BGP in 120 minutes

RIPE89

Wolfgang Tremmel
academy@de-cix.net

About me

- →Wolfgang Tremmel
- →studied Informatik (Uni Karlsruhe)
  - →Degree: Diploma (1994)
- →Network Engineer at
  - →Since 1996 Director NOC
  - →Since 2000 Senior Network Planner DSL at
- →2001 - 2005 Director Network Planning at VIA NET.WORKS
- →2006 - 2016 Manager Customer Support at
- →since 2016: Head of DE-CIX Academy

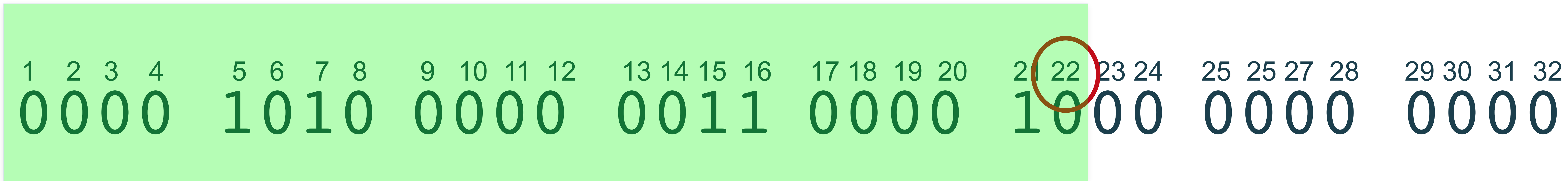
wolfgangtremmel1966

@wtremmel@hessen.social

What is BGP about?

IPv4 Prefixes

10.3.8.0/22



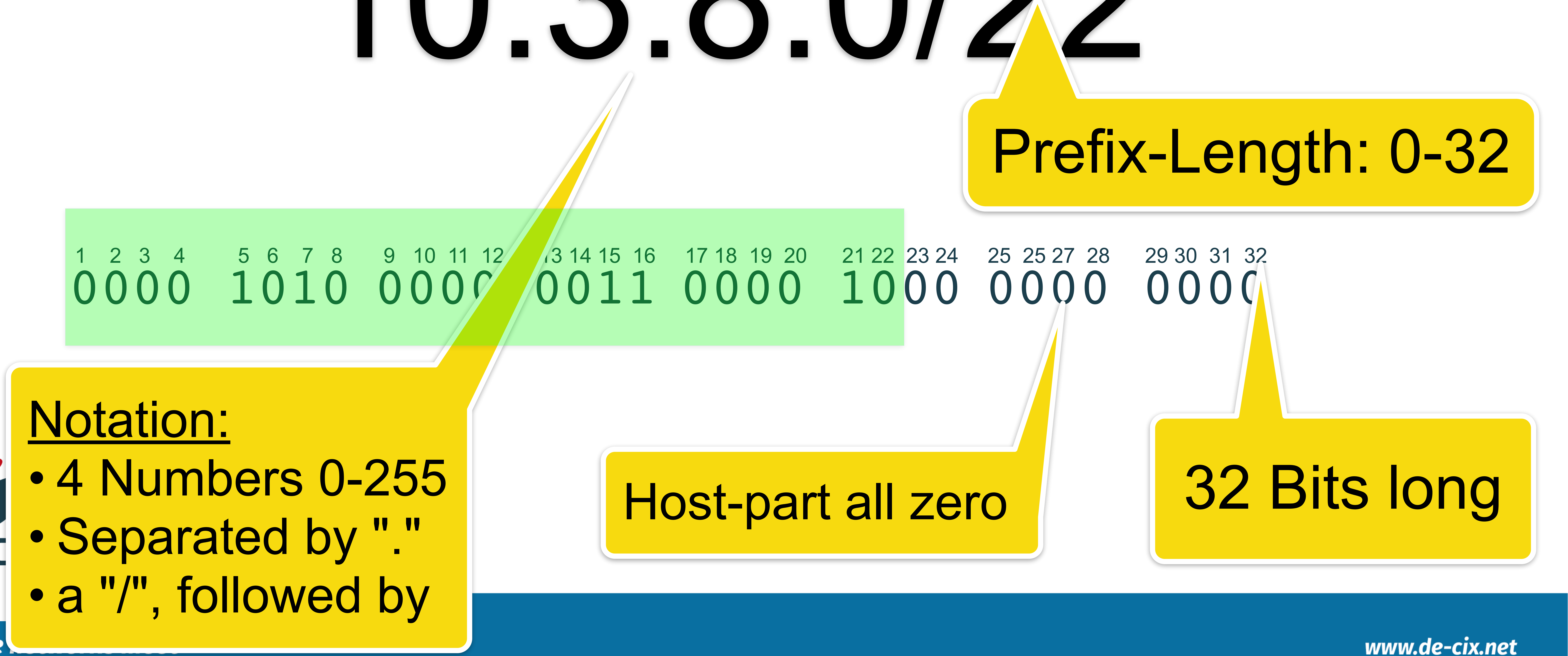
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 25 27 28 29 30 31 32

0000 1010 0000 0011 0000 1000 0000 0000

- → IPv4 and IPv6 addresses have a network and a host part
- → A **prefix** is just the network part
- → Important:
  - **The boundary between network and host can be anywhere!**

Characteristics of Prefixes: IPv4

10.3.8.0/22



Prefix-Length: 0-32

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 25 27 28 29 30 31 32

0000 1010 0000 0011 0000 1000 0000 0000

Host-part all zero

32 Bits long

Notation:
- 4 Numbers 0-255
- Separated by "."
- a "/", followed by

Characteristics of Prefixes: IPv6

Prefix-Length: 0-128

2003:de:274f:400::/64

0 01 02 03 04 05 06 07 08 09 0a 0b 0d 0e 0f 10 11 12 13 14 15 16 17 18 19 1a 1b 1c 1d 1e 1f 20 21 22 23 24 25 26 27 28 29 2a 2b 2c 2d 2e 2f 30 31 32 33 34 35 36 37 38 39 3a 3b 3c 3d 3e 3f 40 41 42 43 44 45 46 47 48 49 4a 4b 4c 4d 4e 4f 50 51 52 53 54 55 56 57 58 59 5a 5b 5c 5d 5e 5f 60 61 62 63 64 65 66 67 68 69 6a 6b 6c 6d 6e 6f 70 71 72 73 74 75 76 77 78 79 7a 7b 7c 7d 7e 7f

Notation:

- 4 digit hex numbers (0-9,a-f)
- Separated by ":"
- "::" = fill up with zeros

Host-part all zero

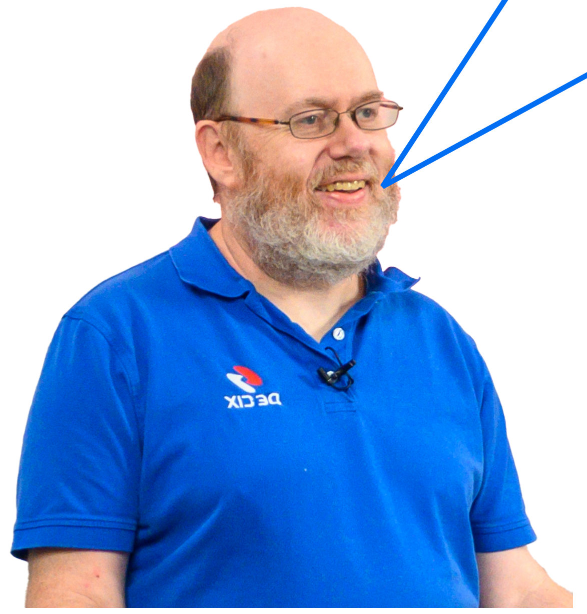
128 Bits long

How does BGP work?

BGP is a protocol to announce prefixes

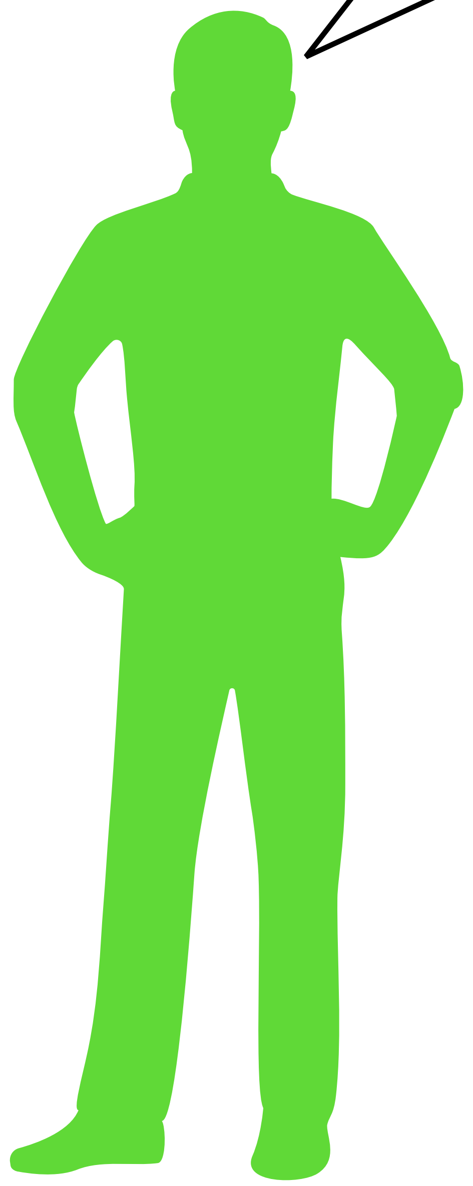
Everybody has Neighbors

I am **AS196610**, DE-CIX Academy, and I announce prefix **2a02:c50:db8::/48**

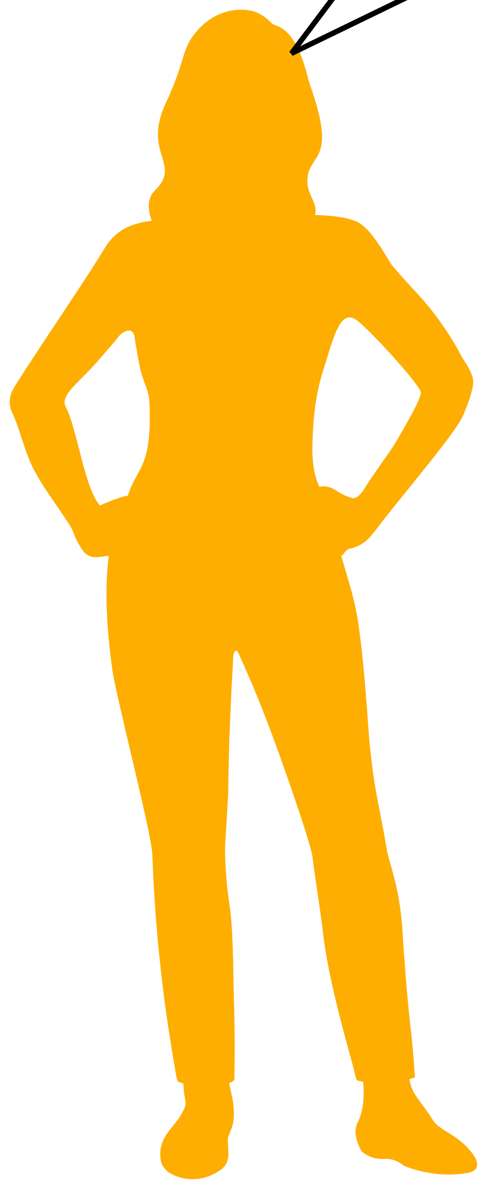


DE-CIX Academy
AS196610

My neighbor AS196610 announces prefix 2a02:c50:db8::/48



My green neighbor told me, his neighbor AS196610 announces prefix 2a02:c50:db8::/48



2a02:c50:db8::/48
2a02:c50:db8::/48
2a02:c50:db8::/48
2a02:c50:db8::/48
192.0.2.0/24
198.51.100.0/24
2a02:c50:db8::/48

DE CIX

BGP announces prefixes

To neighbors

- BGP **announces** IP prefixes to **neighbors**

192.0.2.0/24
198.51.100.0/24
2a02:c50:db8::/48

  - These neighbors have to be **configured**
  - Each BGP speaking device is part of an **Autonomous System**

I am **AS196610**, DE-CIX Academy, and I announc prefix 2a02:c50:db8::/48

  - The path these announcements take is recorded - this is called the **Autonomous System Path**

2a02:c50:db8::/48
🟧🟩🟦

  - The AS Path shows which Autonomous Systems have forwarded the prefix announcement
  - The rightmost AS in the AS Path is called the "**Originator**"

What is an *Autonomous System*?

What is an Autonomous System?

