Taxonomy and impact of throttling policies used worldwide

IETF 122, Bangkok

Kyriakos Zarifis, Meta
March 21, 2025

Throttling background

Communication Service Providers employ throttling to:

- Differentiate service across subscribers
- Enforce data capped plans
- Ensure fair usage
- Manage infrastructure costs
- Mitigate network congestion on busy cells or during busy hours

Throttling implementations

Policers

- Drop excessive packets
- Smooth mean transmission rate, with fluctuations
- High retransmission rates

Shapers

- Buffer excessive packets
- Stable rate near Committed Information Rate
- Inflated RTT

Impact

Communication Service Providers

- CPU and memory resources
- Network overhead (retX)

Content Application Providers

- QoE
- CPU overhead (retX)

End users

- QoE
- Battery consumption (DRX)

Estimating Internet Throttling: An Empirical Perspective

- 2015 (Google)
  - ~2% (NA) - ~7% (Africa) of video segments policed
  - 30% of policed video connections throttled to 0.5 - 2Mbps
  - 20-30% retransmissions, 15% increased rebuffering for policed connections

- 2019 (Northeastern University / several video streaming providers)
  - 30 of 144 analyzed ISPs throttled at least one streaming service
  - Most throttle at around 1.5Mbps
  - Same ISP can throttle different services differently (heuristics)
  - Always-on throttling, regardless of network conditions

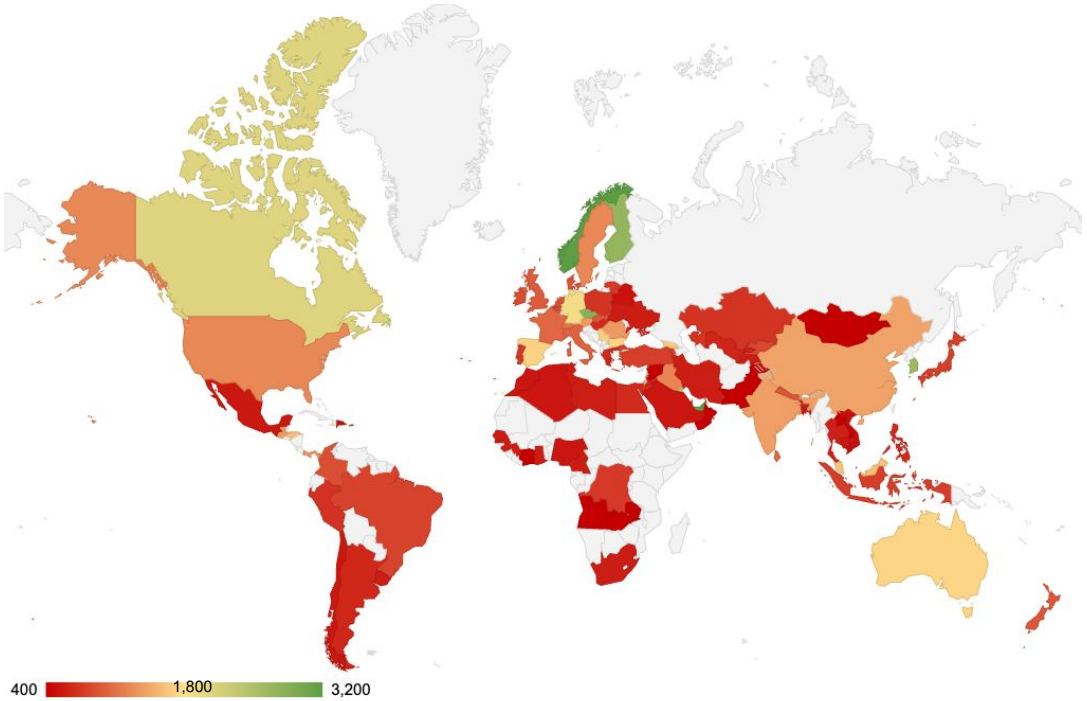
- 2022-2024 (Meta)
  - 200 ASNs outside US, 275 globally
  - (taxonomy of policies below)

% of ASNs in throttling rate range

	<1Mbps	**1-2Mbps**	**2-5Mbps**
2022	60%	30%	7%
2024	68%	30%	2%

Estimating Internet Throttling: An Empirical Perspective

Average detected throttling rates (kbps) (2024)



Average estimated throttling rates across CSPs that were **detected** to throttle Meta services in a country

275 ASNs detected globally

Visualization meant to convey high level regional trends based on some of the detected throttling rates, not to establish ground truth. Does not take into account policies.

Taxonomy of observed throttling policies

- **Constant, Application-Based Throttling**
  - Target video flows from Content Application Providers
  - Consistent bandwidth allocation for certain types of traffic, regardless of network conditions.

- **Time-of-Day (ToD) Based Throttling**
  - Manage peak traffic periods

- **Data Capped User Throttling**
  - Throttle once predefined data limits are exceeded
  - Daily/monthly quotas or manual top-ups
  - App-specific data caps may be enforced

- **User-Specific Policies**
  - Based on individual user profiles or subscription tiers
  - Allows CSPs to offer differentiated services and pricing models

- **Network topology specific policies**
  - Different throttling on 5G vs 4G network.
  - Different throttling on wifi over cellular/satellite/fiber/cable backbone

Combination of above policies are possible

Internet Throttling: An Empirical Perspective

Overall takeaways

- Trend towards more throttling and sometimes towards lower rates as video traffic increases
- Throttling policies vary dramatically across regions and across CSPs in same country/region
- User Data Caps are prevalent globally
- CAPs and CSPs want to collaborate on achieving Fair Usage Policy

Design considerations for throughput advice signalling

- **Frequency of signal**
  - Policies can be constantly active time-of-day based
  - Policies be reactive to user-specific changes, like exceeding data quota
  - Variability implies a requirement of the signaling protocol to alter the advice in step with these changes in policy.

- **Granularity of signal**
  - 2 dimensions:
    - Rates (e.g. fixed options vs arbitrary?)
      - Rates can be based on network usage and demand
      - Rates can be based on controlling video quality for specific application
      - Rates also change over time with no standardization
    - Advice target (e.g. per-user vs network-wide?)
      - Different subscription tiers and data caps require ability to signal rates at a per-user level
  - A solution which seeks to communicate different rates to different users must be able to encode the wide range of policies

Discussion