The "Public" Internet: A Network of Last Resort?

Walter Willinger
NIKSUN, Inc.
wwillinger@niksun.com

TU Berlin

May 16, 2019

Outline

- Plenty of peerings: IXP study from 2012
- Traffic over peerings: A recent data point (Facebook 2017)
- Traffic over peerings: A more recent effort (Akamai 2018)
- A new twist: The large cloud providers

A Look Back: IXP Study 2012

Plenty of Peerings!

- Anja Feldmann's group at TU Berlin/T-Labs obtained high-quality traffic data from on of the largest IXPs in Europe
- *B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, W. Willinger Anatomy of a Large European IXP, ACM Sigcomm 2012*

- This IXP had some 400 active member ASes and handled some10-20 PB traffic on a daily basis (public info)
- Using proprietary IXP data, we showed that at this IXP alone, there were more than 50,000 interconnections (public peerings) that carried actual traffic
- Most of this reported connectivity was not visible in the publicly available BGP and traceroute data

Remark: "Using proprietary IXP data …"

- The end of reproducible networking research!?
- The beginning of reproducible networking research!
  - Use of proprietary data for discovery
  - Knowing what to look for, is it possible to obtain the same findings using publicly available data?
- The IXP work as a textbook example
  - Discovery via proprietary data: our SIGCOMM'12 paper
  - Same with public data: CoNEXT'13 paper by Giotsas et al.

Traffic: A Question and Guess

- Some members of this IXP where we found some 50,000 interconnections (public peerings) also engaged in private peering (buying Private Network Interconnections or PNIs or "cross connections" from the colocation facilities where the IXP operates)

- A intriguing question and an informed guess (Randy Bush)
  - Randy's question: *Of all the traffic that is exchanged in these colocation facilities that house this IXP, what portion traverses public peerings vs private peerings?*
  - Randy's guess: *Probably less the 10% goes over the large number of public peerings, and more then 90% goes over the much smaller number of private peerings!*

Traffic over Peerings: The Problem

- The IXP doesn't have all the information
  - Knowledge of member ASes, RS participants, traffic, …
  - No visibility beyond its own switching fabric (e.g., no information on colocation facility-specific data)
- The colocation facilities don't have all information
  - Knowledge about their tenants and the number and type of interconnection services they have purchased
  - No knowledge about how these interconnections are used (e.g., traffic)

Traffic over Peerings: A Solution(?)

- Third-party/academic researchers are doomed …!
  - Lacking access to adequate vantage points
  - Lacking adequate measurement tools and inference techniques
  - Lacking relevant data
- Are third-party/academic researchers doomed …?
  - Which providers have a global-scale peering fabric?
    - Google, Facebook, Akamai, Microsoft, AWS, …
  - Which providers collect relevant (proprietary) data?
  - A good problem space for collaborations …

A recent Data Point

Facebook Study 2017

- *B. Schlinker, H. Kim, T. Cui, E. Katz-Bassett, H. Madhyastha, I. Cunha, J. Quinn, S. Hasan, P. Lapukhov, H. Zeng. Engineering Egress with Edge Fabric: Steering Oceans of Content to the World, ACM Sigcomm 2017*

- Describes the design of Edge Fabric, an SDN-based system that enables Facebook to best utilize its peering fabric to serve its 2B+ end users
- Does not provide a detailed description and/or quantification of Facebook's existing peering fabric

Facebook Study 2017

- However, the paper uses a subset of 20 of Facebook's PoPs to illustrate selected aspects of traffic over peerings

ACM SIGCOMM, August 21–25, 2017, Los Angeles, CA, USA

PoP ID	1 (EU)		2 (AS)		11 (EU)		16 (AS)		19 (NA)	
	Peers	Traffic	Pe.	Tr.	Pe.	Tr.	Pe.	Tr.	Pe.	Tr.
Private	.12	.59	.25	.87	.02	.24	.21	.78	.13	.73
Public	.77	.23	.39	.04	.45	.45	.54	.13	.85	.07
Rt Srvr	.10	—	.34	—	.52	—	.23	—	0	—
Transit	.01	.18	.01	.10	.01	.31	.02	.08	.01	.20

Table 1: Fraction of peers and of traffic to peers of various types at example PoPs in EUrope, ASia, and North America. A peer with both a private and a public connection will count in both. A peer with a public and a route server connection counts as public (they share an IXP port). Traffic to public and route server peers is combined.

- An empirical confirmation of Randy's informed guess …!?
- Motivation for more recent effort …

A more recent Effort: Akamai Study 2018

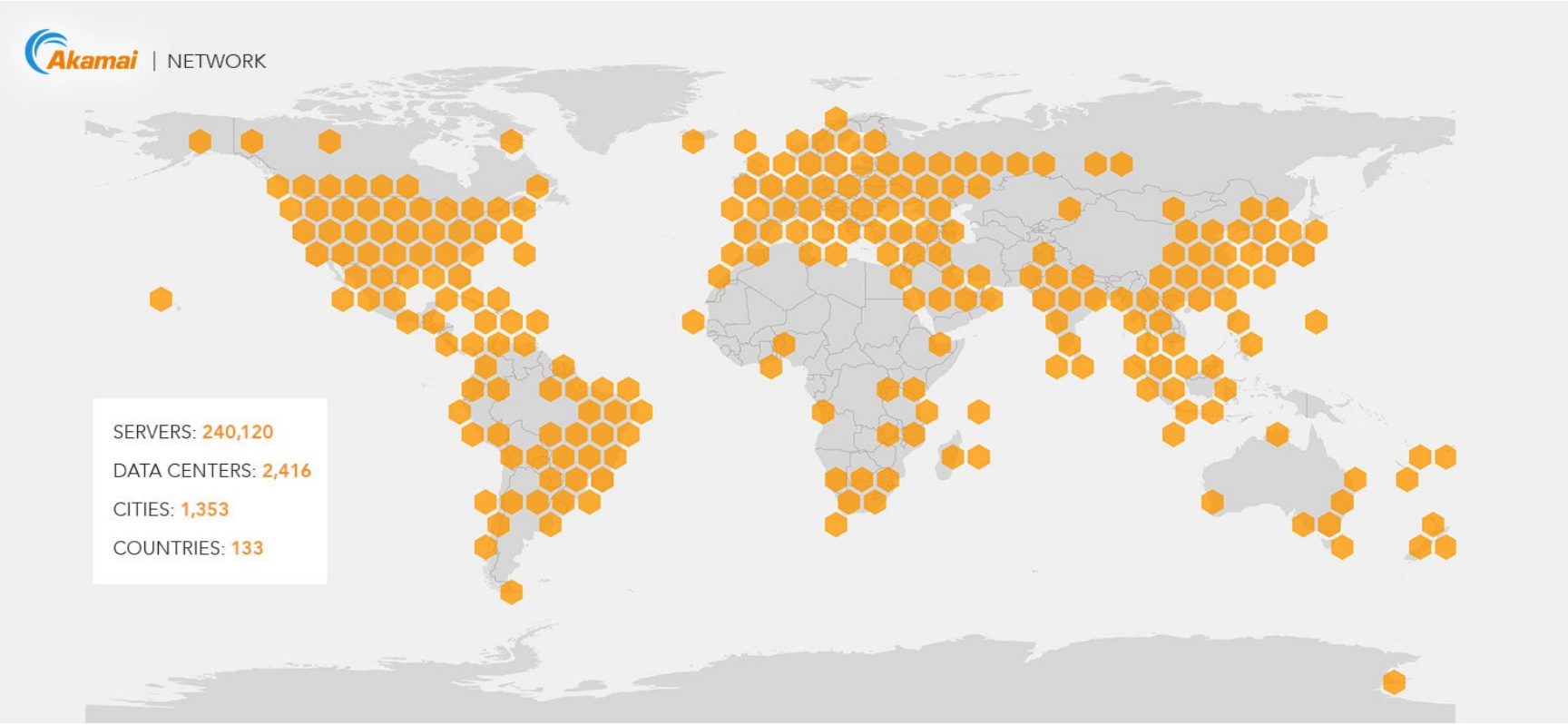
The Serving Infrastructure of Akamai

- *F. Wohlfart, N. Chatzis, C. Dabanoglu, G. Carle, W. Willinger. Leveraging Interconnections for Performance: The Serving Infrastructure of a Large CDN, ACM Sigcomm 2018.*

- What does Akamai's serving infrastructure look like?
  - Footprint (e.g., server clusters or deployments)
  - Peering fabric (i.e., all peerings utilized for serving content)
- How does Akamai leverage its serving infrastructure?
  - What type of peerings carry most of the traffic?
  - Provide an Akamai-specific answer to Randy's question …
- We take Akamai's Mapping System (Sigcomm'13) as a given

