Advanced Computer Networking (ACN)

IN2097 – WiSe 2020-2021

Prof. Dr.-Ing. Georg Carle

Sebastian Gallenmüller, Max Helm,
Benedikt Jaeger, Patrick Sattler, Johannes Zirngibl

Chair of Network Architectures and Services
Department of Informatics
Technical University of Munich

Routing and Forwarding

Routing and Forwarding
Routers vs. Switches

Both store-and-forward devices

- routers: L3/network layer devices (examine network layer headers)
- switches: L2/link layer devices

Routers:

- maintain forwarding tables
- implement routing protocols
- forward IP packets based on forwarding table entries & destination IP address

Switches:

- maintain switching tables
- implement filtering, learning algorithms
- may implement spanning tree algorithm
- switch L2 frames based on switching table entries & destination MAC address

Routing and Forwarding
Network prefix and host number

- L3/IP service goal: forward IP datagrams to destination IP subnet/host/interface
- IP address has role of locator & identifier:
  - network part (network identifier & locator)
  - host part (host identifier)

- Each IP network (often called subnetwork or subnet) has an IP address:
  - IP address of a network = Host number is set to all zeros, e.g., 128.143.0.0

- IP routers are devices that forward IP datagrams between IP networks

- Delivery of an IP datagram proceeds in 2 steps:
  1. Use network prefix to deliver IP datagram to the right network
  2. Once the network is reached, use the host L3 address to deliver to the right interface

Routing and Forwarding
Forwarding and routing

Forwarding: data plane

- Directing a data packet to an outgoing link
- Individual router using a forwarding table

Routing: control plane

- Computing the paths the packets will follow
- Individual router creates a forwarding table

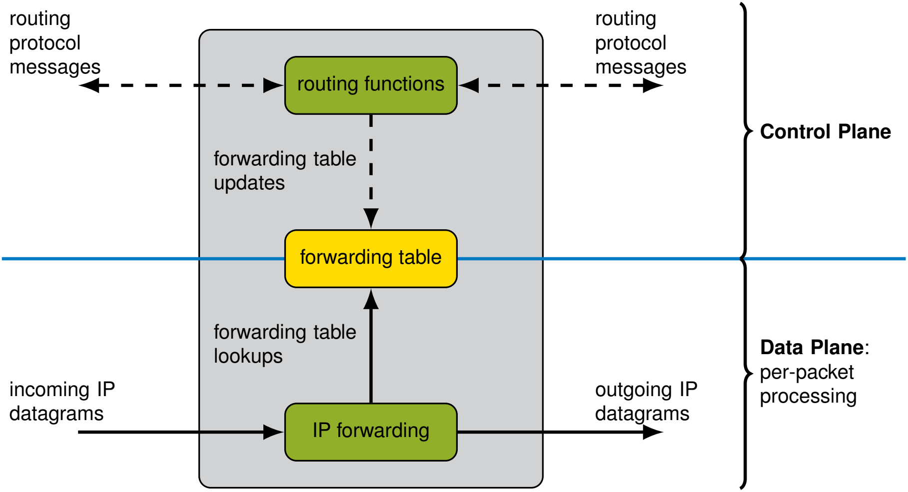
Routing protocols (e.g., RIP, OSPF, BGP):

- Distributed routing protocol entities in routers talking amongst themselves (above IP)

Routing in Software Defined Networking (SDN):

- Router may receive forwarding and/or routing information via SDN control plane
- Control plane might not necessarily be inside the router

Routing and Forwarding
IP router: functional components



Routing and Forwarding
Routing and forwarding

Routing functions include:

- route calculation
- execution of routing protocols
- maintenance of routing state
- maintenance of forwarding table

On commercial routers handled by a single general purpose processor, called routing processor

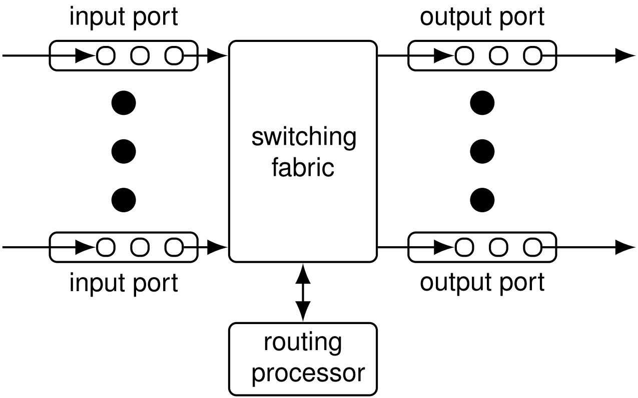
IP forwarding is per-packet processing

- On high-end commercial routers, IP forwarding is distributed (Most work is done on the interface cards)

