Enforcing traffic policies in the streaming era

The communications provider view

Dan Druta, AT&T

IETF #118 SADCDN side meeting

Context

Exponential growth of video traffic over (mobile) networks, evolving patterns and new use cases

- Video makes up much of the internet traffic and streaming content in higher resolution (4K)
- Increased use of ABR (Adaptive Bit Rate)
- Expected Immersive Services (AR/VR and Metaverse) will push for different network requirements (bandwidth and latency). Calls for better understanding of flow characteristics

In-network traffic classification of internet traffic is expensive and inaccurate

- Mixed user-generated content on social media platforms often leads to wrong classification
- Immersive (augmented, virtual) reality applications with mixed content characteristics (such as filters) often lead to wrong classification

Network management requires operators/CSPs to continue to invest in network-based traffic classification tools

- Increased number of content providers makes it difficult to have bilateral agreements between operators and media providers
- Network operators "guess" and "bundle" flows based on inferred classifications
  - Results in potential mistreatment of application traffic that translate in degraded user experience

Problem on hand

- Traffic management is a necessary function for CSPs
- Two major categories of traffic policies:
  - Intentional Management policies include subscription-based limits which may be flow specific
  - Reactive Management policies must be applied to react to congestion events – with very short to very long durations (e.g., varying wireless and mobile air interface conditions)
- Different types of traffic flows may be impacted differently in the face of traffic limits
- Application traffic flows have significantly different network requirements and adaptability mechanisms (such as ABR for streaming)

The industry has not developed a standards-based approach to allow enforcement of traffic management policies while minimizing impacts on application-level QoE

Use Cases

Unlimited Plans With Metered Speed After Threshold - Unlimited plans provide a set amount of high-speed data at the start of the billing cycle, once the subscriber uses this amount, additional traffic management policies are applied

Mobile User On Capped Plan - Once users reach the subscription cap, additional traffic management policies are applied to rate limit user

Tiered or Restricted Streaming Plans - Data plans with a video management policy set to enforce Standard Definition Video or other video options per plan

Intentional policies

Latency Sensitive Plans– Data plans with network attributes that optimizes support for real-time, latency-sensitive traffic, such as AR, VR, remote vehicle control, cloud gaming, etc.