About Akamai's Footprint

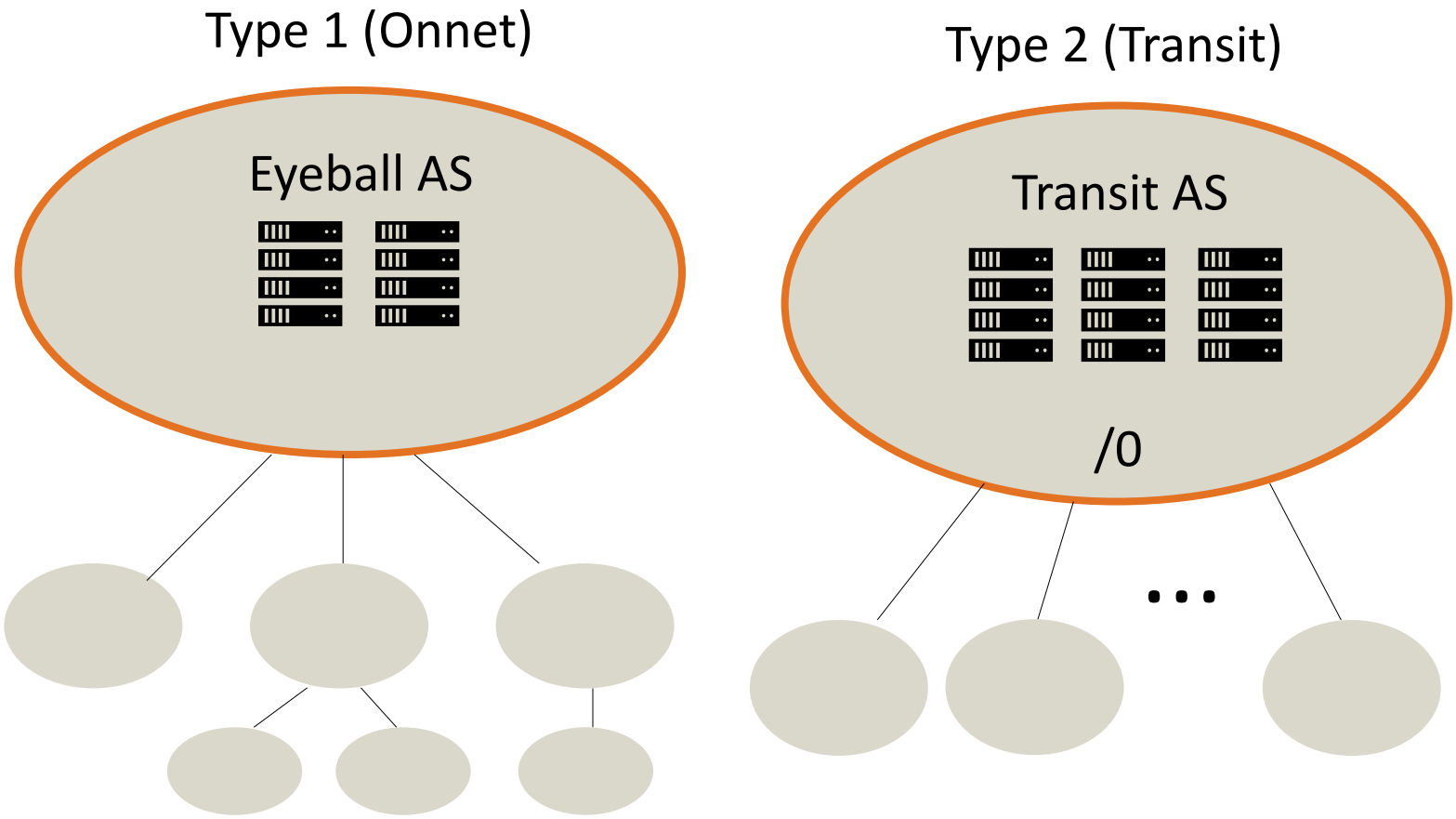
Akamai's Footprint
Generic high-level Description



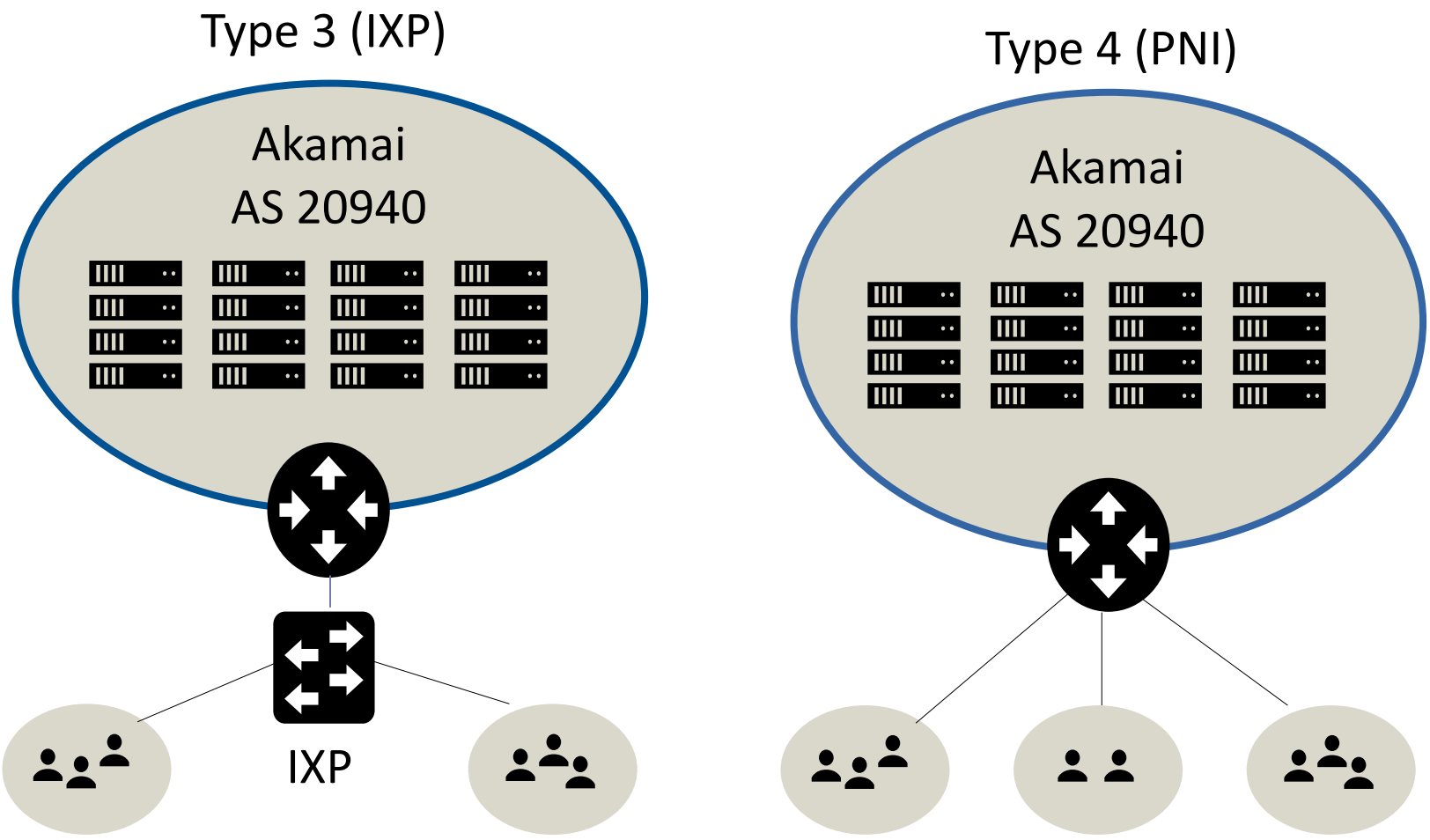
Akamai's Footprint
More detailed Description

- The 200,000+ servers are organized into server-clusters
  - For a clickable map of server clusters/locations, see https://www.akamai.com/us/en/solutions/intelligent-platform/visualizing-akamai/media-delivery-map.jsp
- Server clusters with end user-facing servers are located in some 3,300 different deployments
- These more than 3,300 deployments are located within more than1,700 networks in more than 130 countries around the world

Akamai's Footprint: Deployment Locations: Host Networks



Akamai's Footprint: Deployment Locations: Colocations



Akamai's Footprint: Deployment Characteristics

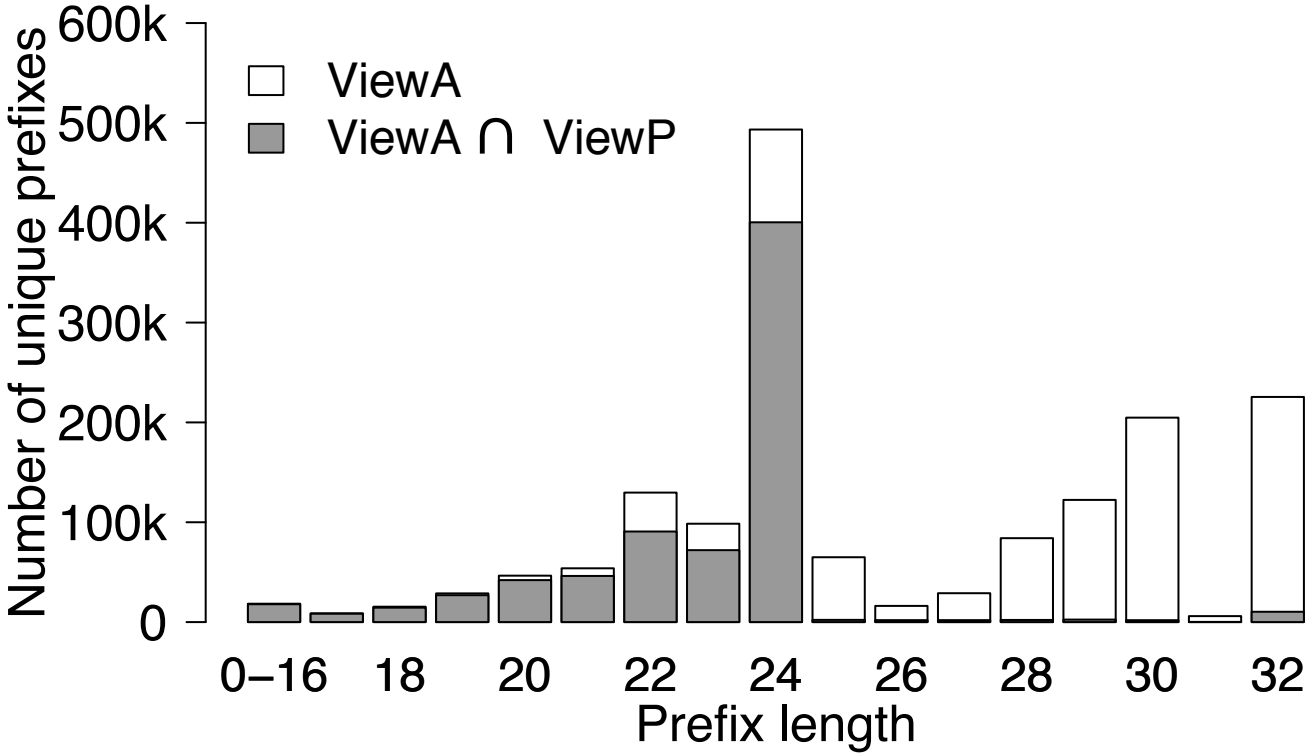
	Deployments w/o Akamai router		Deployments with Akamai router	
	Type 1 (Onnet)	**Type 2 (Transit)**	**Type 3 (IXP)**	**Type 4 (PNI)**
IP space	Host network	Host Network	Akamai	Akamai
ASN	Host Network	Host Network	Akamai	Akamai
Typical Size	Small/Medium	Medium/Large	Large	Large
Can serve all end users	No	Yes	Yes	Yes
Target networks	Host network/ Downstreams	All	Small/Medium	Medium/Large
Proximity to end users	High/Very High	Low/Medium	Medium/High	High
Typical setting	Eyeball Network	Transit Network	IXP	PNIs with Eyeball Networks