Simple Definition

Formal Definition (RFC1930):

```
"An AS is a connected group of one or more IP
prefixes run by one or more network operators
which has a SINGLE and CLEARLY DEFINED routing
policy."
```

- A group of IP prefixes
  - But to route or announce them, you need hardware
  - A router (or multiple routers)
  - This router speaks BGP (to its neighbors)
  - And has an ***Autonomous System Number*** configured
- Another new term: **Autonomous System Number (ASN)**

Router

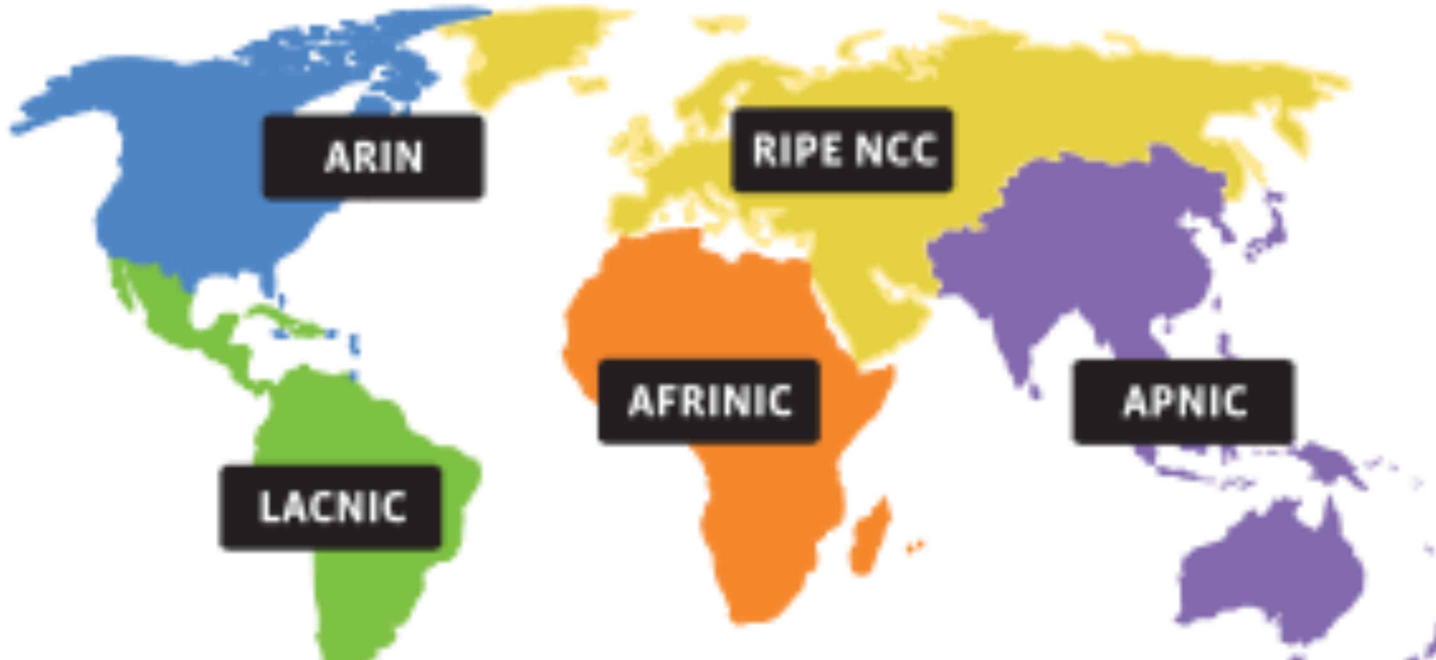
I am **AS196610**, DE-CIX Academy, and I announce prefix 2a02:c50:db8::/48

Autonomous System Number

or AS Number or ASN

> *"An AS has a **globally unique** number (sometimes referred to as an **ASN**, or Autonomous System Number) associated with it;*
> *this number is used in both the exchange of exterior routing information (between neighboring ASes), and as an **identifier of the AS** itself." (RFC1930)*

- Initially 16bit (0...65535) they are now 32bit long (0..."a lot")
- AS numbers are globally unique
- Unique means, somebody has to administrate them
- This is the IANA (Internet Assiged Numbers Authority)
  - But they have delegated that task to the 5 RIRs (Regional Internet Registries)
  - So in Europe: Become a member of the RIPE NCC and request one

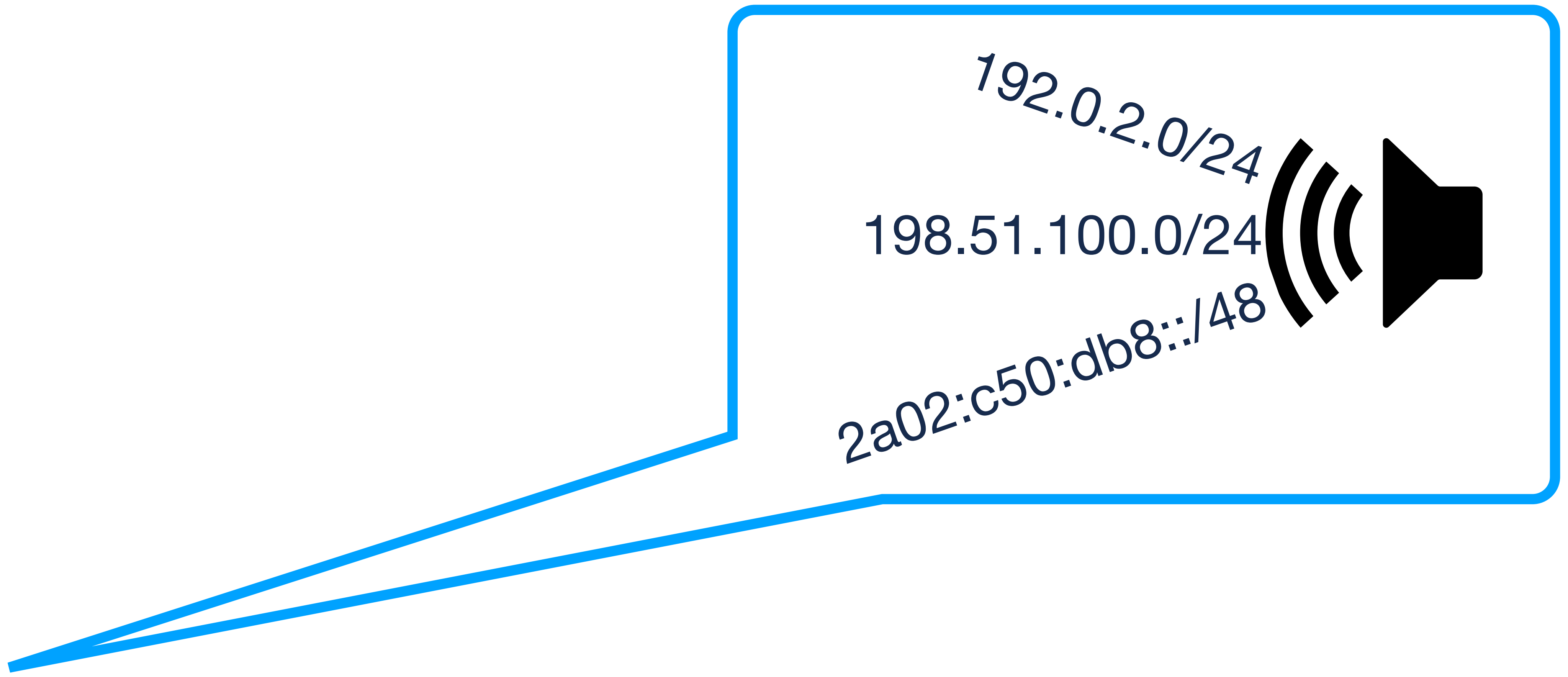


BGP Announcing Prefixes

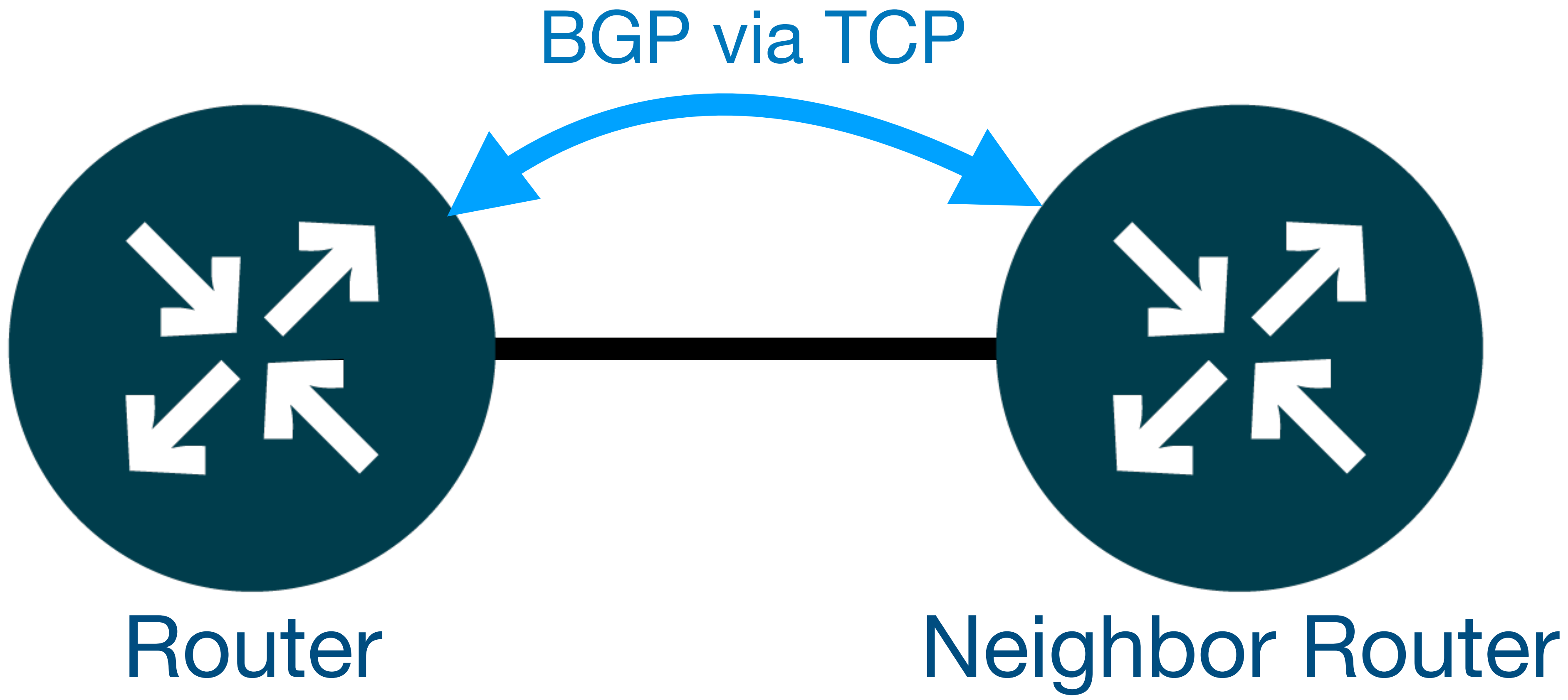
DE CIX

BGP Neighbors

Directly connected neighbors

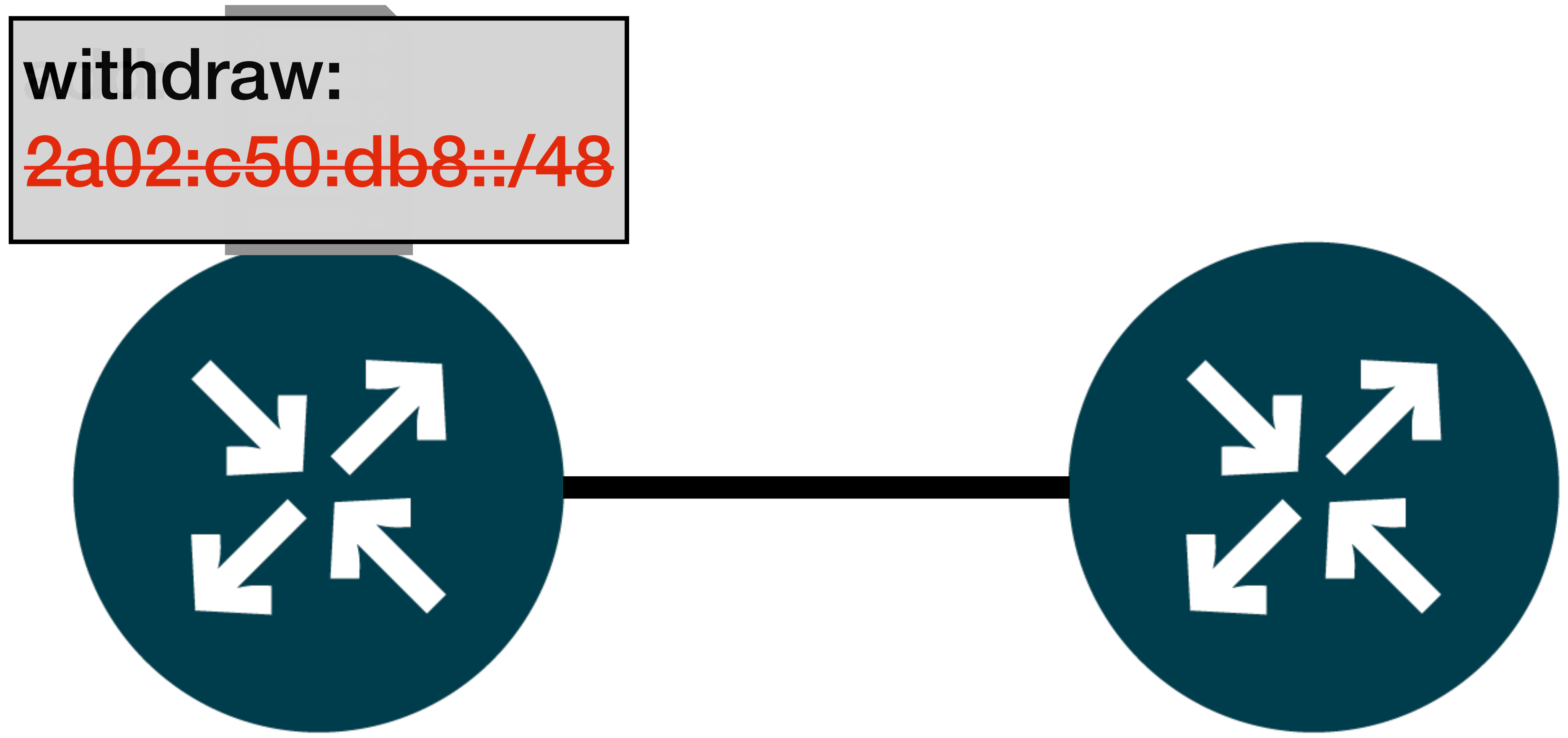


- BGP **announces** IP prefixes to **neighbors**
- These neighbors have to be **configured**
- BGP uses **TCP** to connect to a neighbor
- TCP brings already:
  - **Reliable transport** (sender knows that receiver got it)
  - **Flow control** (do not send faster than the receiver can receive)
  - **Framing** (putting BGP messages into packets)



