CISPA
HELMHOLTZ CENTER FOR INFORMATION SECURITY

VolkswagenStiftung

A Decade Long View of Internet Traffic Composition in Japan

Irina Tsareva

Trinh Viet Doan

Vaibhav Bajpai

IFIP Networking Conference | 14 June 2023

— paper —

Motivation

*What is the **short-** and **long-lasting** impact of these factors on the Internet traffic composition?*

- Internet traffic composition continues to evolve as a consequence of several factors:
  - increasing awareness of security and privacy on the Internet leading to efforts on encrypted Web and encrypted DNS.
  - design of new protocols such as QUIC & H/3
  - pandemics resulting in online communication and online gaming.
  - increasing influence of hyper-giants shaping the structure and function of the Internet.
- Predicting such Internet trends is crucial for ISPs and CDNs to allow them to better manage their network and services.

Datasets

*We analyse **>6.6 TiB** of data collected at a large backbone link in Japan to investigate traffic composition over the **past decade***

I. MAWI Dataset:
   - collected by WIDE project at a backbone in JP.
   - F (link to upstream provider: NTTC GIN)
   - G (link to an IXP in Tokyo, Google is a peer)
   - 15’ anonymised trace at 14:00 on Wed and Sun.

II. RouteViews Archive
   - mapping IP endpoints encompassing IP prefixes to ASNs.

III. PeeringDB
   - classifying ASNs by type of service.

Findings

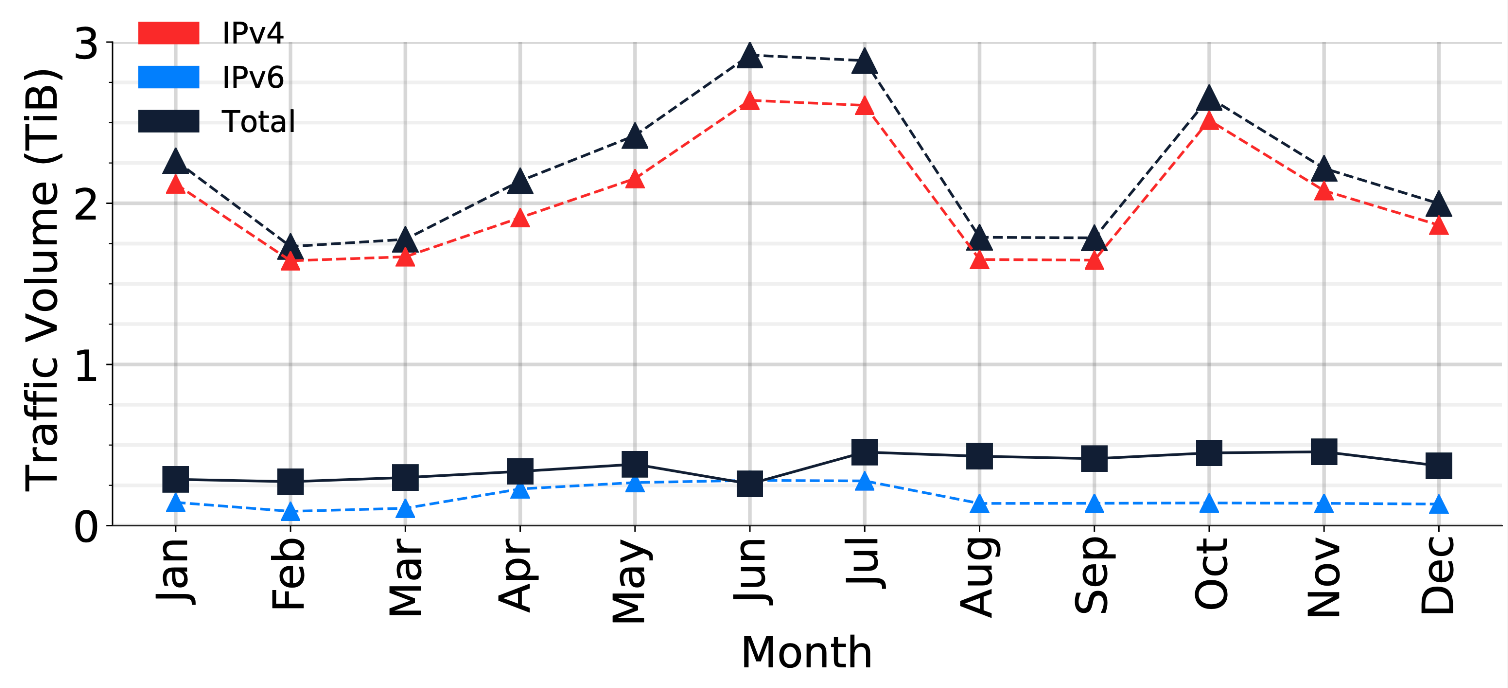
I. Traffic Shares
II. Application Mixes
III. COVID-19 Pandemic
IV. Impact of Peering with Google