About Akamai's Peering Fabric

Akamai' Peering Fabric
Use of Proprietary BGP Information

- BGP information from all of the 80 Akamai BGP collectors
  - Some 4,500 BGP sessions between these collectors and the deployments with Akamai or non-Akamai routers
  - Only routing table info (i.e., "best" paths) not BGP table info (i.e., all paths) is "used"
- A few hourly snapshots from Sept 2017 – May 2018
  - Each snapshot consists of about 1.85M IPv4 and IPv6 prefixes from some 61,000 different Ases
  - Each snapshot consists of about 3.65M AS paths that are used to serve these prefixes

Akamai's Peering Fabric: The Value of Proprietary Control Plane Data

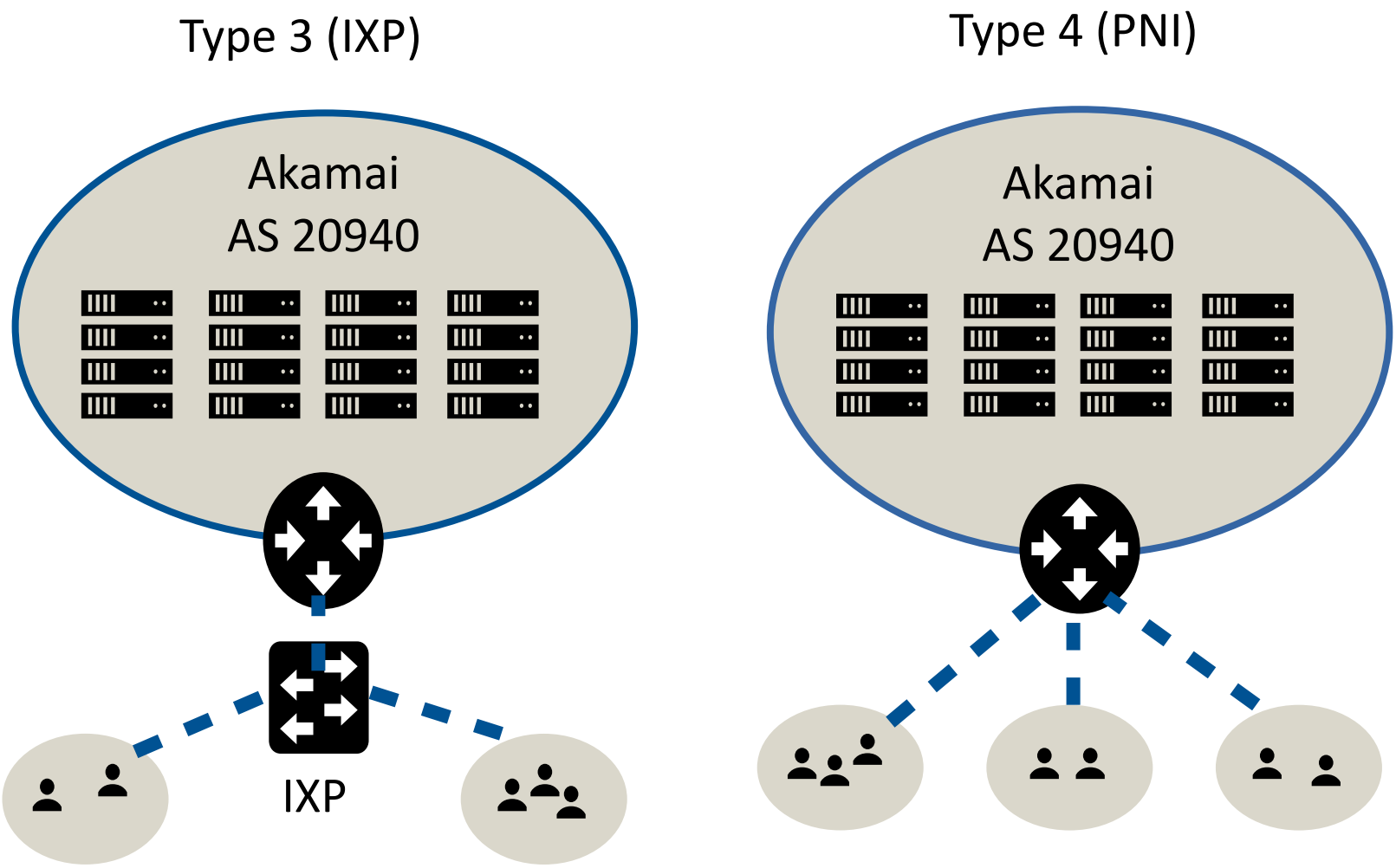


Akamai's Peering Fabric: Explicit and Implicit Peerings

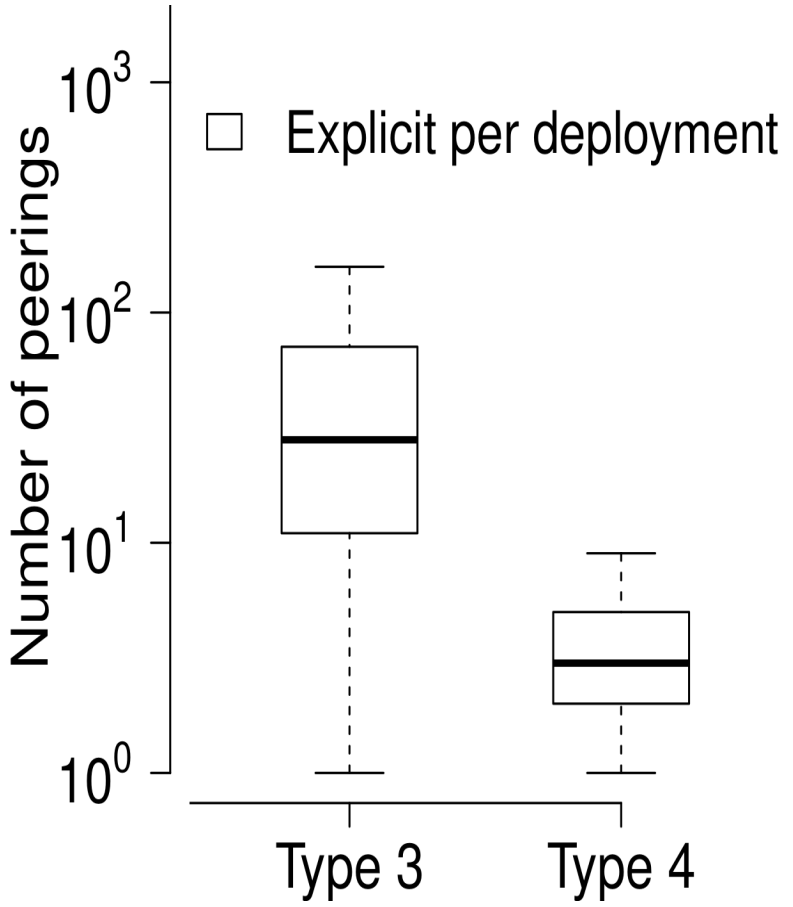
- What peerings is Akamai using to deliver content from its EUF delivery clusters to the prefixes that requested that content?

- Explicit peerings
  - Akamai (ASN 20940) is one of the two peering partners
  - Exist only in Akamai's Type 3 and Type 4 deployments
  - Can be identified using Akamai's BGP data
- Implicit peerings
  - Neither of the two peering partners is Akamai (ASN 20940)
  - Exist only in Akamai's Type 1 and Type 2 deployments
  - "Inherited" by Akamai from the hosting AS
  - Cannot be identified using public BGP data

