ISP Interconnection & Traffic Exchange

Roman Dunaytsev

The Bonch-Bruevich Saint-Petersburg
State University of Telecommunications

roman.dunaytsev@spbgut.ru

Lecture № 2

Outline

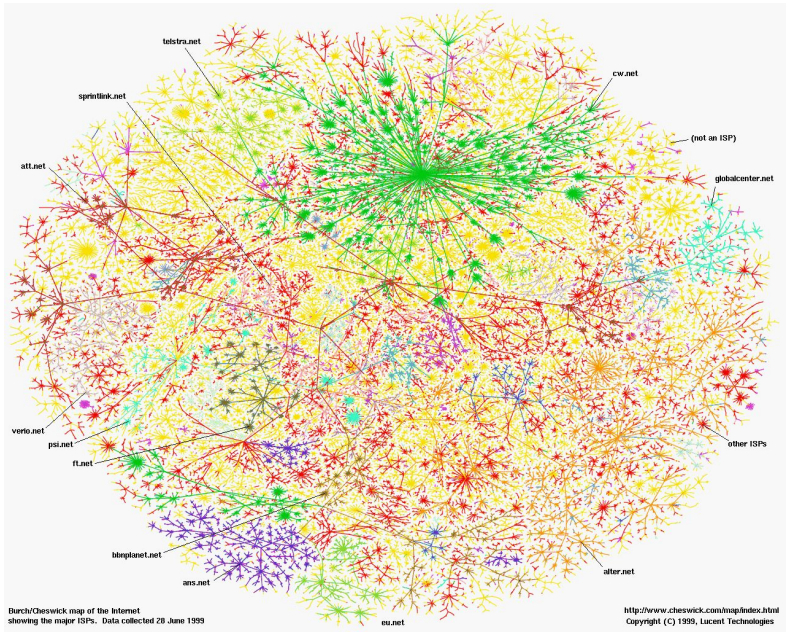
Outline

Introduction

- End users see the Internet as a seamless, global, ubiquitous communication medium
- Behind the scenes, there are many individual networks, owned and operated by different corporate, institutional, and governmental entities
- **Internet Service Providers (ISPs)** connect their networks to each other in order to exchange traffic between their customers and customers of other ISPs
- There is no single ISP that has its cables connected to all computers in the world
- ISP interconnection is the glue that holds the Internet together

Introduction (cont'd)

- **The Internet is not a 'cloud'!**



Introduction (cont'd)

- **Internetworking** – enables networks based on different telecommunication technologies to exchange data
- The binding elements of the global Internet:
  - IP convergence (aka 'everything over IP, IP over everything')
  - Common IP addressing scheme
  - Global BGP routing framework

- **Interconnection** – enables owners and operators of different networks to collaborate as business entities in the provision of seamless end-to-end Internet connectivity to all of their customers

Outline

Types of ISPs

- **2 types of Internet traffic exchange**:
  - Settlement-based (aka **transit**)
  - Settlement-free (aka **peering**)

- **Transit**
  - When one ISP agrees to carry the traffic that flows between another ISP and the rest of the Internet
  - The transit provider receives a fee for the delivery service

- **Peering**
  - When 2 or more ISPs interconnect directly with each other to provide access to each other's customers
  - This is typically done without charging for the delivery service

Types of ISPs (cont'd)

- **3 types of ISPs**:
  - Tier 1
  - Tier 2
  - Tier 3

- **Tier 3 ISP**
  - **Traffic exchange**: might have some peering, but often does not
  - Size: regional or local network
  - Customers: end users
  - Example: AS35807 SkyNet Ltd.
    https://bgp.he.net/AS35807

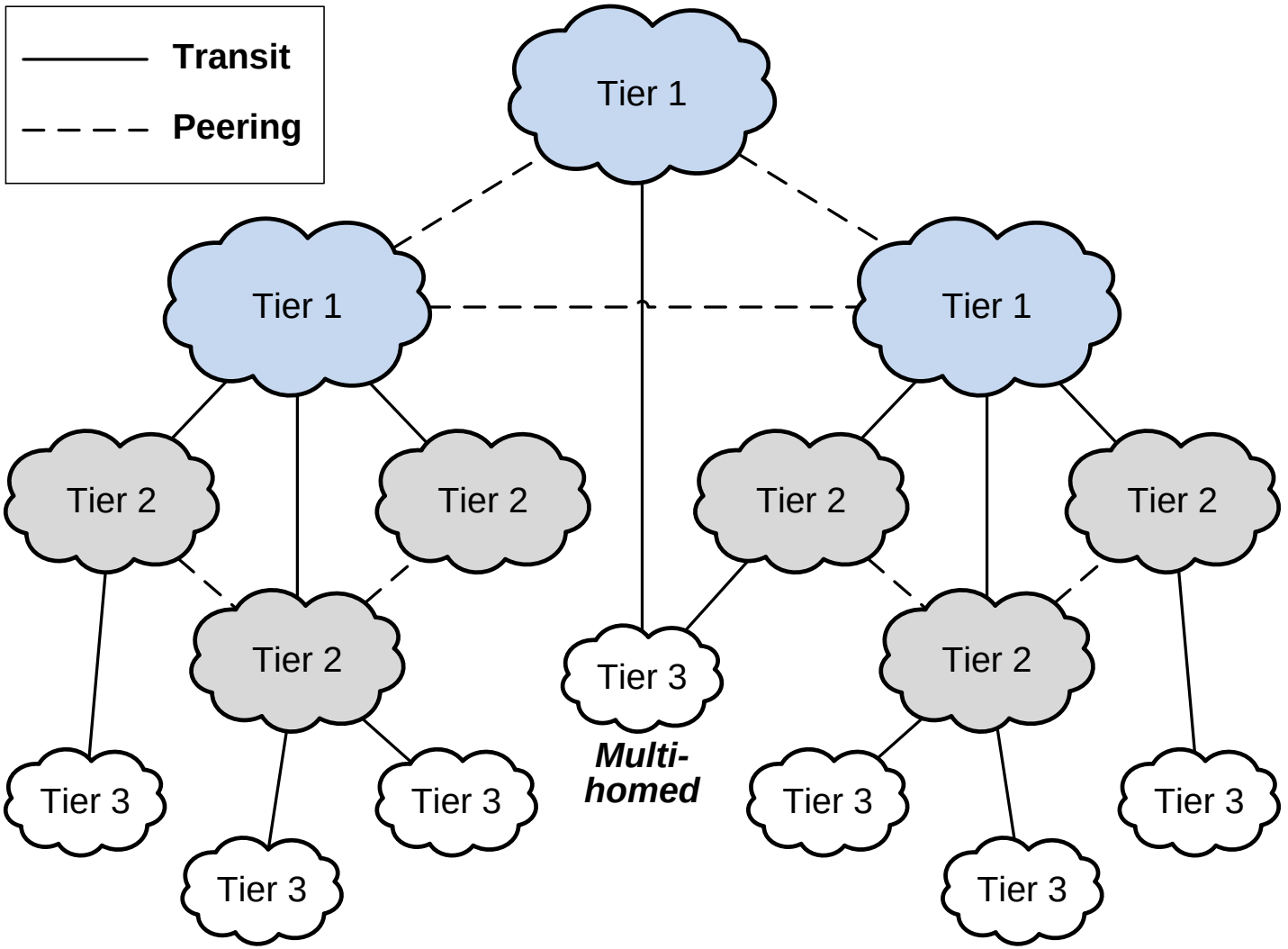
- **Tier 2 ISP**
  - **Traffic exchange**: combination of peering and transit
  - Size: nationwide network
  - Customers: Tier 3 ISPs, end users
  - Example: AS8359 MTS PJSC
    https://bgp.he.net/AS8359

Types of ISPs (cont'd)

- **Tier 1 ISP**
  - **Traffic exchange**: relies completely on peering
  - Size: worldwide or nationwide network
  - Customers: Tier 2 ISPs, Tier 3 ISPs, end users
  - Example: AS3356 Level 3 Communications, Inc. https://bgp.he.net/AS3356
- 'The Tier 1 Club'
  - AT&T
  - Deutsche Telekom
  - Level 3 Communications
  - NTT America
  - Sprint
  - Tata Communications
  - TeliaSonera
  - Verizon Business
  - ??? ('Once a customer, never a peer')

Types of ISPs (cont'd)

- ISP tiers and interconnection



Outline

Types of ASs

- Technically, transit and peering are implemented in the form of configuration information for the **Border Gateway Protocol (BGP)**
- The Internet is divided into administrative domains called Autonomous Systems (ASs)
- **2 types of routing protocols**:
  - Interior Gateway Protocols (IGPs) to perform routing within a single AS
  - Exterior Gateway Protocols (EGPs) to perform routing across ASs
- BGP – the de facto standard EGP for exchanging routing information between ASs in the Internet
  - I.e., BGP is the high-level routing protocol that connects everything

