

# Advanced Computer Networking (ACN)

IN2097 – WiSe 2023–2024

**Prof. Dr.-Ing. Georg Carle**

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

Chair of Network Architectures and Services  
School of Computation, Information, and Technology  
Technical University of Munich

## Autonomous Systems

### BGP

- BGP Overview

- BGP Peering

- Business and Policy Routing

- k-core Algorithm

### Bibliography

Autonomous Systems

BGP

Bibliography

# 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 distributed world-wide, some regional

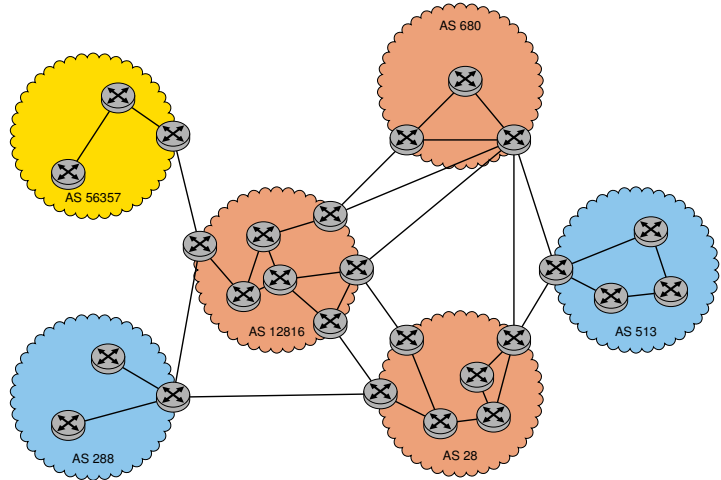
Routing protocols find a path towards a destination and allow to create Forwarding tables  
Routers use Forwarding tables for per-packet lookup to determine outgoing interface

# Autonomous Systems

## Example Network

### Router Types

- Border router (Edge router)
- Core router

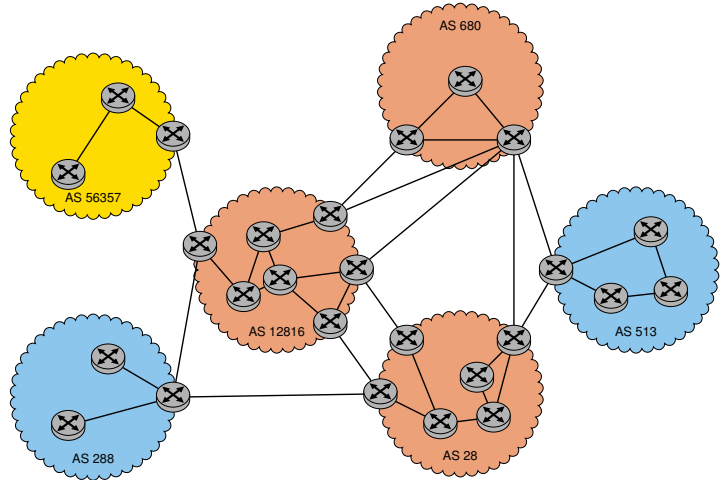


# Autonomous Systems

## 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)

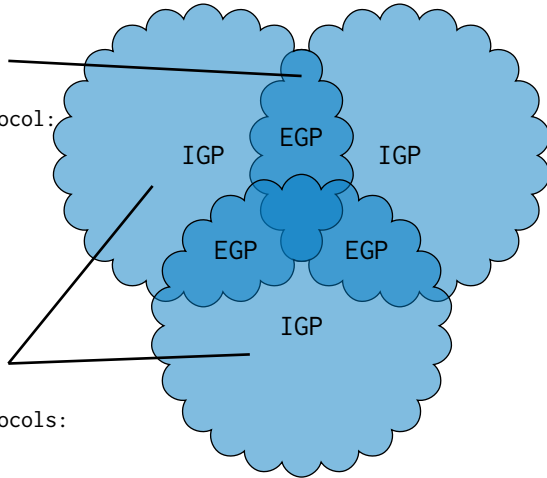
- Used to determine path and routing information inside an AS
- Called **Interior Gateway Protocols (IGPs)**
- Examples: OSPFv2/3, IS-IS, RIP, ...

### What is responsibility of EGP, and what of IGP?

- EGP: Which AS to transfer the packet to?
- IGP:
  - Destination in other AS: Which intra-AS path to reach next AS?
  - Destination inside AS: Which intra-AS path to reach destination?

EGP:  
Exterior Gateway Protocol  
Inter-domain routing protocol:  
BGP v4

IGP:  
Interior Gateway Protocol  
Intra-domain routing protocols:  
RIP, OSPF, IS-IS





