Latency & AQM Observations on the Internet

Jake Holland
Akamai

Outline

- Background & Goals
- Results
  - 2 conclusions with outline of observations
- Next Steps
  - Looking for a collaborator

\+ Backup slides with bonus stuff

Background & Goals

Akamai enabled ECN (RFC 3168) globally in February 2020

Initial Questions:

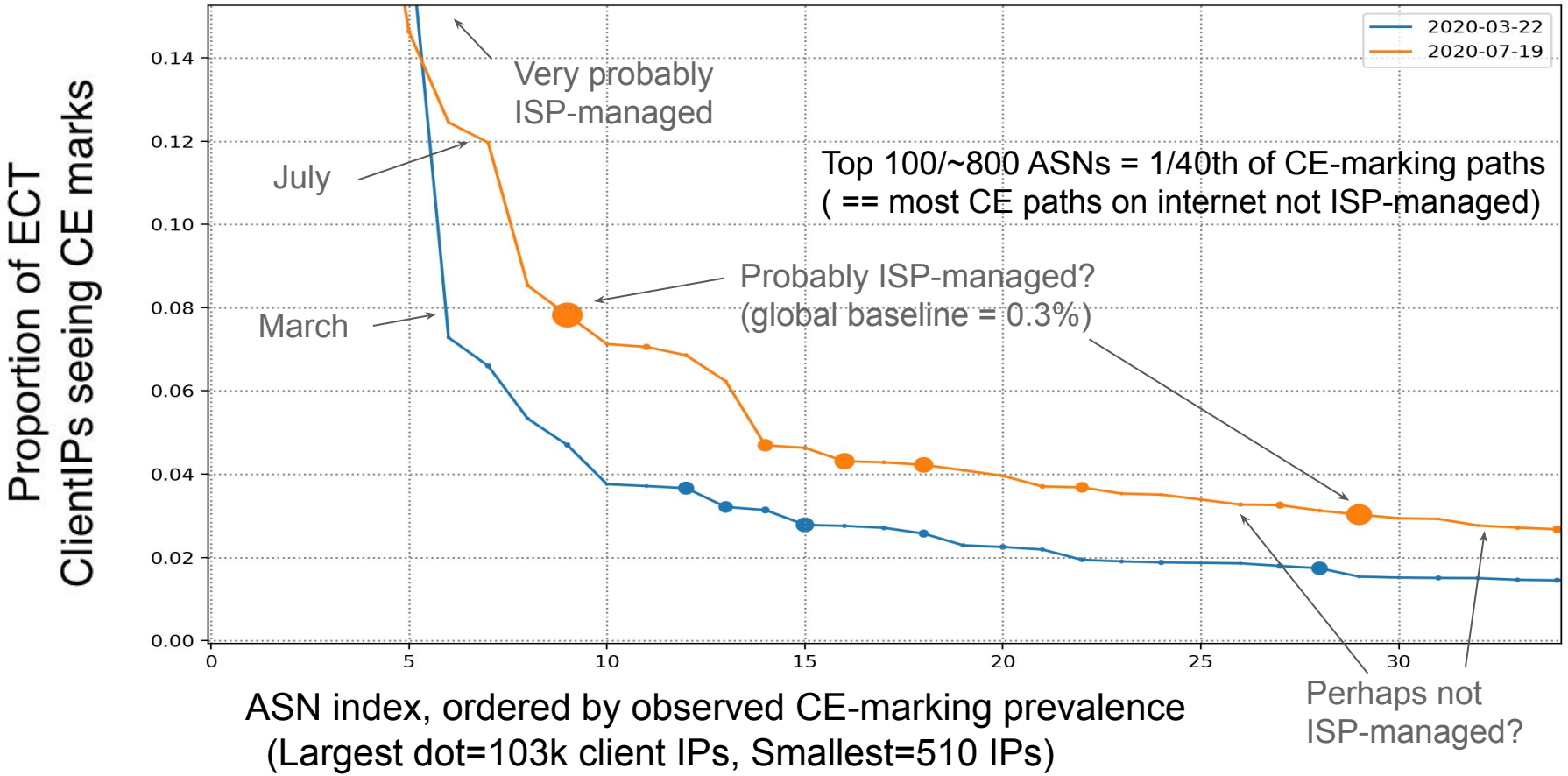
- How prevalent is ECN usage and utilization?
- Can home router solutions fix latency variation?
  - Target: reliable 45ms End-to-End, for gaming

(PS: plz finish in ~3 weeks.)

Conclusion 1: CE-marking low but growing

- Growth pattern suggests some ISPs have picked up marking since March
- Inspired by “Experience Enabling ECN on the Internet” @IETF 98
  - Looks a bit different from the server side
  - Looking Downstream-only (“server ever saw ECE”)
  - Looked by ASN instead of by country
  - (Note: prior 30% Argentina observation was separately discovered as a bug)

March vs. July: CE-marking prevalence, per AS#



Conclusion 2: Home Router AQMs not sufficient :(

- Note: **this study excludes TCP self-congestion** (SYNACK to ACK only)
  - Home routers help a lot for self-congestion!
  - But on small packets, low throughput, in practice problems are still coming from elsewhere, even on CE-marking paths
- CE-marking paths experience ASN-correlated latency variation
  - Usually AQM helps, but not as much as a better ASN