Types of ASs (cont'd)

- **Autonomous System (AS)** – a collection of routers and subnets under the same administrative and technical control, and that all run the same IGP among themselves

- To distinguish and uniquely identify each AS on the Internet, ASs are assigned an **Autonomous System Number (ASN)**
  - ASN assignment is done by the **Regional Internet Registries (RIRs)** or the **Internet Assigned Numbers Authority (IANA)**

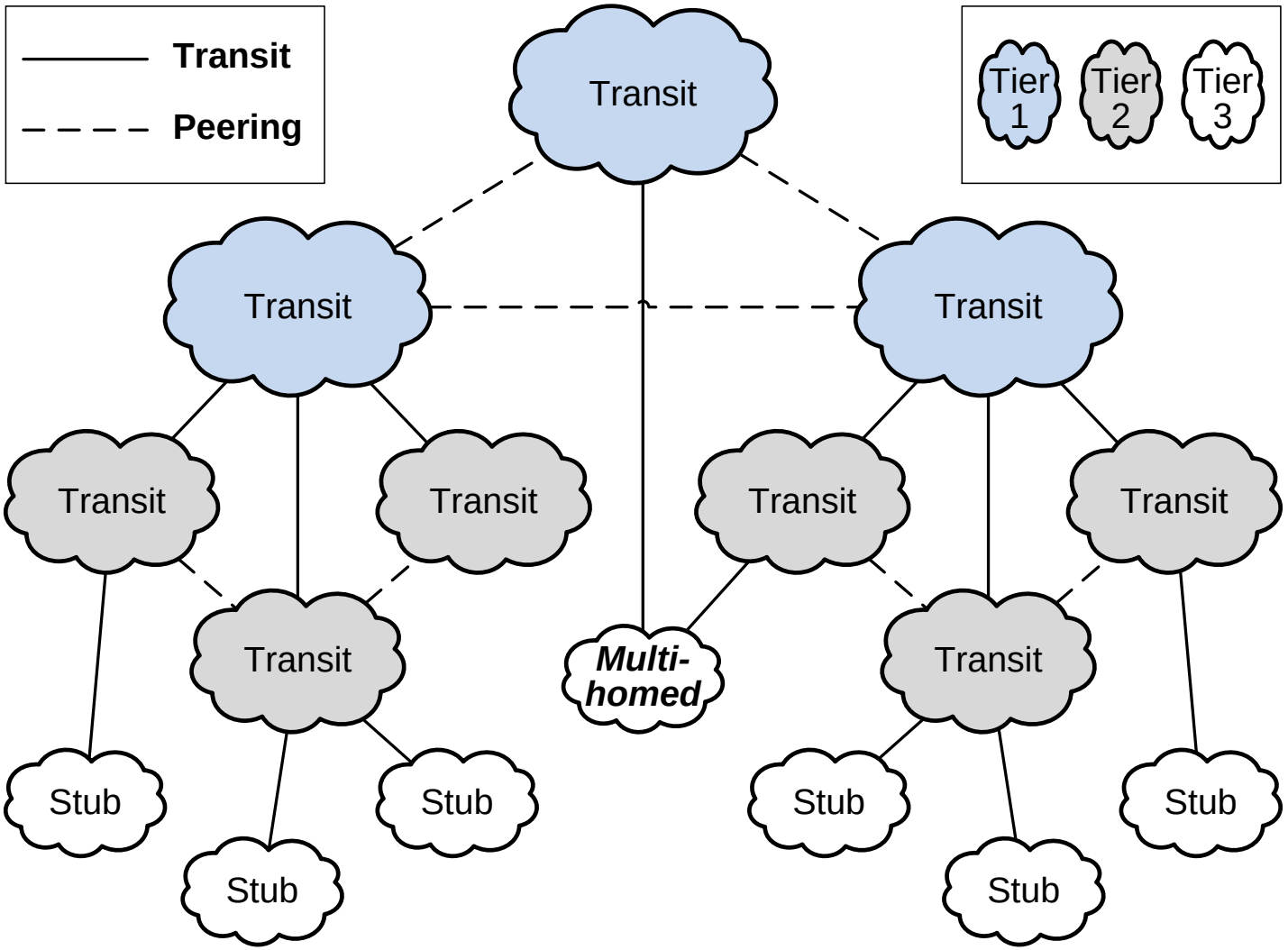
- A large ISP will typically segment its network into multiple ASs

Types of ASs (cont'd)

- **3 types of ASs**:
  - Transit
  - Stub (aka **single-homed**)
  - Multihomed
- **Transit AS**
  - Connectivity: connected to more than 1 AS
  - Traffic: carries both local and **transit** traffic
  - Example: a higher-tier ISP with its downstream customers
- **Stub AS**
  - Connectivity: connected to only 1 AS
  - Traffic: carries only local traffic
  - Example: a Tier 3 ISP that has only a single connection to the Internet
- **Multihomed AS**
  - Connectivity: connected to more than 1 AS
  - Traffic: carries only local traffic
  - Example: a Tier 3 ISP that has multiple connections to the Internet

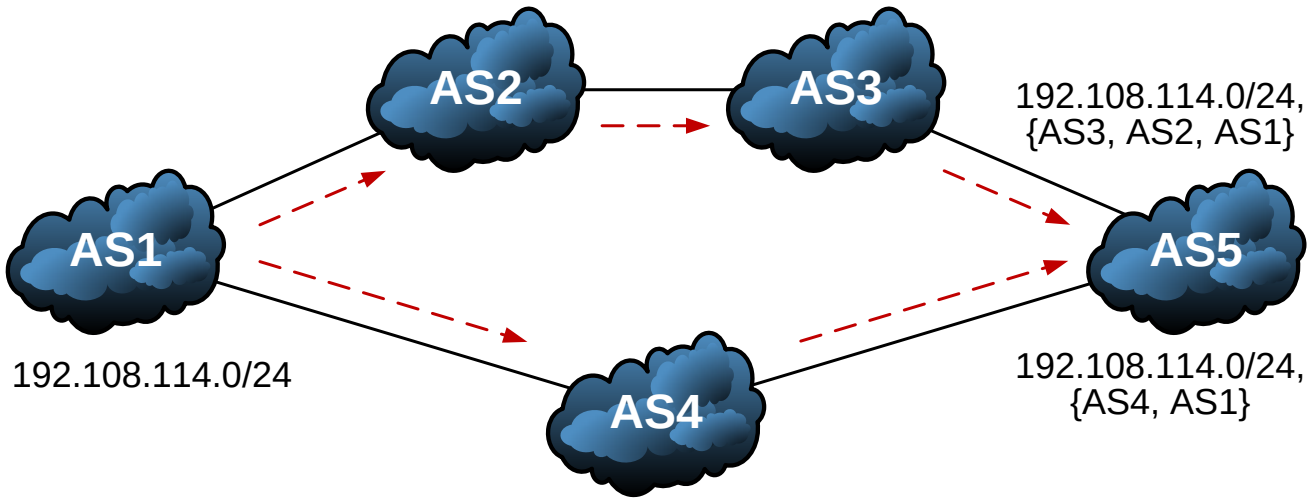
Types of ASs (cont'd)