Routing and Forwarding
Router architecture

Two key router functions

- run routing algorithms/protocol (RIP, OSPF, BGP)
- forwarding datagrams from incoming to outgoing link



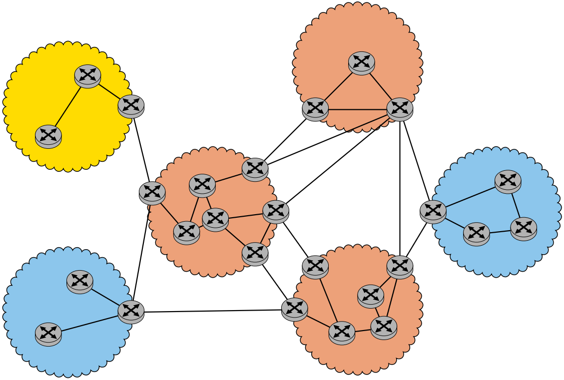
Autonomous Systems
Internet Architecture

How is the Internet built?

- Comprised of independent networks: Autonomous Systems (ASes), of different types
  - Internet Service Providers (ISPs) offering transit services (e.g. Deutsche Telekom, Vodafone, DFN)
  - Campus Networks: Organisations maintaining networks at one or more locations (e.g. universities, companies)
  - Datacenter Networks: Operated by large IT companies (e.g. Facebook) and Hosting providers, e.g., Amazon AWS
  - Content Delivery Networks (e.g., Akamai, Google, Cloudflare)
- Organisations may manage one or more Autonomous System(s)
- Some ASes distribued world-wide, some regional

Routing protocols find a path towards a destination and allow to create Forwarding tables
Routers use Forwarding tables for per-packet lookup to determine outgoing interface

Autonomous Systems
Example Network



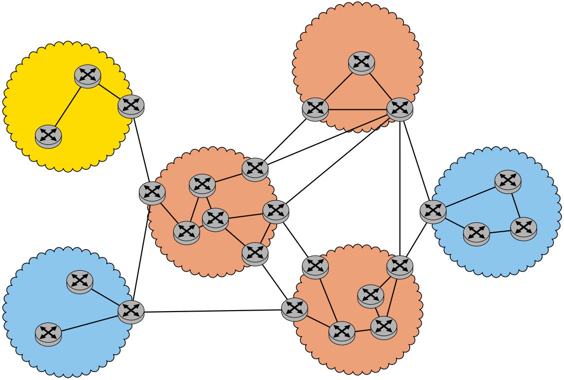
Router Types

- Border router
- Core router
- Edge router

Autonomous Systems
Example Network

Network Types

- Transit AS:
  - Forwards traffic from one AS to another AS (red)
- Stub AS:
  - AS, which is connected to only one other AS (yellow)
- Multi-homed AS:
  - AS, which is connected to multiple ASes, but doesn't forward traffic on their behalf (blue)



Autonomous Systems
Routing Protocol Types

Inter-Domain Protocols (inter AS routing)

- Exchange routing information between ASes
- Called Exterior Gateway Protocols (EGPs)
- In practice, only the Border Gateway Protocol version 4 (BGPv4) is used

Intra-Domain Protocols (intra AS routing)

- Used to determine path and routing information inside an AS
- Called Interior Gateway Protocols (IGPs)
- Examples: OSPFv2/3, IS-IS, RIP, ...

What is IGP, and what EGP?

- EGP: Which AS to transfer the packet to?
- IGP: Which intra-AS route to use to reach this neighboring AS?