BGP works incremental

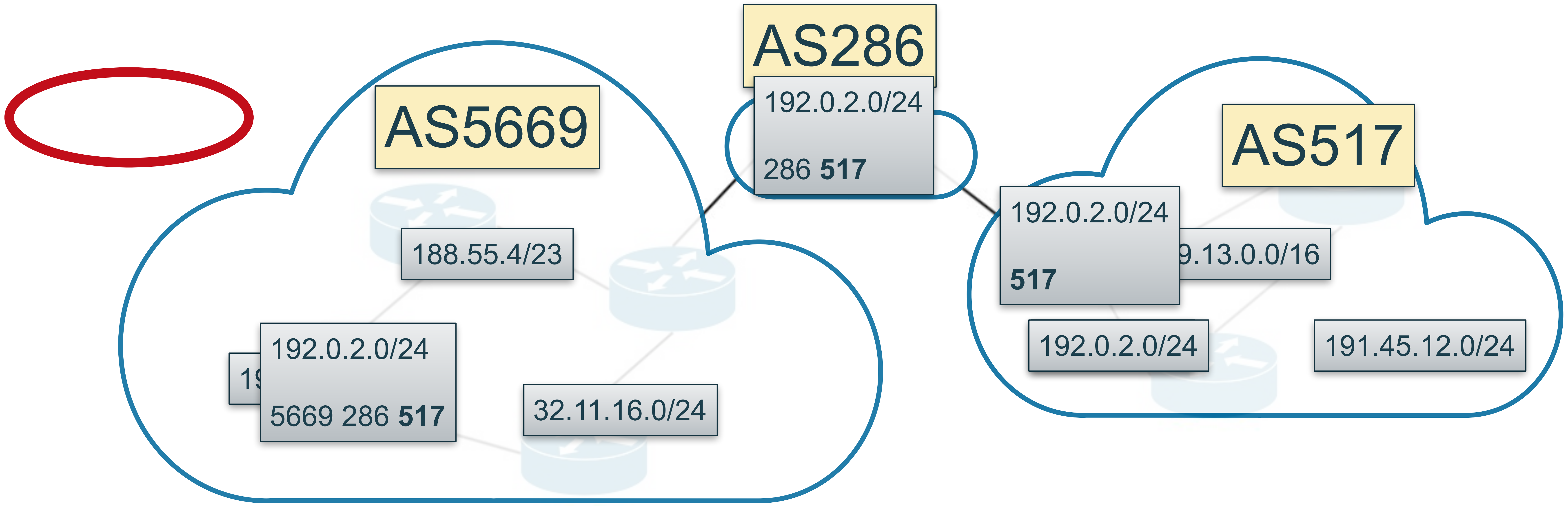
Using add- / withdraw- messages



- At session setup, BGP announces "everything" to its neighbor
- After that, updates are **incremental**:
  - If BGP learns about a new prefix, it sends an **add**-message to neighbors
  - If a prefix goes away, it sends a **withdraw** message to neighbors
- As long as the BGP session is "up", a router assumes its neighbors are "in sync" (= did not forget anything it sent)

BGP Announcing Prefixes

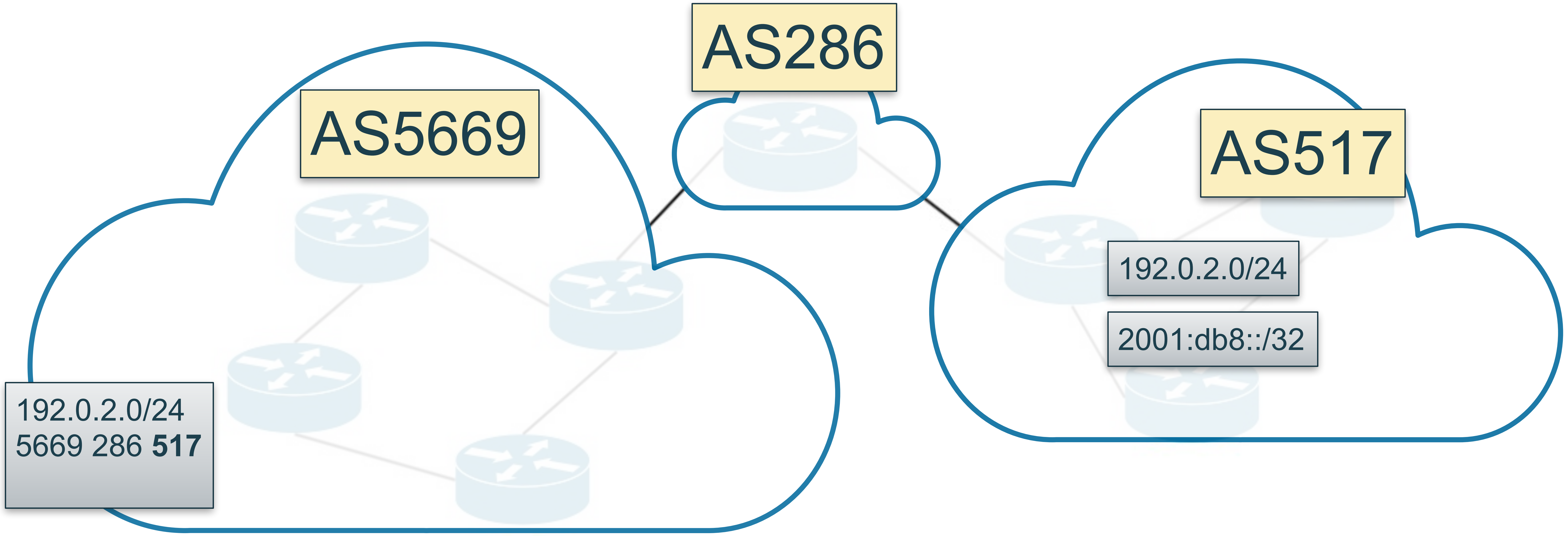
Building the AS path



BGP Announcing Prefixes

- Prefixes
- AS Numbers
- AS Path

Originator AS



192.0.2.0/24
5669 286 **517**

Attributes of BGP prefixes

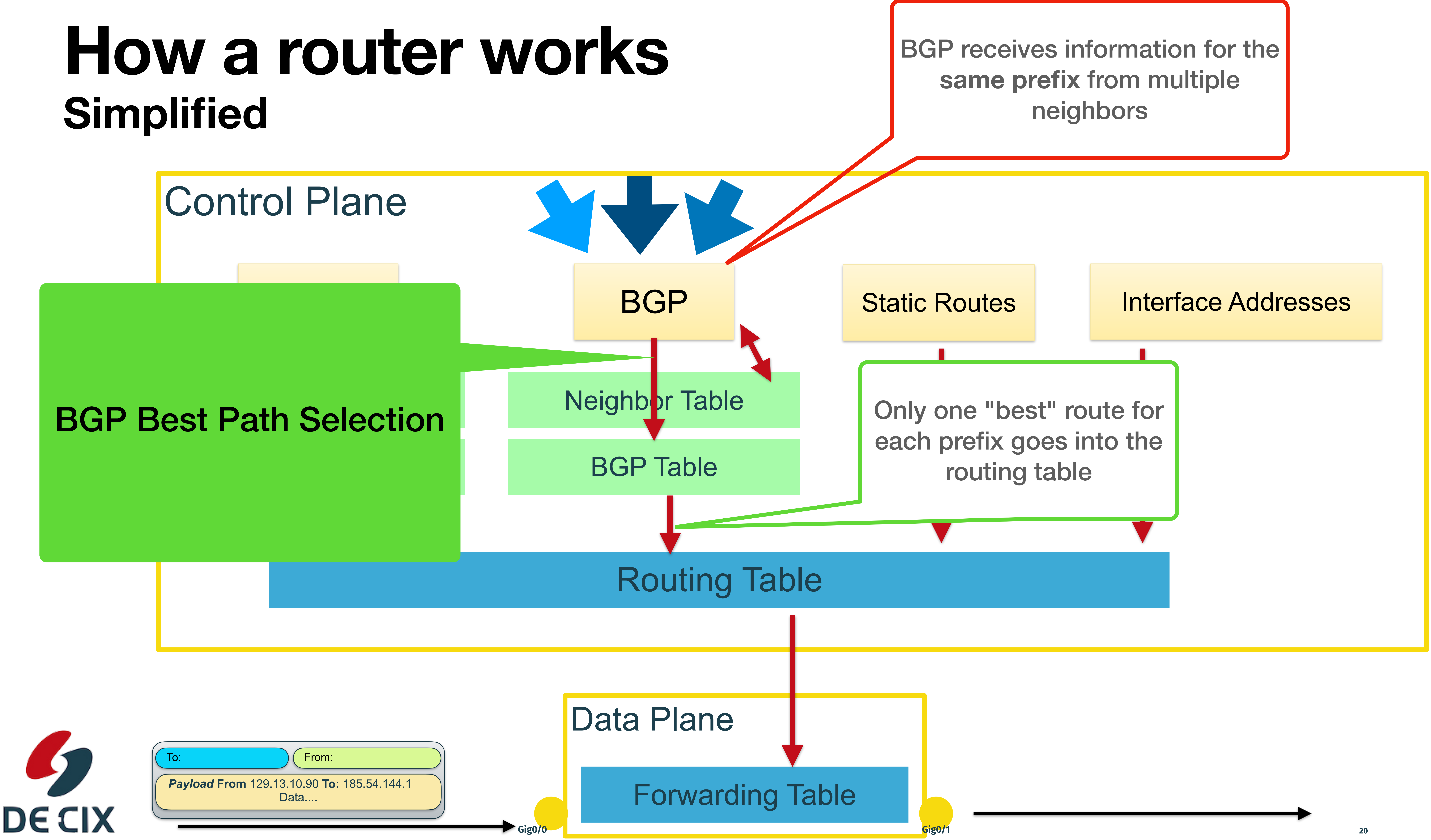
Not only the AS path

- **Mandatory** attributes: have to be there
  - Example: AS-Path
- **Optional** attribute: are, well, optional
  - Example: MED

- **Transitive** attributes
  - are kept on the prefix and forwarded via BGP
- **Non-transitive** attributes
  - are added to a prefix and not forwarded by the receiver

How a router works

Simplified


BGP receives information for the **same prefix** from multiple neighbors
Control Plane
BGP
Static Routes
Interface Addresses
BGP Best Path Selection
Neighbor Table
BGP Table
Only one "best" route for each prefix goes into the routing table
Routing Table
Data Plane
Forwarding Table
To:
From:
Payload **From** 129.13.10.90 **To:** 185.54.144.1
Data....
Gig0/0
Gig0/1

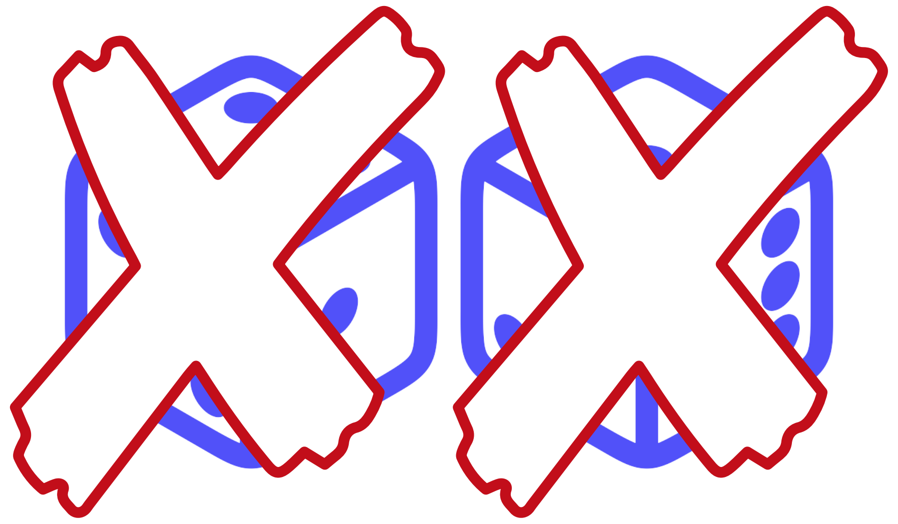
BGP Best Path Selection

BGP Best Path Selection Algorithm

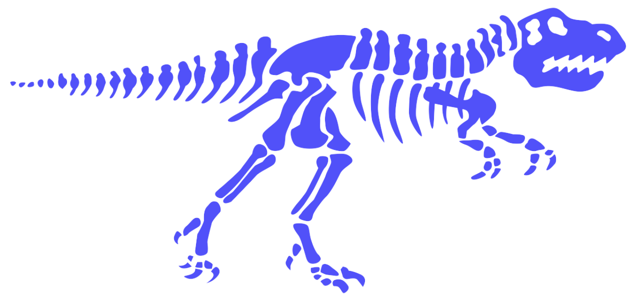
Motivation

1

- Only one single path for each destination is needed (and wanted)
- Decision must be based on attributes
- And must not be random, but deterministic