- Autonomous systems vs. ISP tiers



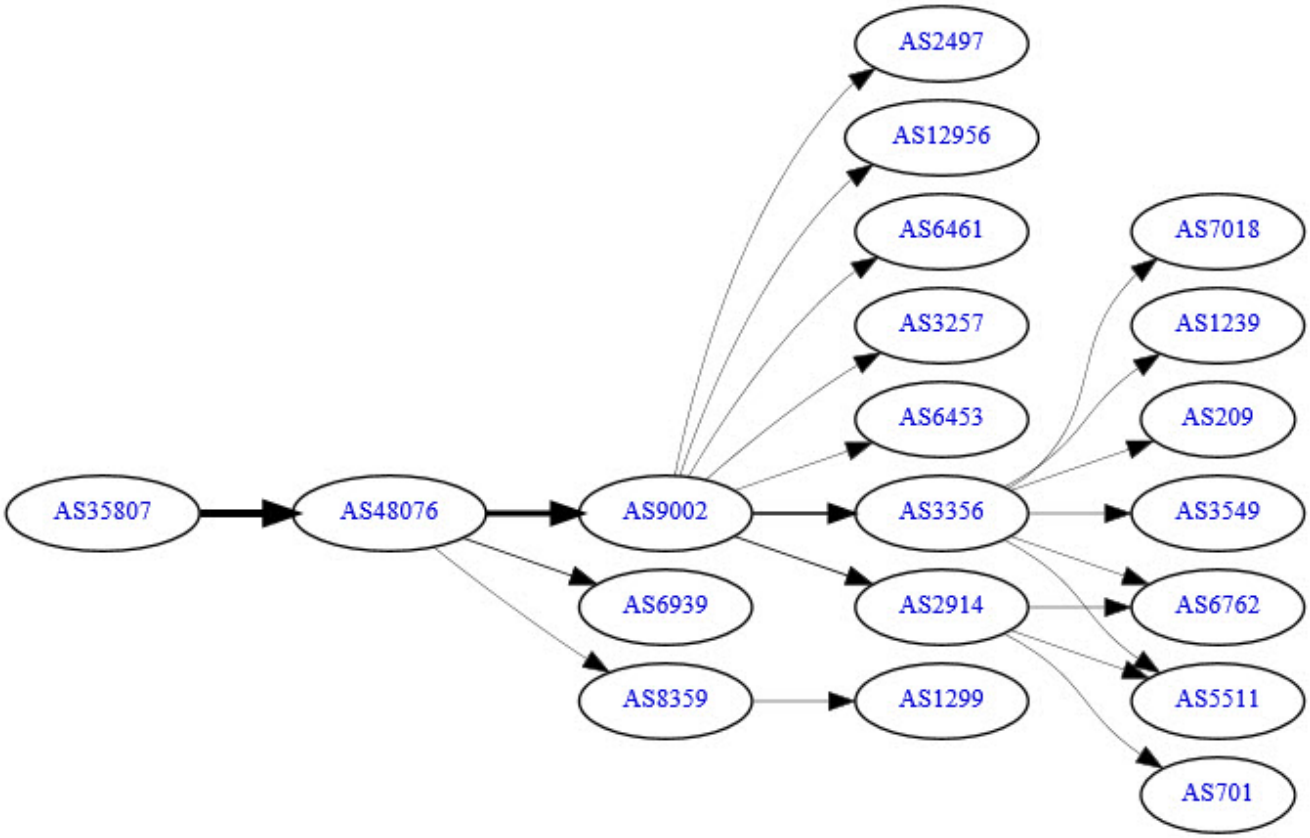
Types of ASs (cont'd)

- **AS path generation in BGP**
  - Every prefix has an origin AS from which reachability for the prefix is propagated across the inter-AS space
  - As the routing advertisement is propagated across the inter-AS space, each prefix accumulates an associated AS path
  - At any point in the network, the AS path describes a sequence of connected ASs that forms a path from the current point to the originating AS
  - The AS path is used as a path metric in the BGP path selection algorithm



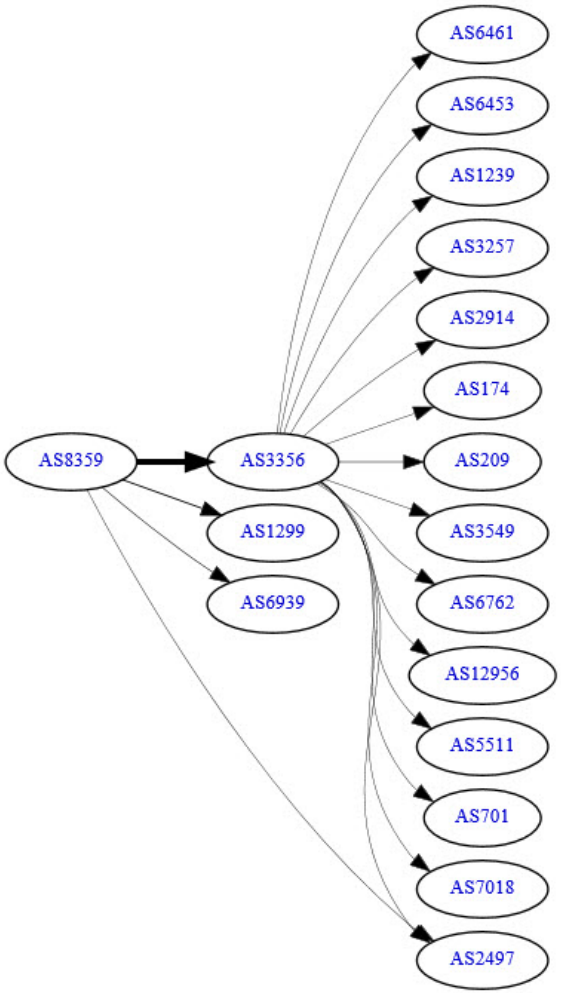
Types of ASs (cont'd)

- AS35807 SkyNet Ltd. (a **Tier 3** ISP):
  https://bgp.he.net/AS35807#_graph4



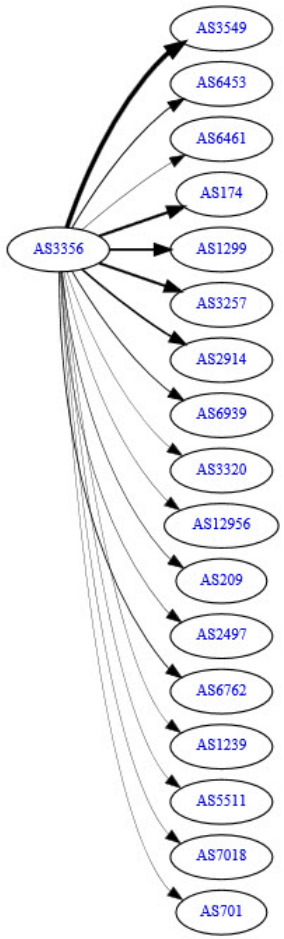
Types of ASs (cont'd)

- AS8359 MTS PJSC (a **Tier 2** ISP): https://bgp.he.net/AS8359#_graph4



Types of ASs (cont'd)

- AS3356 Level 3 Communications, Inc. (a **Tier 1** ISP): https://bgp.he.net/AS3356#_graph4



Types of ASs (cont'd)

- BGP data collected from APNIC's router at DIX-IE, Japan
  http://thyme.apnic.net/current/data-summary
  - A well-connected router that has a good view of the Internet
  - APNIC = Asia Pacific Network Information Centre
  - DIX-IE = Distributed Internet eXchange Point In Edo
- BGP routing table summary (for September 2017):
  - Total ASs = 58,245
  - Stub ASs = 50,337
  - Transit ASs = 7,908
  - Average AS path length (visible in the routing table) = 4.3
  - Prefixes per ASN = 11.32
- **About 86% of all ISPs are stub ASs (i.e., Tier 3)**