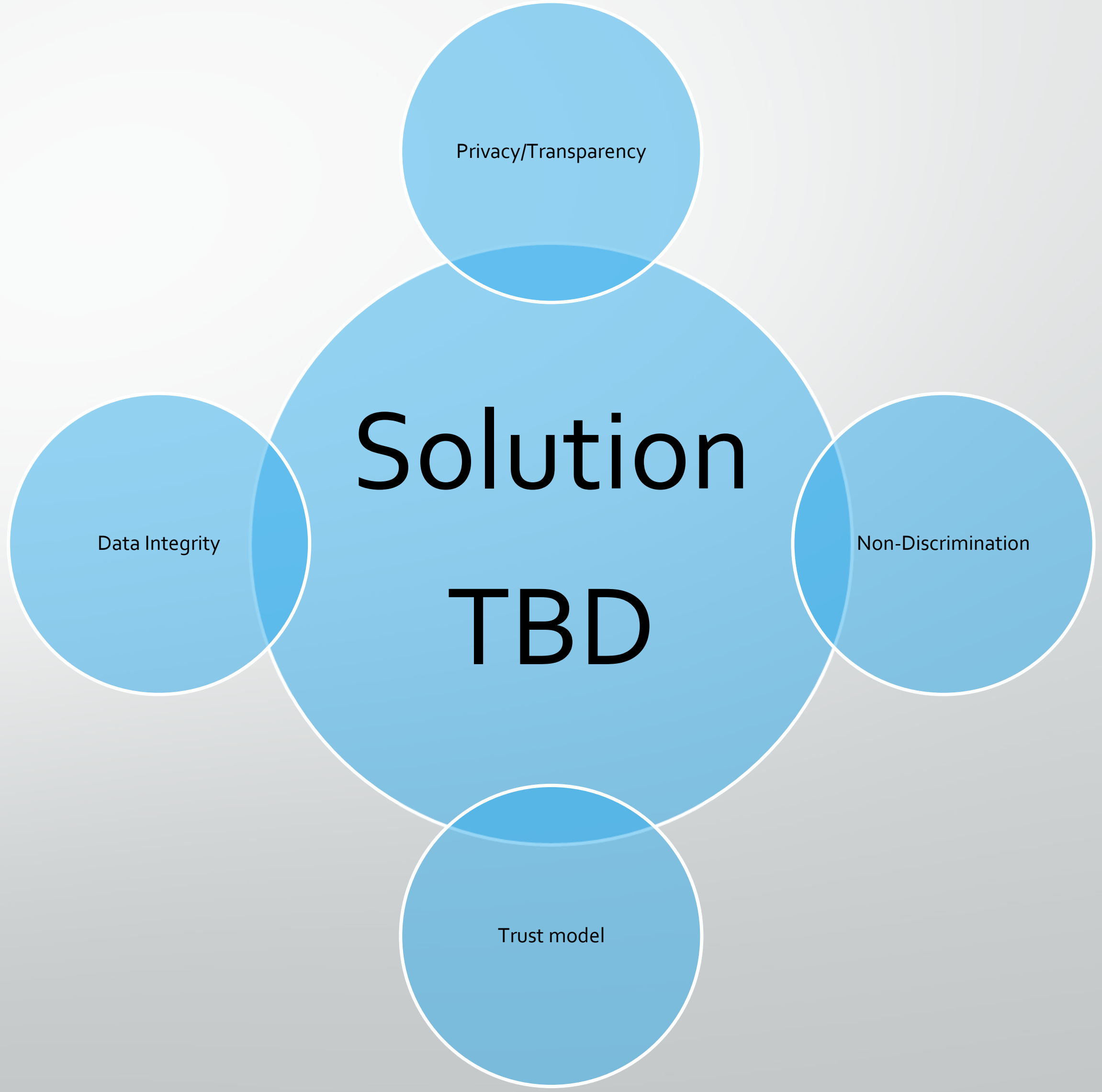
Subscriber Group Management - Apply polices to groups of subscribers to optimize QoE in the face of congestion, examples include prioritizing first responders during an emergency event with congestion

Reactive policies

In all cases, identification/classification of optimizable flows enables traffic management policy with better QoE to the user

Requirements - viable solution characteristics

- Must be win/win/win (for content owners, users and network providers)
- Explicit collaboration between content providers and network providers is required
- Privacy focused with explicit Transparency
- Non-discriminatory – Treat similar content types the same way irrespective of source (content provider) or destination (user, mobile device)
- Broad applicability and availability globally on multiple network technologies
    - Access agnostic – while the challenges are prevalent on mobile networks, we should look at solutions that are access technology agnostic (4G, 5G, WiFi, Fiber, etc.)



Solution TBD

Privacy/Transparency

Data Integrity

Non-Discrimination

Trust model

Nondiscriminatory

- Treat similar content types the same way irrespective of source (content provider) or destination (user, mobile device)
- Non-priority driven classification – the goal is to optimize not to prioritize
- Optional and non-guaranteed

Privacy and Transparency

- Allow the client(user) and the server(content provider) to negotiate what and whom they want to give(or not) visibility into their flows
- Prevent passive interception and prevent man in the middle attacks
- User awareness on information exchange
- Minimize the fingerprinting entropy to avoid user tracking and long-lived sticky identification of device, data flows or users

Trust

- Avoid solutions that rely on provisioned trust relationships between parties as they tend to depend on sophisticated authentication infrastructures that are difficult to maintain
- Cheat-proof mechanisms based on trade-offs discourage parties to communicate misleading info are encouraged
- No guaranteed expectation should be made in the process of bidirectional data exchange between network layer and apps as there is no way to enforce these mechanisms across the internet and all the network operators.

Data Integrity

- The only guarantee required for solutions should be on data integrity
- Encapsulation of data exchange should allow intermediaries in the path add info without tempering with existing flows

Ongoing work in ATIS

ATIS(Alliance for Telecommunications Industry Solutions) started an activity on Content Classification for Traffic Optimization

- Group focused on identifying and evaluating end-to-end solutions for content classification
- The group has looked at available technologies
- New work on this issue in IETF would enable to progress with specific recommendations and industry actions
- White paper with the use cases and requirements available here