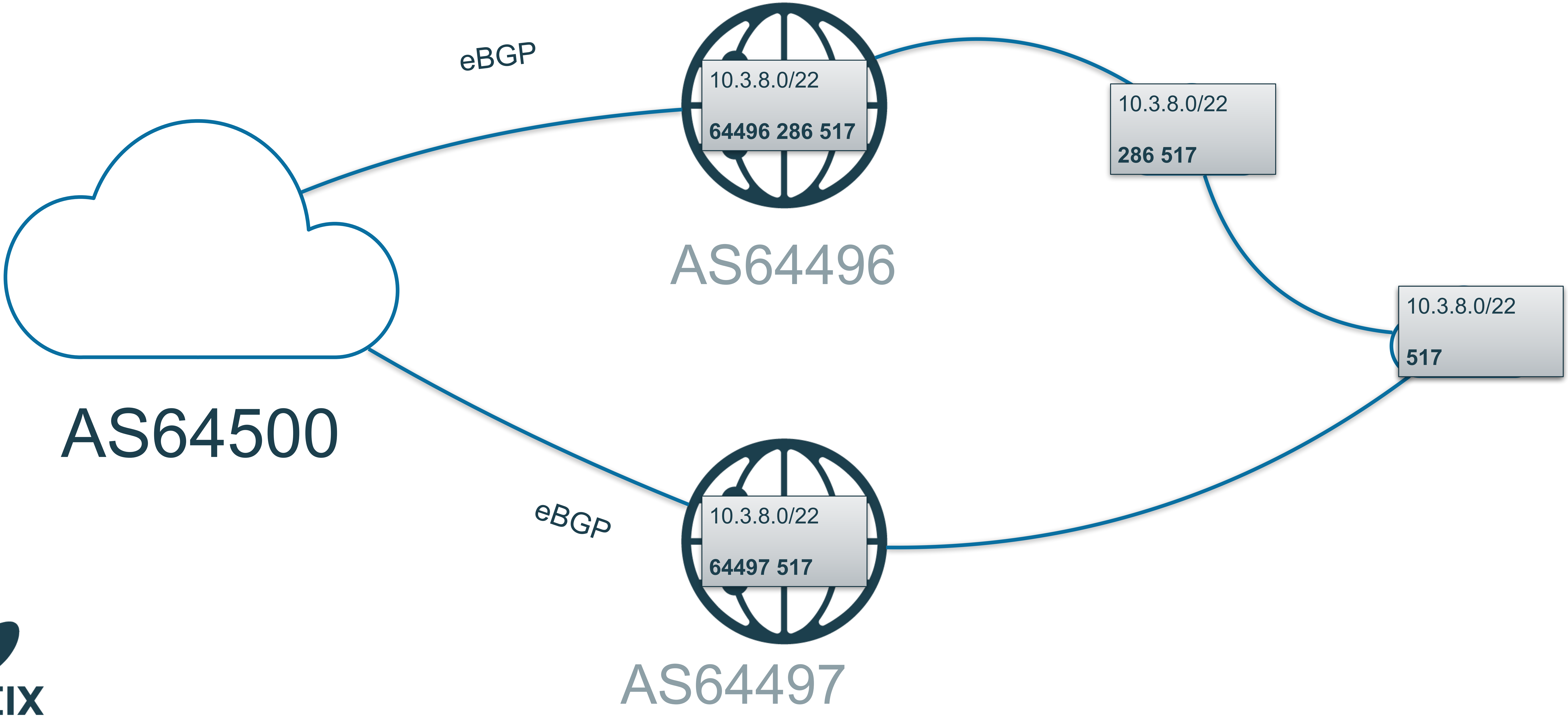


- Some of the criteria will sound strange
- Some are really outdated

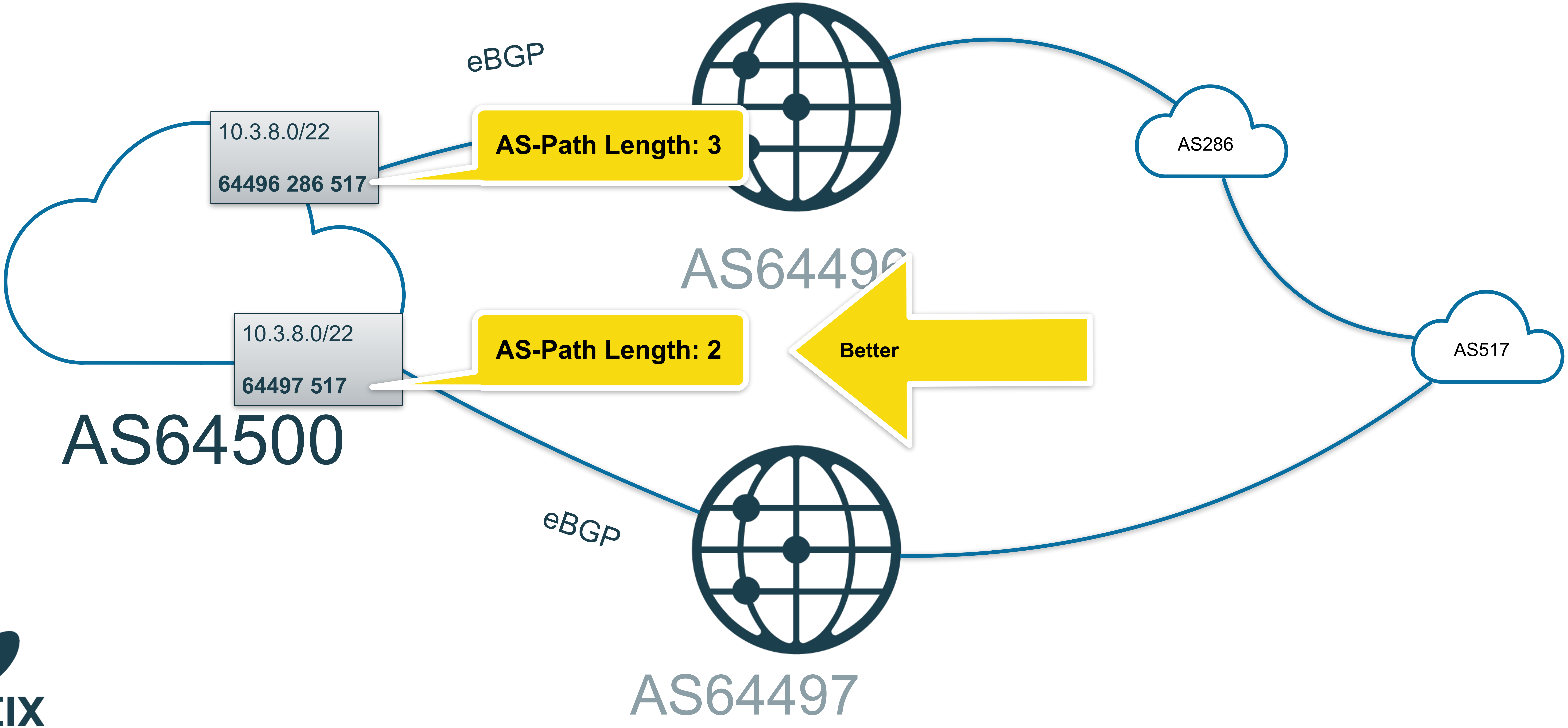


- So lets have a look how this works...

Let's get started.... with two upstreams



Let's get started.... with two upstreams



BGP Best Path Selection

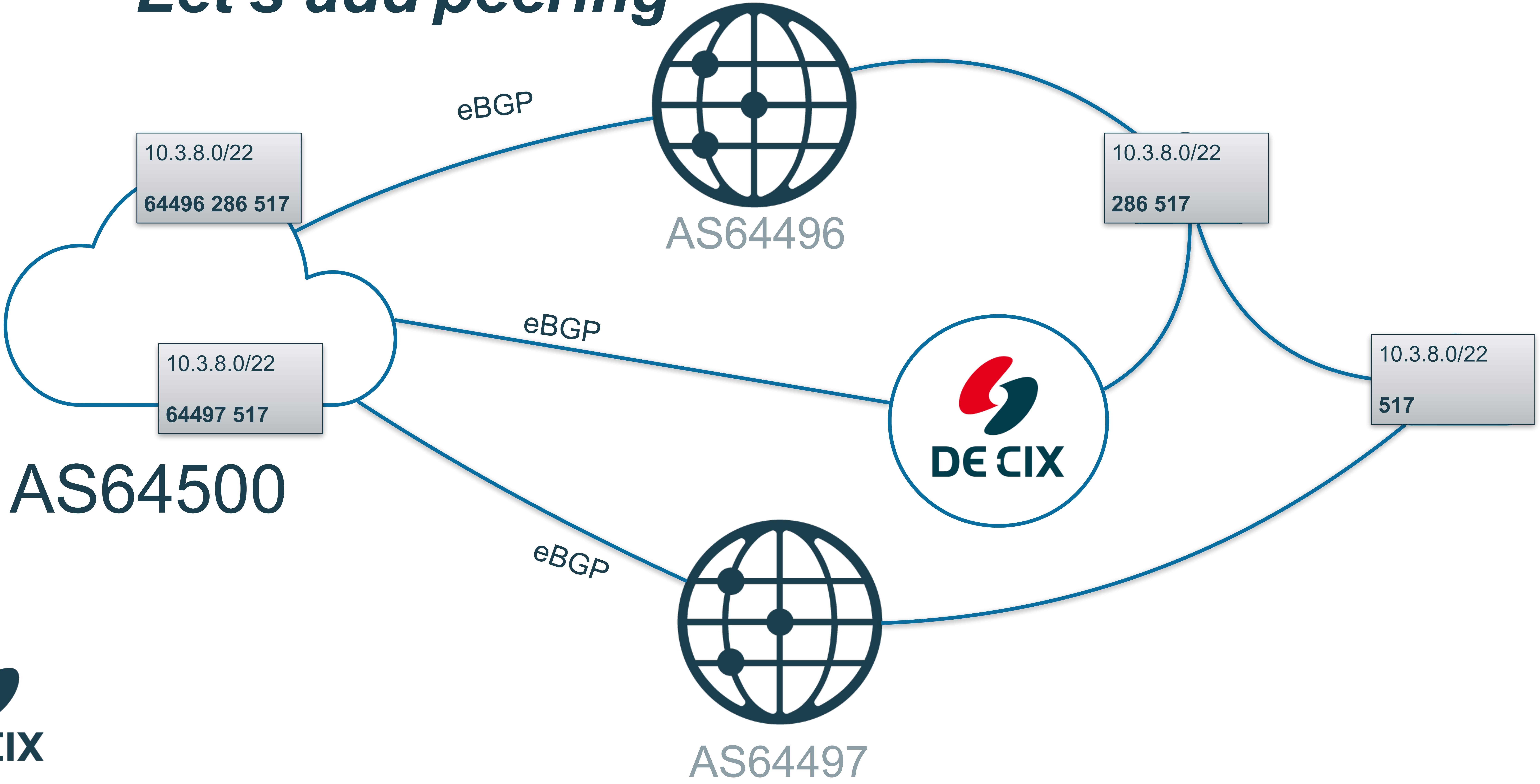
1	NextHop reachable?	Continue if "yes"
2		
3		
4		
5		
6		
7		
8		
9		
10		

AS-Path Length: 3

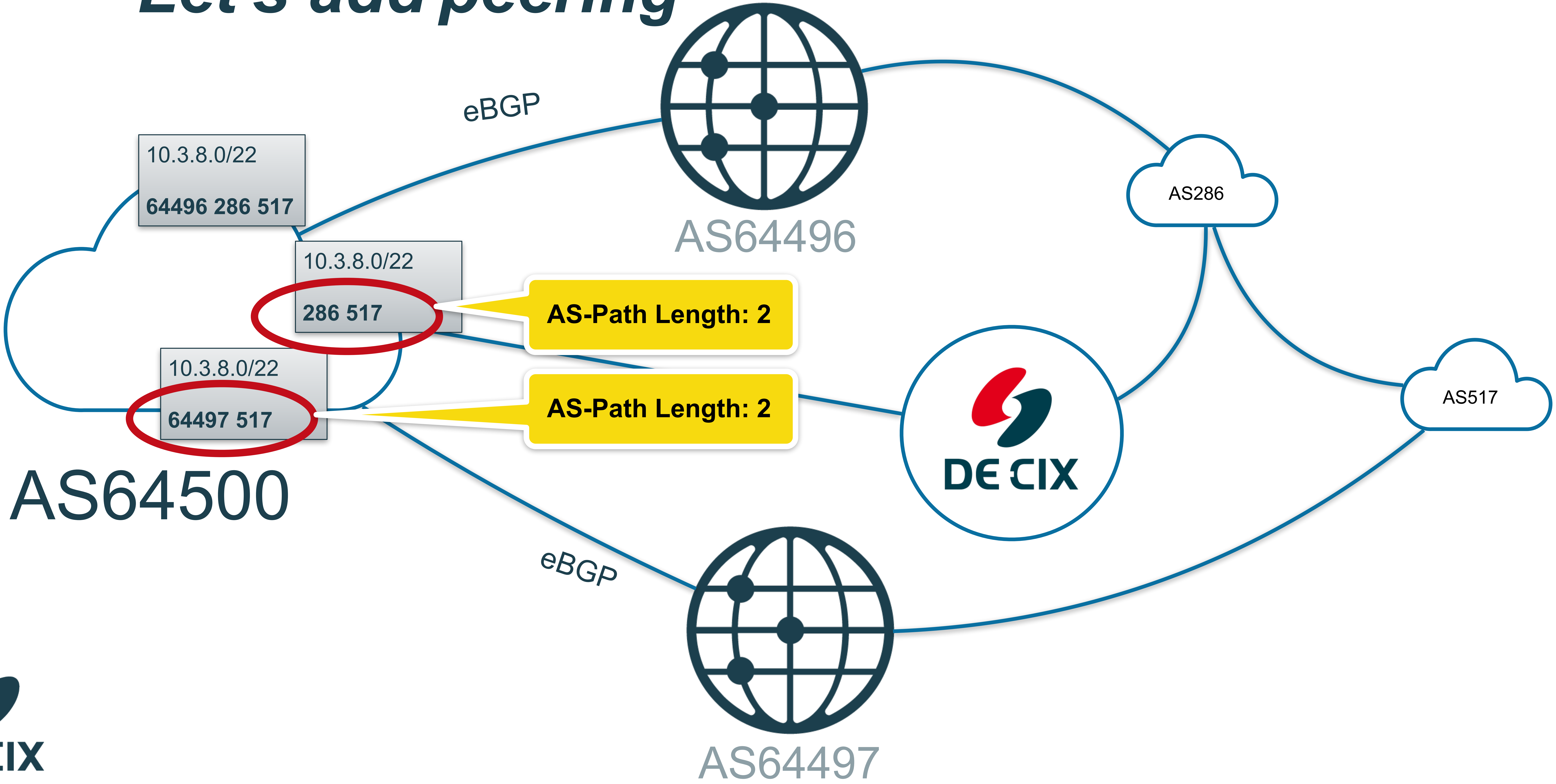
AS-Path Length: 2

Better

Let's add peering



Let's add peering



BGP Best Path Selection

1	NextHop reachable?	Continue if "yes"
2		
3	AS Path Length	shorter wins
4		
5		
6		
7		
8		
9		
10		

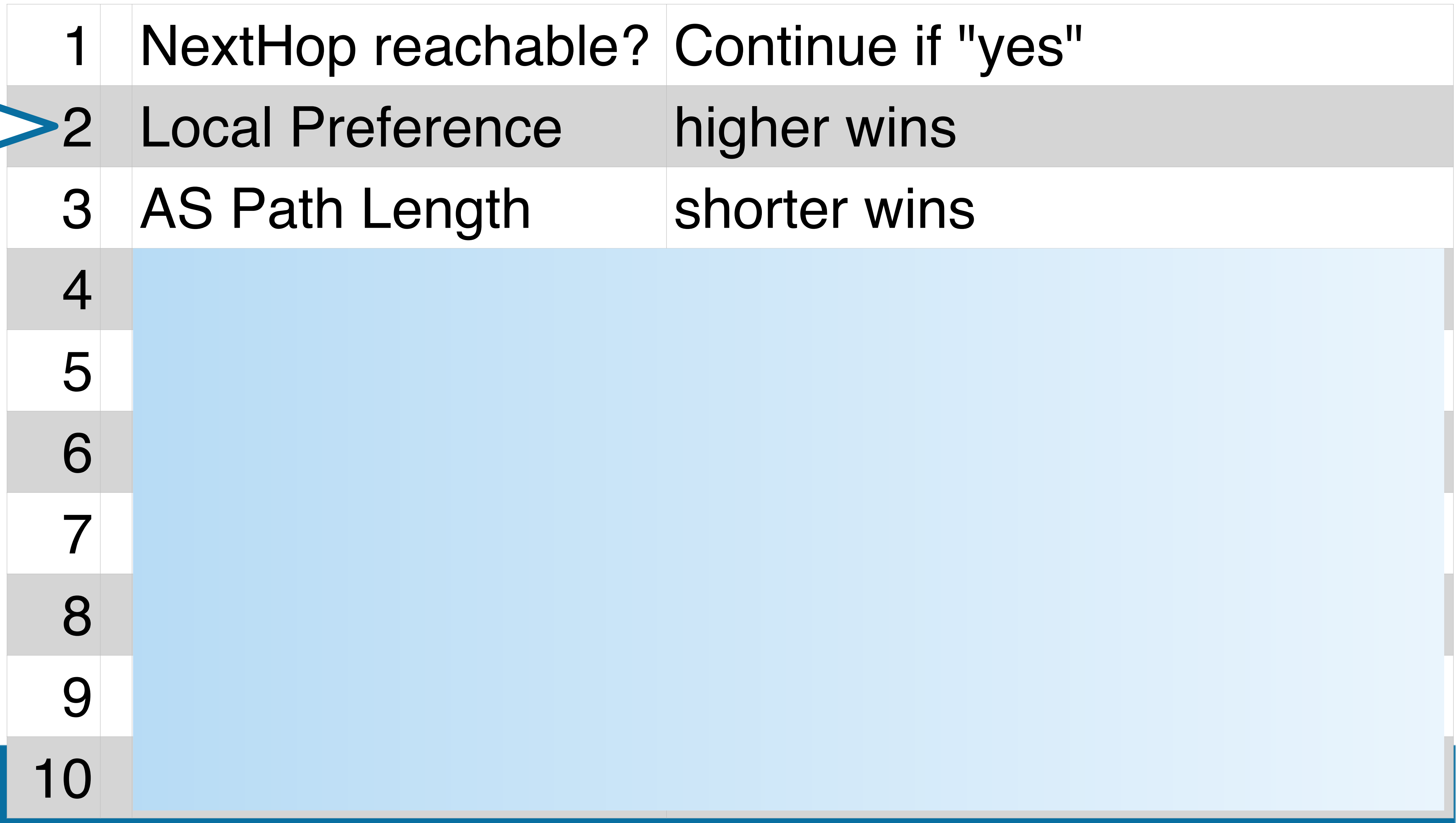
AS-Path Length: 2

AS-Path Length: 2

Local Preference

Why am I not using "100" here?