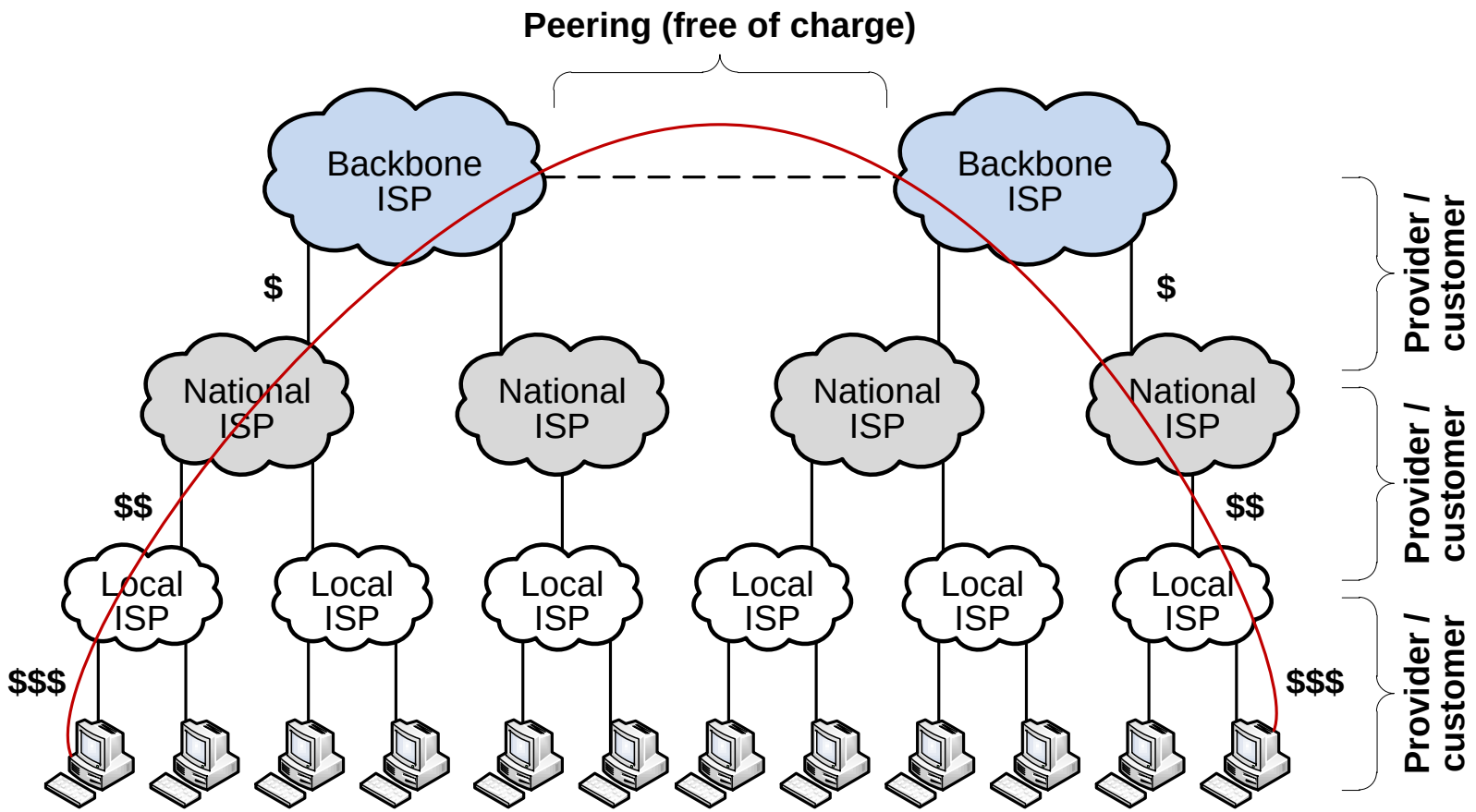
Outline

Transit

- **Transit** – one ISP (aka a **provider** or **upstream ISP**) agrees to carry traffic for another ISP (aka a **customer** or **downstream ISP**) for delivery to/from the rest of the Internet
- The provider charges a fee for carrying the customer's traffic
- To get connectivity outside their local areas, most ISPs buy transit from 1 or more providers with national and/or international coverage
- **Transit is a transitive relationship**
  - I.e., when a downstream ISP purchases transit from an upstream ISP, it gets connectivity to all networks in the routing table of the transit provider
- Transit is the means by which most ISPs obtain global connectivity

Transit (cont'd)

- **The retail model of Internet access services is not end-to-end**
  - The sender only funds the initial path component and the receiver funds the terminating path component



Transit (cont'd)

- **Billing methods**:
  - Flat rate (aka **unmetered**)
  - Average usage
  - Bandwidth quotas (aka **bandwidth caps**)
  - 95th percentile

- **Flat rate** – connectivity is provided at a fixed monthly fee, regardless of the actual bandwidth usage
- **Average usage** – pricing is based on the amount of data transferred per month
- **Bandwidth quotas** – you pay for each GB of data, which exceeds the monthly allocation included in your monthly fee
- **95th percentile** – pricing is based on burstable billing
  - Most ISPs and hosting providers use the 95th percentile billing method

Transit (cont'd)

- E.g., Xifron's transit pricing, http://xifron.com/ip-transit.html
  - Xifron Global LLC is a datacenter service provider operating a transit network with several connections to Tier 1 ISPs https://bgp.he.net/AS24611

- **Xifron's bandwidth quotas**

Traffic - Volume Based	Committed Internet Traffic per GB - Uplink port to our redundant backbone - /29 IP-Subnet included (different subnets upon request) - Beyond the Committed traffic 0,19 € per Gigabyte shall be charged Incoming and outgoing IP-traffic is retained and added at the end of each accounting period (end of month). Charges are according to the price for each Gigabyte or parts thereof.		
	Committed Traffic	**Price**	**Overage Fee**
	500 GB / Month included in all Xifron.com rackspace offers	-	0,19 € / GB

Transit (cont'd)

- **Xifron's flat rate**
  - To compete more aggressively in the market, most ISPs provide pricing discounts for precommitting to certain data rates and volumes of traffic

Committed Data Rate - Capped

Commited Data Rate **(Capped)** Internet Bandwidth

- Uplink port to our redundant backbone
- Limited Data Rate
- /29 IP-Subnet included (different subnets upon request)

Committed Data Rate (CDR) with a cap is very similar to your office DSL connection, except it is the same speed in both directions. Basically, you tell us how large you would like your connection to be, for example 10Mbps, and we will cap the connection to this speed. It means you can use the connection as much as you'd like without incurring extra charges; however the maximum you would be able to throughput would always be fixed to the limit you specify.
We can increase the limit at short notice if demand for your service rises.
Capped connections are monthly charged and provisioned at intervals and are at a contention rate of 1:1, i.e. the entire commitment is guaranteed to be available for 100% of the month.

Committed Data Rate - Capped	Price	Port Speed
5Mbps	120 €	100Mbps
10Mbps	190 €	100Mbps
20Mbps	290 €	100Mbps
50Mbps	490 €	100Mbps
100Mbps	890 €	100/1000Mbps
100Mbps+	price upon request	100/1000Mbps

Transit (cont'd)

- **Xifron's burstable billing**
  - What happens if customers commit to a certain **Committed Data Rate (CDR)** but do not use all of the CDR in a particular month?
  - They are still required to pay as if they had used the full CDR

Committed Data Rate - Burstable

Committed Data Rate **(Burstable)** Internet Bandwidth - **95%**

- Uplink port to our redundant backbone
- /29 IP-Subnet included (different subnets upon request)

Invoiced according to the 95th Percentile method of calculation and burstable within the limit of 200% of the commitment.

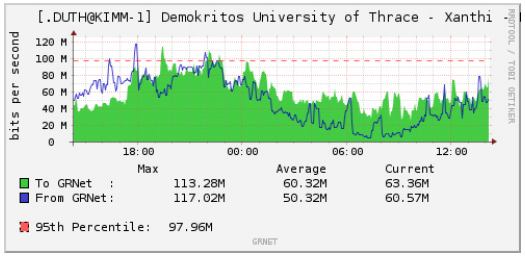
Bandwidth usage from your switch port is sampled at regular intervals, and logged on our system. The top 5% of bandwidth usage for the month is disregarded (roughly 36 hours), the 95th percentile remaining is used as the figure on which you are billed. So if this figure is higher than your commitment, you will be billed for the excess at the average price per Mbps of your CDR.

Commitment	Price	Port Speed
5Mbps	140 €	100Mbps
10Mbps	220 €	100Mbps
20Mbps	330 €	100Mbps
50Mbps	560 €	100Mbps
100Mbps	990 €	100/1000Mbps
100Mbps+	price upon request	100/1000Mbps

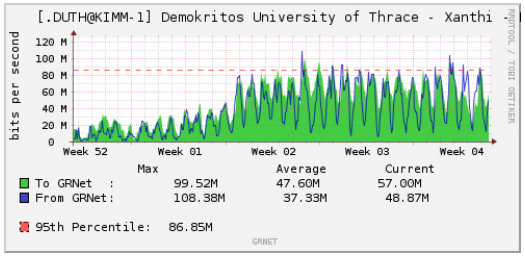
95th Percentile Billing

- **Burstable billing** – a method of measuring bandwidth usage based on peak utilization; it also allows a customer to exceed a specified threshold for brief periods of time without any financial penalties
- **Most data traffic is bursty in a wide range of time scales**

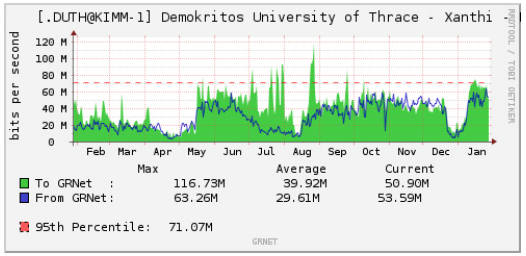
- Day



- Month



- Year



95th Percentile Billing (cont'd)

- 95th percentile billing works as follows:
  1. Transit providers poll customer interface ports on their routers and switches at regular intervals
  2. Each sample contains the number of bytes transmitted to the customer and bytes received from the customer since the sample took place
  3. The provider converts each sample into a data rate (measured in Mbit/s) for the interval
  4. These samples are taken throughout the billing cycle
  5. The provider then sorts measurements from largest to smallest, removes the top 5% of the lines, and bills the customer for the next line which is called **the 95th percentile**
  6. Under this method, 2 separate 95th percentile values are computed: for inbound and outbound traffic
  7. Finally, the cost of transit is computed as follows:

max{95th percentile Inbound; 95th percentile Outbound; Committed Data Rate} $\times$ Transit unit price

95th Percentile Billing (cont'd)