# Autonomous Systems

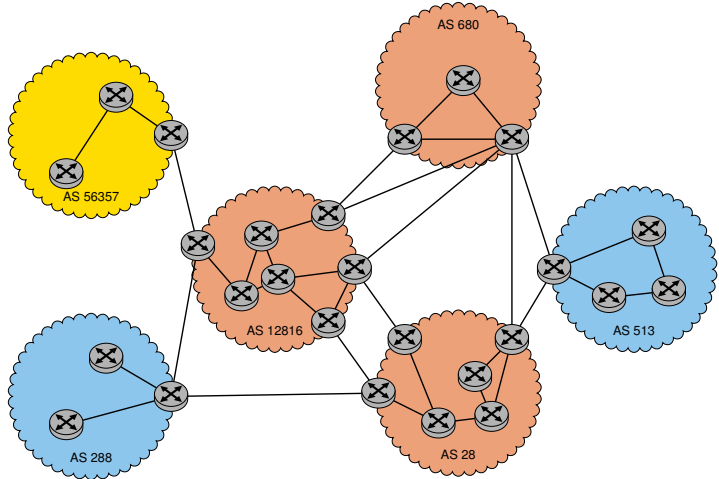
## Terminology

- **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				FIB	
Prefix	Next Hop	Cost Metric		Prefix	Next Hop
10.0.2.128/25	A	90	→	10.0.2.128/25	A
192.168.2.0/24	B	60		192.168.2.0/23	B
192.168.3.0/24	B	30		192.168.0.0/16	D
192.168.0.0/16	C	50		10.0.0.0/8	E
192.168.0.0/16	D	20		0.0.0.0/0	G
10.0.0.0/8	E	70			
0.0.0.0/0	F	100			
0.0.0.0/0	G	20			

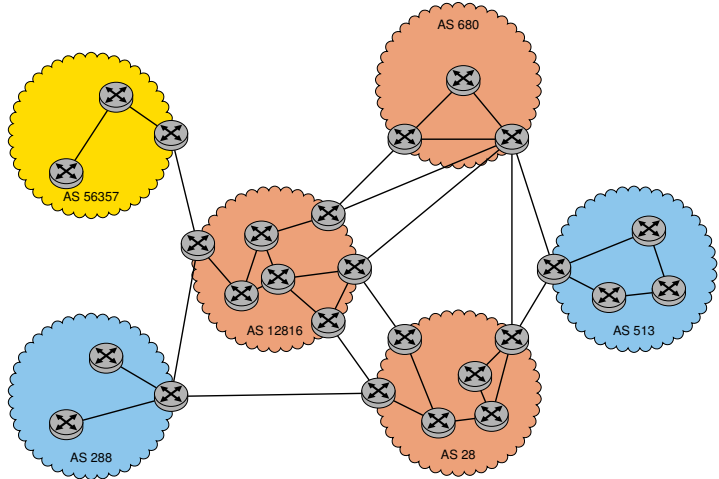
## Forwarding table entries

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



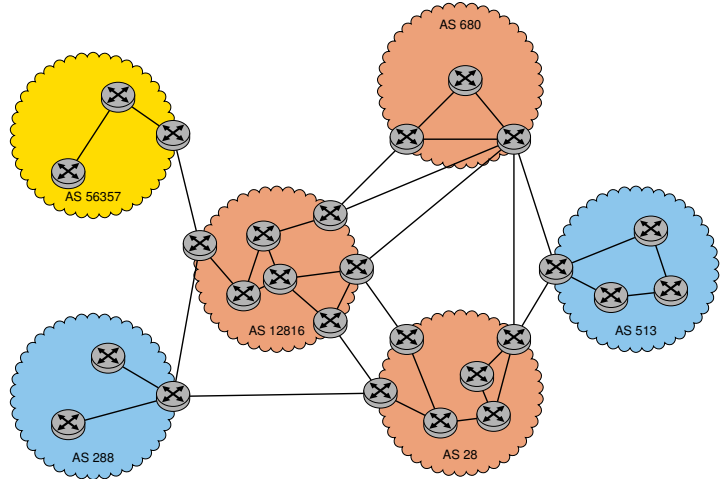
### Inter-AS tasks

- 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



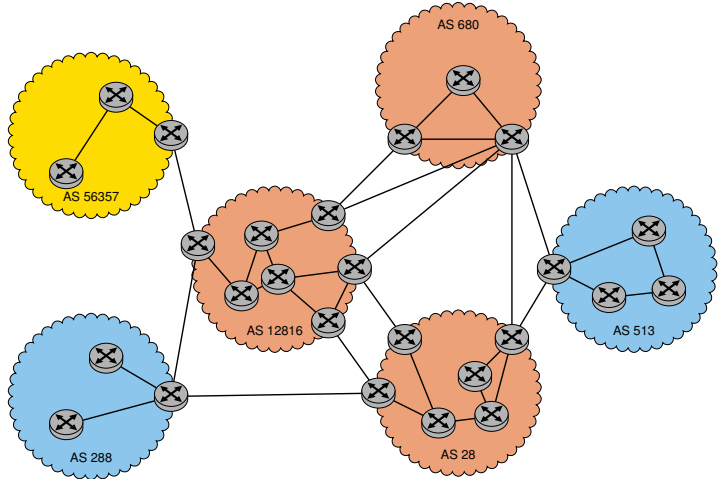
### Setting forwarding table in a router