- → Higher wins
- → Integer value (32bit, 0-4294967295)
- → Propagated via iBGP inside an Autonomous System
- → Usually set using rules when receiving prefixes
- → Typical values:
  - Customer prefixes: 10000
  - Peering prefixes: 1000
  - Upstream prefixes: 10



1	NextHop reachable?	Continue if "yes"
2	Local Preference	higher wins
3	AS Path Length	shorter wins
4		
5		
6		
7		
8		
9		
10		

BGP Route Selection: Origin Type

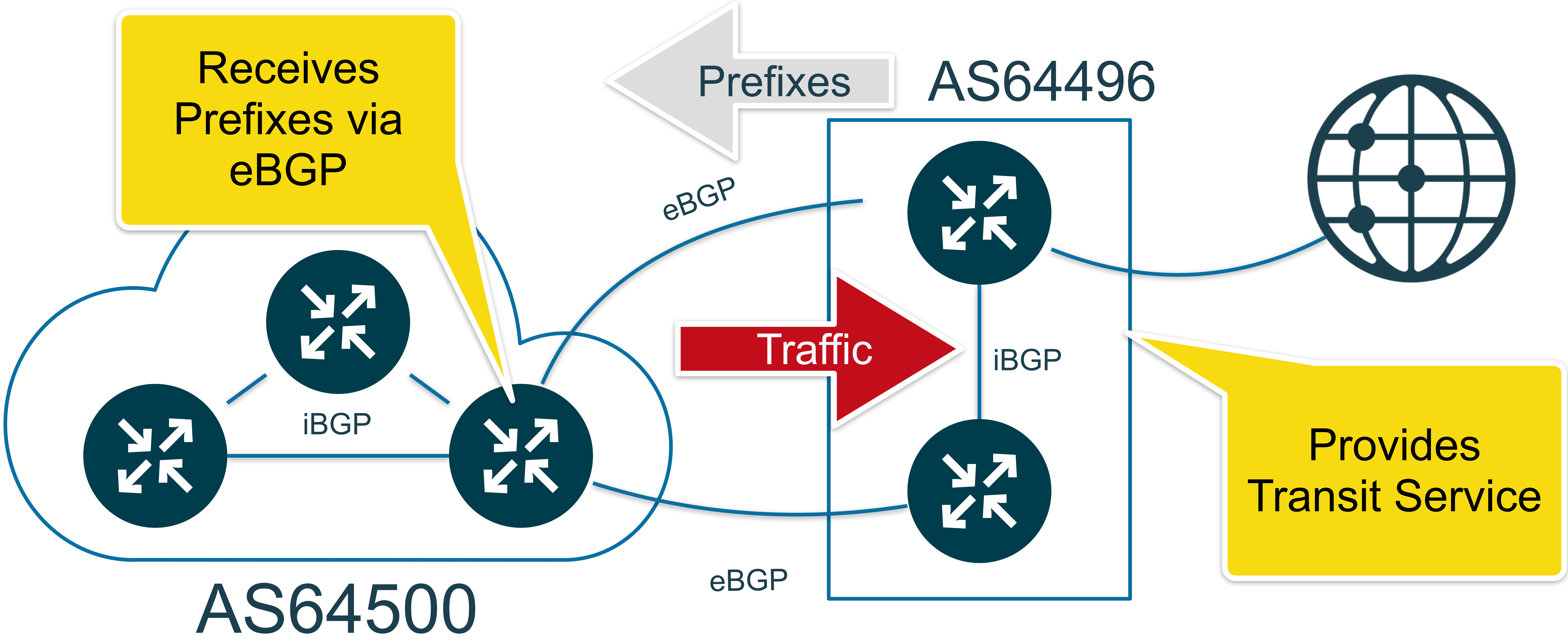
- → Origin Type is a "historical" attribute
- → Three possible values:
  - → IGP - route is generated by BGP network statement - "i"
  - → EGP - route is received from EGP - **"e"**
  - → incomplete - redistributed from another protocol - **"?"** as the "real source" is unknown
- → ***This rule is not really important***
- → Fun fact: There are prefixed in the global routing table marked "e"

Exterior **G**ateway **P**rotocol

Predecessor of BGP which is no longer used

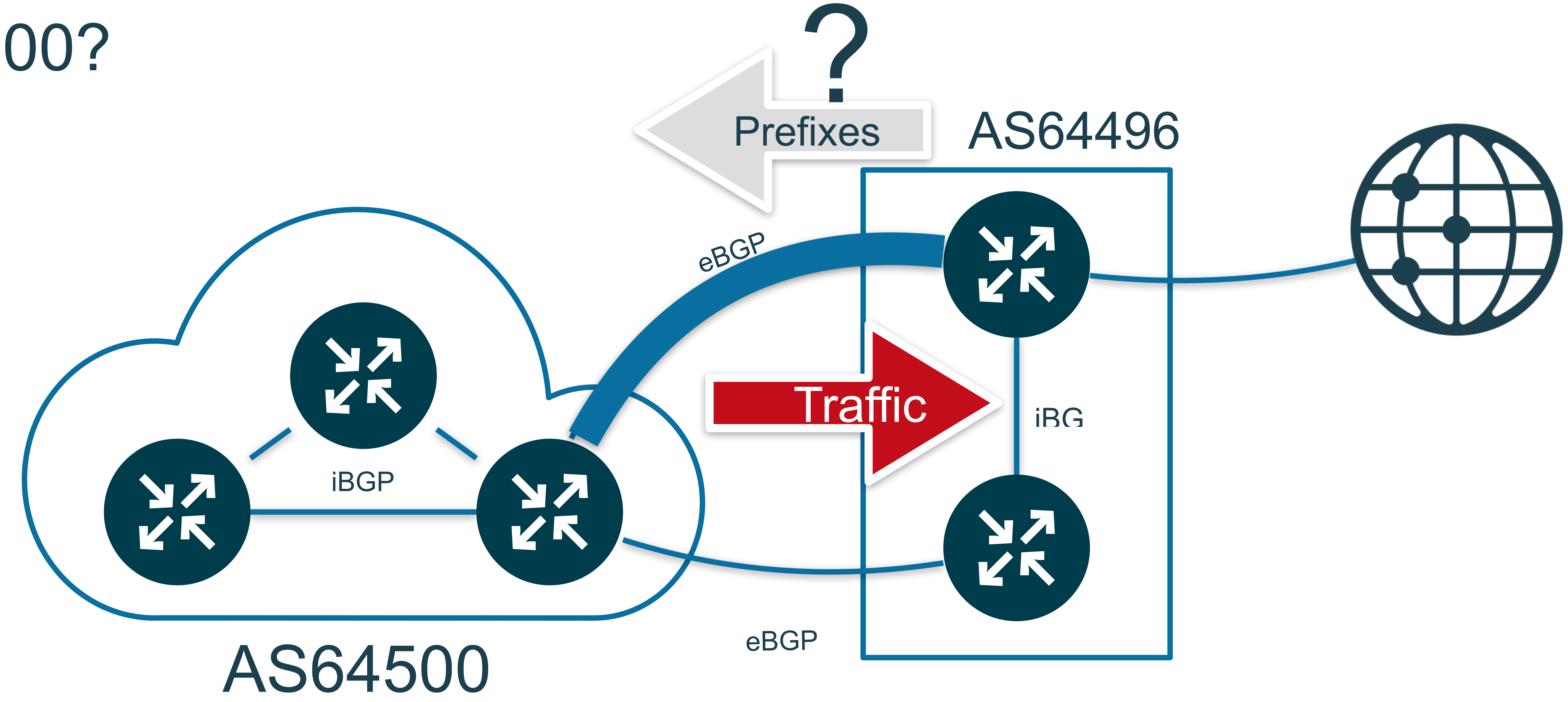
1	NextHop reachable?	Continue if "yes"
2	Local Preference	higher wins
3	AS Path Length	shorter wins
4		
5		
6		
7		
8		
9		
10		

Consider the following network



Consider the following network

- There are two circuits
- AS64496 wants one of them preferred
- How to tell AS64500?

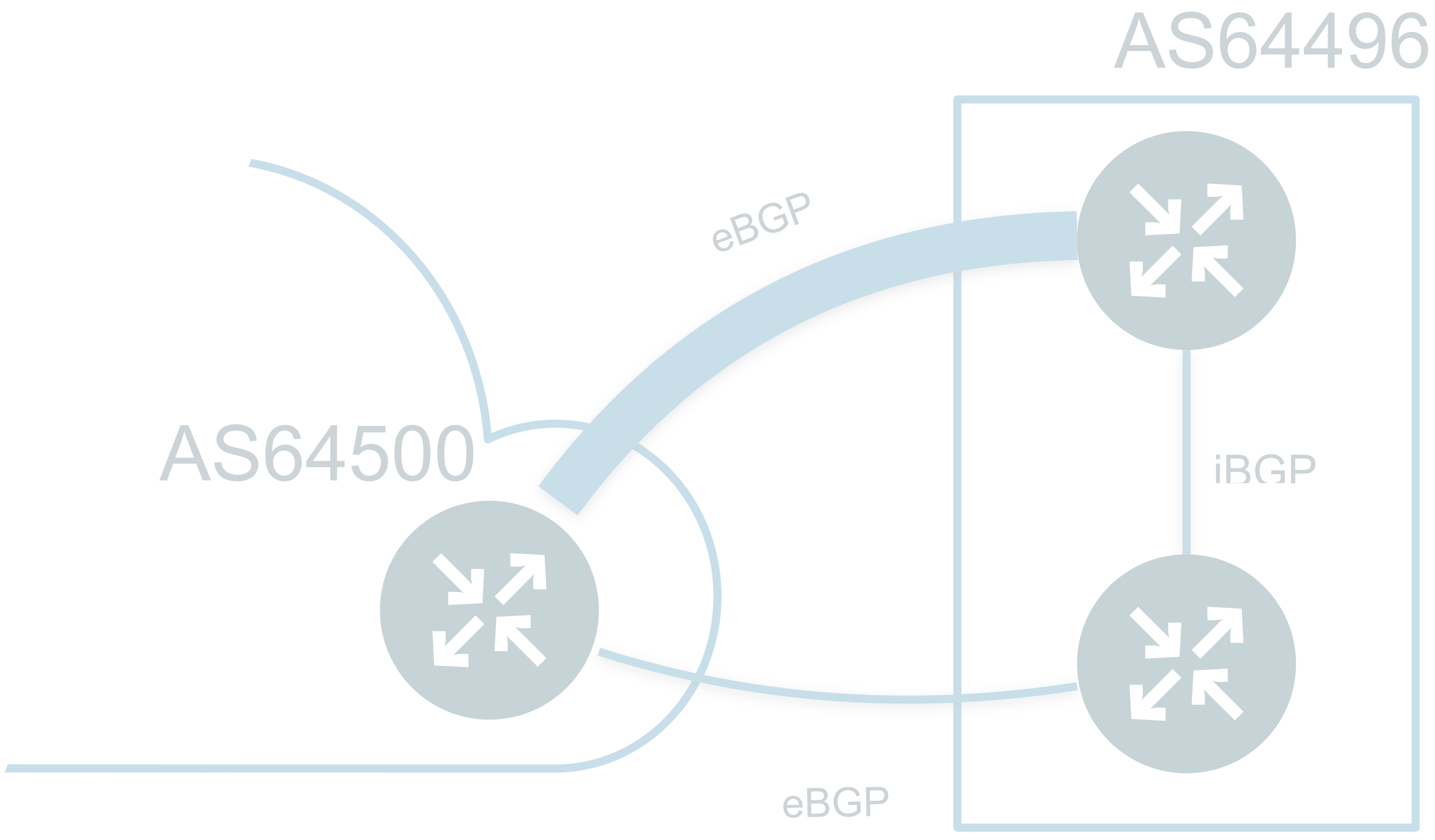


BGP Route Selection Algorithm:

How to tell your neighbor where you prefer traffic?