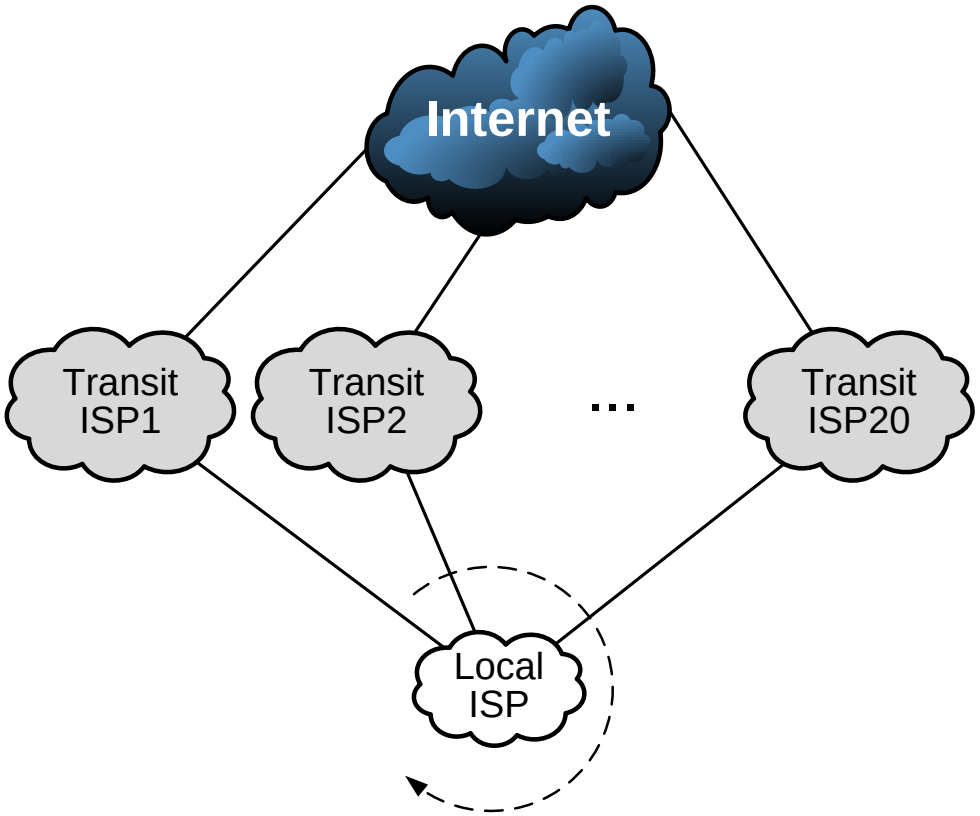
- Sampling interval: **5 minutes**
  - 5 minutes is the most typical value
  - But other values can be used as well (e.g., 30 seconds, 1 minute)
- Billing cycle: 30 days
- Total number of samples: $(60 \times 24 \times 30)/5 = 8,640$
- The samples are sorted from largest to smallest
- Top 5% of the measurements: $8,640 \times 0.05 = 432$
  - I.e., the top 432 measurements are ignored
  - Thus, there is no billing for $(432 \times 5)/60 = 36$ hours per month
  - You can transmit as fast as possible for 36 hours per month free of charge
- Bottom 95% of the measurements: $8,640 - 432 = 8,208$

95th Percentile Billing (cont'd)

- Sampling interval: **1 minute**
- Billing cycle: 30 days
- Total number of samples: $(60 \times 24 \times 30)/1 = 43,200$
- The samples are sorted from largest to smallest
- Top 5% of the measurements: $43,200 \times 0.05 = 2,160$
  - I.e., the top 2,160 measurements are ignored
  - Thus, there is no billing for $(2,160 \times 1)/60 = 36$ hours per month
  - You can transmit as fast as possible for 36 hours per month free of charge – **similar to the case with 5-minute sampling!**
- Bottom 95% of the measurements: $43,200 - 2,160 = 41,040$
- **What effect does the sampling interval have?**
  - I.e., 30 seconds, 1 minute, 5 minutes, ...

A Bit of Tactics

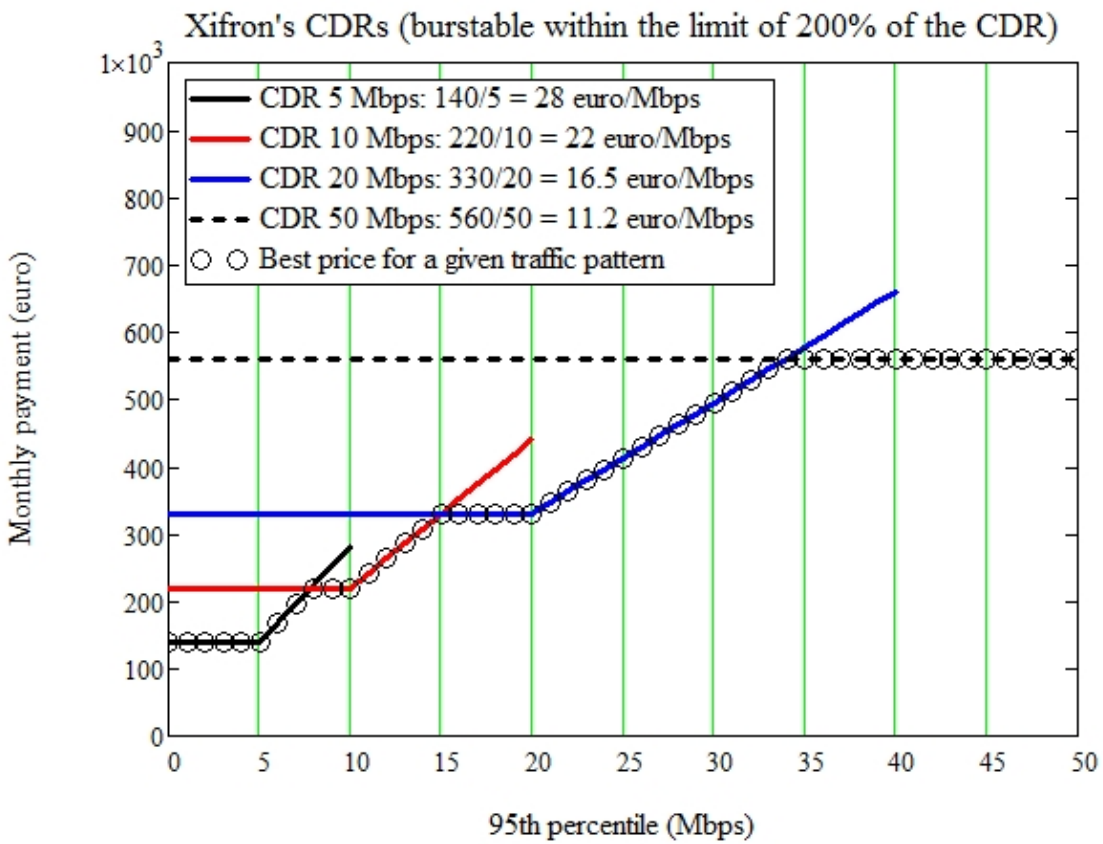
- **Bad tactics:** Taking advantage of the 95th percentile billing method
  - **Background:** At the end of the month, the samples are sorted from largest to smallest and the top 5% is discarded ($24 \times 30 \times 0.05 = 36$ hours of peak traffic per 30-day billing cycle)
  - **Requirement:** Identical billing methods without committed cost



Transit	Used, h	Top 5%
ISP1	36	36 (skip)
ISP2	36	36 (skip)
ISP3	36	36 (skip)
...	...	...
ISP20	36	36 (skip)
95th percentile (to pay): 0		

A Bit of Tactics (cont'd)

- **Good tactics:** Switching between Committed Data Rates
  - **Background:** The higher the CDR, the lower the price per Mbps of traffic being carried
  - **Requirement:** Predictable traffic patterns