- 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



## 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

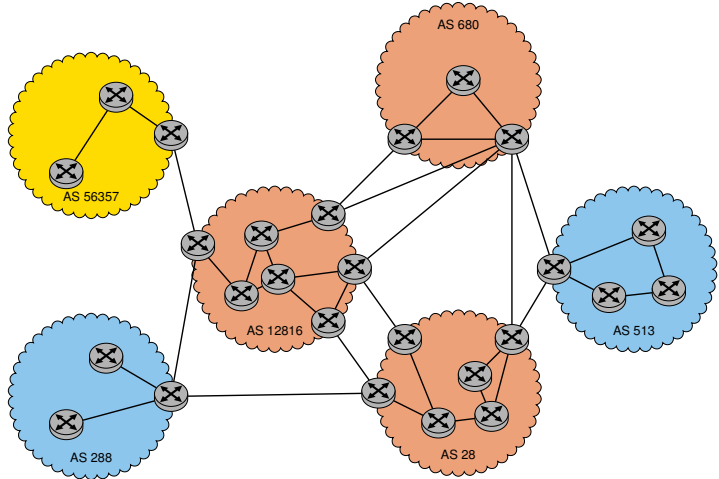


# Autonomous Systems

## Interconnected ASes

### 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

## Autonomous Systems

### BGP

BGP Overview

BGP Peering

Business and Policy Routing

k-core Algorithm

## Bibliography

## 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



How does an organization get a subnet?

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

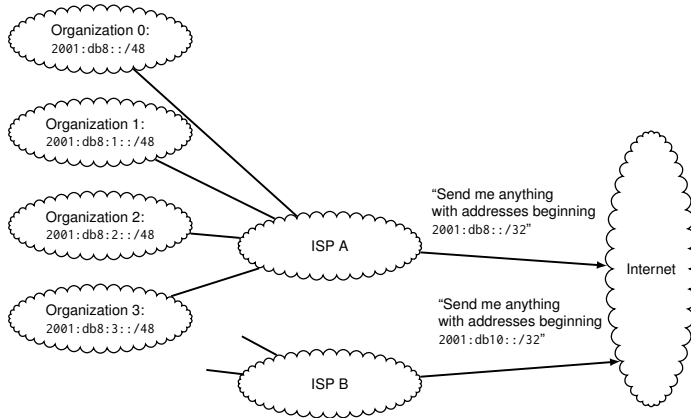
Example:

- 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

## BGP Overview

## Hierarchical prefix announcement using BGP

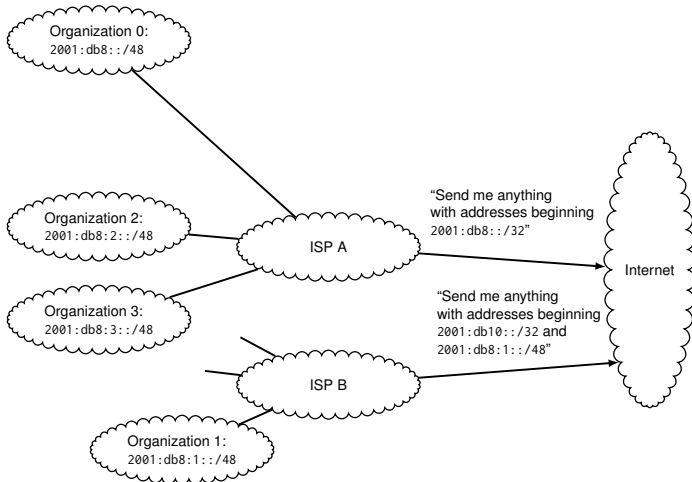
Hierarchical addressing allows for **efficient advertisement** of routing information:



## BGP Overview

## Hierarchical prefix announcement using BGP

ISP B has a **more specific** route to Organization 1:



### 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

## BGP Overview

### Message Types

BGP uses different message types, transmitted over TCP (see RFC 4271)

- **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

## BGP Overview

### 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)

## BGP Overview

### Example Update Message

#### Information

Prefix: 185.86.232.0/22

AS Path: 202109 33891 48918 56357

Next Hop: 5.101.111.2

Origin: EGP

#### 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 the EGP

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)

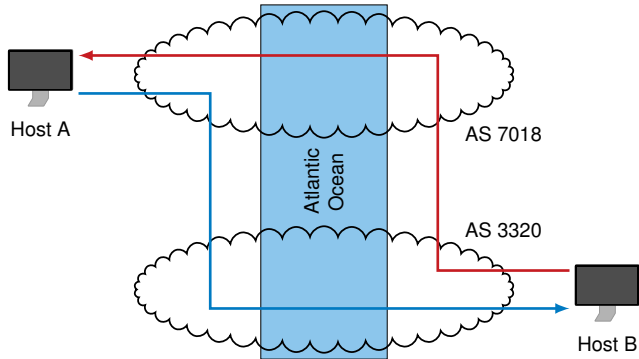


## BGP Overview

### 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



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, the primary meaning (technical viewpoint: *protocol viewpoint*) of 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. Here, more narrowly peering implies that there is no financial compensation.*

## BGP Peering

### 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
- 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

## BGP Peering

### 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

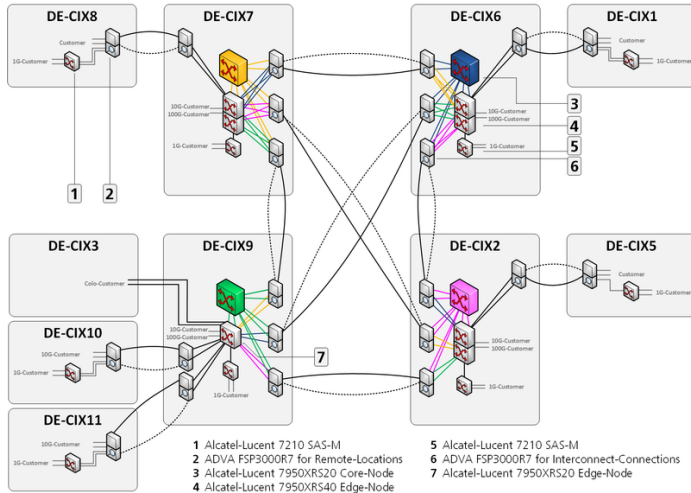
Examples:

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

Public peering use cases:

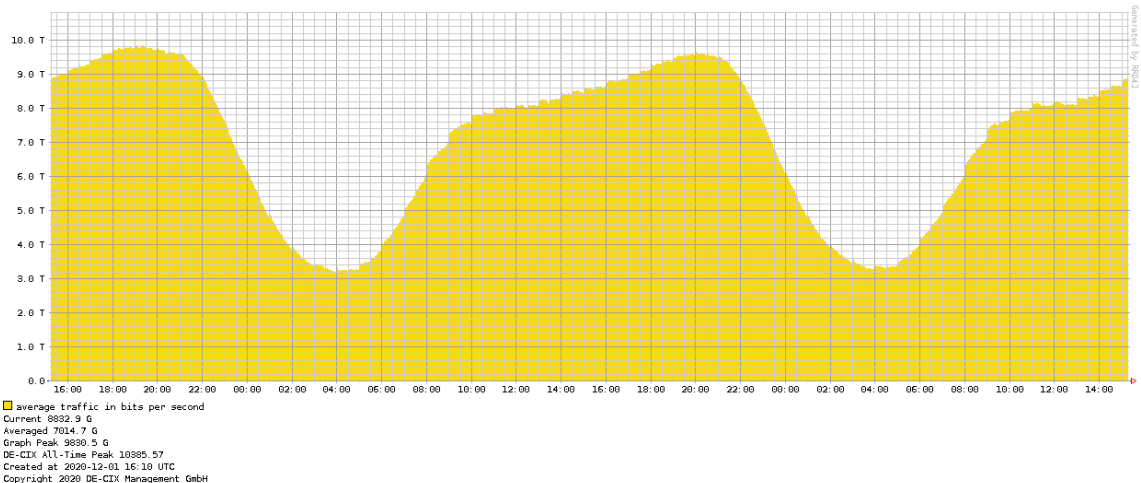
- Peer with as many ASes as possible
- Reduce the traffic you send to your upstream provider

# BGP Peering DE-CIX Topology



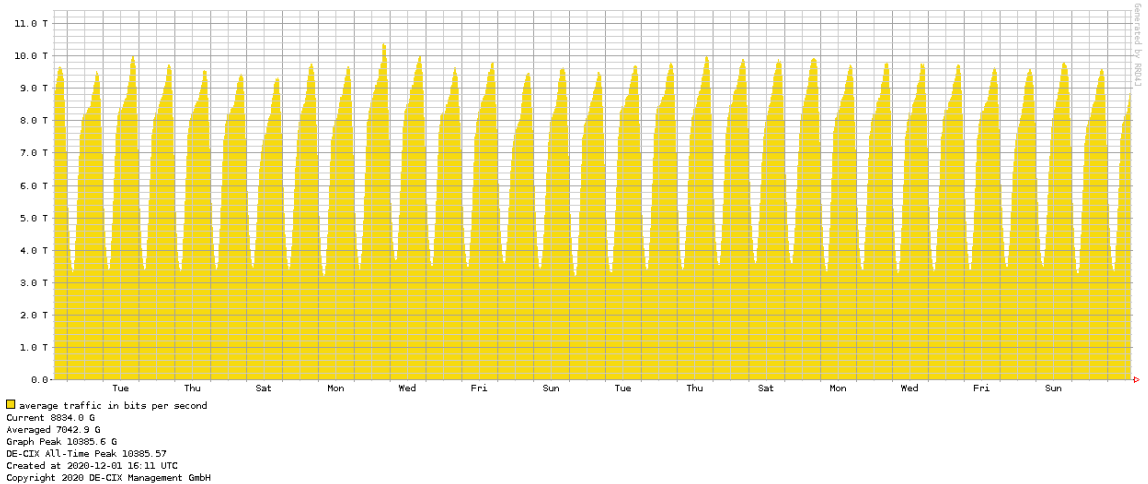
Source: <https://de-cix.net/about/topology/> (2015)