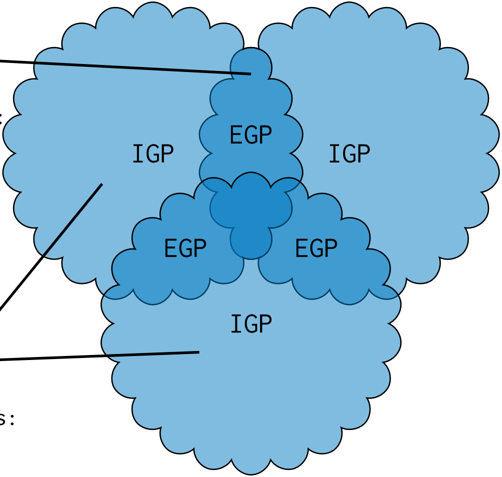
Autonomous Systems

EGP:

Exterior Gateway Protocol

Inter-domain routing protocol:

BGP v4



IGP:

Interior Gateway Protocol

Intra-domain routing protocols:

RIP, OSPF, IS-IS

Autonomous Systems
Terminology

- Routing Information Base (RIB): All routing information a router can gather from updates of neighboring routers
  - May contain multiple routes to the same destination
  - Path selection also depends on business considerations (policy routing)
- Forwarding Information Base (FIB): Mapping from a destination IP network address (prefix) to outgoing interface or next hop router IP address
  - Unique entry for each destination
  - Uses Longest-Prefix-Matching (LPM)
- Forwarding Decision: Algorithm uses the FIB to decide how to forward individual packets

RIB

Prefix	**Next Hop**	**Cost Metric**
10.0.2.128/25	A	90
192.168.2.0/24	B	60
192.168.3.0/24	B	30
192.168.0.0/16	C	50
192.168.0.0/16	D	20
10.0.0.0/8	E	70
0.0.0.0/0	F	100
0.0.0.0/0	G	20

⟶

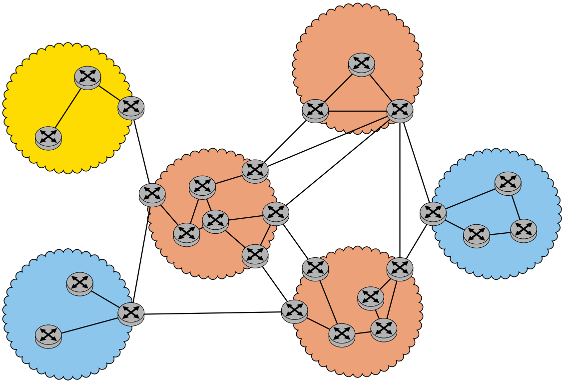
FIB

Prefix	**Next Hop**
10.0.2.128/25	A
192.168.2.0/23	B
192.168.0.0/16	D
10.0.0.0/8	E
0.0.0.0/0	G

Autonomous Systems
Interconnected ASes

Forwarding table entries

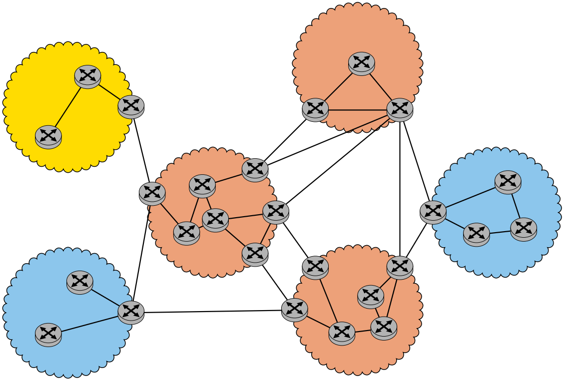
- Intra-AS routing algorithm sets entries for internal destinations
- Inter-AS and intra-AS routing algorithms set entries for external destinations



Autonomous Systems
Interconnected ASes

Inter-AS tasks

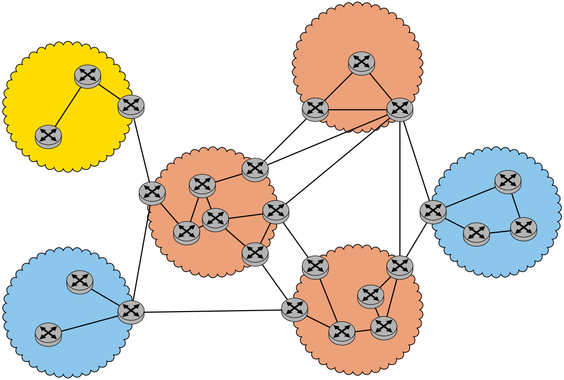
- Suppose core router in a red AS receives datagram destined outside of AS
- Core router should forward packet to border router, but which one?
- Red ASes must learn which destinations are reachable through which neighboring AS
- Red ASes must propagate this reachability info to all routers within AS (i.e., not just the border routers)
- This is job of inter-AS routing protocol BGP