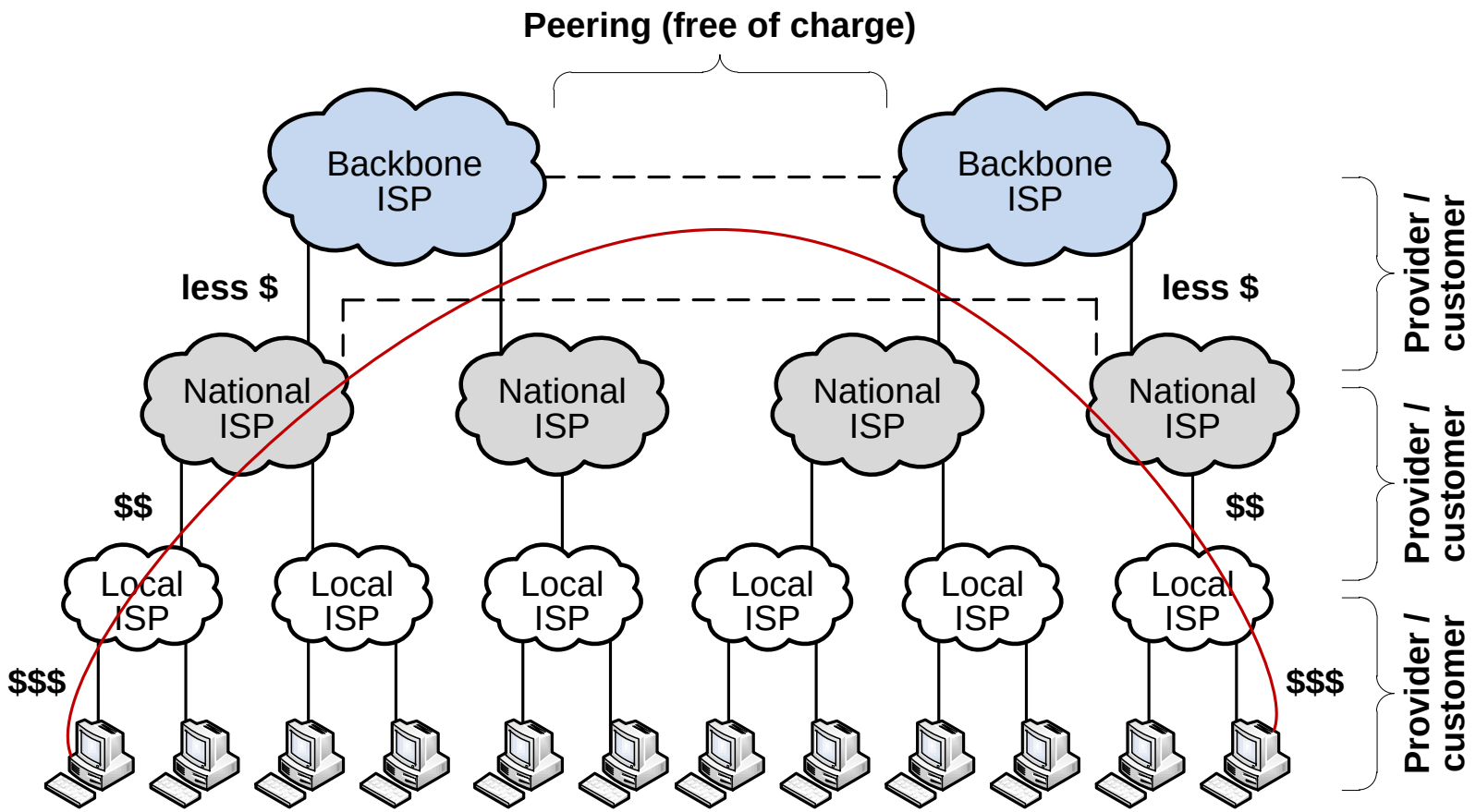
Outline

Peering

- **Peering** – 2 or more ISPs interconnect with each other to exchange traffic of their customers (and customers of their customers)
- Peering is generally done without settlement and, hence, can be considered as a barter transaction
  - I.e., with each ISP bearing the cost of the other ISP's use of its network in exchange for the benefit of the use of the other's network
- Peering is often (but not always) based on the **'Sender Keeps All' (SKA)** accounting arrangement
  - I.e., the sender of the traffic does not compensate the next network in line but keeps the entire sum (aka 'Bill and Keep')
  - In fact, both sender and receiver keep their part
- **Peering is not a transitive relationship**
  - The fact that ISP A is peering with ISP B and ISP B is peering with ISP C does not imply that ISP A customers can reach ISP C customers

Peering (cont'd)

- **Peering is not a perfect substitute for transit**
  - Transit is a service that provides access to the global Internet, while peering simply provides a more direct path to some networks



Peering (cont'd)

- Why do ISPs peer?
  - **Reduced transit costs** – you are no longer paying a transit provider to deliver some portion of your traffic
  - **Reduced latency** – by eliminating the transit hops and taking a more direct route
  - **Improved quality of service (QoS)** – the overall performance of the network is increased, including better availability, the ability to provide greater security, etc.
  - **Improved scalability** – by distributing traffic over interconnections with many different ISPs, you can potentially improve your ability to scale
  - **Improved market position** – ISPs announce their well-peered backbones, shorter paths, lower latency, etc. in their promotional material

Peering (cont'd)

- Why don't ISPs peer?
  - **Lack of information** – some ISPs either do not realize peering exists, do not realize it is available to them, or have no idea how to get started
  - **Work to be done** – coordinating peering means figuring out who to peer with and pestering them until they turn up the sessions, at which point the sessions need to be maintained
  - **Questionable performance gains** – if your transit provider is peering locally, performance gains from your peering may be minimal or they may be entirely fictional
  - **Investments and peering costs** – peering can be more expensive than transit, so some ISPs decide it does not make financial sense or that they can not afford it

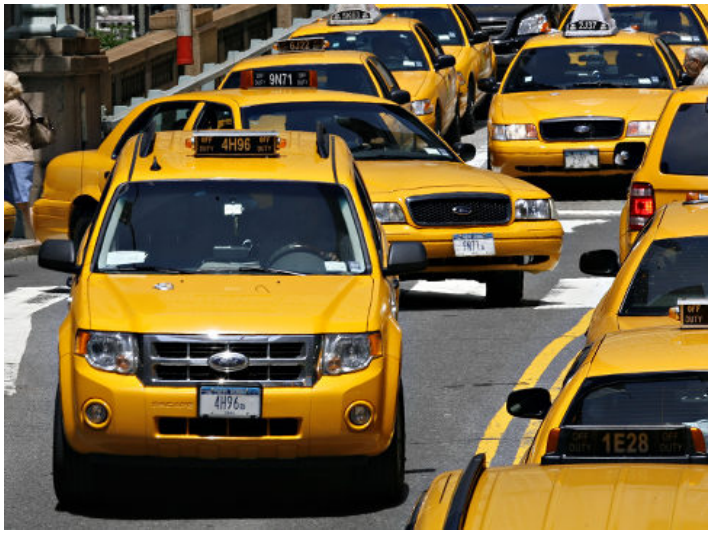
- While peers do not generally charge each other for peering, that does not make it free!

Peering (cont'd)

- While peering can happen anywhere where 2 ISPs can connect to each other, it is usually done at an Internet exchange point
- **Internet eXchange Point (IXP)** – a facility where ISPs interconnect
  - It is generally an Ethernet switch that all the participants plug into and use to establish BGP sessions between their networks
  - Many IXPs also offer private cross connects, cables going directly between networks in their facilities that the ISPs can use for interconnections
- **Why peering is not for free?**
  - IXP operators often charge for use of the facility and ports
  - If the traffic to be handed off via peering is originating somewhere other than the IXP location, there are extra circuits and extra routers to buy, rack space to rent, etc.

Peering (cont'd)

- Transit vs. peering as to go by taxi vs. to buy a car and go by it
  - In both cases, your ride is not for free – it costs you some money, but in different ways
  - Choosing which one is better depends on your personal needs and budget
  - The overhead associated with peering (owning a personal car) can easily become higher than the cost of purchasing transit (taking a taxi)



Peering (cont'd)

- The Internet Exchange Map
  https://www.internetexchangemap.com/

Peering (cont'd)

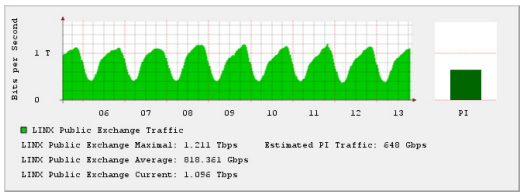
- **2 types of IXPs**:
  - Commercial (typically, in the USA)
  - Non-commercial (typically, in Europe)
- Technological progress:
  - Shared Ethernet / FOIRL (Fiber-Optic Inter-Repeater Link) Ethernet
  - Switched 10BaseT Ethernet
  - Shared FDDI (Fiber Distributed Data Interface)
  - Switched FDDI
  - 100BaseT Ethernet / 100BaseFX Ethernet
  - Gigabit Ethernet (1GE)
  - 10-Gigabit Ethernet (10GE)
- Features:
  - Location and available space
  - Power per cabinet and UPS (Uninterruptible Power Supply)
  - Cooling and fire suppression
  - Security and technical support

Peering (cont'd)

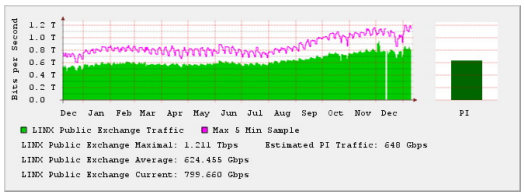
Peering (cont'd)

- The London IXP (LINX) – a non-commercial IXP with over 390 members connecting from over 50 different countries worldwide https://www.linx.net
- Membership fee = £1,200 per year (1 GBP ≈ 1.2 EUR)
- Public peering port fee (100 Mbps) = £80 per month
- Installation fee (ISP's equipment to the switches) = £200 per port
- Graphs show the aggregated traffic that flows across all LINX switches and sites
  - LINX public exchange maximal (for September 2017) = 3.58 Tbit/s