- Substantial change in volume and usage of IPv6 over the last decade. On the influence of QUIC on UDP shares.
- How the HTTPs-to-HTTP ratio has evolved in favour of the encrypted Web, and the development of DNS beyond UDP.
- Impact of COVID-19 on weekly traffic patterns, changing application mixes.
- Impact on Google on traffic volumes, on QUIC and on IPv6 traffic distribution.

Traffic Shares

How did the Internet traffic composition evolve over the last (2007→2019) decade?

- Decrease in traffic during academic breaks, with weekday-weekend patterns during the semester.
- Average monthly traffic has **increased by >480%** in the last decade.
- IPv6 traffic has increased by 18x, is now **comparable to overall traffic** 10 years ago.



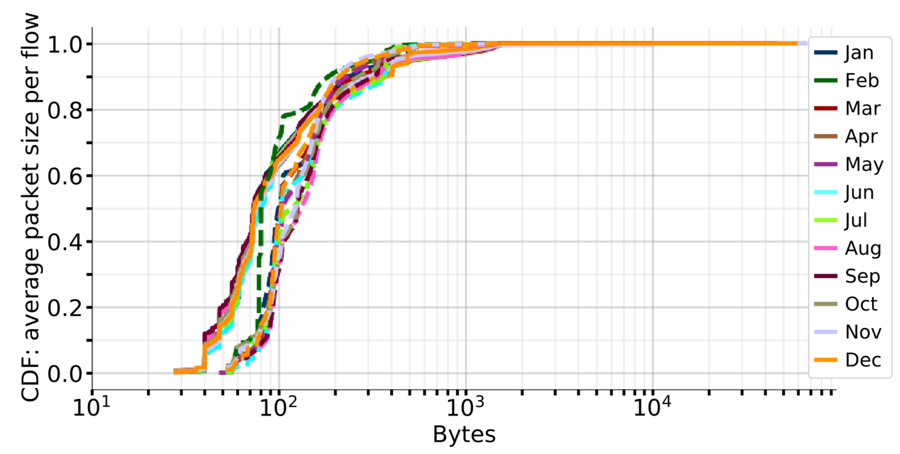
The annual traffic has increased significantly; IPv6 traffic in 2019 is now similar to the total traffic in 2007.

Traffic Shares

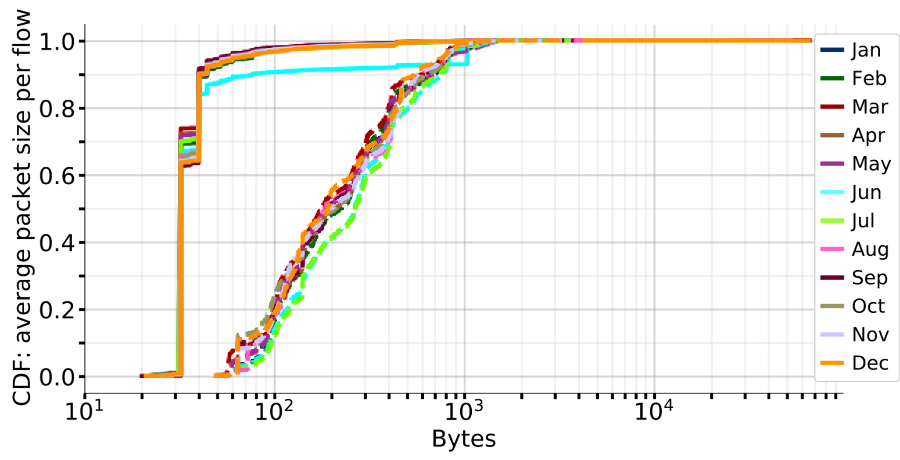
How did the Internet traffic composition evolve over the last (2007→2019) decade?