Autonomous Systems
Interconnected ASes

Setting forwarding table in a router

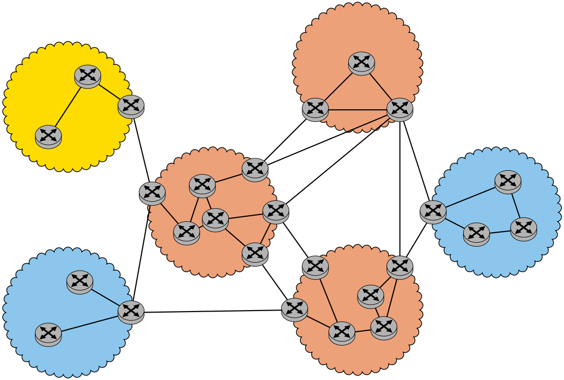
- Suppose AS1 learns (via inter-AS protocol - BGP) that subnet x is reachable via one neighbor AS2 but not via another neighbor AS3.
- Inter-AS protocol propagates reachability information to all internal routers.
- Core router determines from intra-AS routing information which interface is on the least cost path to border router towards AS2.
- Core router installs forwarding table entry for that subnet x



Autonomous Systems
Interconnected ASes

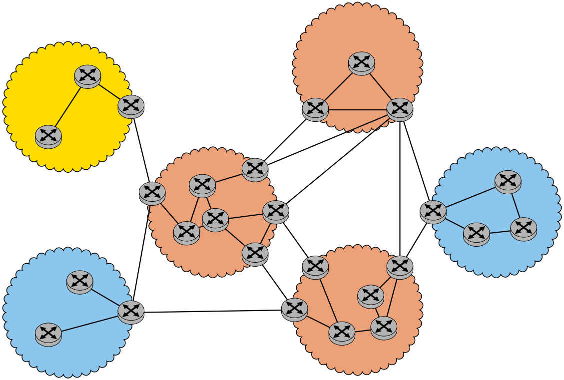
Choosing among multiple ASes

- Now suppose AS1 learns from inter-AS protocol that subnet x is reachable from AS3 and from AS2.
- To configure forwarding table, core router must determine towards which border router it should forward packets for destination x.
- This is also job of inter-AS routing protocol BGP



Inter-AS and intra-AS routing

- Inter-AS routing
  - Only for destinations outside of own AS
  - Used to determine gateway router
  - Also: steers transit traffic (from AS i to AS j via our own AS)
- Intra-AS routing
  - Used for destinations within own AS
  - Used to reach gateway router for outside destinations



Routing and Forwarding

Border Gateway Protocol [1]

Basics

- Each AS has a unique AS Number (ASN)
  - e.g., Vodafone UK (AS6847), Vodafone DE (AS6751), LRZ (AS12816), I8 net.in.tum (AS56357)
- BGP is based on the Path-Vector model
- BGP router exchange information derived from their routing table entries
- Which next AS (and as a consequence which next router) to choose is a policy (business) decision
- Path-Vector: UPDATE Messages contain all ASes on the path towards a destination network (prefix)
- AS-level Loops can be noticed if an ASN is contained multiple times in a path

There are two BGP variants:

- Internal BGP (iBGP): BGP exchanges information with routers in the same AS
- External BGP (eBGP): BGP exchanges information with routers of neighboring ASes

Hierarchical prefix announcement using BGP

How does an organization get a subnet?