1	NextHop reachable?	Continue if "yes"
2	Local Preference	higher wins
3	AS Path Length	shorter wins
4	Origin Type	IGP over EGP over Incomplete
5		
6		
7		
8		
9		
10		

BGP Route Selection Algorithm: MED

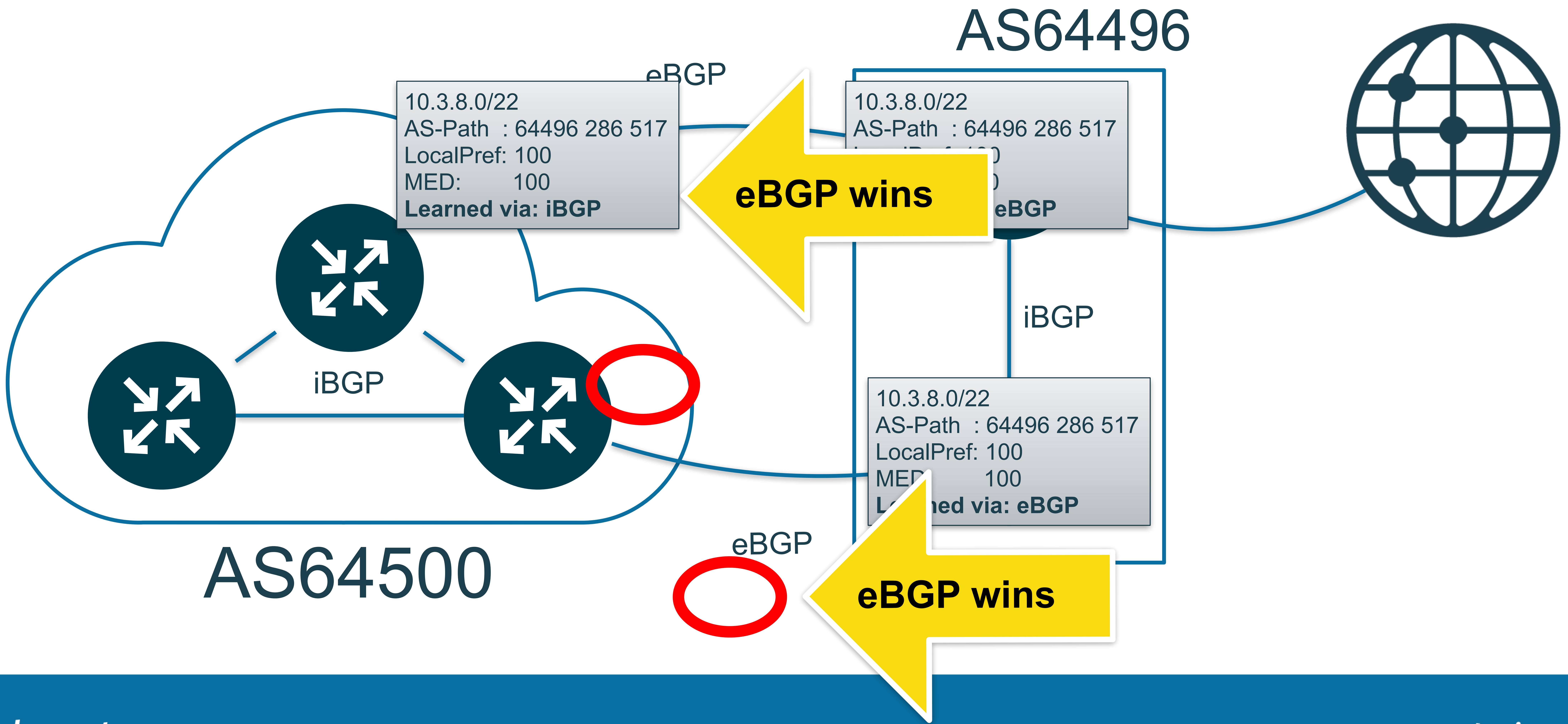


- → MED = **M**ulti-**E**xit **D**iscriminator
- → Only compared if next-hop AS is the same
- → 32bit value (0..4294967294)
- → Lower wins
- → Optional (does not have to be there), non-transitive (does not get forwarded)
- → A missing MED can be treated as "best" (=0, default) or "worst" (=4294967294)
- → And of course you can override whatever you receive

BGP Route Selection : Hot Potato Rules

1	NextHop reachable?	Continue if "yes"
2	Local Preference	higher wins
3	AS Path Length	shorter wins
4	Origin Type	IGP over EGP over Incomplete
5	MED	lower wins
6		
7		
8		
9		
10		

BGP Route Selection : eBGP wins

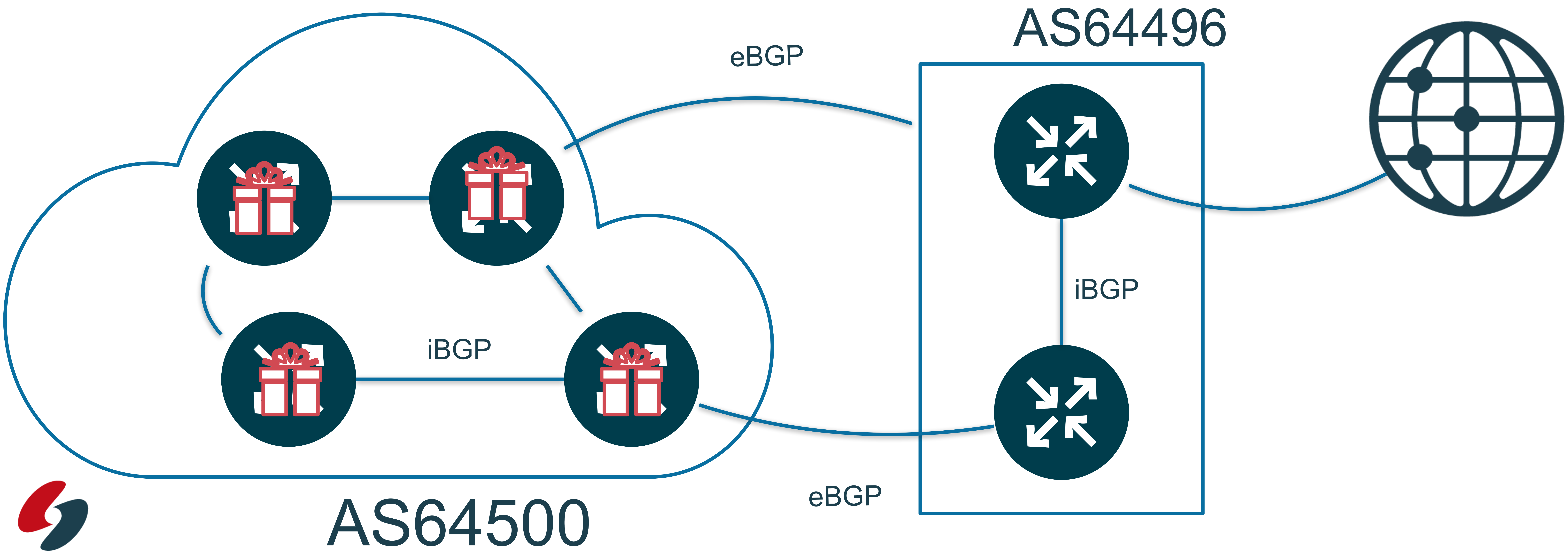


10.3.8.0/22
AS-Path : 64496 286 517
LocalPref: 100
MED: 100
Learned via: iBGP

10.3.8.0/22
AS-Path : 64496 286 517
eBGP

10.3.8.0/22
AS-Path : 64496 286 517
LocalPref: 100
MED 100
L ned via: eBGP

BGP Route Selection : nearest exit wins

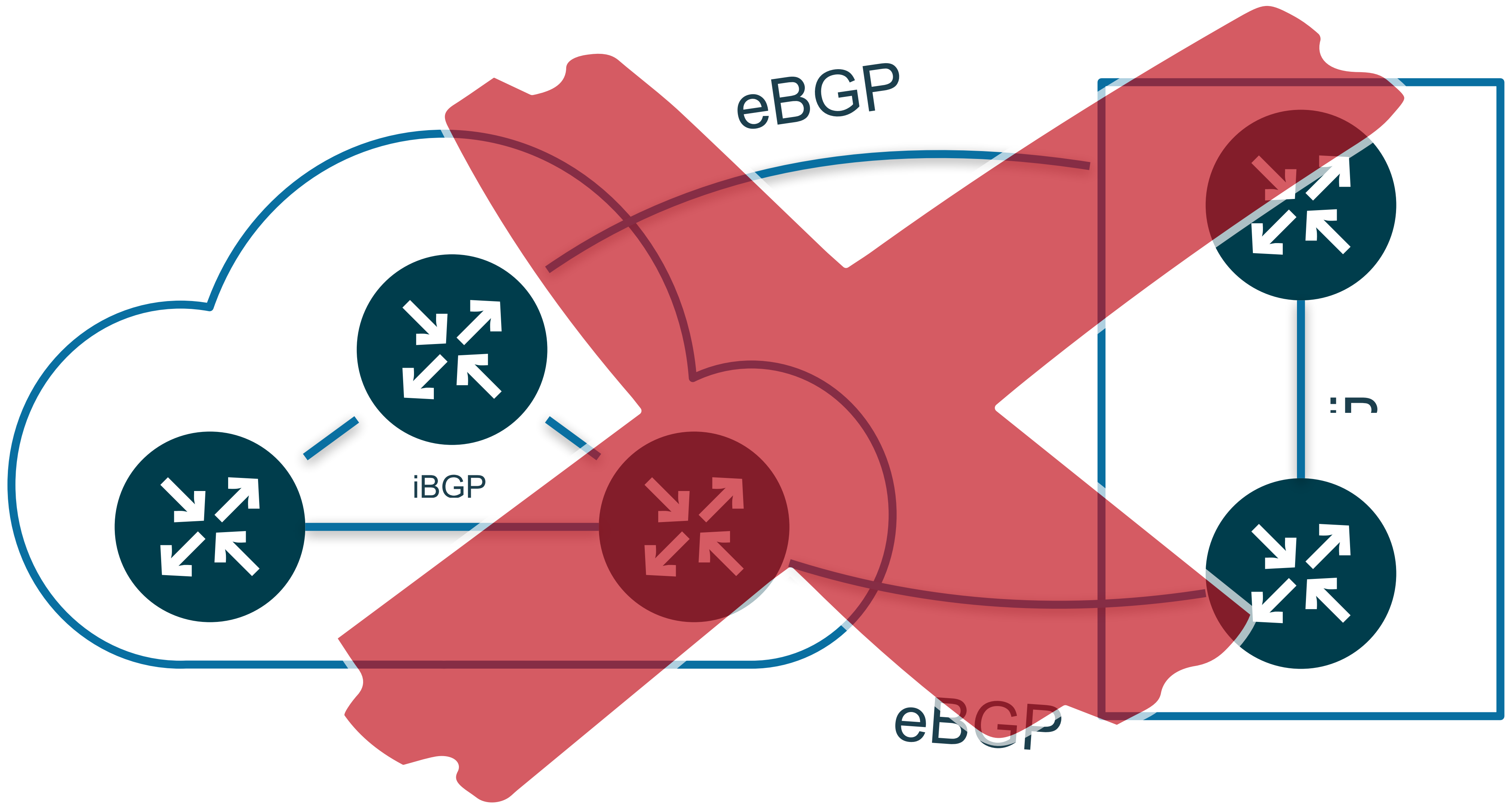


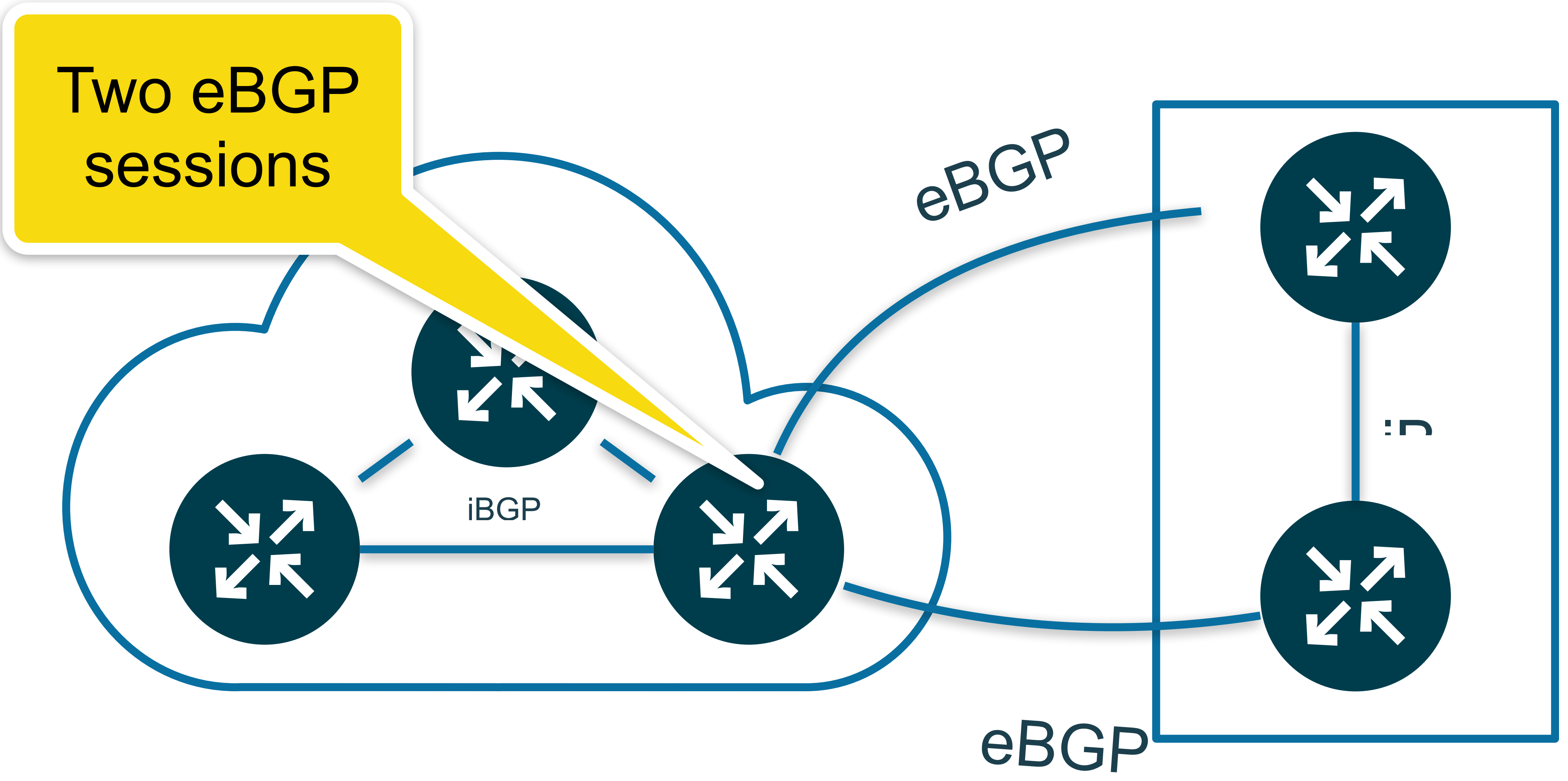
BGP Route Selection : Age / Stability

1	NextHop reachable?	Continue if "yes"
2	Local Preference	higher wins
3	AS Path Length	shorter wins
4	Origin Type	IGP over EGP over Incomplete
5	MED	lower wins
6	eBGP, iBGP	eBGP wins
7	Exit	nearest wins
8		
9		
10		

BGP Route Selection : Age / Stability

- Exact phrasing is (Cisco):
  "When both paths are external, prefer the path that was received first"
- So this applies only if a router has two (or more) eBGP sessions
- Which happens quite often when connecting to Internet Exchanges




Two eBGP sessions

BGP Route Selection : Last Resort

1	NextHop reachable?	Continue if "yes"
2	Local Preference	higher wins
3	AS Path Length	shorter wins
4	Origin Type	IGP over EGP over Incomplete
5	MED	lower wins
6	eBGP, iBGP	eBGP wins
7	Exit	nearest wins
8	Age of route	older wins
9		
10		

BGP Route Selection : Last Resort

- → Router ID: lower wins
- → Neighbor IP: lower wins
- → Rules of last resort
- → ...because at the end one and only one best path has to be selected
- → Usually path selection stops before it gets to these two rules.