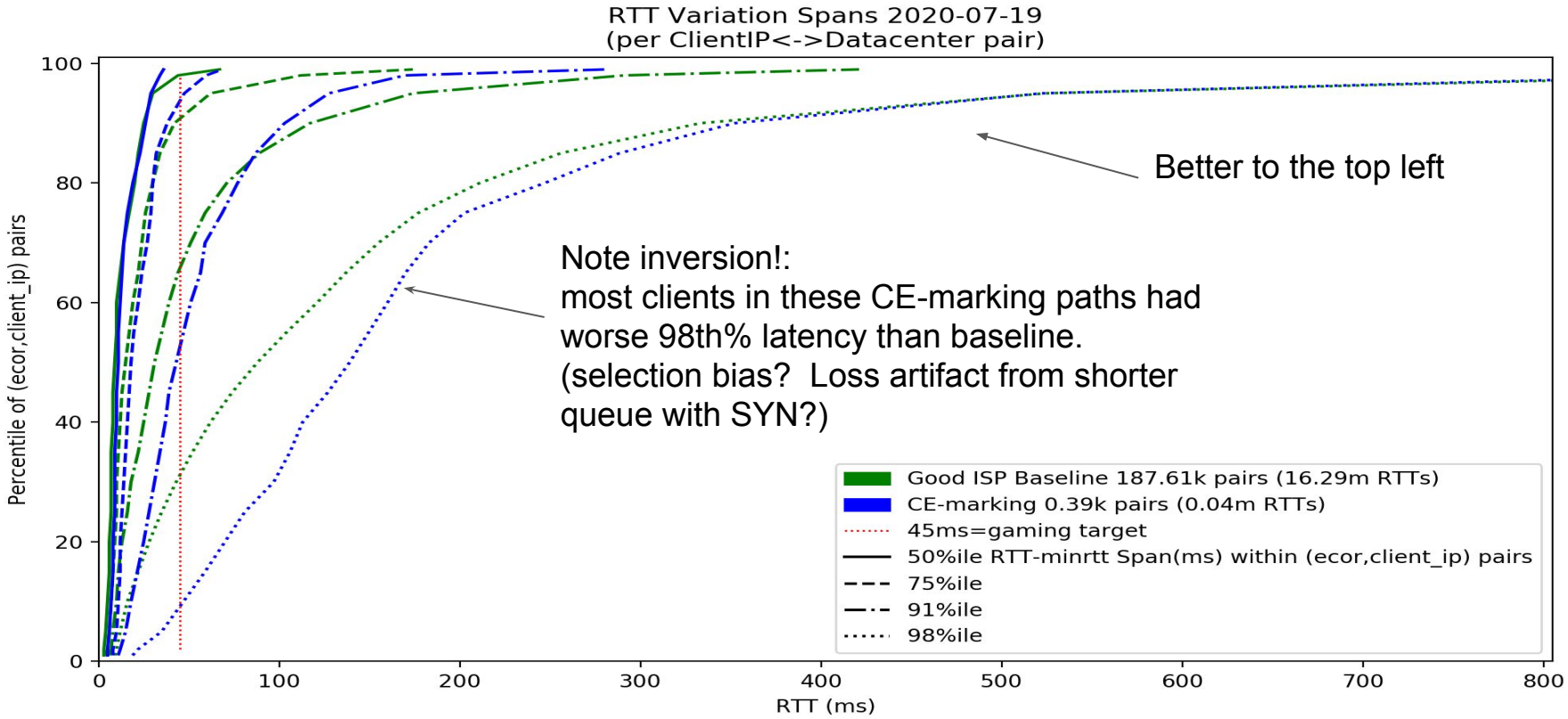
Data Set: Latency Span over 1 day

- Latency Span within Client IP over 1 day
- Grouped by ASN
  - I chose 2 particular ASNs for illustration
  - “Good” and “Middle”, according to 91st and 98th%ile of latency for median client IP
- Inspired by “Measuring Latency on the Internet” @IETF 99’s maprg
  - Server-side passive measurements on production traffic
  - Latency Sample = Delay between SYN+ACK and ACK

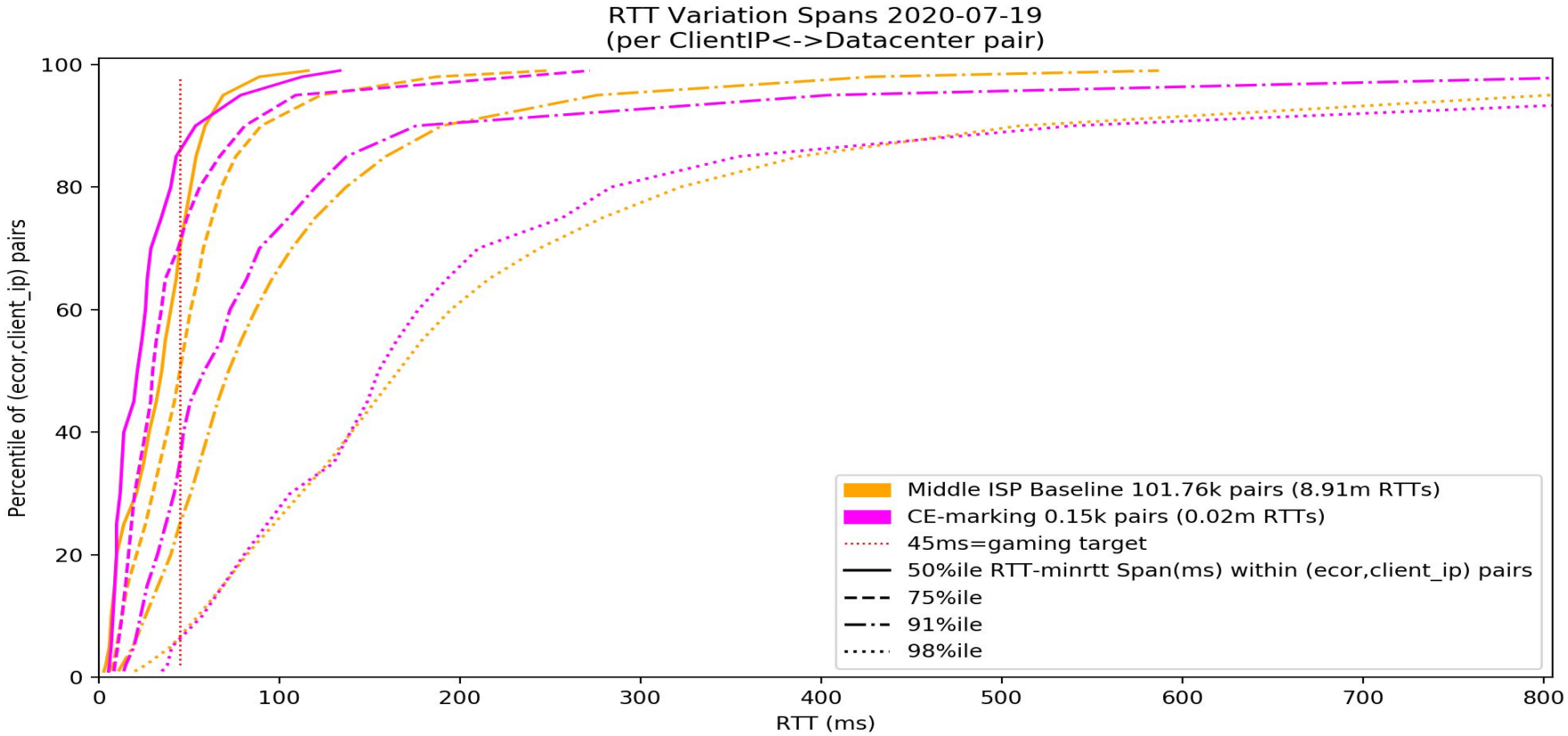
Data Filters: isolate access path to end users