Explict Peerings: Deployments with Akamai Router

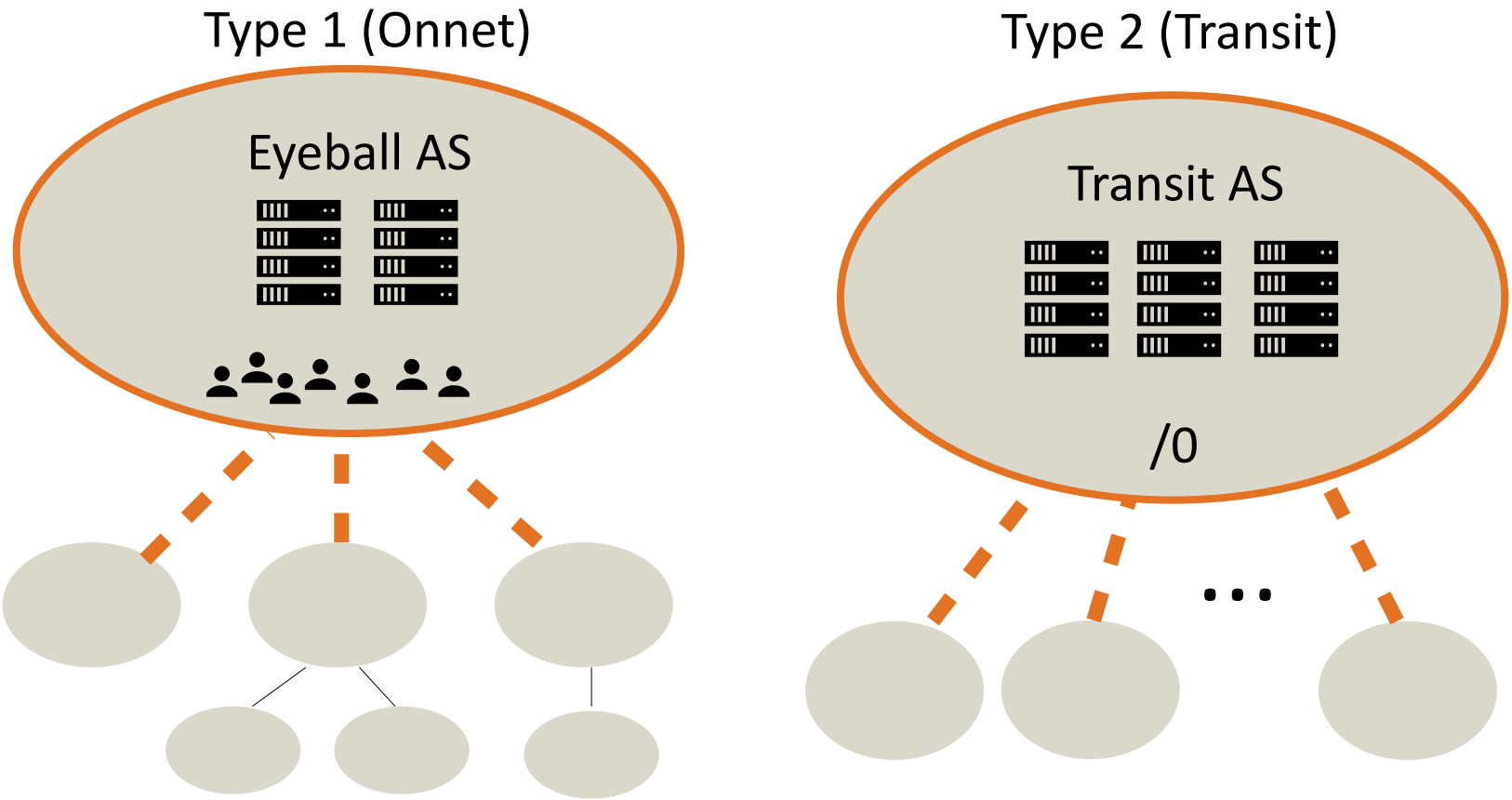


Explicit Peerings by the Numbers …

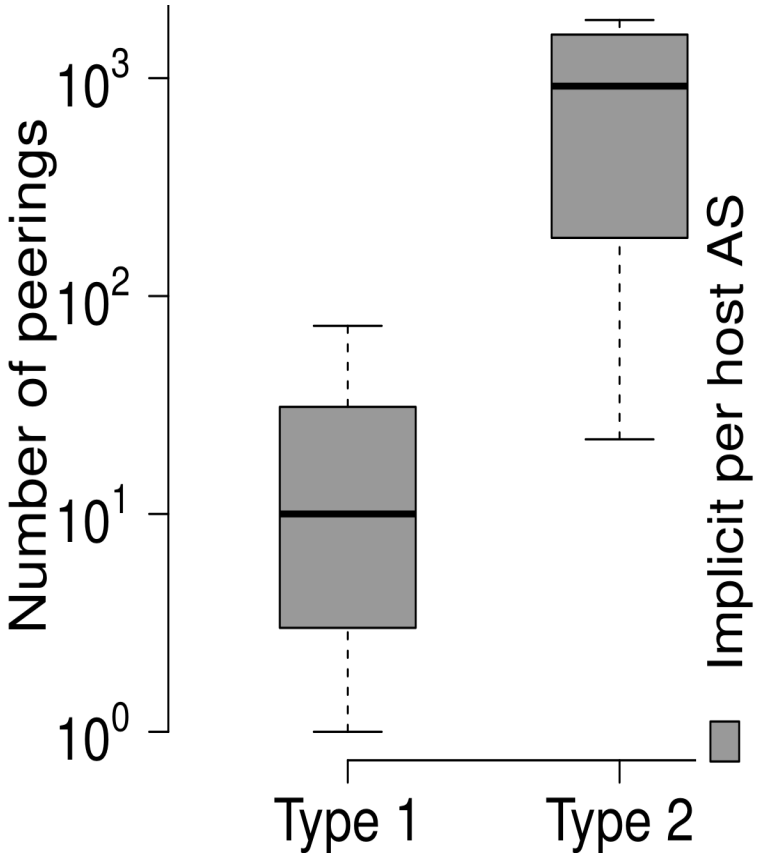


- Number of explicit peerings (IPv4 and IPv6)
  - Type 3: 6,075
  - Type 4: 227
- Total number of (unique) explicit peerings
  - 6,111
  - 50 at 25 or more deployments
  - 859 at 5 or more deployments

Implicit Peerings: Deployments without Akamai Router



Implicit Peerings by the Numbers …



- Number of implicit peerings (IPv4 and IPv6)
  - Type 1: 26,429
  - Type 2: 7,322
- Total number of (unique) implicit peerings
  - 28,353
  - Some 10 per hosting eyeball AS
  - Some 1,000 per hosting transit AS
  - # Type 1 deployments >> # Type 2 deployments

Akamai's Serving Infrastructure: Summary

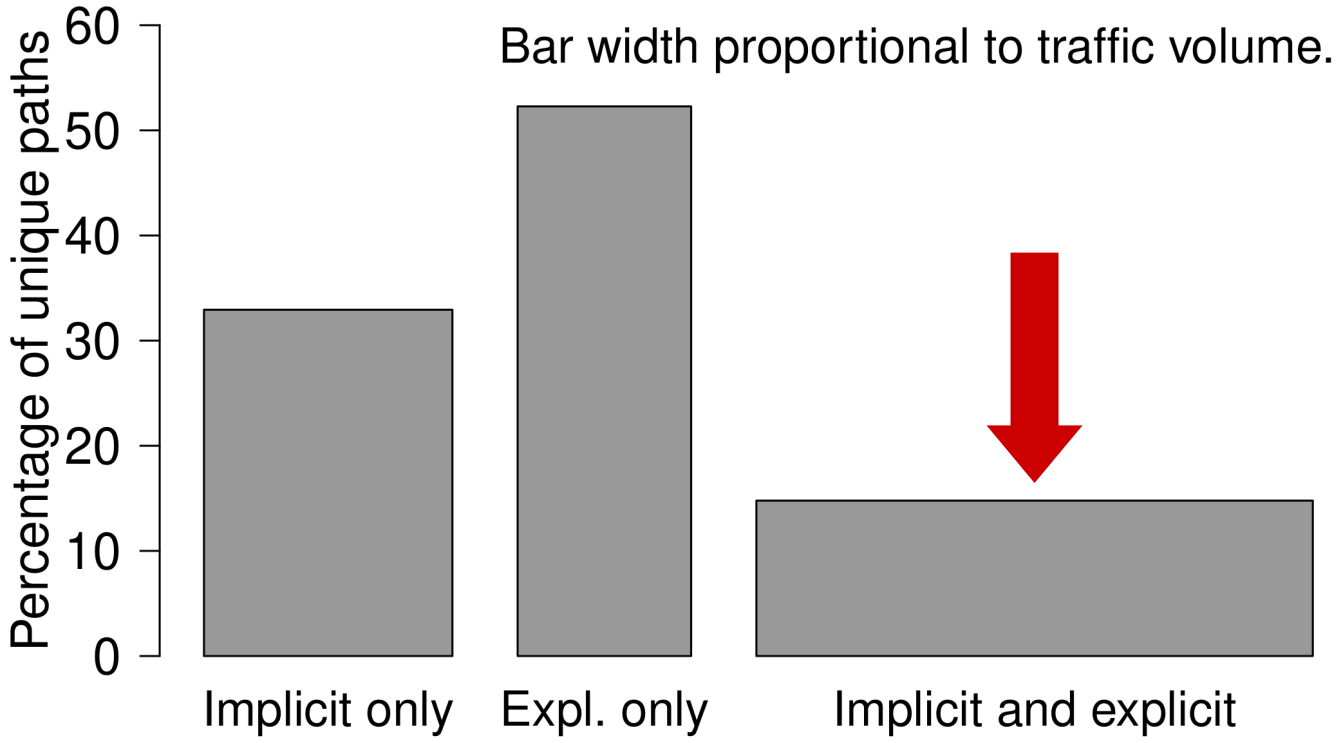
- Footprint
  - Some 3,300 deployments with EUF delivery clusters
  - More than 80% of the deployments are of Type 1-4
  - Deployments w/o EUF delivery clusters exist (e.g., BGP collectors)
  - The 3,300 deployments serve some 1.75M (100K) unique IPv4 (IPv6) originating prefixes in some 61,000 different Ases
  - These prefixes can be served via some 3M unique AS paths (prefixes /25 or longer have typically only a single AS path)
- Peering Fabric
  - Some 6,000 explicit peerings and more 28,000 implicit peerings
  - None of the implicit peerings can be recognized in public BGP data
  - Some of the explicit peerings are visible in the public BGP data