## BGP Peering DE-CIX Statistics - 2 Days



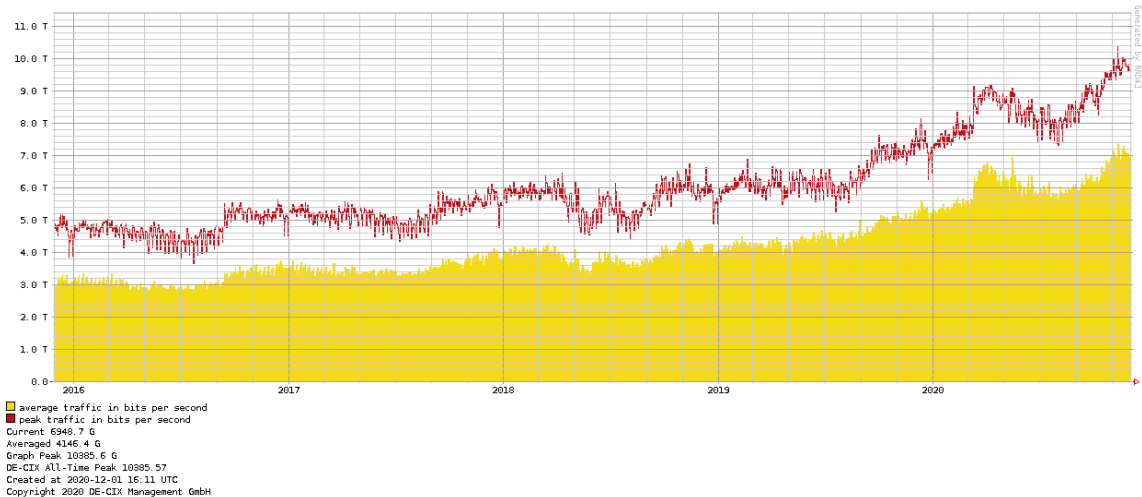
Source: <https://www.de-cix.net/en/locations/germany/frankfurt/statistics>

# BGP Peering DE-CIX Statistics - 1 Month



Source: <https://www.de-cix.net/en/locations/germany/frankfurt/statistics>

## BGP Peering DE-CIX Statistics - 5 Years



Source: <https://www.de-cix.net/en/locations/germany/frankfurt/statistics>



## Business and Policy Routing

### Basic Principles

#### Routing: Prefer routes that incur financial gain

1. Route via a customer (financial gain)
2. Route via a peer (no financial gain or loss)
3. Route via a provider (financial loss)

#### Route Announcement

- Announce routes that incur financial gain if others use them
  - Others = customers
- Announce routes that reduce costs if others use them
  - Others = peers
- Do not announce routes that incur financial loss  
(... as long as alternative paths exist)

## Business and Policy Routing

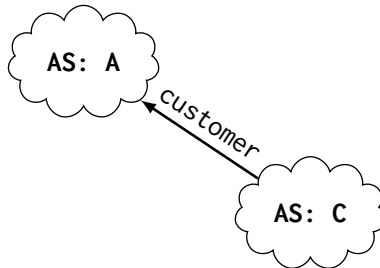
### Situation: Stub ASes

#### Remember:

- Stub ASes have exactly one BGP relationship (with their provider)

#### Rules

- Provider AS announces routes for reaching the whole internet to customer AS
- Customer AS announces routes for reaching its prefixes to provider AS
- The more traffic the customer sends to the provider or receives from the provider, the more money the provider makes



## Business and Policy Routing

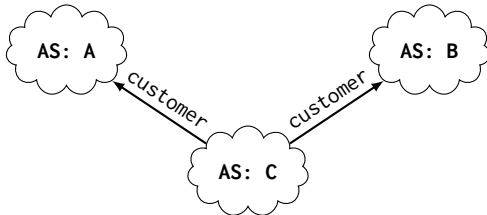
### Situation: Multi-Homed ASes

#### Remember:

- Multi-Homed ASes have multiple providers

#### Rules

- Several provider ASes announce routes for reaching the whole internet to customer AS
- Customer AS announces routes for reaching its prefixes to several provider ASes
- The more traffic the customer exchanges with one provider, the more money that provider makes
- Customer chooses the cheapest / best-quality provider
- Customer only announces its own prefixes to the providers
  - In case customer (by mistake) announces other prefixes, providers would exchange traffic over the customer network: No benefit for the customer, but financial loss
- Customer may use AS path prepending, by adding his AS number several time to path in BGP announcement, to discourage the use of expensive providers for incoming traffic



## Business and Policy Routing

### Path Prepending

#### Trying to influence other peoples routing decisions

- Router of the source AS may use “Shortest AS-PATH” as criterion for choosing where to hand over the packet to the destination AS

#### Example:

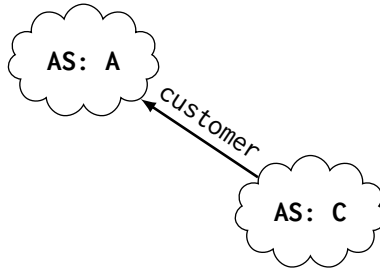
- The destination AS has one router in Munich, and one router in London
- The destination datacenter is also in Munich
- The source AS can choose to hand over the packet in London or Munich
- It is cheaper for the destination AS if the source chooses the Munich router
- The destination AS includes its own ASN multiple times in the BGP updates it sends in London
- The AS-path to Munich is shorter than the AS-path to London
- This may (or may not) convince a source AS to hand over traffic in Munich

## Business and Policy Routing

### Business Routing Example (1)

A tells C routes to all reachable prefixes

- The more traffic comes from C or vice versa, the more money A makes

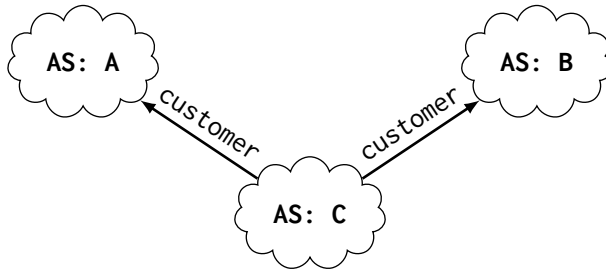


## Business and Policy Routing

### Business Routing Example (2)

A and B tell C routes to all reachable prefixes

- The more traffic flows from C to A or vice versa, the more money A makes
- The more traffic flows from C to B or vice versa, the more money B makes
- C will pick the one with the cheaper offer / better quality / ...