- Filtered for “nearby home/office consumer line”:
  - At least 50 samples/client IP from same datacenter in the same day
  - At most 300 samples/client IP in the same day (exclude CGNAT/VPN)
  - At most 20ms for the minimum sample
- Post-filter:
  - 5.3b Latency samples in 31m client IPs
  - 33k clients ever saw markings out of 11m that ever negotiated ECN
- Actually Datacenter <-> ClientIP Pair
  - Same client has different latency span to different datacenter
  - So the same client IP could appear up to 6 times after filter

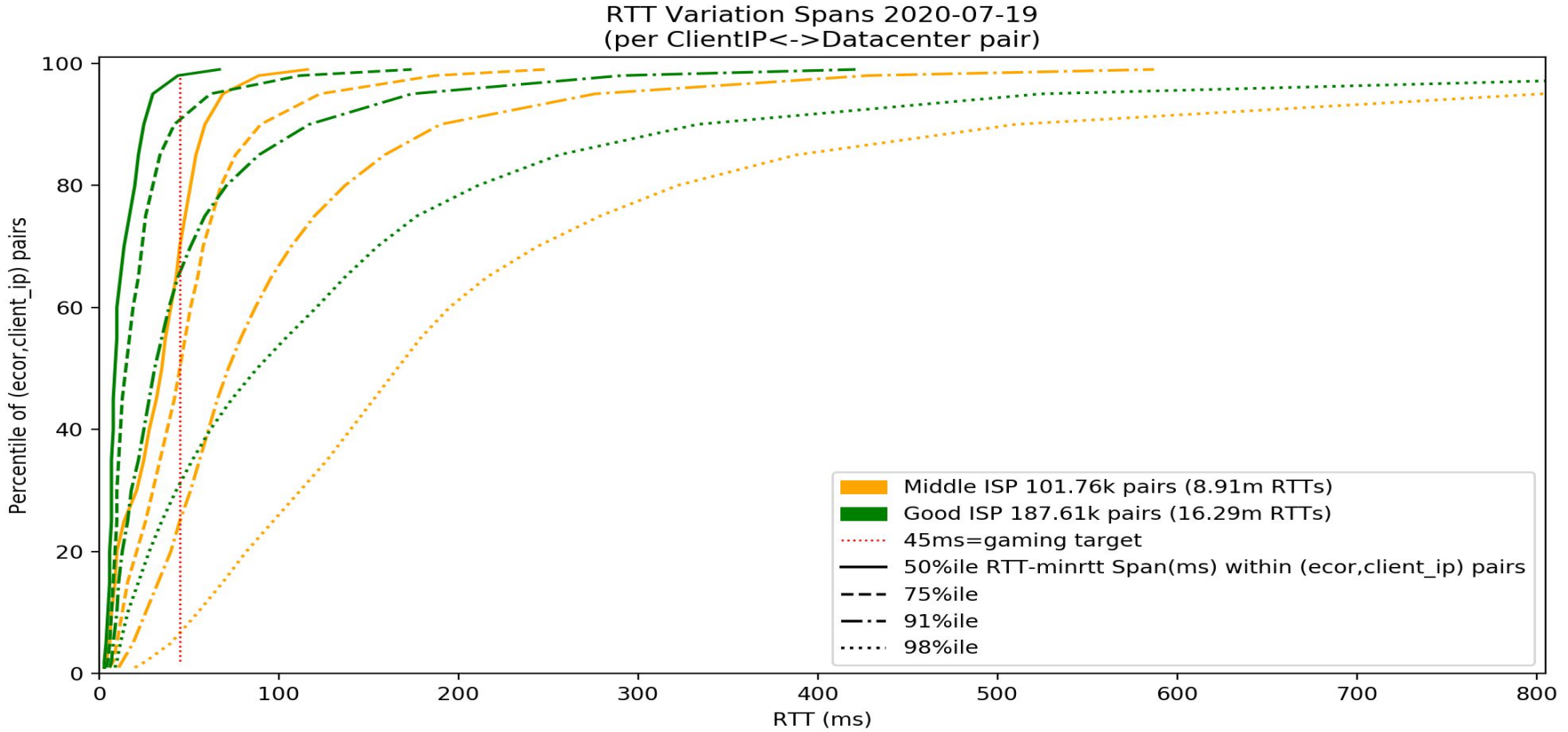
“Good” ISP Latency Spans: CE-marking vs Overall



“Middle” ISP Latency Spans: CE-marking vs Overall



“Good” ISP vs “Middle” ISP



Next Steps

- Find a collaborator
  - Interesting stuff, but out of my scope

Contact jholland@akamai.com if you want to give this data the analysis and write-up it deserves.