1	NextHop reachable?	Continue if "yes"
2	Local Preference	higher wins
3	AS Path Length	shorter wins
4	Origin Type	IGP over EGP over Incomplete
5	MED	lower wins
6	eBGP, iBGP	eBGP wins
7	Exit	nearest wins
8	Age of route	older wins
9	Router ID	lower wins
10	Neighbor IP	lower wins

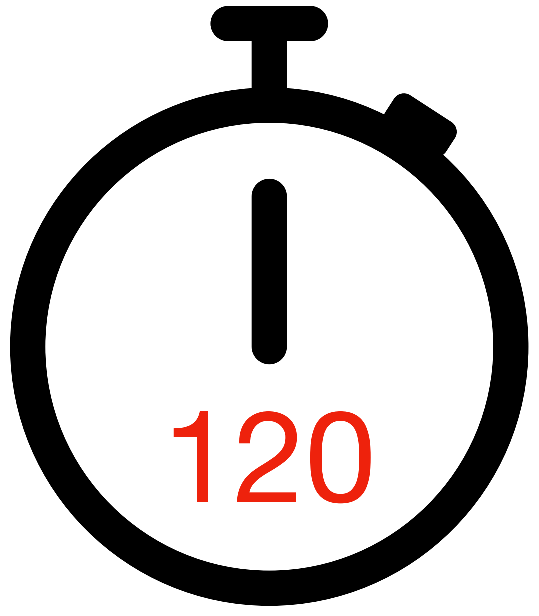
BGP Route Selection : Summary

1	NextHop reachable?	Continue if "yes"
2	Local Preference	higher wins
3	AS Path Length	shorter wins
4	Origin Type	IGP over EGP over Incomplete
5	MED	lower wins
6	eBGP, iBGP	eBGP wins
7	Exit	nearest wins
8	Age of route	older wins
9	Router ID	lower wins
10	Neighbor IP	lower wins

Other versions of this presentation

BGP in 120 minutes

What we did today



- Length: 90-120 minutes
- Features:
  - me talking
  - you asking questions
- Covers:
  - The very basics of BGP
  - Up and including BGP best path selection / more depending on time

BGP 4-5 hour workshop

Not just the basics...

- Length: 4-5 hours, including at least one break
- Happened a number of times at workshop Sunday at DENOG
- Features:
  - Me talking
  - You asking questions
  - Limited number of **lab experiments** using FRRouting
- Covers:
  - The very basics of BGP
  - Up and including BGP best path selection
  - BGP Communities if time permits

3.5 Day BGP Seminar

All and everything

- Length: 3.5 days, starting Monday noon, finishing Thursday late afternoon,
- Classroom seminar, max. 14 attendees
- Features:
  - Me talking
  - You asking questions
  - Extensive number of lab experiments using FRRouting
- Covers:
  - All of BGP
  - Including BGP Security, Traffic Engineering, Peering Relationships
  - Tools useful for BGP and peering

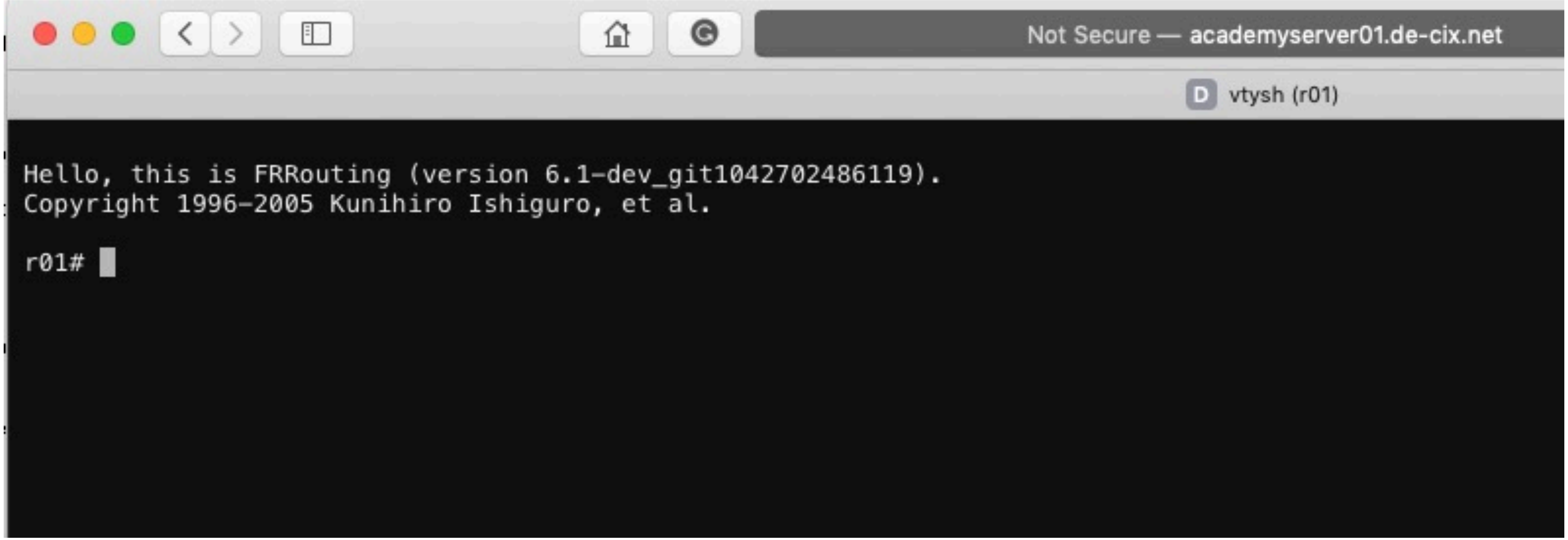
Experiment time!

Lets play with a BGP router

You just need your browser

https://bgplab.as196610.net:7000/

https://bgplab.as196610.net:7000/



```
Hello, this is FRRouting (version 6.1-dev_git1042702486119).
Copyright 1996-2005 Kunihiro Ishiguro, et al.

r01#
```

Things to try:

- show bgp summary
- show bgp ipv4
- show bgp ipv6

DE-CIX Academy BGP lab

- The lab is open source and available for download
- Get it here:

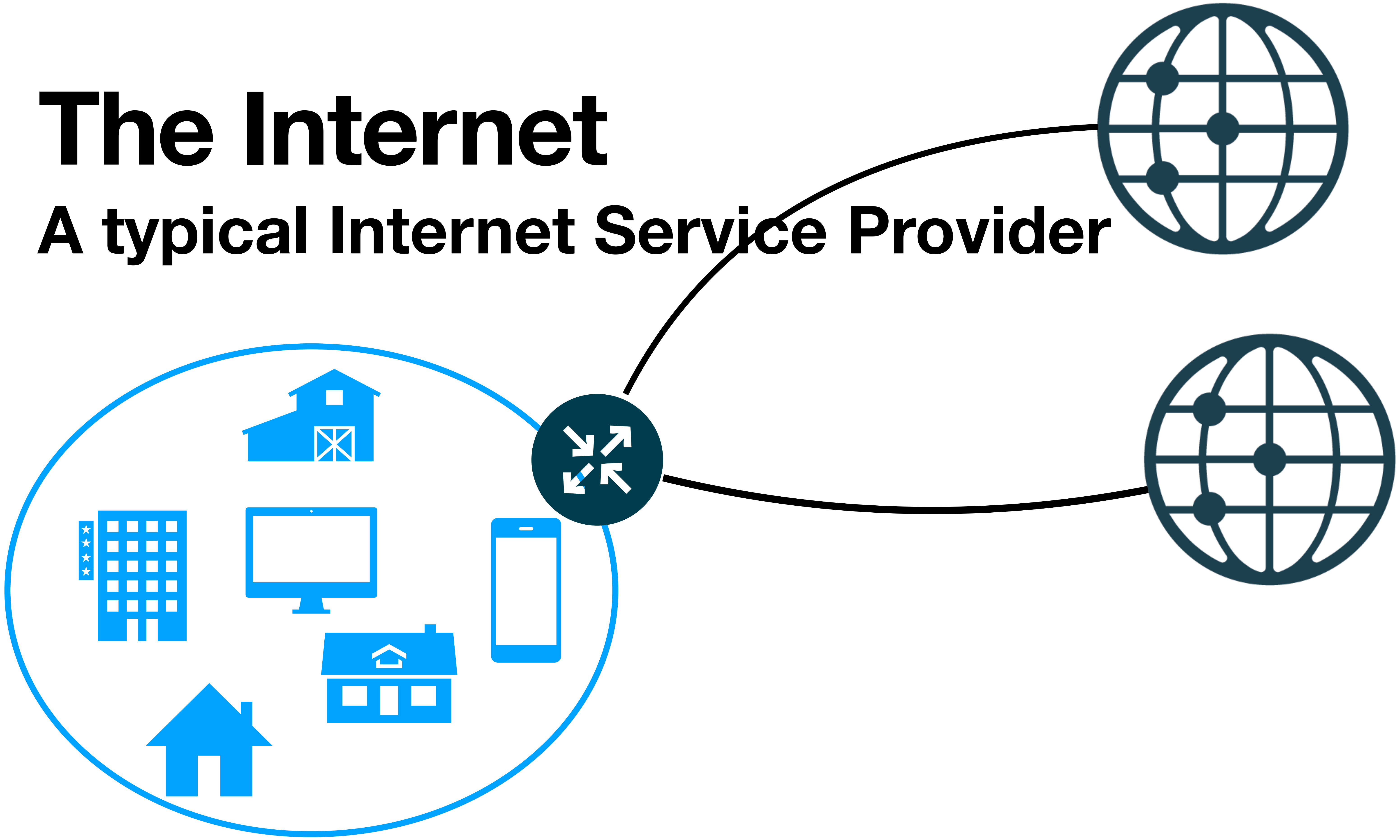
  https://gitlab.com/de-cix-public/team-academy/bgp/BGPLab

Network relationships

DE CIX

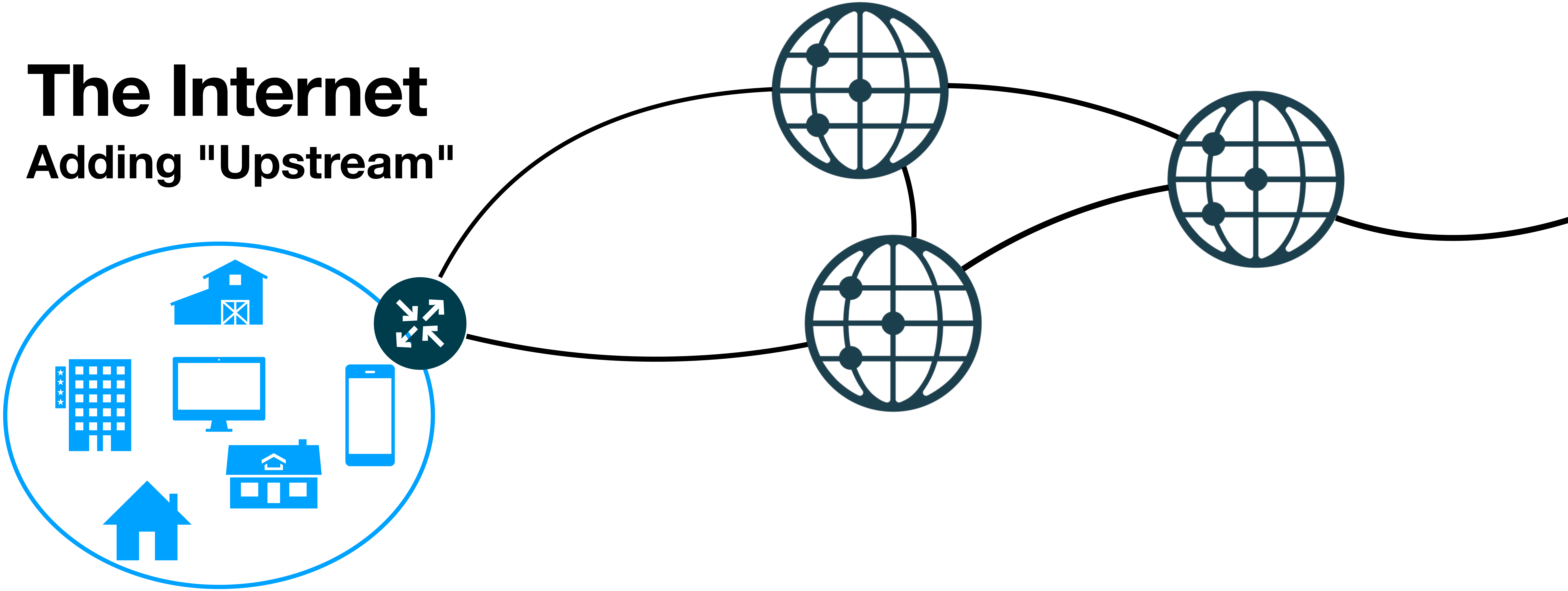
The Internet

A typical Internet Service Provider



The Internet

Adding "Upstream"



The Internet

Adding a 2nd ISP

The Internet

Data transport via upstreams

The Internet

More direct via "peering"