## Business and Policy Routing

### Business Routing Example (3)

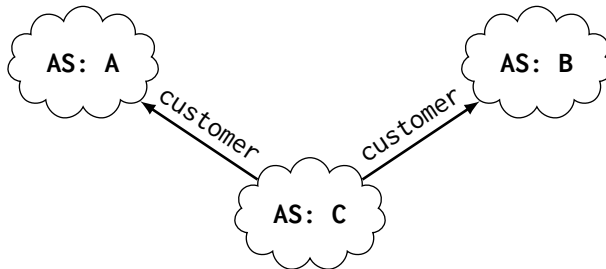
C tells A its own prefixes, C tells B its own prefixes

- C wants to be reachable from outside

C does not tell A routes learned from/via B

C does not tell B routes learned from/via A

- C does not want to pay money for traffic  $\dots \leftarrow A \leftarrow C \leftarrow B \leftarrow \dots$



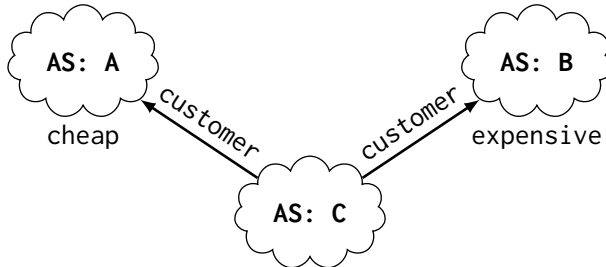
## Business and Policy Routing

### Business Routing Example (4)

C tells A its own prefixes

C may tell B its own prefixes

- ... but inserts "C" multiple times into AS path (Path Prepending)
- Result: Route available, but longer path = less attractive



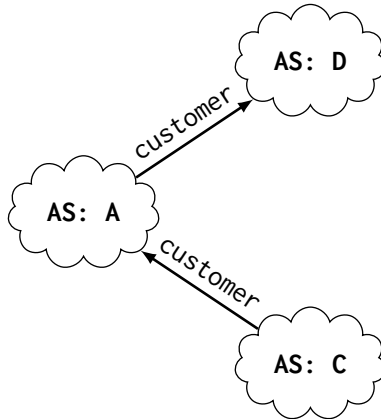


## Business and Policy Routing

### Business Routing Example (5)

What should A announce here

- C tells A about its own prefixes
- A tells D about its route to C's prefixes: pays money to D, but gains money from C

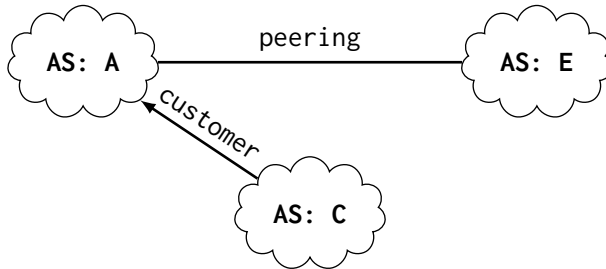


## Business and Policy Routing

### Business Routing Example (6)

What should A announce to E

- A tells peering partner E about its own prefixes and route to C's prefixes
- no cost on link to E, but gains money from C

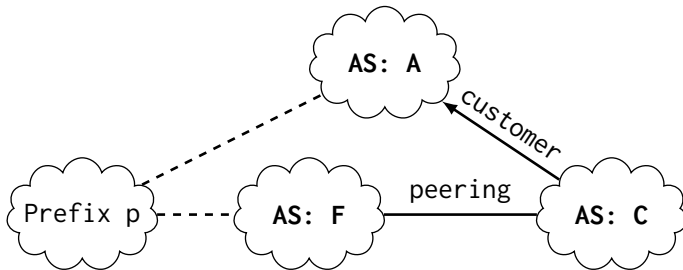


## Business and Policy Routing

### Business Routing Example (7)

Which route to prefix p does C receive, which should C select?

- A tells C about route to prefix p (C loses money)
- F tells C about route to prefix p (no cost involved)
- C prefers route via F

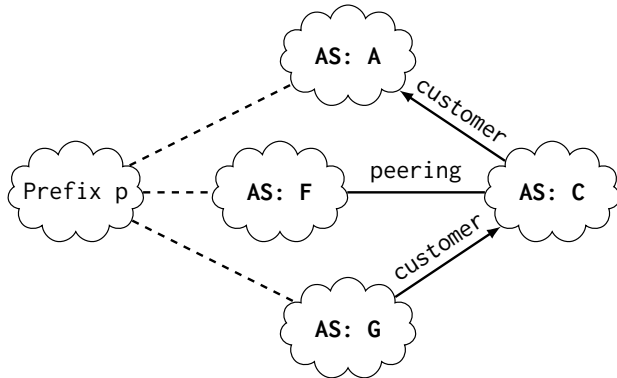


## Business and Policy Routing

### Business Routing Example (8)

Which route to prefix p does C receive, which should C select?