Backup Slides
and Supplemental Data

Single day global counts, 2020-07-19

Pre-filter totals

Connections	32.7b
ECT Connections	1.25b
Connections with CE marks	694k
Total Client IPs	36.7m
Client IPs that ever used ECT	15.1m
Clients that ever saw CE	55k

Post-filter totals

Connections	5.3b
ECT Connections	163m
Connections with CE marks	157k
Total Client IPs	30.7m
Client IPs that ever used ECT	11.1m
Clients that ever saw CE	33.2k
Client IPs in 100 top prevalence ASNs	538k
ECT Client IPs in top 100	252k

Single day global counts, 2020-03-22

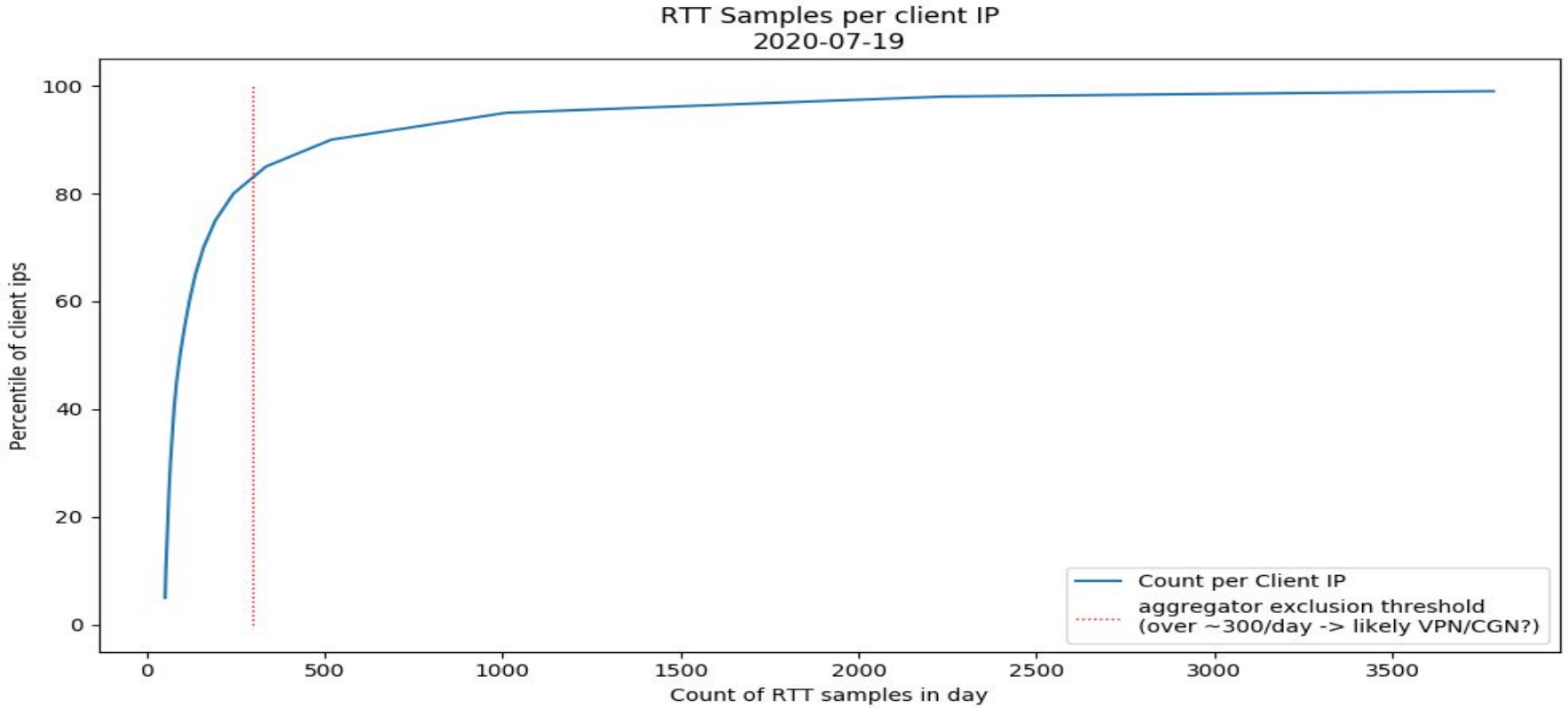
Pre-filter totals

Connections	31.2b
ECT Connections	1.07b
Connections with CE marks	245k
Total Client IPs	22m
Client IPs that ever used ECT	9.1m
Clients that ever saw CE	22.5k

Post-filter totals

Connections	3.5b
ECT Connections	97m
Connections with CE marks	54k
Total Client IPs	17.1m
Client IPs that ever used ECT	6.4m
Clients that ever saw CE	12k
Client IPs in 100 top prevalence ASNs	409k
ECT Client IPs in top 100	210k

Why cut at 300 for CGNAT/VPN?



Precluded Objections

I’ve heard these raised, but I think they’re filtered:

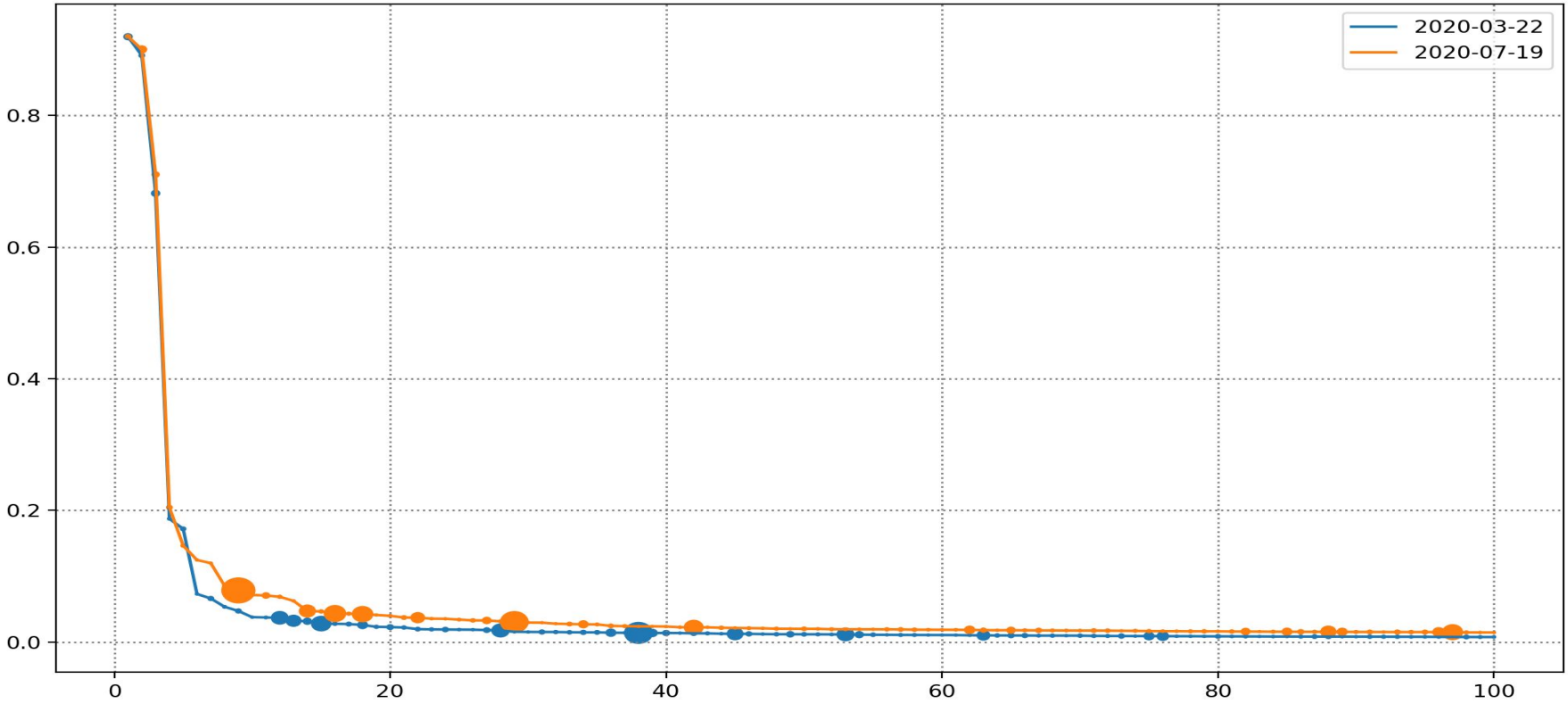
- Variation is from geographic diversity within ISP
  - No: limited to Client IPs that saw min sample <20ms that day
- Variation is from dissimilar paths behind VPNs
  - Minor noise: excluded Client IPs with high sample count (VPN/CGNATs)
- Variation is wi-fi
  - No: major differences between ISPs, even though everyone uses wi-fi

ASN Filter Criteria

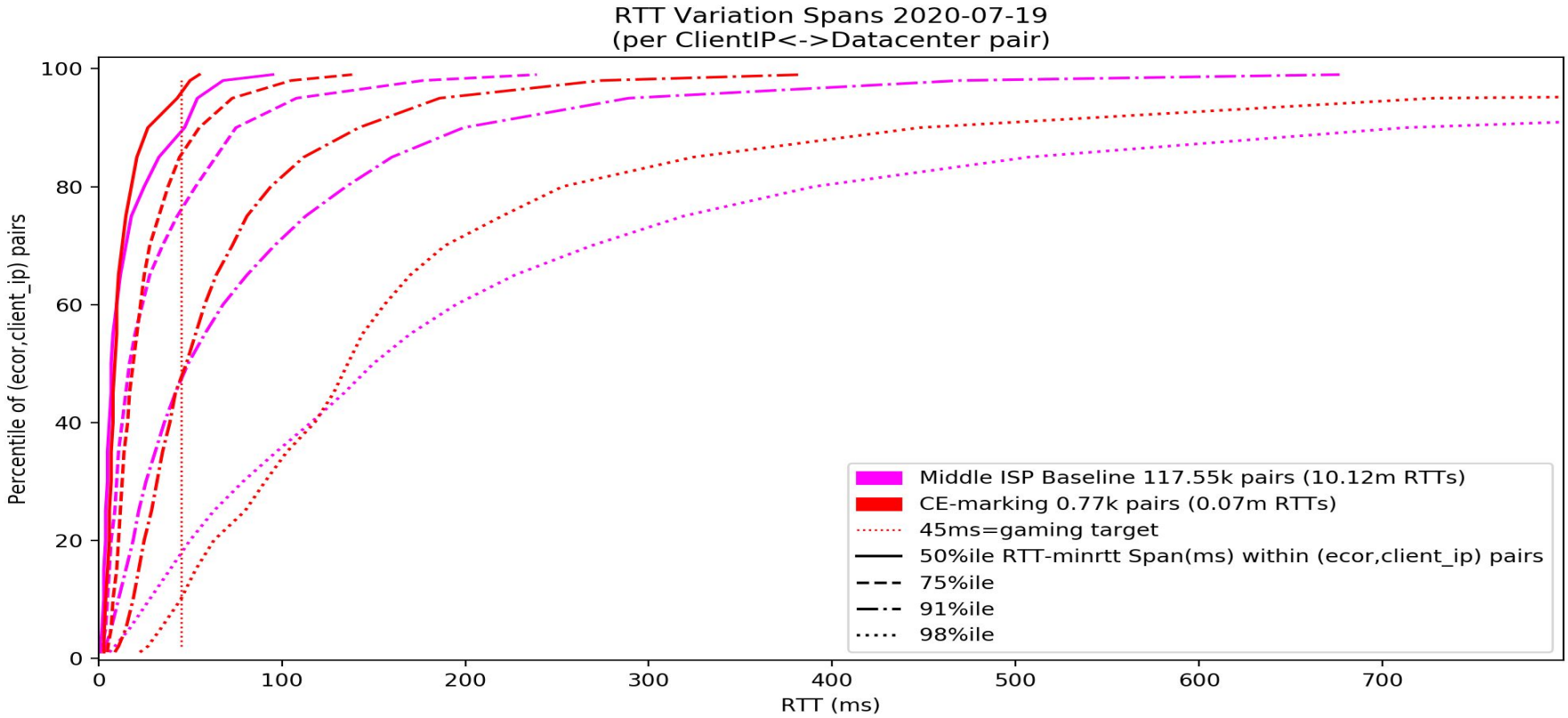
Exclude networks too small to use

- >= 450 Client IPs (post-client filters)
  - Sounds small, but by manual check includes many eyeball networks
- >= 100 ECN-using client IPs
  - May include some false positives. (Note that 3% CE-marking prevalence on 100 ECN-using clients is only 3 clients.)
  - However, low counts of CE-marking on CE-capable paths, and likely have false negatives as well.
    - CE marking observations only give **lower bound** on prevalence
    - Many very tiny connections won’t see congestion (there are lots)

