

The 50/50 Vision for Internet Traffic

Version 1.0, August 2023

Introduction

The 50/50 Vision defines the Internet Society's strategy to rally multi-stakeholder efforts and international and national resources to ensure that at least 50% of all locally generated traffic in selected economies remains local by 2025. Reaching this ambitious target will strengthen Internet quality and reduce access costs for individuals.

A [2012 study](#) established a baseline at Internet Exchange Point (IXP) hubs in Kenya and Nigeria, and a [follow-up study in 2020](#) showed that levels of local traffic jumped from 30% to 70%. The effort helped increase understanding of the impact of peering on the local infrastructure. It was found that the increase in local traffic led to significant cost savings for participating networks and put these two countries in a stronger position to participate in the digital economy. Meanwhile, South Africa localized over 80% of its local traffic, and now enjoys stable, resilient, high quality, and affordable Internet.

The aim of 50-50 vision study is to quantify the extent to which these Over-The-Top (OTT) services fetch their content from a local server/cache rather than externally (out of the country). This document describes the Internet Society's methodology to quantify the level of local traffic at a country-level.

Definitions

- OTT: Over-The-Top services e.g., Netflix, Whatsapp, YouTube, etc
- Local traffic: refers to traffic that stays local and does not leave the country.
- Local content: Internet content which is hosted within the country (e.g., news website, e-government services, etc.) and

- RIPE Atlas: is an Internet measurement framework operated by the RIPE NCC which consists of probes (vantage points) and anchors (targets). The probes can be used to run several measurements (latency, traceroute, DNS, SSL, HTTP) from the edge.
- OONI (Open Observatory of Network Interference): is a platform used for censorship measurement. OONI tests are carried out by users with the OONI app installed on their phone/desktop.
- CDN: Content Delivery Networks are responsible for delivering content to the edge.
- Content cache: is a content hosting equipment placed by a content provider close to the end-users.
- Vantage point: where the measurement is being conducted from.
- Edge network: refers to access network where eyeballs (consumers) are located.

Data on Traffic Volume

Measuring local vs non-local traffic levels for a country is not straightforward. Access providers are usually aware of the most used services and which ones are the biggest consumers of bandwidth within their network. It is often known that some Internet Service Providers (ISPs) perform traffic engineering by throttling the bandwidth to (de)prioritize access to some services based [on consumer demand](#). However, there is no distinction made between local and non-local traffic, as the ISPs simply act as a conduit to the content. Data about the most used services are therefore known to the ISPs, but they are rarely shared publicly.

In some rare cases and based on prior agreement, ISPs provide *s-flow* traffic data (sample flows), from which the destination IP can be extracted. The underlying services (e.g., YouTube, Facebook, Netflix, etc.) can be inferred from the destination IP of TCP packets flowing through the network. However, it would be impractical to perform such a study at a large scale, notwithstanding the fact that operators are generally reluctant to share traffic data to third parties.

In the absence of any ground truth, we shall rely on aggregated reports such as the one provided by Sandvine (released yearly), which provides data about the most popular content by traffic volume on a regional basis.

Table 1: Example of traffic volume by region ([Sandvine 2022](#))

Americas			APAC			EMEA		
Application	Volume	Normalized	Application	Volume	Normalized	Application	Volume	Normalized
Netflix	19.85%	30.53	Facebook	18.80%	24.19	Netflix	16.10%	24.39
YouTube	15.02%	23.10	Facebook Video	17.37%	22.35	YouTube	12.75%	19.32
Facebook	7.32%	11.26	YouTube	13.10%	16.86	Xbox Live update	7.41%	11.23
Generic QUIC	6.50%	10.00	TikTok	12.86%	16.55	Playstation.net	5.69%	8.62
Disney+	3.62%	5.57	Instagram	6.67%	8.58	Disney+	5.37%	8.14
HTTP media stream	3.31%	5.09	Google	2.18%	2.81	Amazon Prime	4.52%	6.85
Amazon Prime	2.71%	4.17	HTTP media stream	1.90%	2.44	HTTP media stream	4.12%	6.24
TikTok	2.67%	4.11	Generic QUIC	1.89%	2.43	Facebook video	3.80%	5.76
Twitch	2.13%	3.28	Snapchat	1.53%	1.97	TikTok	3.35%	5.08
Steam	1.88%	2.89	Netflix	1.41%	1.81	Instagram	2.89%	4.38
Total	61.71%	100%		75.81%	100%		61.88%	100%

For the three regions of the world (Americas, APAC and EMEA), video and social media apps generate most of the traffic. The top 10 apps represent more than 60% of all traffic within these geographic regions. The rest of the traffic can be attributed to different services reachable on the Internet – and without further information, the remaining traffic count will not be considered in this study.

Identifying Location Using Geo-Hints

Most of the content served by the OTT services in Table 1 are delivered by CDNs or content caches placed at the edge. When accessing content on social media platforms or streaming websites, users are usually taken to the closest content cache. This redirection is performed in different ways including HTTP/DNS redirection or using anycast. In case of HTTP redirection, a custom URL is generated and the content is fetched using an HTTP request. Sometimes the URL is geo-tagged (unique based on the destination address) or it is generic, in which case, DNS-based redirection is used.

E.g., Accessing fast.com (Netflix speedtest service) provides a URL with the following domain:
 ipv4-c002-flo001-mauritius-telecom-isp.1.oca.nflxvideo.net

How We Will Determine Content Locality

For each application listed in Table 1, we perform the following steps:

1. Access the service from a local vantage point and gather the HTTP Response.
2. Extract geo-hints and image/object domains.
3. Run additional tests such as ping or traceroute to the domains collected and infer locality.

The above is performed on all major ISPs (ASNs) where measurement probes are available for a given country.

Traffic share formula

In this model, we will make use of three main variables for a country c : (1) the traffic volume per application (2) whether the application content is locally hosted and (3) the market share of the ASN from which the measurement is run. We define the local traffic level for a country c as the following:

$$T_c = \sum_{asn} (M_{c,asn} * \sum_{app} (V_{c,app} * L_{asn,app}))$$

Variable	Description	Data source
c	Target country	-
app	Application	Sandvine
$V_{c,app}$	Normalized Traffic Volume for country c and application app	Sandvine (%)
$M_{c,asn}$	Market share of ASN for country c	APNIC (%)
T_c	Traffic Level of country c	Calculated (%)
$L_{asn,app}$	% of content hosted locally	Calculated (%)

Output

We plan to produce a detailed report on how much traffic is staying local based on the above methodology by the end of 2023. We will also develop a dashboard on our Pulse platform to visualize the data in Q2 2024.