000
1500
1000
500
0
2019-07
2019-10
2020-01
2020-04
2020-07
2020-10
2021-01
2021-04
2021-07
Month
3320
5607
3209
2856
6830


Germany
Final Responsiveness (RPM)
1750
1500
1250
1000
750
500
250
0
3320
3209
6830
8881
6805
2019-07
2019-10
2020-01
2020-04
2020-07
2020-10
2021-01
2021-04
2021-07
Month


France
3215.0
12322.0
5410.0
15557.0
nan
Final Responsiveness (RPM)
0
1000
2000
3000
4000
5000
6000
2019-07
2019-10
2020-01
2020-04
2020-07
2020-10
2021-01
2021-04
2021-07
Month

Internet’s performance from Facebook’s edge

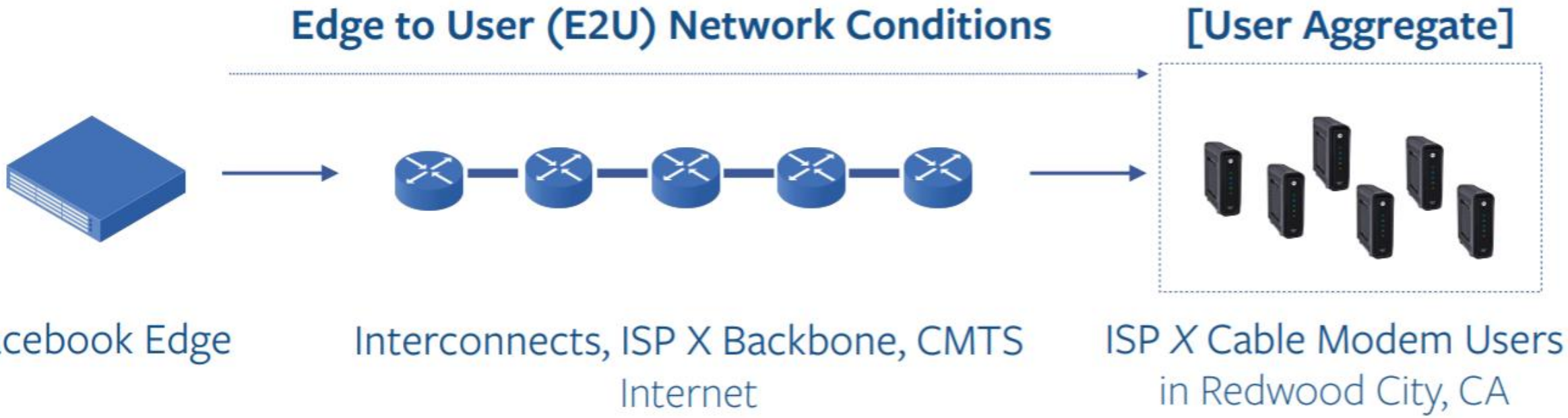
Brandon Schlinker

Internet Performance from Facebook's Edge

Brandon Schlinker
Facebook

Previously published and presented at the *2019 ACM Internet Measurement Conference* with
Ítalo Cunha, Yi-Ching Chiu, Srikanth Sundaresan, and Ethan Katz Bassett

When are network conditions (QoS) between our edge and end-users the barrier to quality of experience (QoE)?



When are network conditions (QoS) between our edge and end-users the barrier to quality of experience (QoE)?

For instance, when do network conditions….

1. Prevent stall-free playback of video on demand with acceptable MOS from beginning within 1 second of request?

2. Result in a user waiting more than 500 ms for an image to load?

When are network conditions (QoS) between our edge and end-users the barrier to quality of experience (QoE)?

Why do we want to know this?

1. Characterize product network requirements
2. Quickly evaluate how changes in network conditions impact QoE
3. Identify improvements at product, network, and transport layers

Measuring Network Conditions: Propagation Delay

Minimum network time for any request, regardless of response size

What does it capture?

Link propagation + interleaving + transmission + persistent queues

How is it measured?

p50(minRTT) over aggregate connections in time window

How is it used?

Ascertain when propagation delay alone is barrier to QoE

Variations *may* indicate congestion at shared links (but not always!)

Measuring Network Conditions: Goodput Success Rate

Probability that a connection can deliver a stream of bytes at a given rate

What does it capture?