- A tells C about route to prefix p (C loses money)
- F tells C about route to prefix p (no cost involved)
- G tells C about route to prefix p (C gains money)
- C prefers route via G

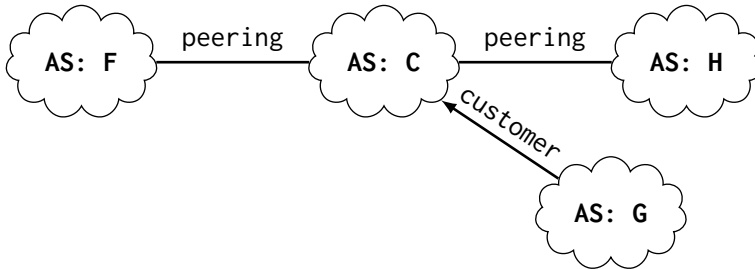


## Business and Policy Routing

### Business Routing Example (9)

What should C announce here?

- C announces to F and H: its own prefixes and G's prefixes
- C does **not** announce to H: routes to prefixes it learned from F
  - Otherwise: H could send traffic towards F but would not pay anything. F would not pay either. C's network would be loaded with additional traffic.
- C does **not** announce to F: routes to prefixes it learned from H
  - Same reason

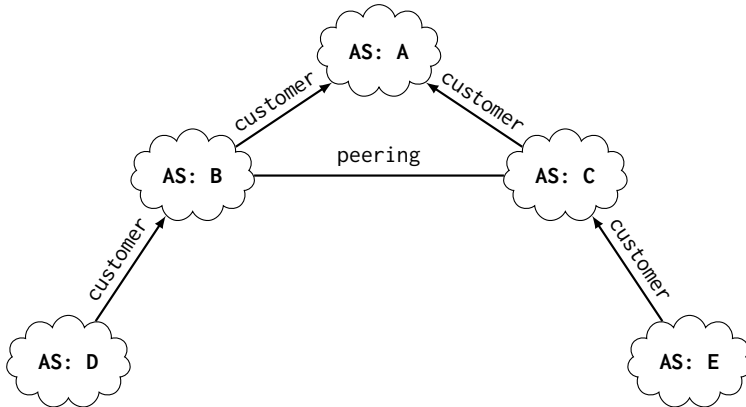


## Business and Policy Routing

## Policy Routing: Valley-Free Routing (Idealized)

Result: Packets always travel ...

1. upstream: sequence of Customer  $\rightarrow$  Provider links (possibly length = 0)
2. then possibly one peering link
3. the downstream: sequence of Provider  $\rightarrow$  Customer links (possibly length = 0)



## Business and Policy Routing

### Siblings

Not everything is provider/customer or peering

Sibling = mutual transit agreement

- Provide connectivity to the rest of the Internet for each other
- $\approx$  very extensive peering

#### Examples

- Two small ASes close to each other that cannot / do not want to afford additional Internet services
- Merging of two companies
  - Merging two ASes into one = difficult
  - Keeping two ASes and exchanging everything for free = easier
- Example AT&T has five different ASNs (7018, 7132, 2685, 2686, 2687)

## Business and Policy Routing

### Tier-1, Tier-2, Tier-3, etc...

#### Providers can be categorized into Tiers

- **Tier-1 / Default-Free-Zone**: only peerings, no providers
- **Tier-2**: only peerings, and one or more Tier-1 provider
- **Tier-3**: at least one Tier-2 as a provider
- **Tier-n**: at least one Tier-(n-1) provider
  - defined recursively
  - $n \leq 4$ : Rare in Western Europe, North America, East Asia
- **"Tier-1.5"**: almost a Tier-1, but pays money for some links
  - Example: "Deutsche Telekom" used to pay money to Sprint, but is now Tier-1
  - Marketing effect: Tier-1 sounds better



## Business and Policy Routing

### Tier-1 Providers

1. AT&T
2. Cogent Communications (formerly PSINet)
3. Deutsche Telekom AG (ICSS)
4. Global Telecom & Technology (GTT) (formerly Tinet & nLayer)
5. KPN International
6. Liberty Global
7. Lumen Technologies (formerly CenturyLink formerly Level 3 and Global Crossing)
8. NTT Communications (America) (formerly Verio)
9. Orange (OpenTransit)
10. PCCW Global
11. Sprint
12. Tata Communications (INDIA) (Acquired Teleglobe)
13. Telecom Italia Sparkle (Seabone)
14. Telia Carrier
15. Telxius (subsidiary of Telefonica)
16. Verizon Enterprise Solutions (formerly UUNET and XO Communications)
17. Zayo Group (formerly AboveNet)

1. Receive BGP update
2. Apply import policies
  - Filter routes
  - Tweak attributes (advanced topic ...)
3. Best route selection based on attribute values
  - Policy: Local preference settings and other attributes
  - Install forwarding tables entries for best routes
  - Possibly transfer to Route Reflector RR as alternative to logical full mesh of iBGP sessions
4. Apply export policies
  - Filter routes
  - Tweak attributes
5. Transmit BGP updates

## Business and Policy Routing

### BGP Business Relationship Summary

#### Import Policy: Which routes to use

- Select path that incurs most money
- Special/political considerations (e.g. Iranian AS does not want traffic to cross Israeli AS, other kinds of censorship)