Akamai's use of its Serving Infrastructure

Traffic over Peerings (1): Use of Proprietary Traffic Information

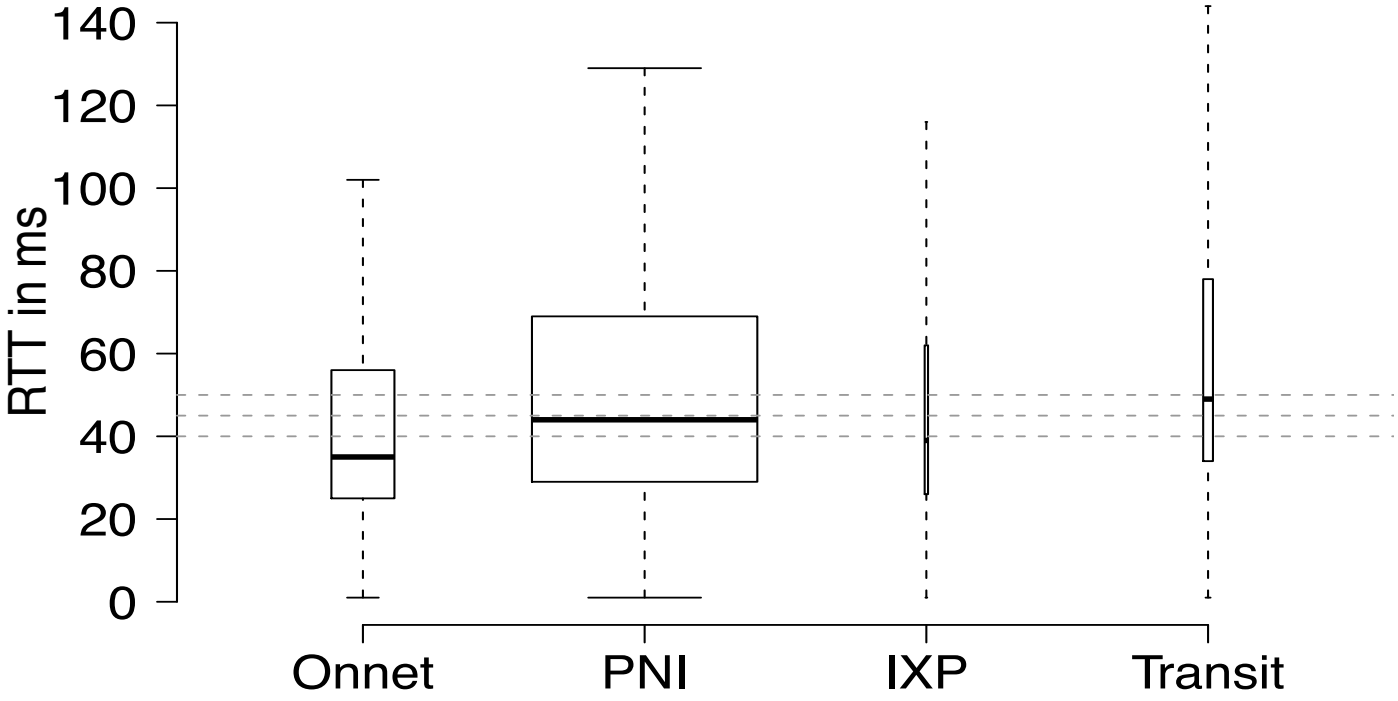
- Server logs of the HTTP/S sessions between EUF delivery clusters and request-generating clients
  - Sampling rate of 5% (1-in-20 sessions)
  - Each record includes
    - Client IP address
    - Server IP address
    - Total number of bytes sent to client
    - Transfer time
- A couple of days' worth of logs, some 11B records per day
  - Consider only objects large than 300K bytes
  - Consider two metrics
    - Mean throughput of session
    - Smoothed RTT value

Traffic over Peerings (2): The Value of Proprietary Data Plane Data



Bar width proportional to traffic volume

Traffic over Peerings (3): Case: Serving a Country with Multiple ISPs



Bar width proportional to traffic volume

Traffic over Peerings (4): Case: Serving Different Metros around the World

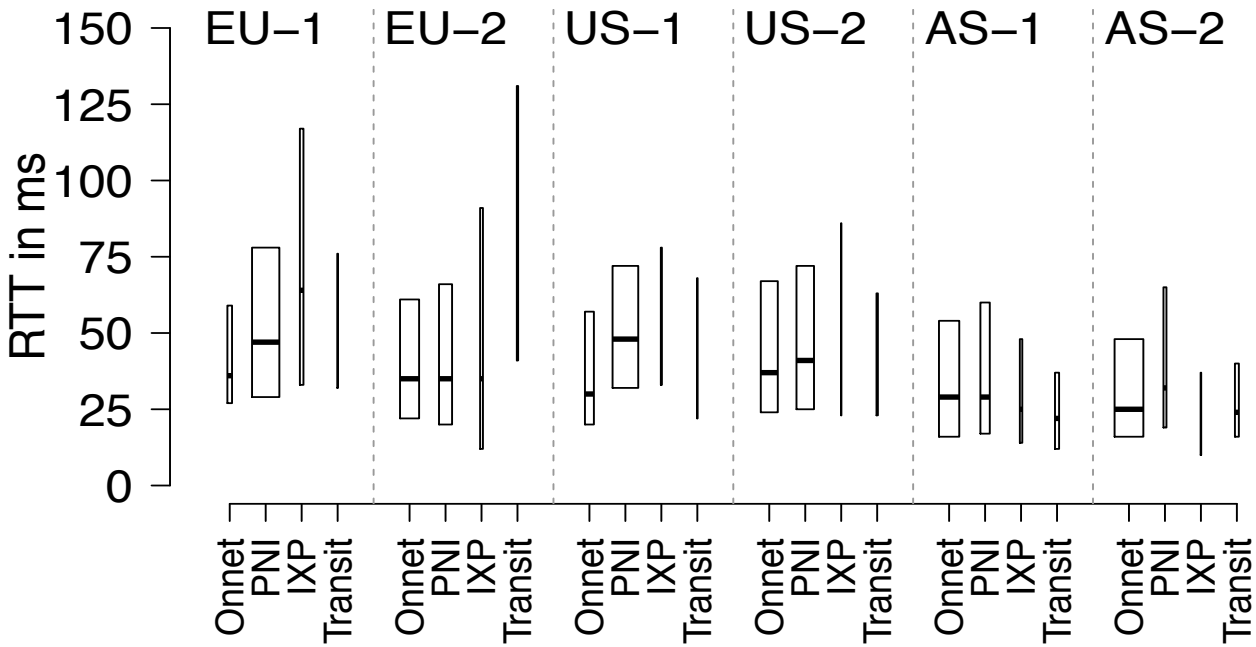
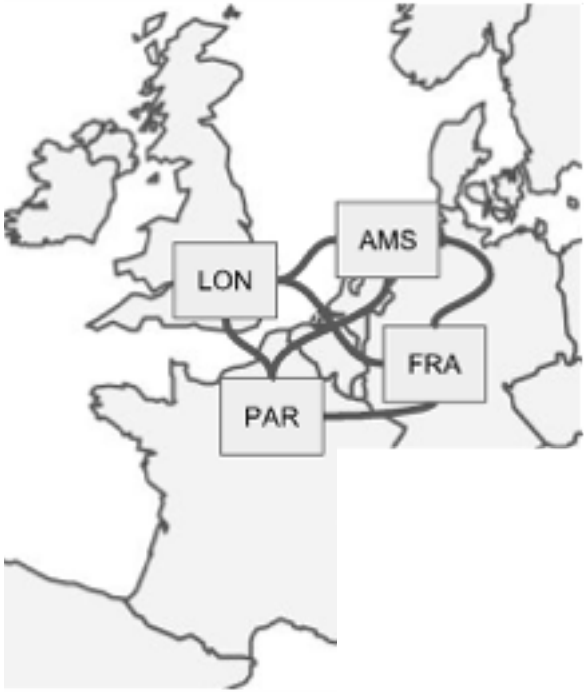


Figure 8: RTT by link type and metro.

Bar width proportional to traffic volume

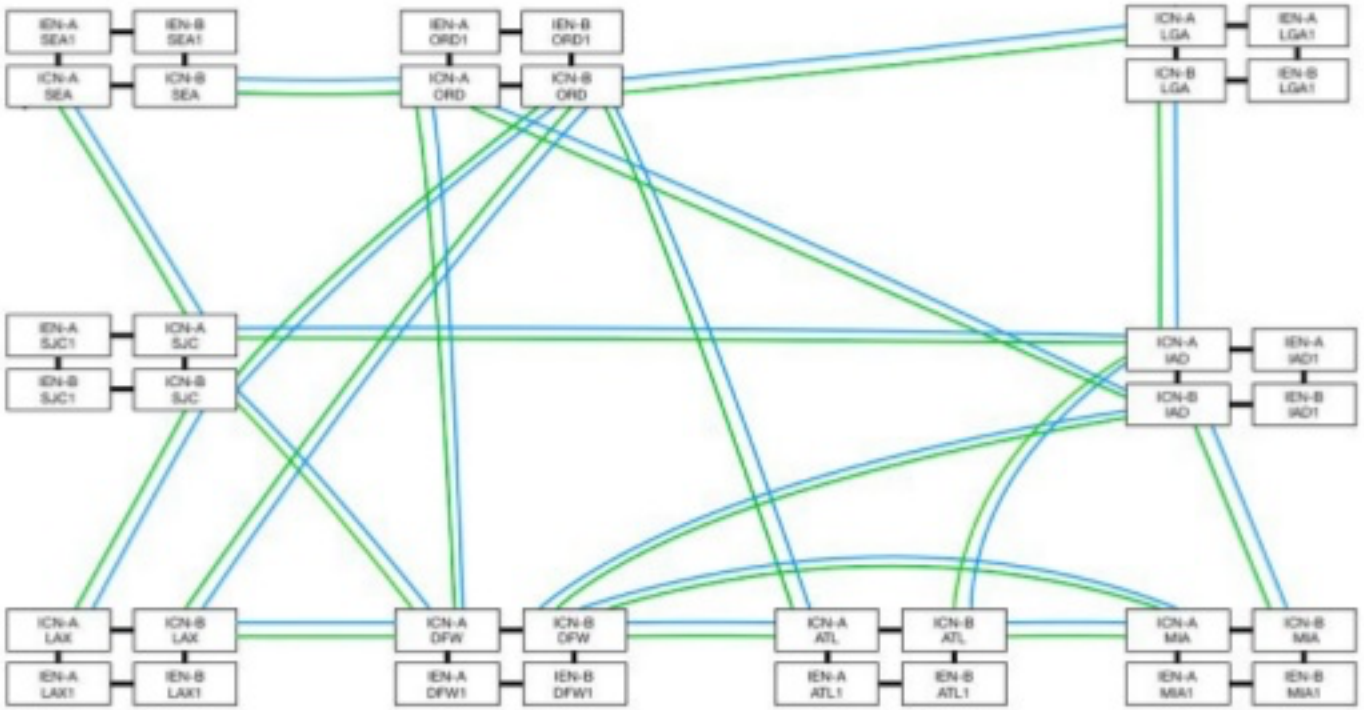
Addendum: As of late 2017, Akamai has a Private Backbone (ICN)



https://pc.nanog.org/static/published/meetings/NANOG71/1532/20171003_Kaufmann_Lightning_Talk_Akamai_v1.pdf