- **IPv4** traffic has shifted from larger average packet sizes per flow **to smaller average packet sizes** per flow 10 years later.
- The average packet size per flow in **IPv6** have **increased by >68%** in last 10 years.
- The relative **UDP** byte traffic share has grown for every month by **up to 4 times** in the last 10 years.

IPv6 flow sizes have increased, due to increased Web traffic. QUIC causes UDP traffic to increase by 4 times in last 10 years.



(a) *2007*



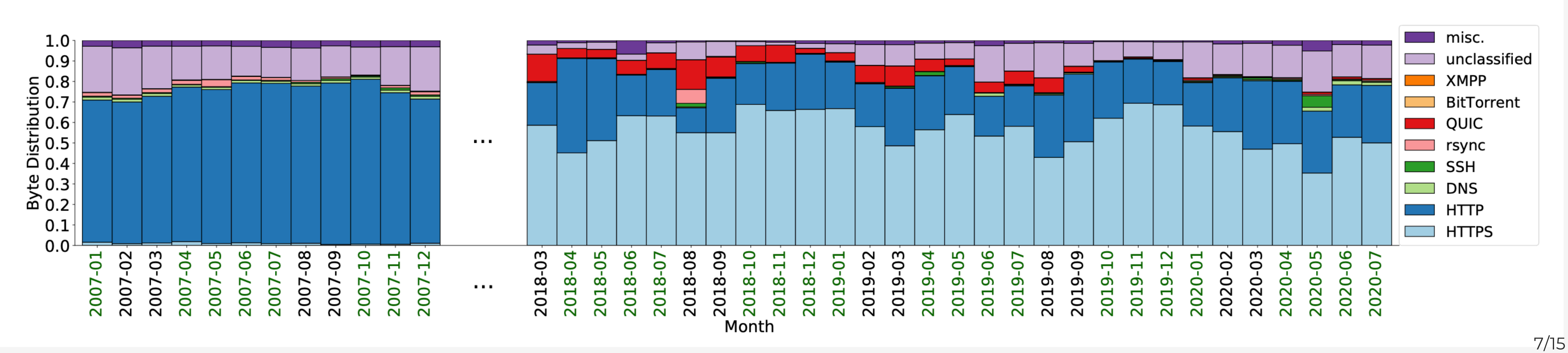
(b) *2019*

Application Mixes — IPv4

How did the application mix evolve over the last (2007→2019) decade?

- The share of HTTP Web traffic over IPv4 used to be >70%. 10 years later, HTTP-to-HTTPS ratio has increased to roughly **1-to-2.**
- The HTTPS share tends to increase during the academic term.
- There is an **increase in DNS traffic** share (0.2% → 1%)

Over a decade, a significant shift towards an encrypted Web is visible.



Application Mixes — IPv6

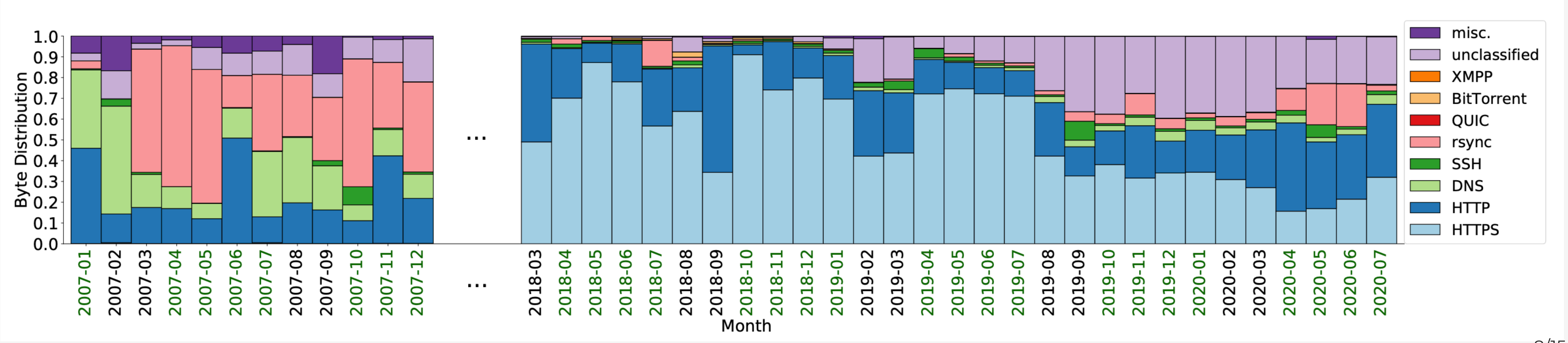
How did the application mix evolve over the last (2007→2019) decade?