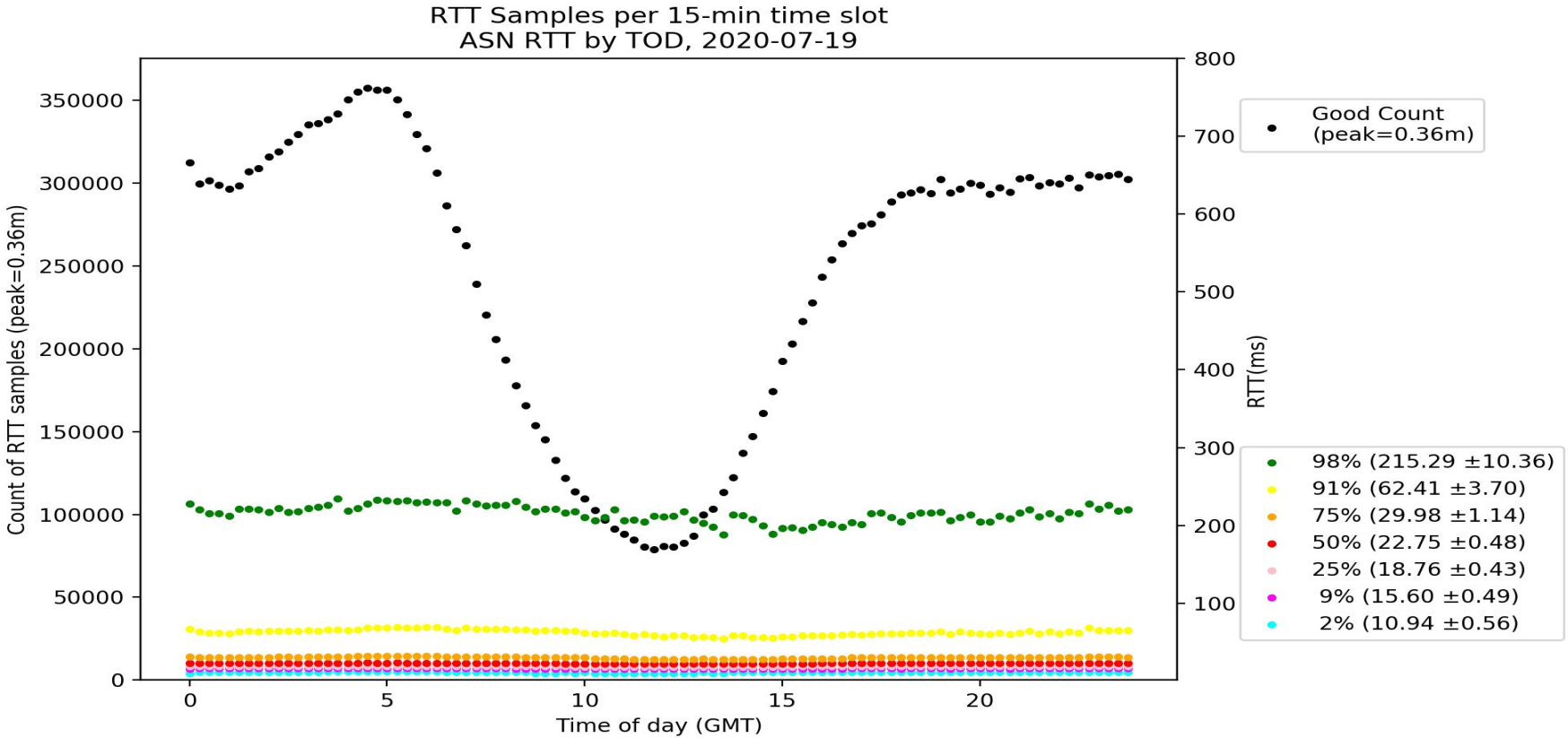
Top 100 ASNs by CE-prevalence full view



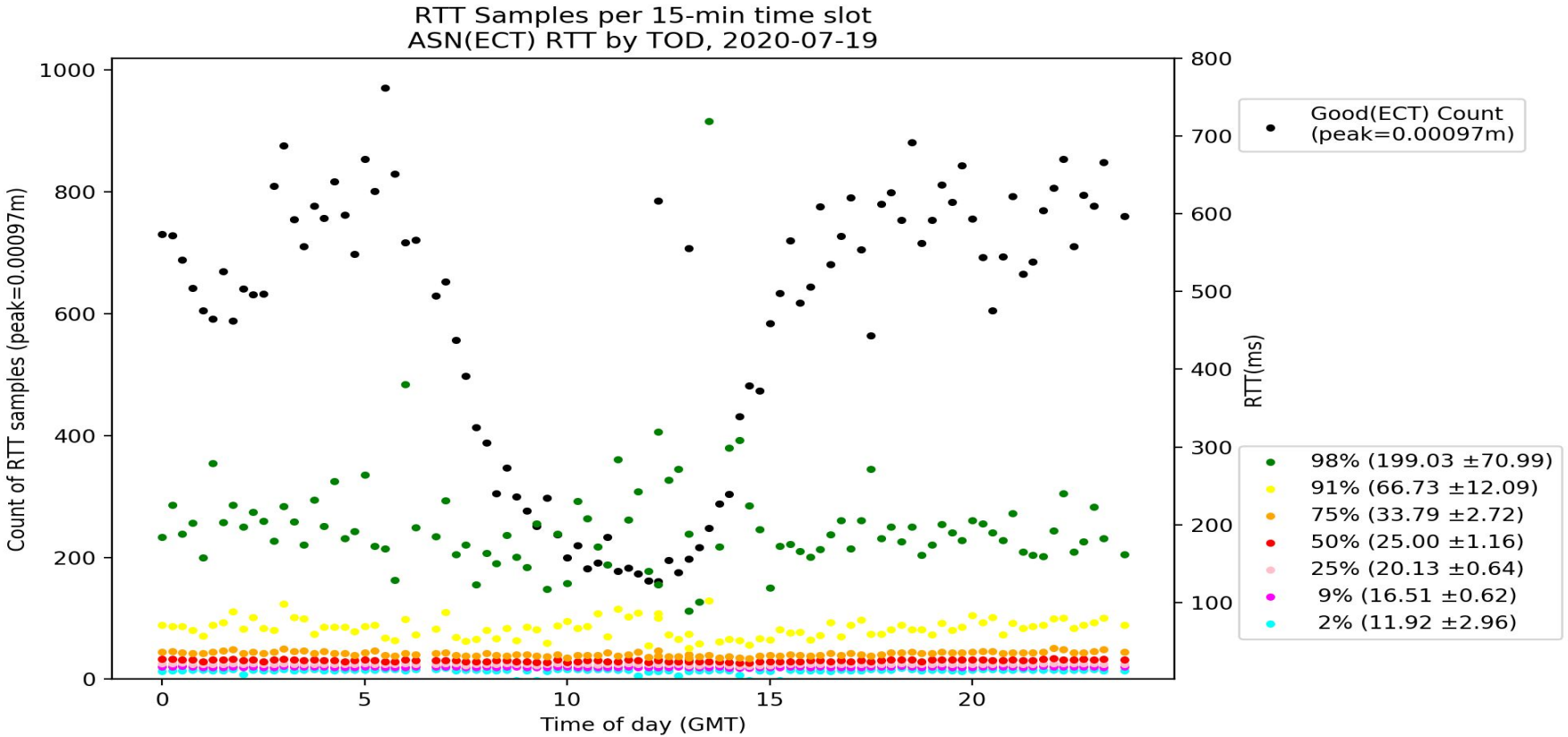