ICN Phase 1 (2017):
9 cities in US, 15 worldwide

City Views – North America



Akamai's Network Architecture

- Private backbone (ICN)
  - ICN connects Akamai's deployments together
  - ICN transports only Akamai traffic
  - ICN handles the ever-growing "midgrass" or cluster-to-cluster traffic

- Global-scale serving infrastructure
  - Highly-distributed platform in 130+ countries and 1,700+ networks
  - Lets Akamai peer closer to the source and handle the transport of traffic to the destination
  - Optimized traffic delivery for more localized end user traffic
  - Optimize cache fill over network links that are under Akamai's control

Familiar Look: Google

- Private backbone (B4, ~2012) for inter-datacenter traffic (~10 nodes)
- Global-scale serving infrastructure
  - Edge PoPs (some 100 locations)
  - Google Global Cache (GGC) servers (some 1,000 deployments)
- https://peering.google.com/#/infrastructure

Familiar Look: Facebook

- Private backbone (Express Backbone EBB, 2017)
- https://code.fb.com/data-center-engineering/building-express-backbone-facebook-s-new-long-haul-network/

- Global-scale serving infrastructure
  - PoPs in dozens of locations worldwide
  - Deployment of Facebook-owned and supplied servers known as Facebook Network Appliances (FNA) in hosting networks

Observation 1: Networks over time …

- The large content providers/CDNs have arrived at a similar architecture …
- Mid-gress (internal cluster-to-cluster) traffic is growing much faster than egress (end user-facing) traffic
- None of the voluminous private backbone traffic is seen in the public Internet

Observation 2: Traffic over Peerings

- Indications that a small number of PNIs carry the bulk of today's Internet traffic
- The bulk of peerings is used for the "long tail" of the traffic
- Randy's informed guess holds (however, there are geographic variations) …

Observation 3: Network Operations

- The large content providers/CDNs have built their own systems for leveraging their vast serving infrastructures
  - Google's Espresso
  - Facebook's Edge Fabric
  - Akamai's Mapping System
- Within their private backbone networks, the large content providers/CDNs can deploy their own protocols
  - Implement and deploy first, standardize later …
  - SPDY (~2010), QUIC (~2013), …

What about Cloud-related Traffic?

The Cloud Eco-system

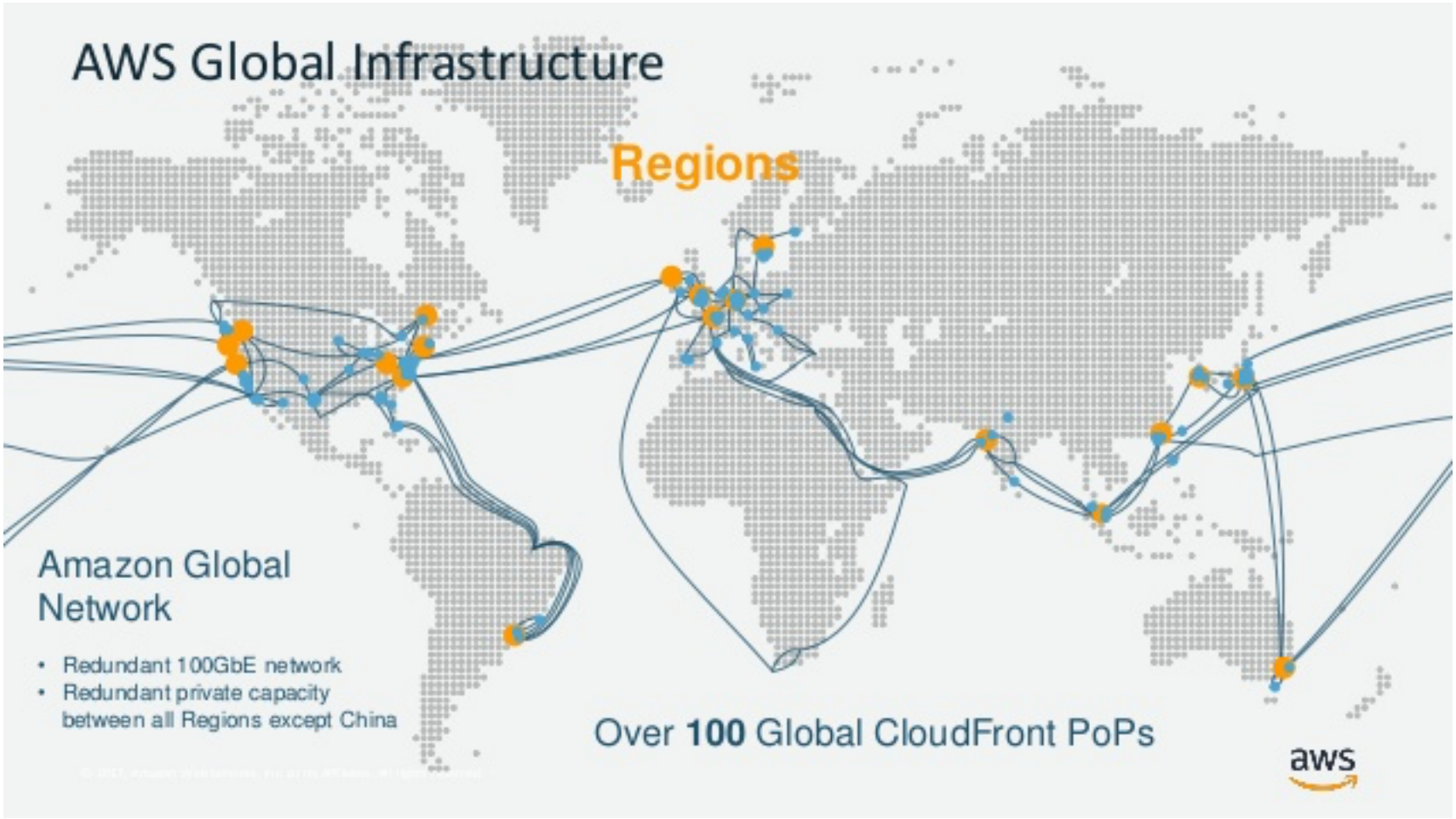
- The (large) cloud providers
  - Amazon (AWS), Microsoft (Azure), Google (GCP)
  - Oracle Cloud, IBM Cloud, Verizon Cloud, …
- The (large & medium) data center provider & interconnection services provider companies
  - Equinix, CoreSite, EdgeConneX
  - Digital Reality Trust, DuPont Fabros Technology, CyrusOne
- The large number of (large, medium & small) enterprises

The (large) Cloud Providers

The (large) Cloud Providers

- The typical architecture of a large cloud provider
  - Private global-scale backbone network
  - Some 10-100 locations around the globe where the cloud provider has presence and where enterprise customers can access its services
- All the large cloud providers are aggressively expanding the number of locations where enterprise customers can access their cloud services
- A common selling point of all the large cloud providers
  - A customer's cloud traffic is sent over direct, private network links as opposed to the public Internet
  - As a result, better performance, improved security, more reliability

Amazon and AWS



Amazon and AWS

- 53(+14) Availability Zones within 18(+4) geographic Regions and 1 Local Region around the world

- Each region contains several data centers
- Private AWS fiber links interconnect all major regions
- Once the traffic reaches the cloud provider's private backbone, it stays there and does not flow over the public Internet