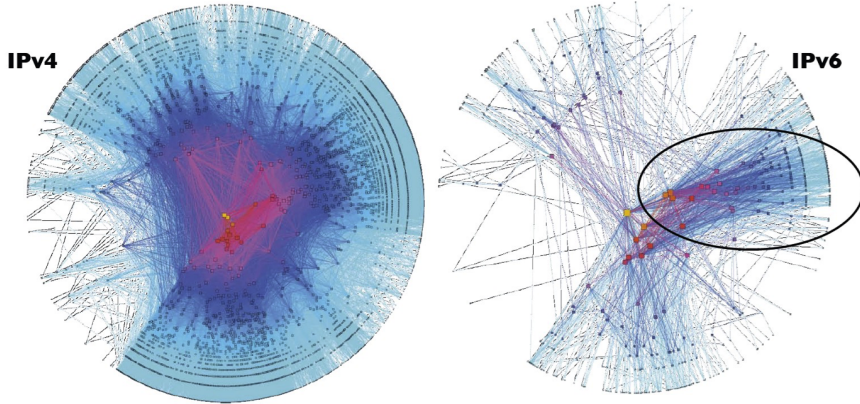
#### Export Policy: Which routes to propagate to other ASes

- Not all known routes are advertised
- Export only, ...
  1. ... if it incurs revenue
  2. ... if it reduces cost
  3. ... if it is inevitable

#### Policy routing: Money, Money, Money, ...

- Route import and export driven by business considerations
- But **not** driven by technical considerations
  - Example: Slower route via peer may be preferred over faster route via provider

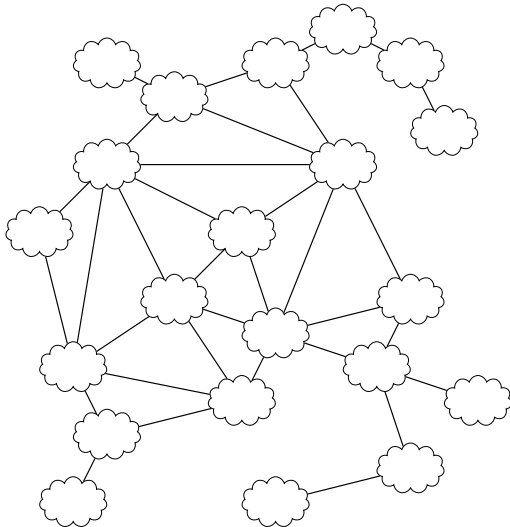
## k-core Algorithm IPv4 vs. IPv6 AS Graphs



Source: [http://www.caida.org/research/topology/as\\_core\\_network/](http://www.caida.org/research/topology/as_core_network/)  
ASes are ranked here by their customer cone size (number of their direct and indirect customers)  
Alternative ranking approach: k-core algorithm

## k-core Algorithm

### k-core Algorithm



Input: Full Graph (if all vertices are connected its a 1-core)

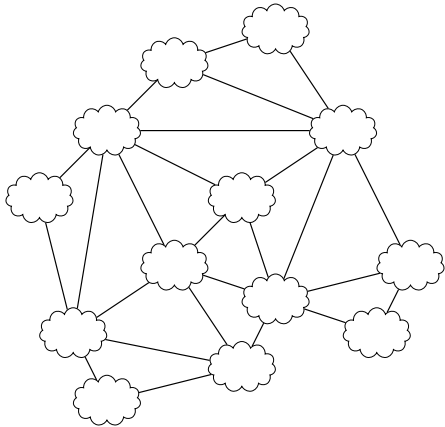
Algorithm steps:

1. Remove all nodes of degree  $\leq 1$  recursively
2. Did you remove any node?
  - YES: goto step 1
  - NO: return

Result: 2-core subgraph

## k-core Algorithm

### k-core Algorithm

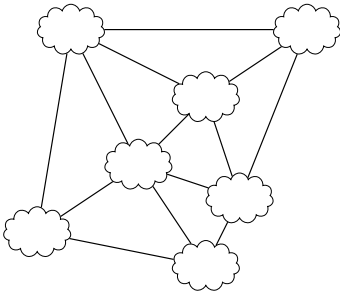


Input: 2-core subgraph

Algorithm steps:

1. Remove all nodes of degree  $\leq 2$  recursively
2. Did you remove any node?
  - YES: goto step 1
  - NO: return

Result: 3-core subgraph



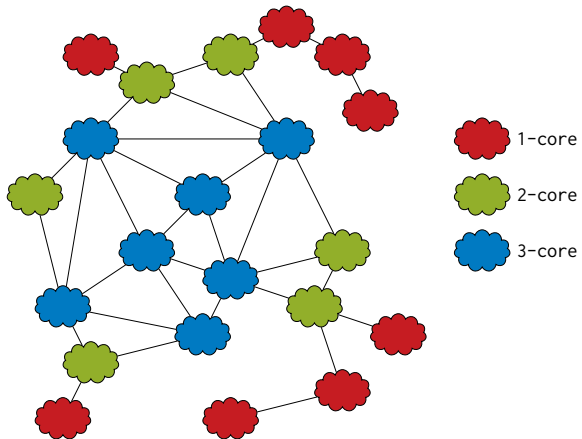
Input: 3-core subgraph

Algorithm steps:

1. Remove all nodes of degree  $\leq 3$  recursively
2. Did you remove any node?
  - YES: goto step 1
  - NO: return

In this example, this would remove all remaining nodes. Therefore, the constellation on the left is the core of this example network.

k-core Algorithm  
k-core Algorithm





Autonomous Systems

BGP

Bibliography

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