Not All Middle ISPs--some get decent AQM benefits



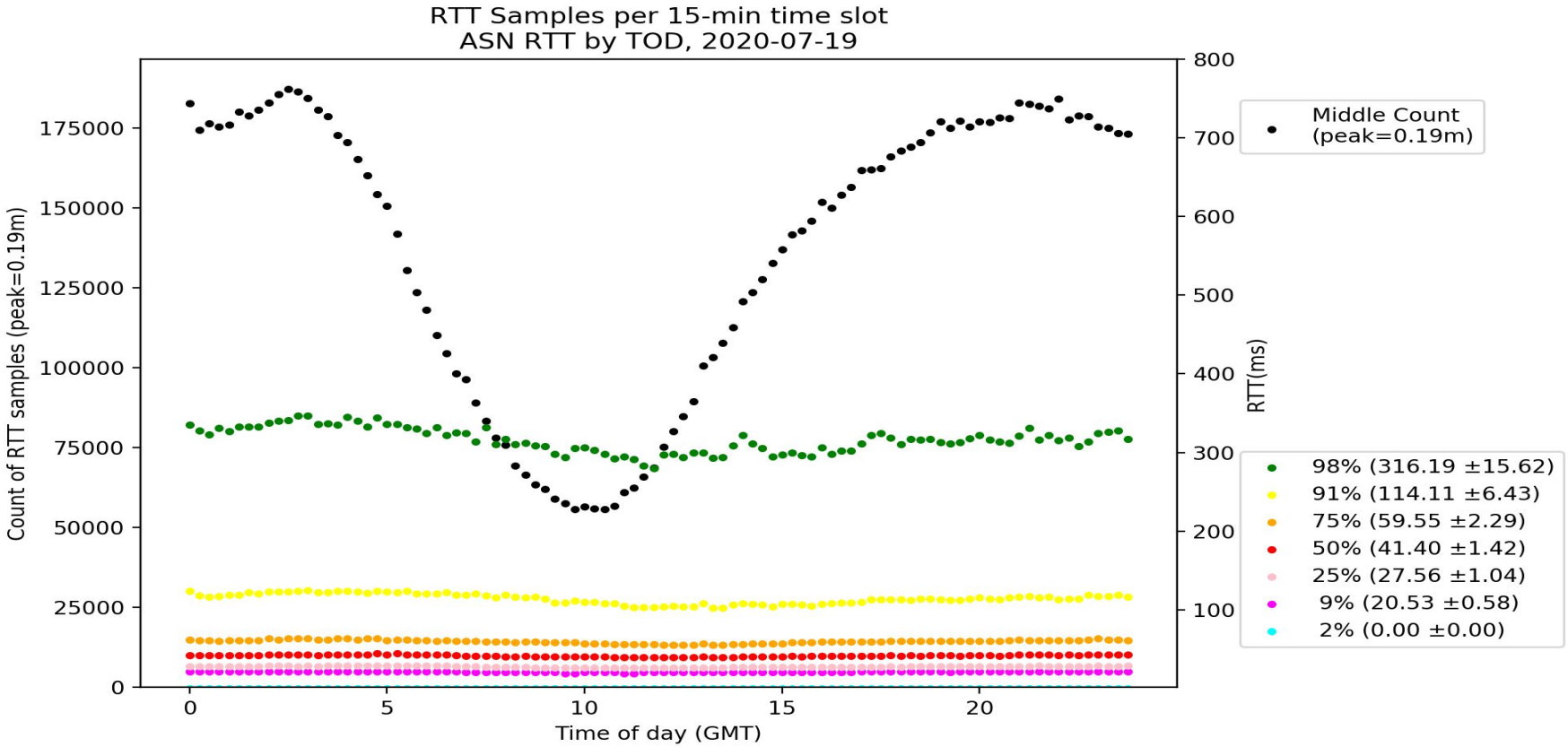
Time of Day Latency vs. Sample Count (Good)