- HTTP (22%), DNS (20%), and rsync (36%) together used to have a share of >70%. One decade later, **IPv6 carries majorly Web traffic** (together >60%).

- The HTTPS share increases significantly during the academic term.

- There is a trend towards **more rsync traffic** (<1% → 9%)

A decade later, the application mix over IPv6 resembles more closely with IPv4.

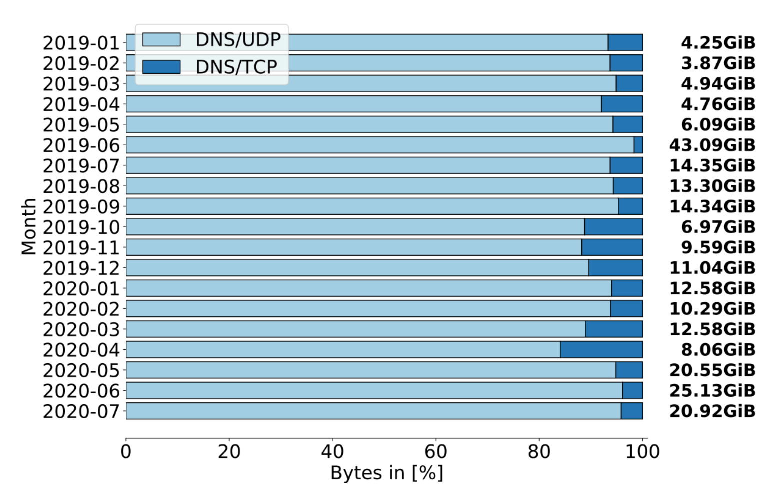


Application Mixes — DNS

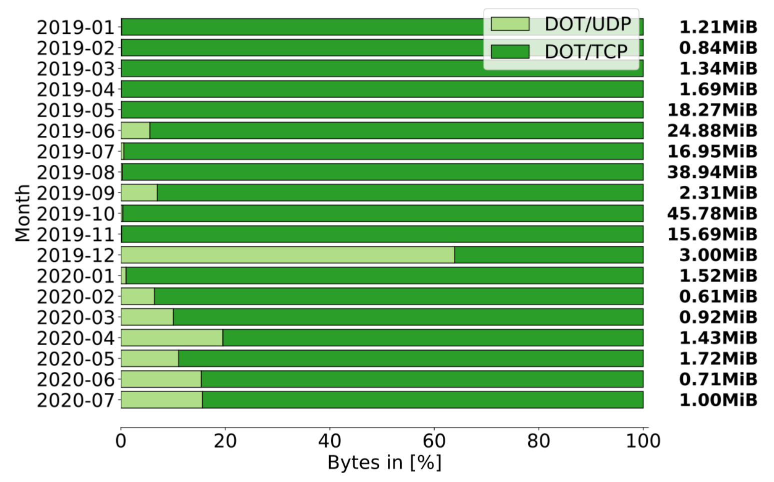
How did the application mix evolve over the last (2007→2019) decade?

- DNS traffic **(>90%)** is dominated by DNS/UDP. DNS/TCP share is roughly 6-7%.
- **DNS over TCP** traffic is increasing. DNS over TLS share is negligible.
- We observe that TLS 1.2 (+1.3) are now dominant TLS protocols for **DNS over TLS.** Most DNS over TLS traffic originates from or is sent to **Quad9 and Cloudflare servers.**

Over a decade, DNS traffic remains majorly unencrypted.