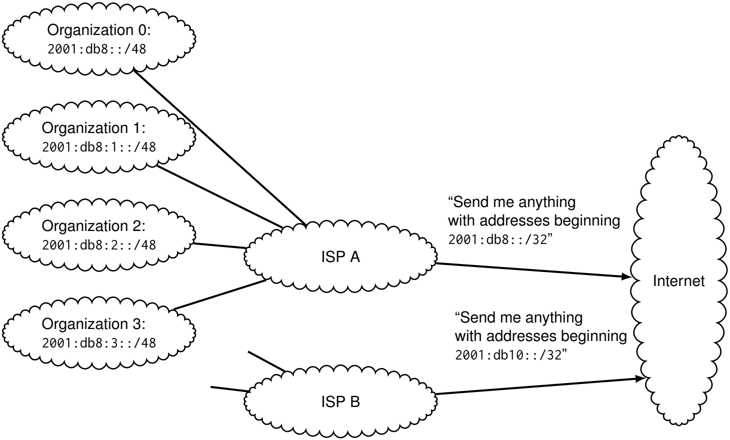
- Gets allocated portion of its provider ISP's address space

Example:

- ISP: `2001:db8::/32`
- Organization 0: `2001:db8::/48`
- Organization 1: `2001:db8:1::/48`
- Organization 2: `2001:db8:2::/48`
- Organization 3: `2001:db8:3::/48`
- ...
- Organization 7: `2001:db8:7::/48`

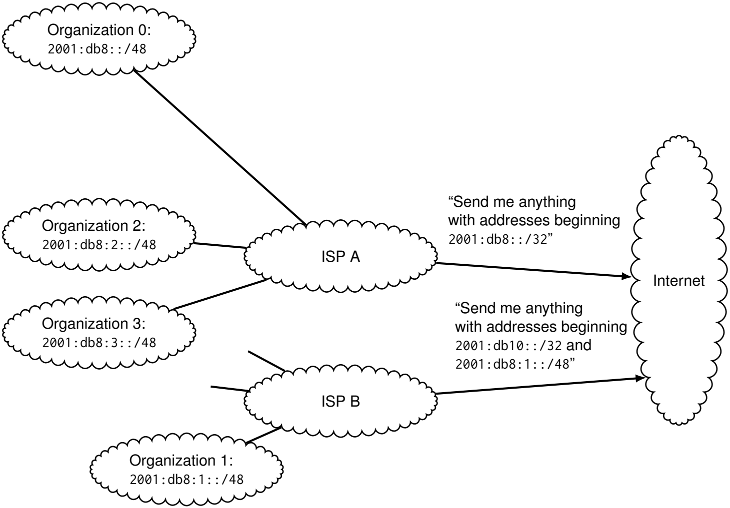
Hierarchical prefix announcement using BGP

Hierarchical addressing allows for efficient advertisement of routing information:



Hierarchical prefix announcement using BGP

ISP B has a more specific route to Organization 1:



BGP Overview
iBGP

Difference between iBGP and eBGP

- iBGP: Both routers have the same ASN
- eBGP: Routers have different ASNs
- iBGP: Propagates information on externally reachable prefixes to routers within AS

Message Types

c.f. RFC 4271 BGP uses different message types, transferred over TCP

- OPEN: Opens a connection between two routers
- TEARDOWN: Close the connection
- NOTIFICATION: Send error codes
- UPDATE: Announce a new route, or un-reachability of an old one

Update Messages

Message contains:

- Destination prefix (a.b.c.d/x)
- AS path - list of ASes
- Next Hop - IP address of the router sending the update
- Origin - Learned via IGP/EGP/other

More attributes can be added:

- LOCAL_PREF “Local preference”: Used to prefer one gateway over another
- MULTI_EXIT_DISC “Multiple exit discriminator”: If multiple entries (i.e., border routers) into an AS exists, says which is preferred

Local information is used on top of the information provided in the update messages, in order to make a routing decision.

Each AS writes its own ASN at the beginning of the AS path. This is important for the loop detection.

BGP is an “information hiding protocol” (quote from Randy Bush)

Example Update Message

Information

Prefix: 185.86.232.0/22

AS Path: 202109 33891 48918 56357

Next Hop: 5.101.111.2

Origin: IGP

Explanation:

- This update concerns itself with the destination subnet 185.86.232.0/22
- Packets will be routed through 4 ASes
- The next hop is 5.101.111.2
- The next hop learned this route through an IGP

BGP route selection

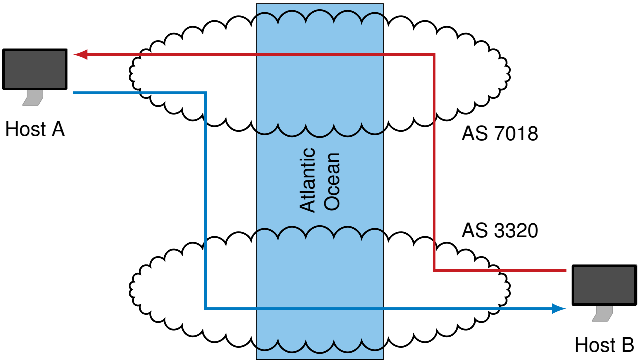
Router has to select which route to choose (multiple may be available)