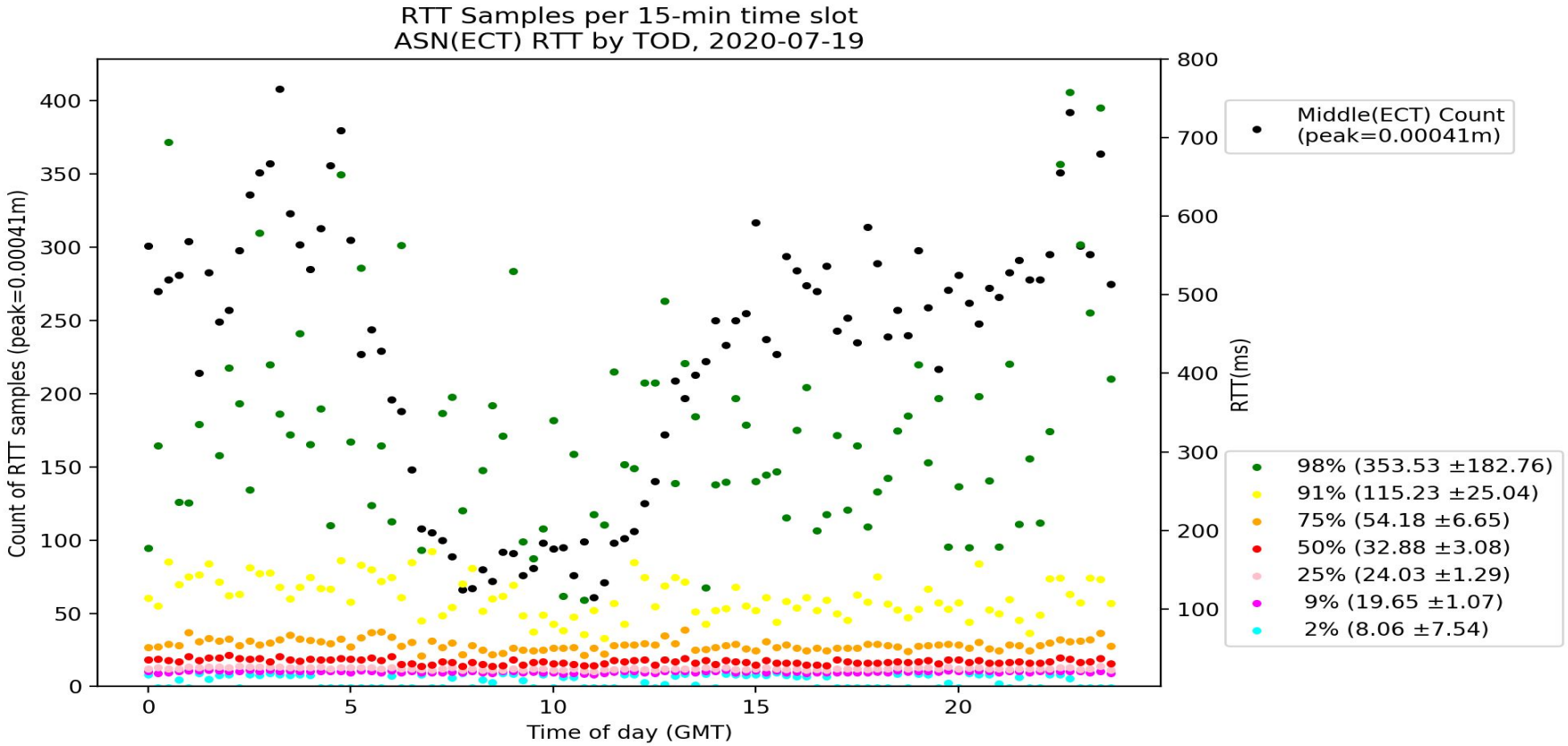
Time of Day Latency vs. Sample Count (Good CE-marking)



Time of Day Latency vs. Sample Count (Middle)



Time of Day Latency vs. Sample Count (Middle CE-marking)



Known potential sources of error #1

- Most flows not seeing CE marks even on CE-marking paths
  - Most common count of sessions with CE marks per client is 1. 2nd most common is 2, etc. thru 8.(of minimum 50 sessions)
  - Likely under-counting, by unknown amount
  - Many of our flows are tiny and may miss CE when cross-congested
  - Our pacing/CC strategies may often avoid causing congestion
- Downsampled input stream
  - Some odd and complex downsampling before I get the data
  - Possible source of unknown biases.
- Inter-day trends and comparisons
  - Different days have different traffic profiles depending on events
  - Not automated, no systematic inter-day comparisons.
  - Spot checks for robust conclusions hold up well so far, but are anecdotal

Known potential sources of error #2

- Latency variation due to wi-fi and other effects not known
  - Would be nice to know what's achievable with ISP access AQM
  - Gamers often run non-wi-fi. Would be nice to differentiate somehow.
- Client IP remapping
  - Thought to be uncommon, but adds unknown amount of noise
- Other Weaknesses in SYNACK->ACK dataset
  - Intra-flow variation would be nice to add
  - Especially with instrumentation measuring pre- vs. post- congestion events, like Toke's other experiment in maprg 99 talk
- Count of CE-marking paths too small for good numbers in some cases

Known potential sources of error #3

- Non-marking AQM provides same latency
  - CE-marking not a great proxy for AQM, only a lower bound
  - Especially with DOCSIS3.1.
- Unintended side-effects from filtering (trying “close”, did it pick up more?)
- “Experienced a CE” may be biased relative to “did not experience CE”
  - Could explain inversions? (where CE worse @98th %ile)
- Server-side latency (within datacenter, e.g.)
  - Believed but not proven to be negligible. If not, could corrupt results.