(a) *DNS (w/o DoT) traffic composition.*

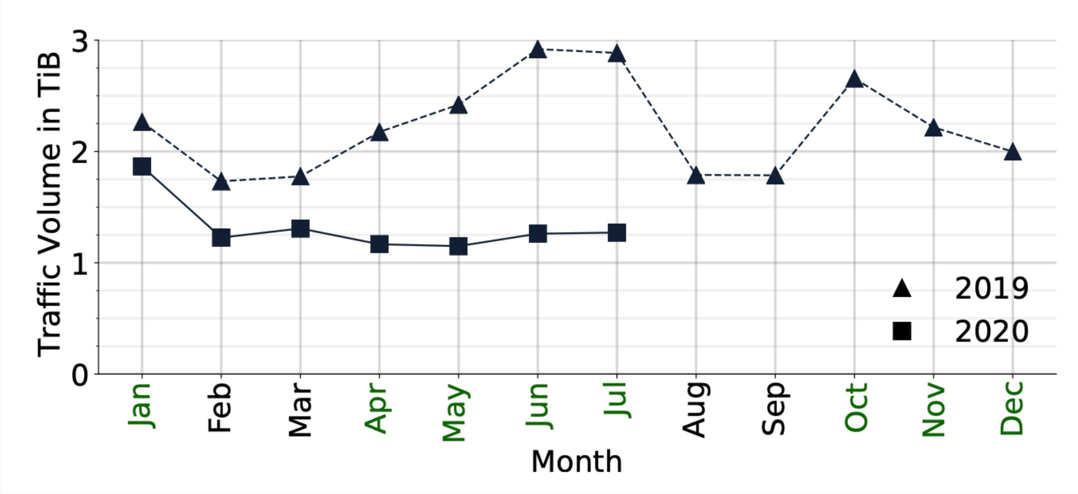


(b) *DoT traffic composition.*

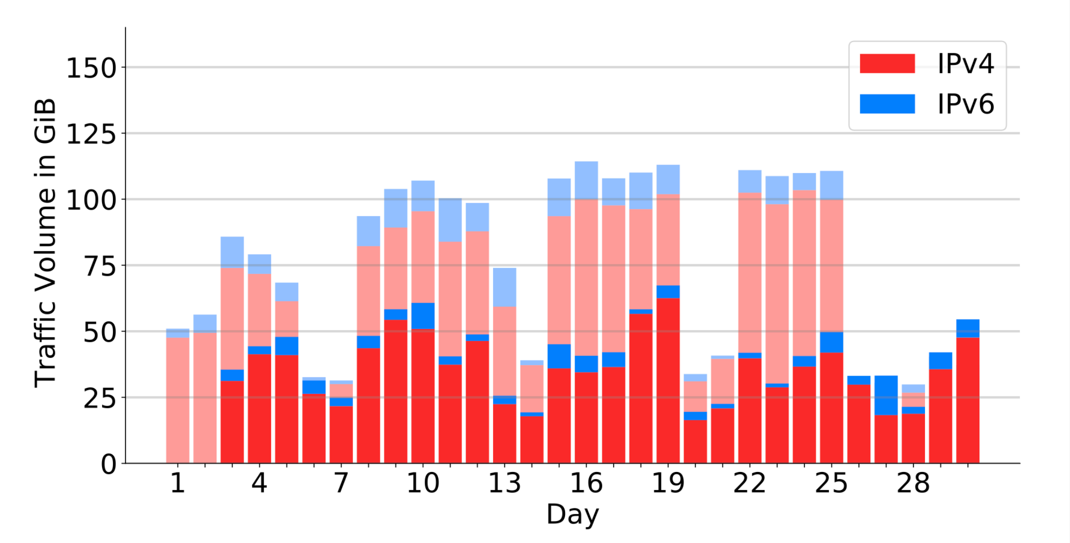
COVID-19 Pandemic

How did the Covid-19 pandemic (2020–) impact the Internet?

- During the lockdown, the **traffic volume** during spring term 2020 is even lower than during the summer break 2019.



- The **weekday- weekend pattern** **disappears** and the daily traffic volume stays as low as on weekends, even after the restriction were lifted.