1. Local preference value attribute: policy decision
2. Shortest AS-PATH
3. Closest NEXT-HOP router outside AS: hot potato routing
4. Additional criteria (e.g., lowest Multi-Exit Discriminator MED when different border routers to same neighboring AS exist)

Hot potato routing

Large ASes connect at different sites

- Where to hand over traffic destined for the other AS?
- Always hand it over as fast as possible
- The longer a packet is transferred inside an AS, the more is costs
- Choosing the “nearest” connection site minimizes the cost
- Leads to asymmetric routing



Peering Types

There are different types of peering relationships

- Private: Direct connection to (frequently large) AS
- Public: Exchanging traffic with other ASes at an Internet Exchange Point (IXP)

Peering is preceded by a peering agreement, which may be made by two sysadmins talking to each other, or using lengthy contracts.

There exist multiple definitions of peering. In this course, peering is defined as follows:

- Two ASes peer with each other, if they have some kind of BGP relationship, i.e., two ASes are directly connected
- This reflects the *protocol viewpoint: It is irrelevant if one party pays the other party*
- *Alternative viewpoint - the* policy viewpoint: It is very relevant if one party pays the other party

Private Peering

Private peering can be accomplished in a variety of ways:

- Install a cable from the server room of AS 1 to the server room of AS 2
- Using multi-tenant data center providers operating facilities for servers and communication systems that offers redundant electricity, connectivity, cooling, fire suppression, and security services
- Peering possible at colocation center operated by a carrier-neutral data center provider
  - Examples: Interxion (pronounced “interaction”), e-shelter
- Peering possible at colocation center operated by a carrier
  - Example: AT&T, Verizon, Level 3 Communications

Keep in mind: Different rules exist regarding connections between customers, ranging from no charge to significant monthly fee

Private peering use cases:

- Exchange a large amount of traffic with a single AS
- Attractive setup for upstream providers
- Interconnection of, within, and inbetween data centers

Public Peering

Public peering is done, by meeting at peering locations

- “A room full of switches that many providers connect to”
- Configure L2 (VLAN) connections in switch, instead of having to put in $\mathcal{O}(n^2)$ separate wires
- Payment per “switch port”, priced by connection speed

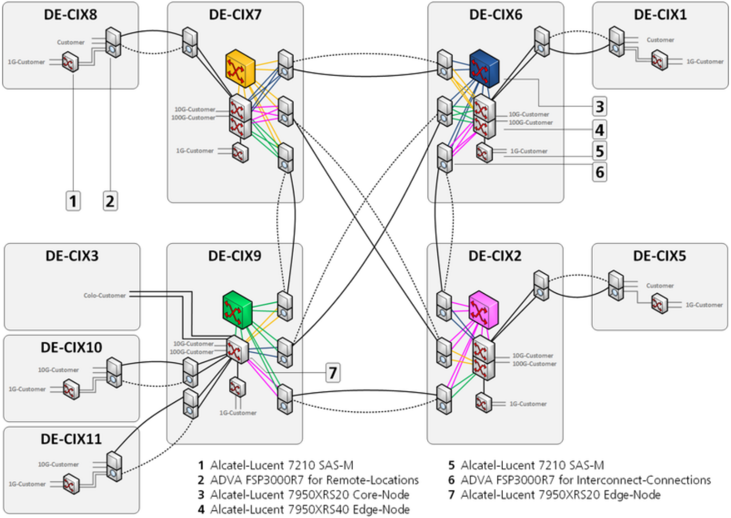
Examples:

- DE-CIX, Frankfurt (by peak traffic largest in world)
- AMS-IX, Amsterdam
- LINX, London
- MSK-IX, Moscow

Public peering use cases:

- Peer with as many ASes as possible
- Reduce the traffic you send to your upstream provider