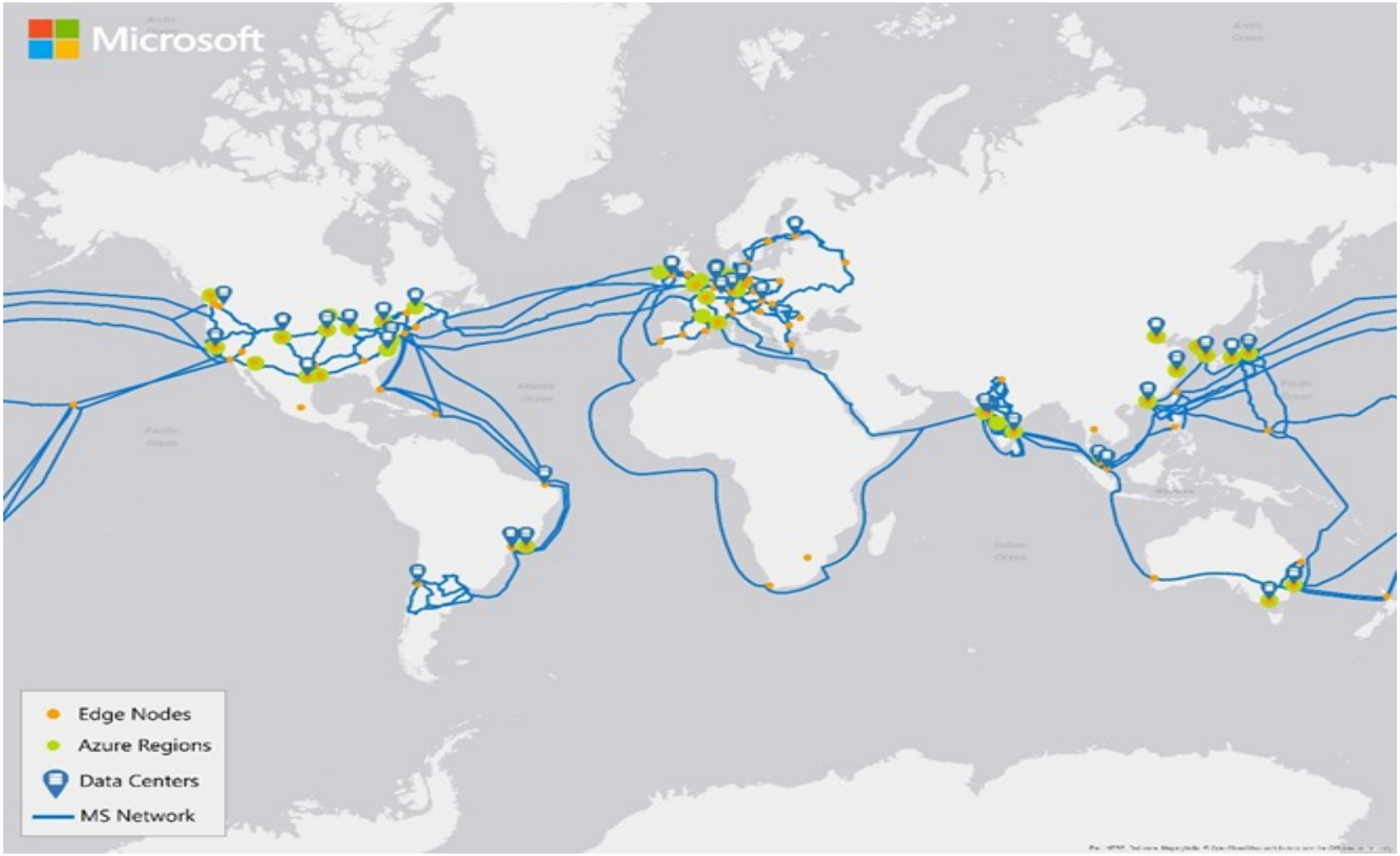
- Unknown (moving targets):
  - Number of AWS Edge network locations (~100?)
  - Number of locations where AWS Direct Connect services are available (e.g., EdgeConneX in Portland)

AWS Direct Connect

- AWS Direct Connect is a cloud service solution that enables enterprises to establish a dedicated network connection from their premises to one of AWS Direct Connect locations

- This dedicated connection can be partitioned into multiple virtual interfaces
  - The enterprise can use the connection to access public resources such as objects stored in Amazon S3 using public IP address space
  - The enterprise can use the same connection to access private resources such as Amazon EC2 instances running within an Amazon virtual private network using private IP space
  - The enterprise can maintain network separation between the public and private environments
  - Virtual interfaces can be reconfigured at any time to meet the enterprise's changing needs

Microsoft and Azure



Microsoft and Azure

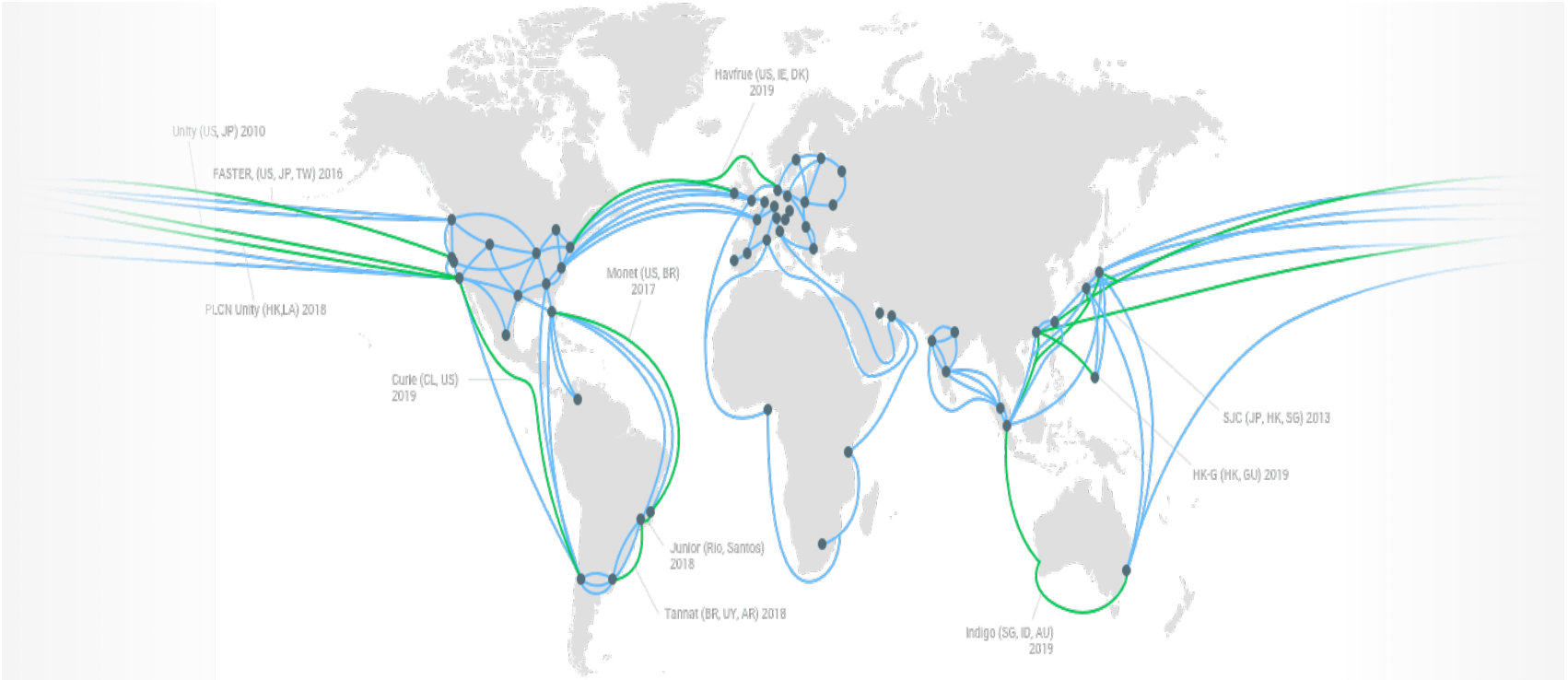
- Azure traffic enters Microsoft's global network its edge nodes, or points of presence.
- These edge nodes are directly interconnected to more than 2,500 unique Internet partners through thousands of connections in more than 130 locations.
- Azure traffic between Microsoft's datacenters stays on Microsoft's network and does not flow over the public Internet. This includes all traffic between Microsoft services anywhere in the world.

Azure ExpressRoute

- ExpressRoute enables enterprises to create private network connections to Azure
- ExpressRoute connections bypass the public Internet and offer more reliability, faster speeds, and less latency than typical Internet connections.

- With ExpressRoute, enterprises can connect to Azure
  - at an ExpressRoute location at specific Microsoft edge sites (e.g., IXP location)
  - or directly from their existing corporate WAN by using, for example, a MPLS VPN provided by a network service provider

The Google Cloud Platform (GCP)



Google and GCP

- The Google Cloud Platform (GCP) has 17 regions, 52 zones, and over 100 points of presence across 35 countries

- Google's global fiber-optic network consists of some 100,000s of miles of fiber optic cable

Google and Direct Interconnect

- Google's way for enterprises to establish a private network connection to the Google Cloud Platform is called Dedicated Interconnect
- It enables enterprises to extend their corporate datacenter network into the Google Cloud as part of a hybrid cloud deployment, if their network can physically meet Google's network in one of its supported colocation facility
- Google has currently some 69 Dedicated Interconnect locations around the globe
- Once connected, Google's network provides access to all GCP regions using a private fiber network

The (large) Colocation Providers

The (large) Colocation Providers

- Operating new infrastructure
  - IXPs and Cloud exchanges
- Providing new connectivity options
  - Virtual private interconnections
- Tapping into new pool of customers
  - Enterprises that don't participate in BGP (no ASN)

CoreSite

CoreSite delivers secure, reliable, high-performance data center and interconnection solutions to a growing customer ecosystem at 21 operating data centers across eight key North American Markets.