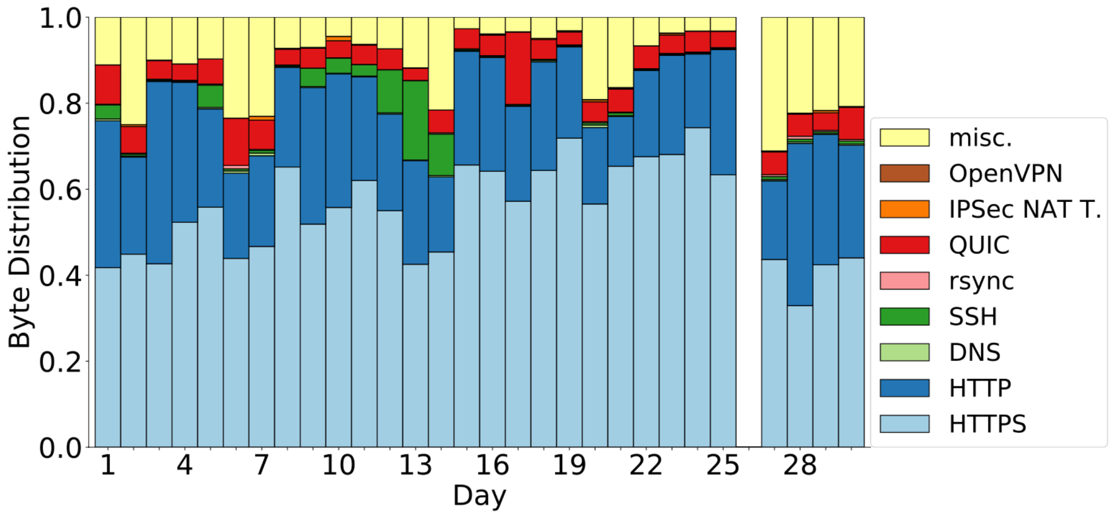
Covid-19 significantly alters traffic composition with decrease in traffic volume and a disappearance of the characteristic weekend-weekday traffic pattern.

COVID-19 Pandemic

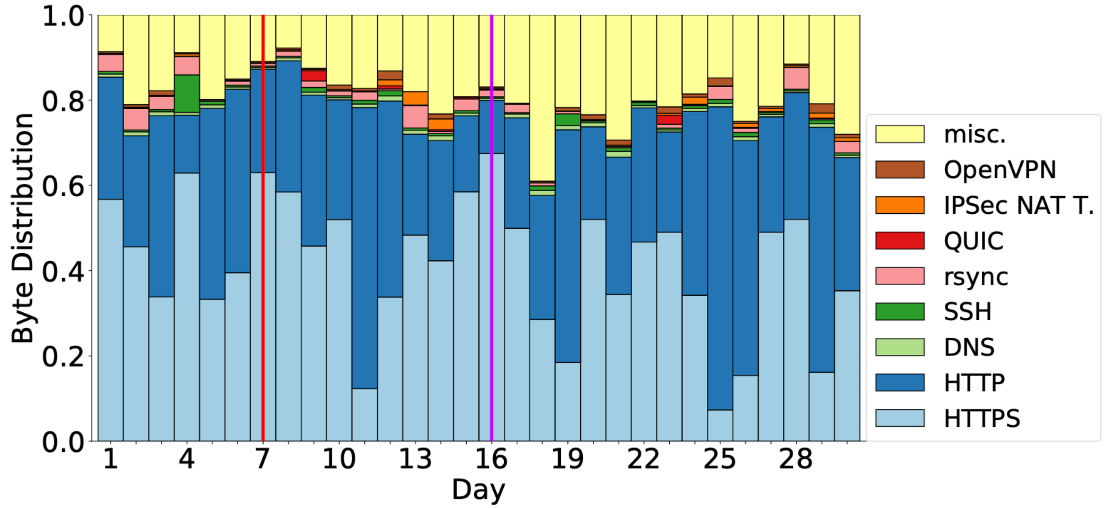
How did the Covid-19 pandemic (2020–) impact the Internet?

- The daily share of **OpenVPN** increases on average by a **factor of 67.2.**
- The share of **rsync** increases for every month; >28% of the daily traffic. There is a minor increase of IPSec NAT Traversal as well.
- The daily traffic volume of **BitTorrent** increases, on average by **>5.6 times.**

OpenVPN and remote shell protocols are more frequently used during COVID-19.



(a) *April 2019*



(b) *April 2020*

COVID-19 Pandemic

How did the Covid-19 pandemic (2020–) impact the Internet?

- The traffic volume to top 10 destination ASes reduced by **>30%**
- Outbound traffic drops by **>52%** compared to inbound traffic.
- Traffic to **Dropbox** increases, while towards **Apple** decreases by a **factor of 9.**