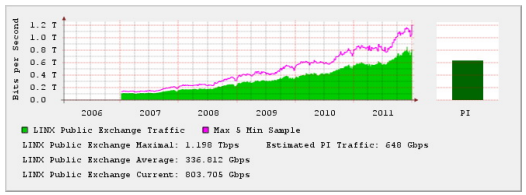
- Week



- Year

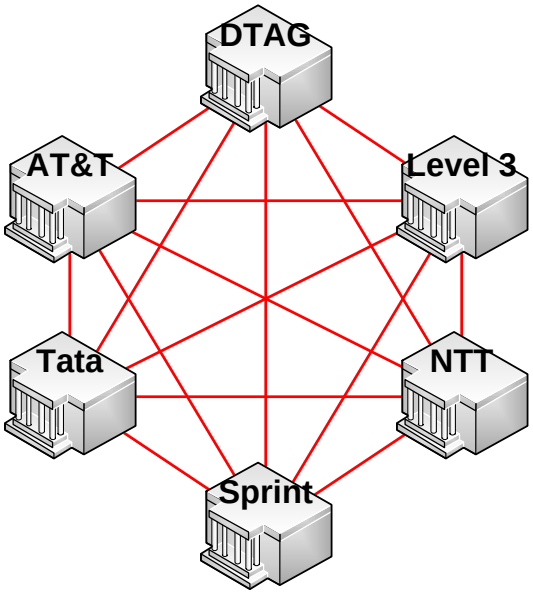


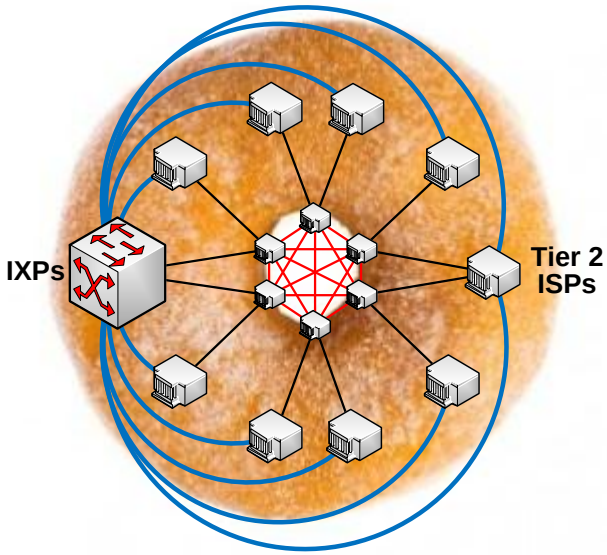
- Decade



Types of Routing

- **Donut peering model**
- **Inner core** – the Tier 1 Club, connected in a full-mesh topology
  - Aka the **default-free zone (DFZ)** – the core routers of Tier 1 ISPs maintain a complete routing table and do not use default routes
  - All other ISPs' routers use a default route pointing up the hierarchy
- **Outer rim** – lower-tier ISPs peer with each other around Tier 1 ISPs
  - In order to reduce transit costs and achieve performance gains



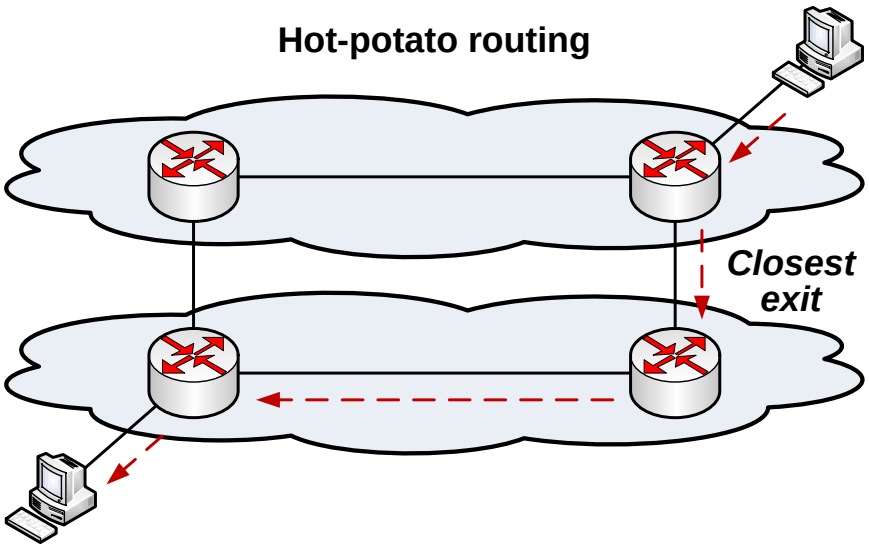


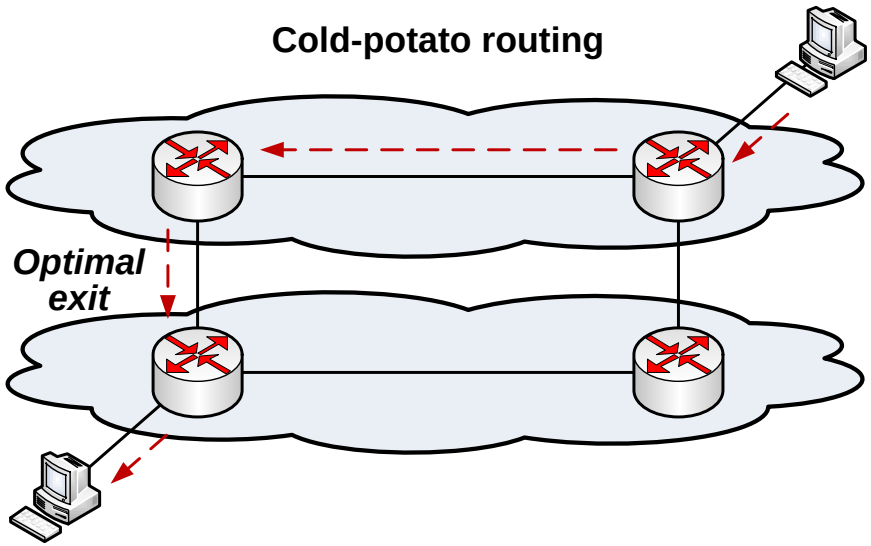
Types of Routing (cont'd)

- **2 types of routing**:
  - Hot-potato (aka **closest-exit**)
  - Cold-potato (aka **optimal-exit**)
- **Hot-potato routing** – the practice of passing traffic off (i.e., 'the hot potato') to another ISP as quickly as possible, thus using their network for wide-area transit
- **Cold-potato routing** – the practice of holding onto traffic until it is as near to the destination as possible, so that customer's costs of carrying it are reduced

Types of Routing (cont'd)

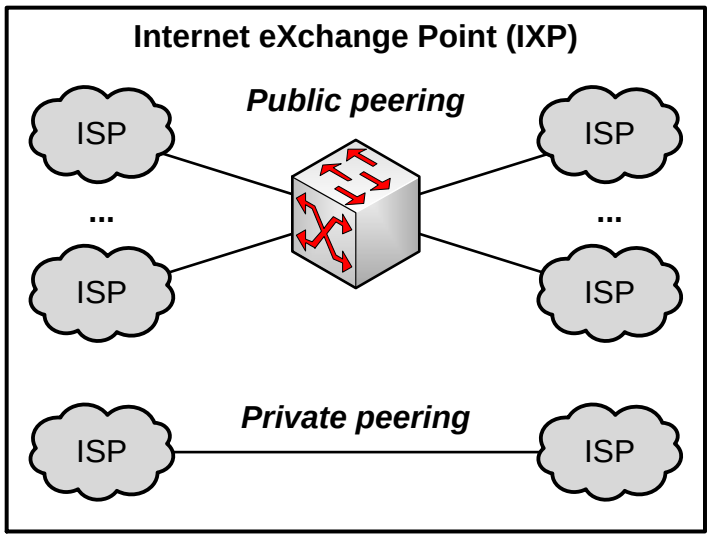
- In **hot-potato** routing, the ISP gets rid of traffic as quickly as possible (more precisely, as inexpensively as possible)
- In **cold-potato** routing, the ISP carries traffic as far as possible on its own network before handing it off to another ISP
- In general, an ISP's routing policy would lie somewhere in between
  - ISPs that offer QoS guarantees often use cold-potato routing for their premium traffic





Types of Peering Agreements

- **2 types of peering agreements**:
  - Public
  - Private
- **Public peering** – an interconnection **between many** ISPs across a Layer 2 technology, generally called a shared peering fabric
- **Private peering** – a direct interconnection **between only 2** ISPs across a Layer 1 or 2 technology that offers dedicated capacity
  - I.e., not shared by any other parties



Types of Peering Policies

- **3 types of peering policies**:
  - Open
  - Selective
  - Restrictive
- **Open policy** – the ISP is willing to peer with anyone (or almost anyone), without imposing specific requirements
- **Selective policy** – the ISP is willing to peer with those who meet a specific set of requirements, but may not peer with everyone
- **Restrictive policy** – the ISP is inclined not to add any new peers
  - Restrictive ISPs may list a set of specific requirements, but with the bar so high that the intention is for no one to reach it
- **Paid peering** – when 2 ISPs agree that one of them derives greater value from peering than the other and should compensate for it

Types of Peering Policies (cont'd)

- **Peering is very much like dating...**
  - Find an attractive partner
  - Get in touch with
  - Make a good impression
  - Establish relations
  - Your relations should be mutually 'beneficial' (how much? equally?)
  - Otherwise, get/give a divorce (**depeering**)
  - etc.