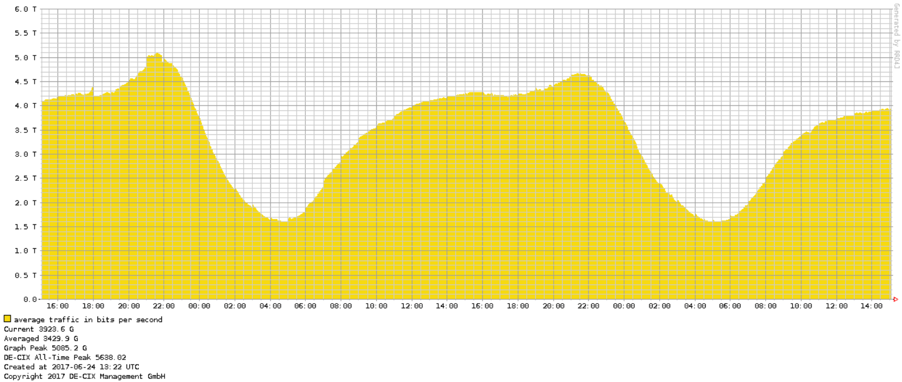
BGP Peering
DE-CIX Topology



1 Alcatel-Lucent 7210 SAS-M
2 ADVA FSP3000R7 for Remote-Locations
3 Alcatel-Lucent 7950XRS20 Core-Node
4 Alcatel-Lucent 7950XRS40 Edge-Node
5 Alcatel-Lucent 7210 SAS-M
6 ADVA FSP3000R7 for Interconnect-Connections
7 Alcatel-Lucent 7950XRS20 Edge-Node

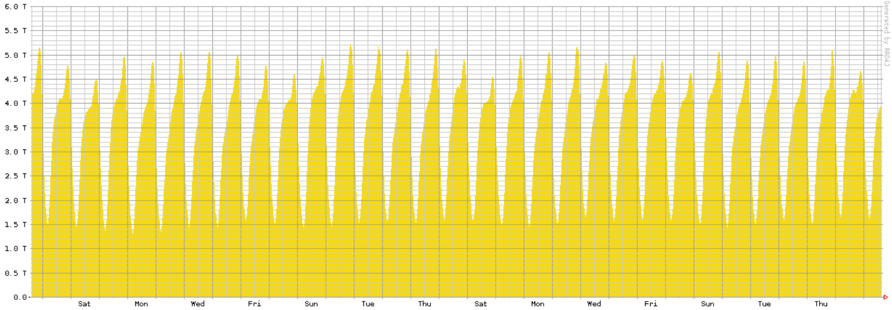
Source: de-cix.net/about/topology/ (2015)

BGP Peering
DE-CIX Statistics - 2 Days



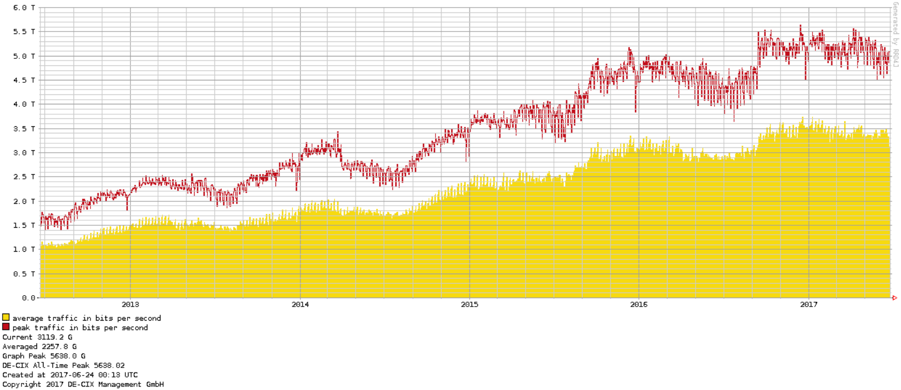
BGP Peering

DE-CIX Statistics - 1 Month



average traffic in bits per second
Current 3923.6 G
Averaged 3362.2 G
Graph Peak 5229.2 G
DE-CIX All-Time Peak 5638.02
Created at 2017-06-24 13:04 UTC

BGP Peering
DE-CIX Statistics - 5 Years



[1] Y. Rekhter, T. Li, and S. Hares. A Border Gateway Protocol 4 (BGP-4). https://tools.ietf.org/html/rfc4271. 2006.