April 2019			April 2020		
Destination ASes		Bytes[GiB]	Destination ASes		Bytes[GiB]
AS4538	ERX-CERNET-BKB	119.91	AS17676	GIGAINFRA	121.40
AS4134	CHINANET-BACKBONE	88.86	AS4766	KIXS-AS-KR	78.48
AS2907	SINET-AS	88.21	AS1659	ERX-TANET-ASN1	55.50
AS2500	WIDE-BB	56.25	AS17816	CHINA169-GZ	37.34
AS5609	ASN-CSELT	52.98	AS17512	JAL	35.14
AS9462	BOLEH-NET-AP	52.42	AS8803	MIGROS	34.78
AS4637	ASN-TELSTRA-GLOBAL	50.64	AS4782	GSNET	32.34
AS17676	GIGAINFRA	49.20	AS2830	MCI-DUAL-HOMED-CUSTOMERS	30.38
AS9667	HOSTWORKS-AS-AP	44.35	AS4837	CHINA169-BACKBONE	28.91
AS55552	NETWORK-BOX-HK	44.27	AS2500	WIDE-BB	27.35

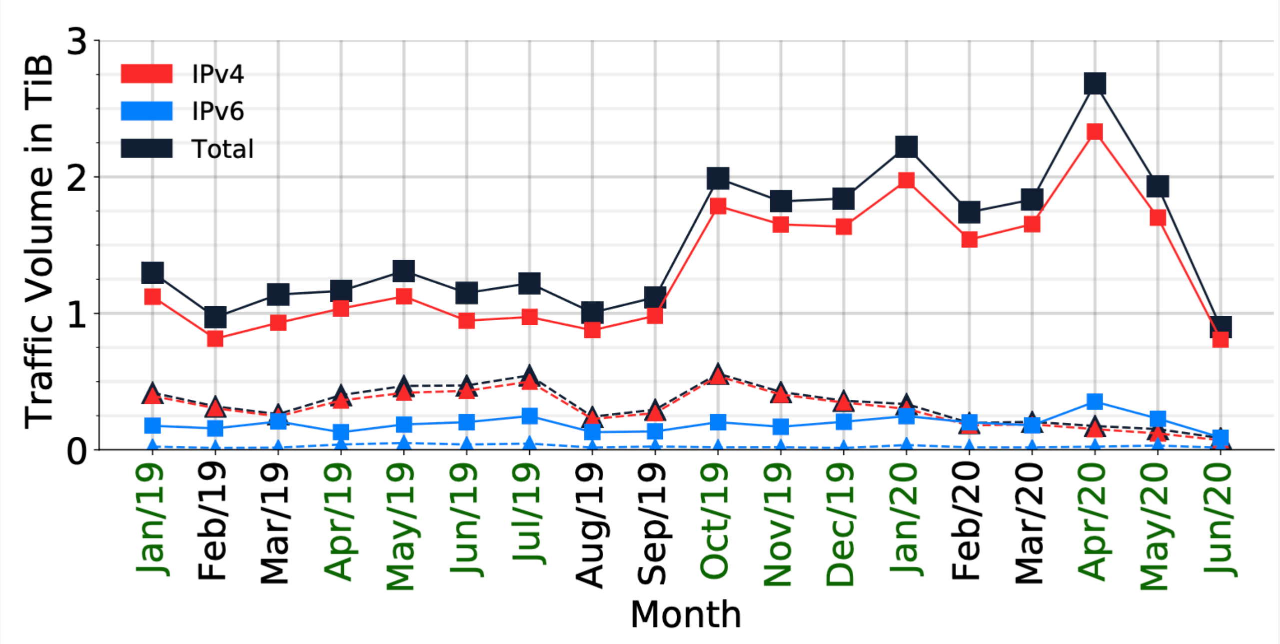
April 2019			April 2020		
Source ASes		Bytes[GiB]	Source ASes		Bytes[GiB]
AS714	APPLE-ENGINEERING	210.54	AS714	APPLE-ENGINEERING	75.94
AS701	UUNET	75.46	AS6319	MARRIOT-ASN	39.32
AS7018	ATT-INTERNET4	52.51	AS3462	HINET	37.29
AS16625	AKAMAI-AS	40.80	AS16625	AKAMAI-AS	30.01
AS10796	TWC-10796-MIDWEST	33.82	AS4662	QTCN-ASN1 GCNet	28.75
AS3320	DTAG	33.77	AS21928	T-MOBILE-AS21928	23.97
AS4134	CHINANET-BACKBONE	31.05	AS38676	FLEXNET-AS-KR	23.51
AS16509	AMAZON-02	29.82	AS4782	GSNET	21.84
AS20940	AKAMAI-ASN1	27.11	AS19679	DROPBOX	20.69
AS2500	WIDE-BB	27.02	AS16509	AMAZON-02	19.87

There is a shift towards file hosting services such as Dropbox, and Apple services (iTunes) are less used from the educational backbone.