Types of Peering Policies (cont'd)

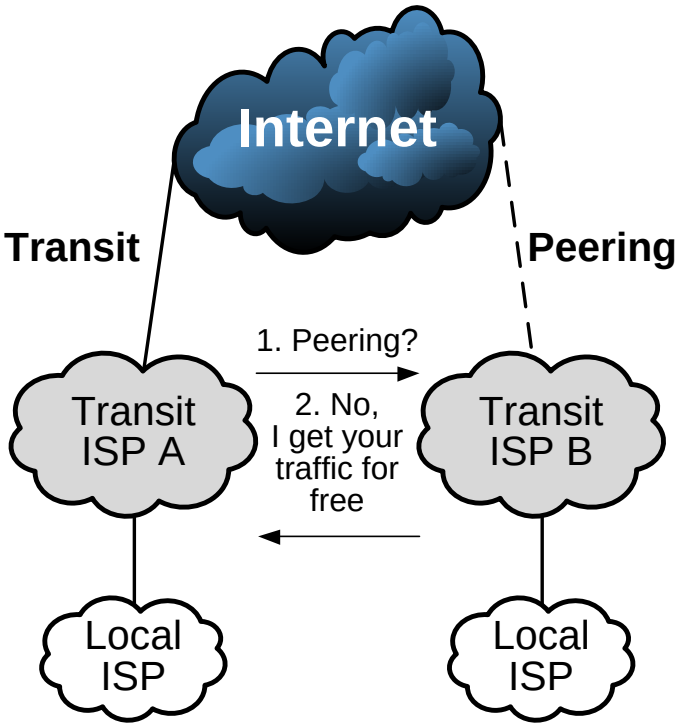
- **. . . in many ways, except monogamy**
  - To hell with monogamy!
  - Peer as much as it provides savings and performance gains!

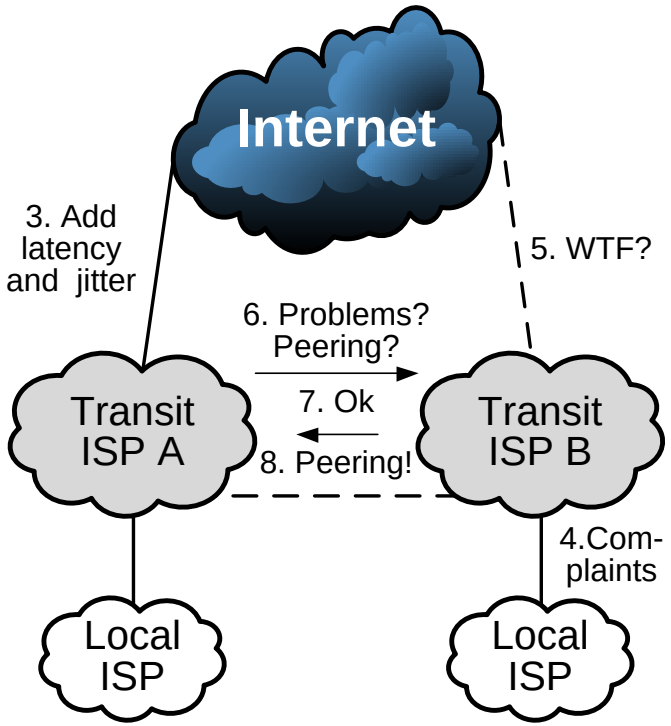
Peering Coordinators

- **Peering coordinator** – a role in charge of developing the peering strategy for an ISP
  1. Analyze the traffic trends
  2. Look for ISPs that are interesting for peering
  3. Contact them (e.g., e-mail to: peering@isp.domain)
  4. Reach an agreement for peering in specific locations and capacities
- Peering coordinators are a new role that combines different skills:
  - Network planning & development
  - BGP expertise
  - Project management
  - Selling
- Exchange of information related to peering: https://www.peeringdb.com

A Bit of Tactics

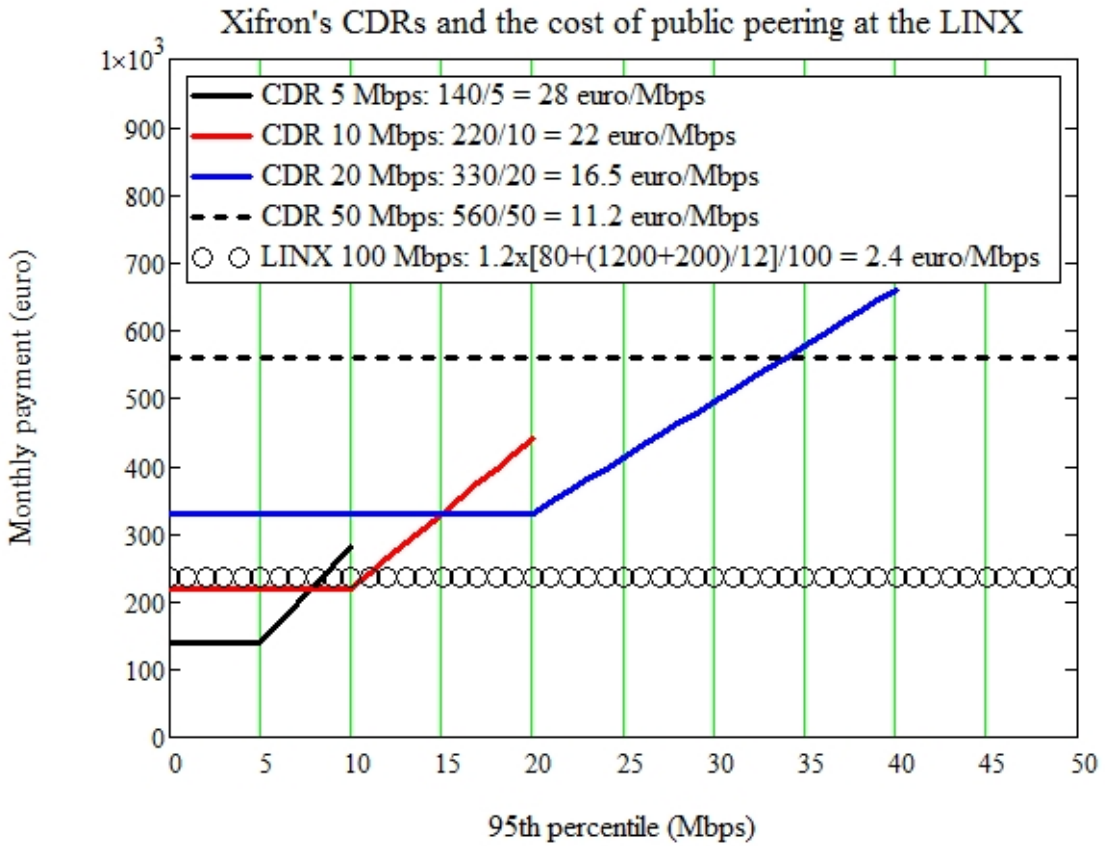
- **Bad tactics:** Bluffing performance problems
  - **Background:** An ISP can be motivated to peer if the other party can demonstrate a significant performance problem that can be solved by peering
  - **Requirement:** No persistent performance monitoring by the target ISP





A Bit of Tactics (cont'd)

- **Good tactics:** Switching between transit and peering
  - **Background:** The more traffic is exchanged between peers, the lower the transit costs
  - **Requirement:** A large amount of traffic to exchange and populated IXPs around



Outline

Summary

	Transit	**Peering** (except paid peering)
ISP interconnection	Settlement-based	Settlement-free
Participants	Providers (higher-tier ISPs), customers (lower-tier ISPs)	Equal peers (peering ISPs)
Payer	Downstream ISP (customer)	No
Payee	Upstream ISP (provider)	No
Billing method	95th percentile	“Sender Keeps All” (SKA)
Traffic exchange	Gives access to the entire Internet	Gives access only to the peer’s customers
Routing (**in theory**)	Cold-potato	Hot-potato
QoS and SLA (**in theory**)	Expectations and penalties are defined in the SLA	No formal SLA and typically “best-effort” delivery

Bibliography

- DrPeering.net, http://drpeering.net/HTML_IPP/index.html
- William B. Norton, 'The Internet Peering Playbook: Connecting to the Core of the Internet', DrPeering Press, 2014