- 21 data centers in 8 markets
- IXP (Any2Exchange)
- Open Cloud Exchange
- 27,000+ interconnections

420+ network service providers
325+ cloud service providers
450+ enterprises

The CoreSite Open Cloud Exchange

- Established in 2013, it provides a single port into a layer-2 Ethernet switching product, enabling private virtual connections to multiple providers (i.e., a single port allows users to establish multiple virtual interfaces with other cloud, network or service providers)
- Provisioning is done in real time through a private online portal or an API
- Cloud providers can reach remote enterprises via network service providers or SDN providers (e.g., Packet Fabric, Megaport)
- "Carrier neutral" as well as "cloud-neutral"

CoreSite LA (CoreSite's Largest Facility)

CoreSite One Wilshire Campus – Los Angeles

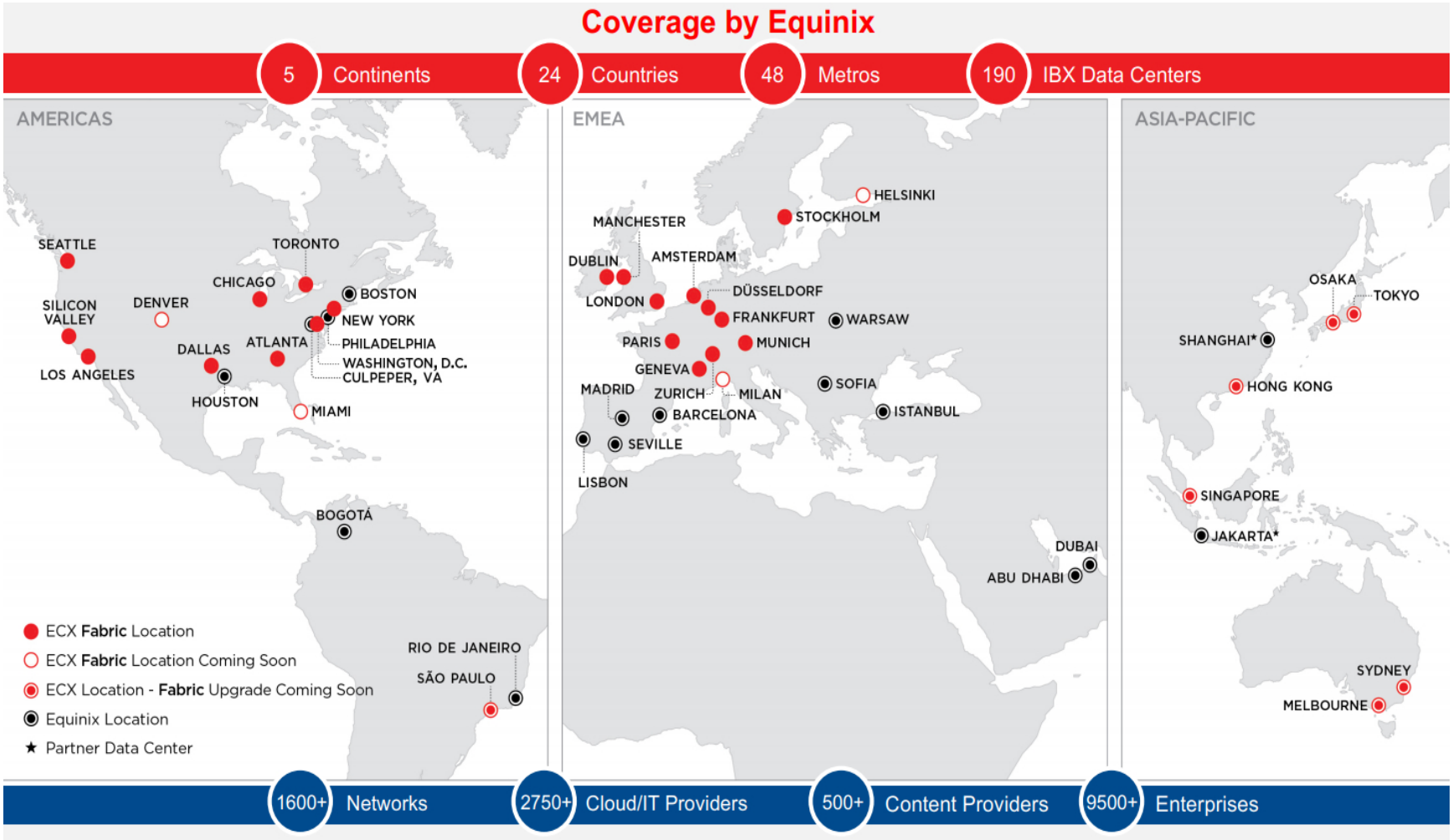
A Virtual Campus Connected via Dark Fiber

Our Entertainment and Gaming ecosystem in Los Angeles is interconnected using more than 2,000 cross connections

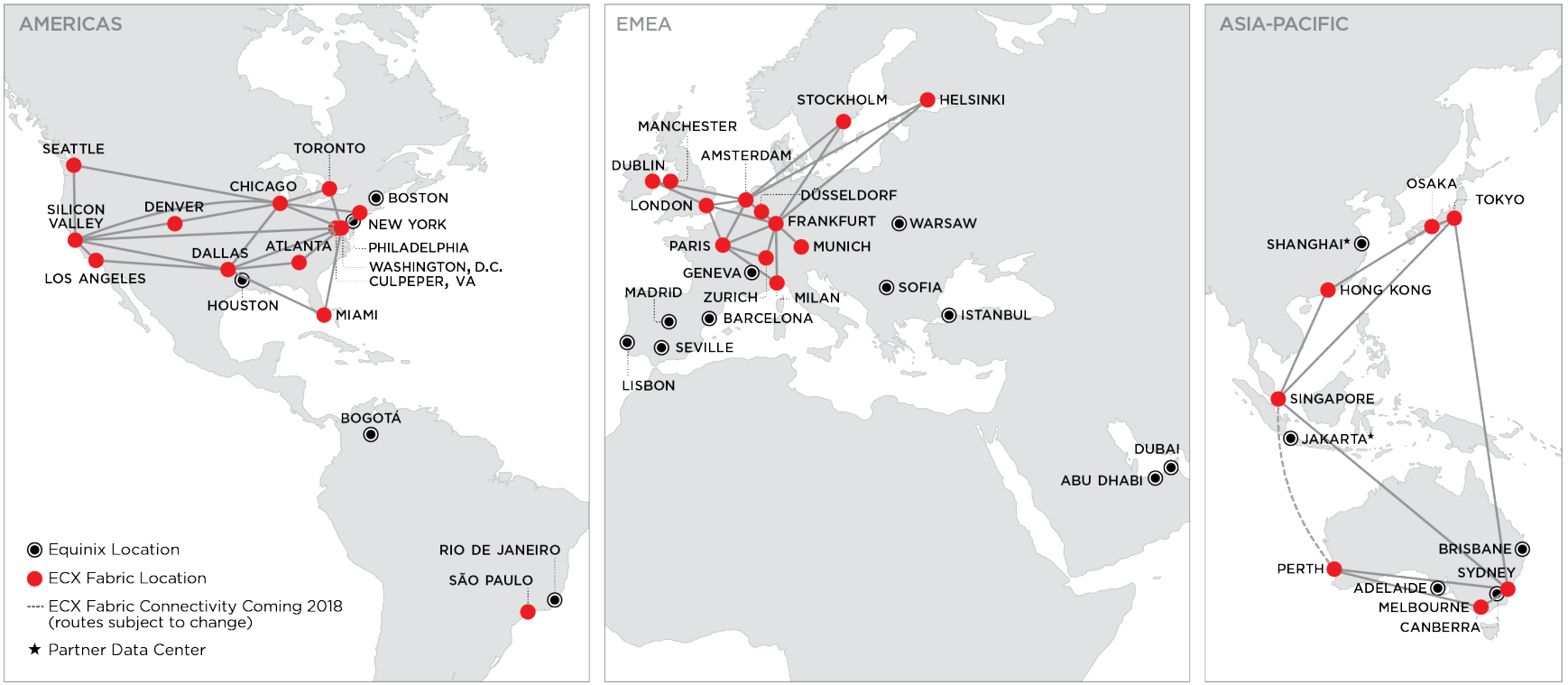
- 240+ Networks
- 110+ Cloud & IT
- 150+ Enterprise & Digital Content

Note: One Wilshire campus not entirely representative of total operating portfolio.

Equinix (founded in 1998)



Equinix (founded 1998)



- 190 data centers in 48 markets in 24 countries in 5 continents
- IXP (Equinix Internet Exchange)
- ECX (Equinix Cloud Exchange Fabric)
- 240,000+ interconnections

1,600+ network providers
2,500+ cloud IT providers
500+ content providers
9,000+ enterprises

Equinix Cloud Exchange Fabric (ECX)

- "Software-Defined Interconnection"
- "ECX is doing to interconnection provisioning what AWS did to installing servers in a data center …"
- ECX Fabric directly, securely and dynamically connects distributed infrastructure and digital ecosystems on Platform Equinix via global, software-defined interconnection.
- Available across 25+ locations, ECX Fabric is designed for scalability, agility and connectivity over a self-service portal or API. Through a single port, Equinix customers can connect directly to any other Equinix customer on demand, locally or across metros or across regions. Customers pay only for the amount of time they use the ECX Fabric.
- Note: With ECX Fabric, enterprises can connect directly and privately to their customers, suppliers, and cloud providers through Equinix – they don't need to have a separate connection with a service provider.

EdgeConneX

- Established in 2013 and specializes in data center solutions at the edge of the network (in Tier-2 markets rather than in metro areas)
- Since late 2013, EdgeConneX has built 31 Edge Data Centers (EDC) across North America and Europe (and South America)
- In select EDC locations, enterprise customers can plan, manage, and provision interconnection capacity in real time through fully automated SDN platforms.
- Enterprises can instantly access their cloud, network carriers, and content service providers through an online portal, allowing elastic, on-demand consumption of cloud services. Virtual cross connects are enabled in minutes.

Enterprises (of all Sizes)

Enterprises (of all Sizes)

- Large new pool of customers for colocation providers
  - No need for ASN (don't have to participate in BGP)
  - New revenue sources
    - Placing equipment (i.e., rent cabinet space) directly in colocation facility
    - Connecting to colocation facility by service provider (i.e., purchase interconnection service)
    - Connecting to cloud providers (i.e., purchase of VPIs)
- Enterprises as vast customer pool for cloud services
  - Connecting to cloud providers has to be easy and quick
  - Using business-critical cloud services requires reliable, performant and secure connectivity

Apropos “Public Internet” …

Some Final Observations

- Colocation providers such as CoreSite are morphing into US-wide network service providers …
- Colocation providers such as Equinix are starting to compete with global-scale network service providers …
- More data center activity in medium- to smaller-sized cities (at the edge of the network); e.g., EdgeConneX
- With the growing number of virtual private interconnections that can be established in (close-to) real-time, interconnectivity becomes highly dynamic
- Most cloud traffic (traverses VPIs and) bypasses the public Internet
- What's left for the public Internet?

Thank you!

Questions?