Impact of Peering with Google

How and to what extent does peering with Google change the traffic composition?

- Link that peers with Google (-G) contains on average >5 times **more packets** and >10 times **more bytes**, but roughly 6 times **less flows** than the traffic collected on another link (-F)



- The UDP traffic at -G is mostly composed of **QUIC.**

- The absolute share of **IPv6** at -G is **7 times larger** >8 times than on -F.

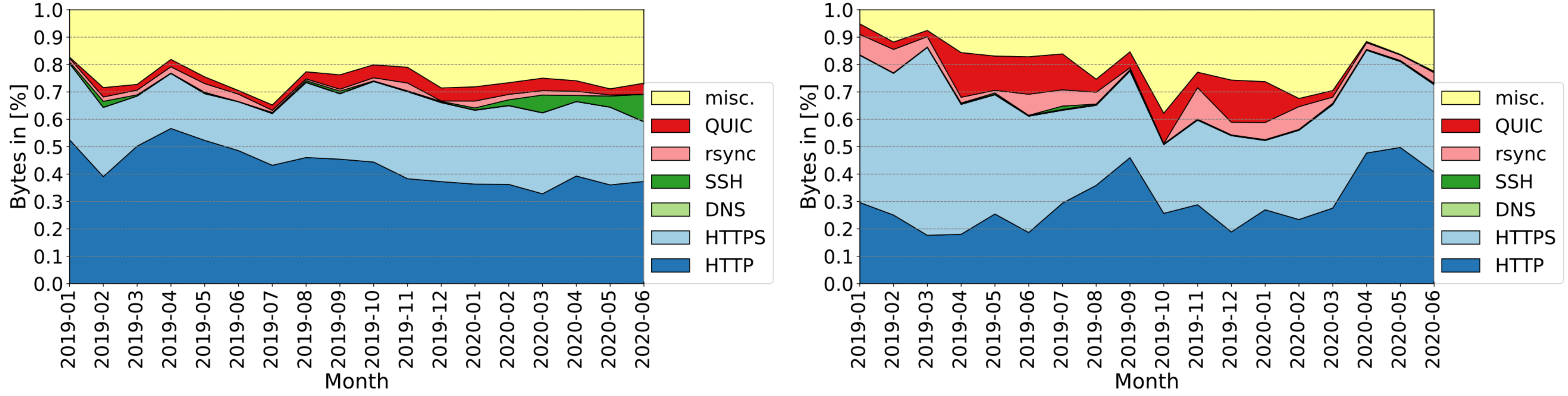
Peering with Google increases QUIC traffic; the relative share over IPv6 is even larger than over IPv4.

Impact of Peering with Google

How and to what extent does peering with Google change the traffic composition?

- The HTTP share is slightly larger than HTTPS over IPv4 (9-to-5) while **HTTPs share is larger over IPv6** (9-to-10).
- QUIC share is more than **twice as large,** rsync is 4 times as large as over IPv4.

When peering with Google, HTTPs, QUIC and rsync share is larger over IPv6 than over IPv4.



(c) *IPv4 traffic.* (d) *IPv6 traffic.*

see paper for comparison with -F

Takeaways

A Decade Long View of Internet Traffic Composition in Japan

Irina Tsareva*, Trinh Viet Doan*, Vaibhav Bajpai†
*Technical University of Munich, Germany
[irina.tsareva | trinhviet.doan]@tum.de
†CISPA Helmholtz Center for Information Security, Germany
vaibhav.bajpai@cispa.de

I. Traffic Shares
II. Application Mixes
III. COVID-19 Pandemic
IV. Impact of Peering with Google

— Paper —

— Slides —

www.vaibhavbajpai.com
bajpai@cispa.de | @bajpaivaibhav

I. IPv6 traffic in 2019 is comparable to that of the total traffic in 2007 and is (now) increasingly used to carry Web traffic.

II. Over IPv4, there significant growth in encrypted Web traffic with HTTPS-to-HTTP ratio evolving to 2-to-1 in 2019 compared to less than 2% HTTPS share in 2007.

III. COVID-19 led to a vanishing weekday-weekend pattern and a shift towards increased usage of OpenVPN and rsync traffic.

IV. Peering with Google significantly increases traffic share of QUIC over both address families, and overall a larger HTTP-to- HTTPS ratio.