Cong. control (CC) + transport behavior + loss + shaping at bottleneck link

Challenges in measuring from FB's production traffic

Cannot use object transfer times as they are a function of application behavior (multiplexing), CDN conditions, CC startup behavior, prop delay

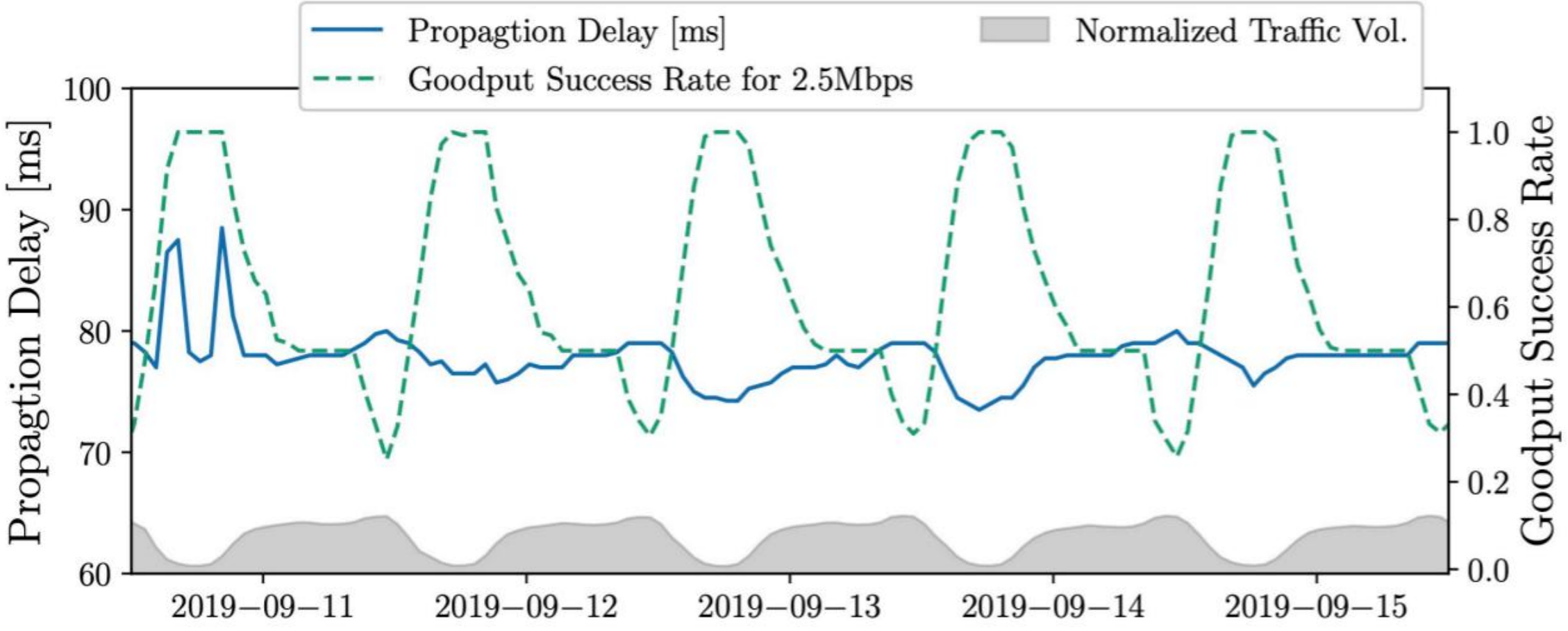
How is it measured?

By measuring transport's ability to reliability deliver bytes (goodput)

[See paper for details, approach accounts for CC behavior, multiplexing, and other aspects.]

Example of Variation in Network Conditions

Likely cause: Cellular RAN congestion



Key Insights: Variations and Opportunities

How often does network performance degrade for an aggregate?

Not often. 97% of global traffic saw at most a 10ms degradation in propagation delay during the 10-day study.

Can performance-aware routing provide tangible benefit?

Not often. Only 2% of global traffic would see any improvement in either metric suggesting access network is often bottleneck.

More opportunity in emerging markets, but difficult to take advantage of it due to potential for self-induced congestion when shifting traffic.

Key Insights: Network Metrics

Why not use other metrics, such as sRTT or loss / retransmission rate?
Difficult to interpret in isolation — they are functions of congestion control and sender behavior. For instance, in some environments CUBIC has a lower retransmission rate than BBR, but also has lower goodput.

Goodput more holistically captures network conditions relevant to these applications: the ability to get bytes to users.

How do you determine network conditions required for "good" QoE?
Ongoing work; key challenge is that it depends on how application makes use of network resources (e.g., prefetch, accuracy of ABR BWE).

Temporal Analysis of Performance

Group by { PoP, Client IP Prefix, Client Country }, 15 minute windows

10ms latency degradation threshold

No degradation for 97% of traffic

Continent	Traffic with +10ms diurnal degradation
Africa	9.2%
Asia	4.1%
South America	2.3%
Europe	1.2%
North America	0.9%

Discussion

To be enqueued, please write ‘+q’ in the chat

To cancel being enqueued, please write ‘-q’ in the chat

The duration of each comment is limited by 60 seconds