The Internet

Peering on multiple levels

The Internet

Peering on multiple levels

Peering Hierarchy

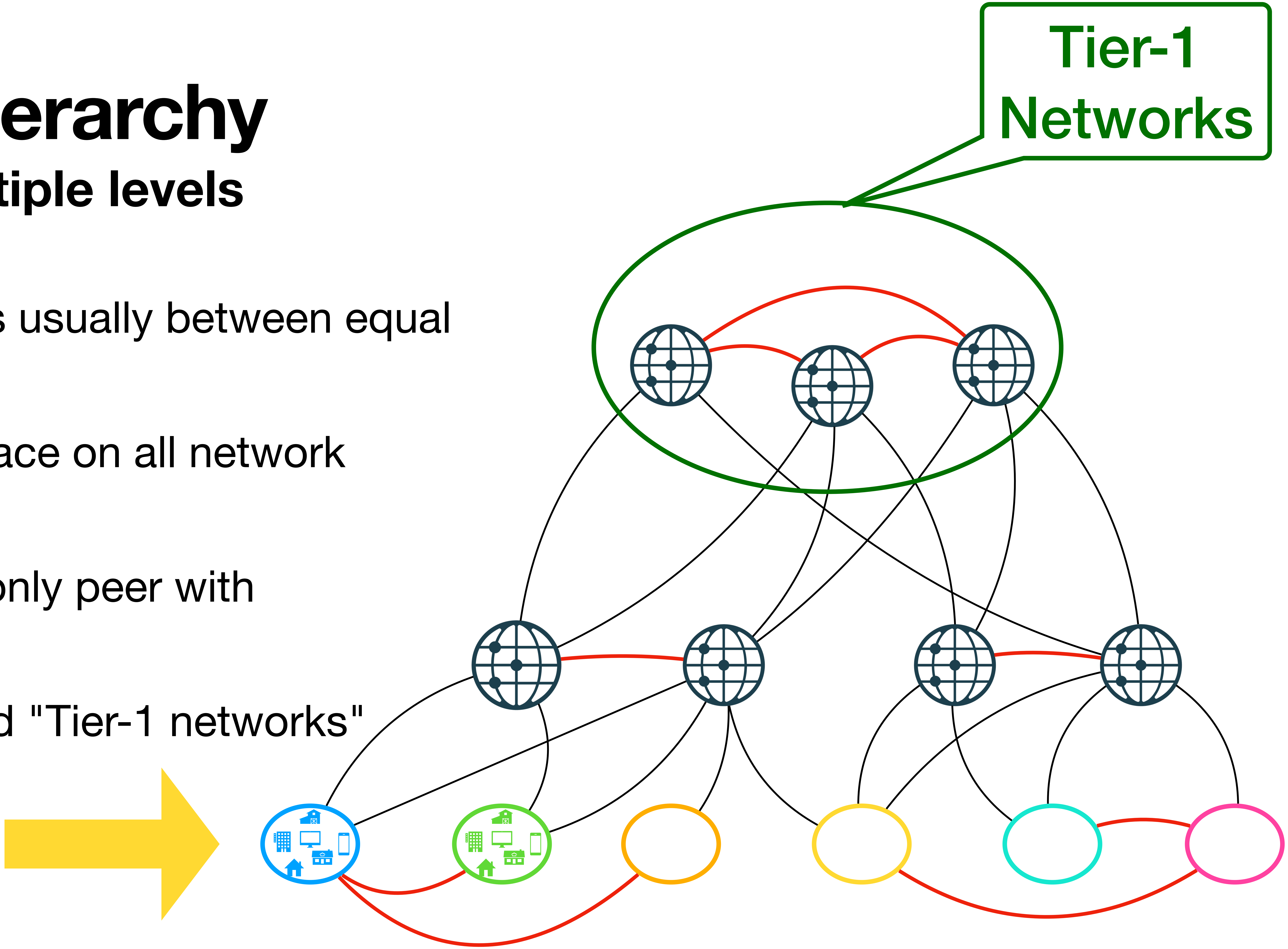
Peering on multiple levels

- Peering happens usually between equal size networks

- Peering takes place on all network levels

- The "top ones" only peer with each other

  - They are called "Tier-1 networks"

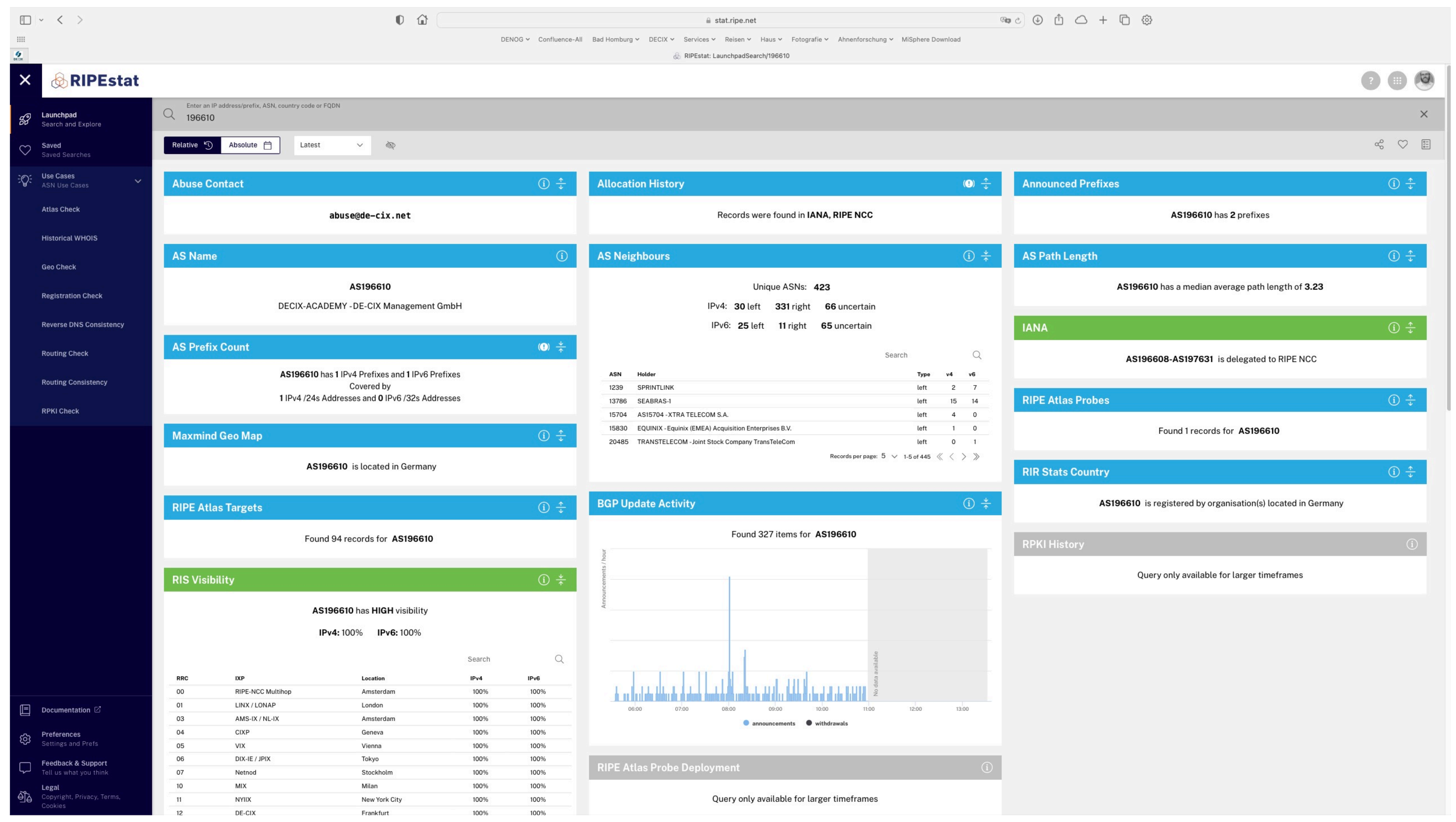


Public tools for BGP

Public tools for BGP

RIPE Stat

- Operated by the RIPE NCC (same entity handing out AS numbers in this region)
- Details about prefixes, ASes and more
- just check it out at https://stat.ripe.net



Enter an IP address/prefix, ASN, country code or FQDN: 196610

Abuse Contact

abuse@de-cix.net

AS Name

AS196610

DECIX-ACADEMY -DE-CIX Management GmbH

AS Prefix Count

AS196610 has **1** IPv4 Prefixes and **1** IPv6 Prefixes
Covered by
1 IPv4 /24s Addresses and **0** IPv6 /32s Addresses

Maxmind Geo Map

AS196610 is located in Germany

RIPE Atlas Targets

Found 94 records for **AS196610**

RIS Visibility

AS196610 has **HIGH** visibility

IPv4: 100% **IPv6:** 100%

Search

RRC	IXP	Location	IPv4	IPv6
00	RIPE-NCC Multihop	Amsterdam	100%	100%
01	LINX / LONAP	London	100%	100%
03	AMS-IX / NL-IX	Amsterdam	100%	100%
04	CIXP	Geneva	100%	100%
05	VIX	Vienna	100%	100%
06	DIX-IE / JPIX	Tokyo	100%	100%
07	Netnod	Stockholm	100%	100%
10	MIX	Milan	100%	100%
11	NYIIX	New York City	100%	100%
12	DE-CIX	Frankfurt	100%	100%

Allocation History

Records were found in **IANA, RIPE NCC**

AS Neighbours

Unique ASNs: **423**

IPv4: **30** left **331** right **66** uncertain

IPv6: **25** left **11** right **65** uncertain

Search

ASN	Holder	Type	v4	v6
1239	SPRINTLINK	left	2	7
13786	SEABRAS-1	left	15	14
15704	AS15704 -XTRA TELECOM S.A.	left	4	0
15830	EQUINIX -Equinix (EMEA) Acquisition Enterprises B.V.	left	1	0
20485	TRANSTELECOM -Joint Stock Company TransTeleCom	left	0	1

Records per page: 5 1-5 of 445

BGP Update Activity

Found 327 items for **AS196610**

RIPE Atlas Probe Deployment

Query only available for larger timeframes

Announced Prefixes

AS196610 has **2** prefixes

AS Path Length

AS196610 has a median average path length of **3.23**

IANA

AS196608-AS197631 is delegated to RIPE NCC

RIPE Atlas Probes

Found 1 records for **AS196610**

RIR Stats Country

AS196610 is registered by organisation(s) located in Germany

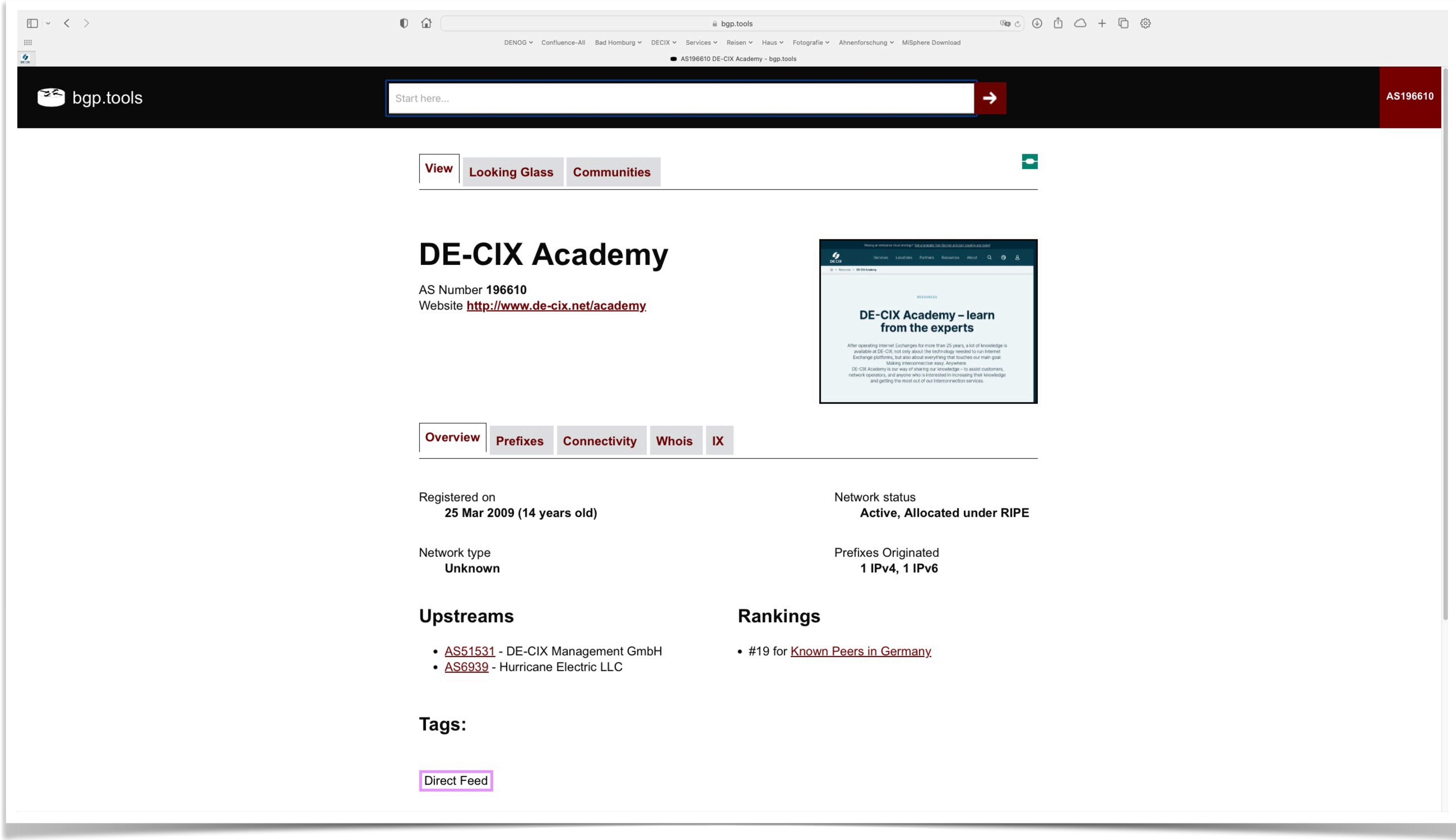
RPKI History

Query only available for larger timeframes

Public tools for BGP

bgp.tools

- Private initiative
- Free, offer premium monitoring service for a fee
- just check it out at https://bgp.tools



DE-CIX Academy

AS Number **196610**
Website http://www.de-cix.net/academy

DE-CIX Academy – learn from the experts

After operating Internet Exchanges for more than 25 years, a lot of knowledge is available at DE-CIX: not only about the technology needed to run Internet Exchange platforms, but also about everything that touches our main goal: Making interconnection easy. Anywhere.
DE-CIX Academy is our way of sharing our knowledge – to assist customers, network operators, and anyone who is interested in increasing their knowledge and getting the most out of our interconnection services.

Registered on
25 Mar 2009 (14 years old)

Network status
Active, Allocated under RIPE

Network type
Unknown

Prefixes Originated
1 IPv4, 1 IPv6

Upstreams

- AS51531 - DE-CIX Management GmbH
- AS6939 - Hurricane Electric LLC

Rankings

- #19 for Known Peers in Germany

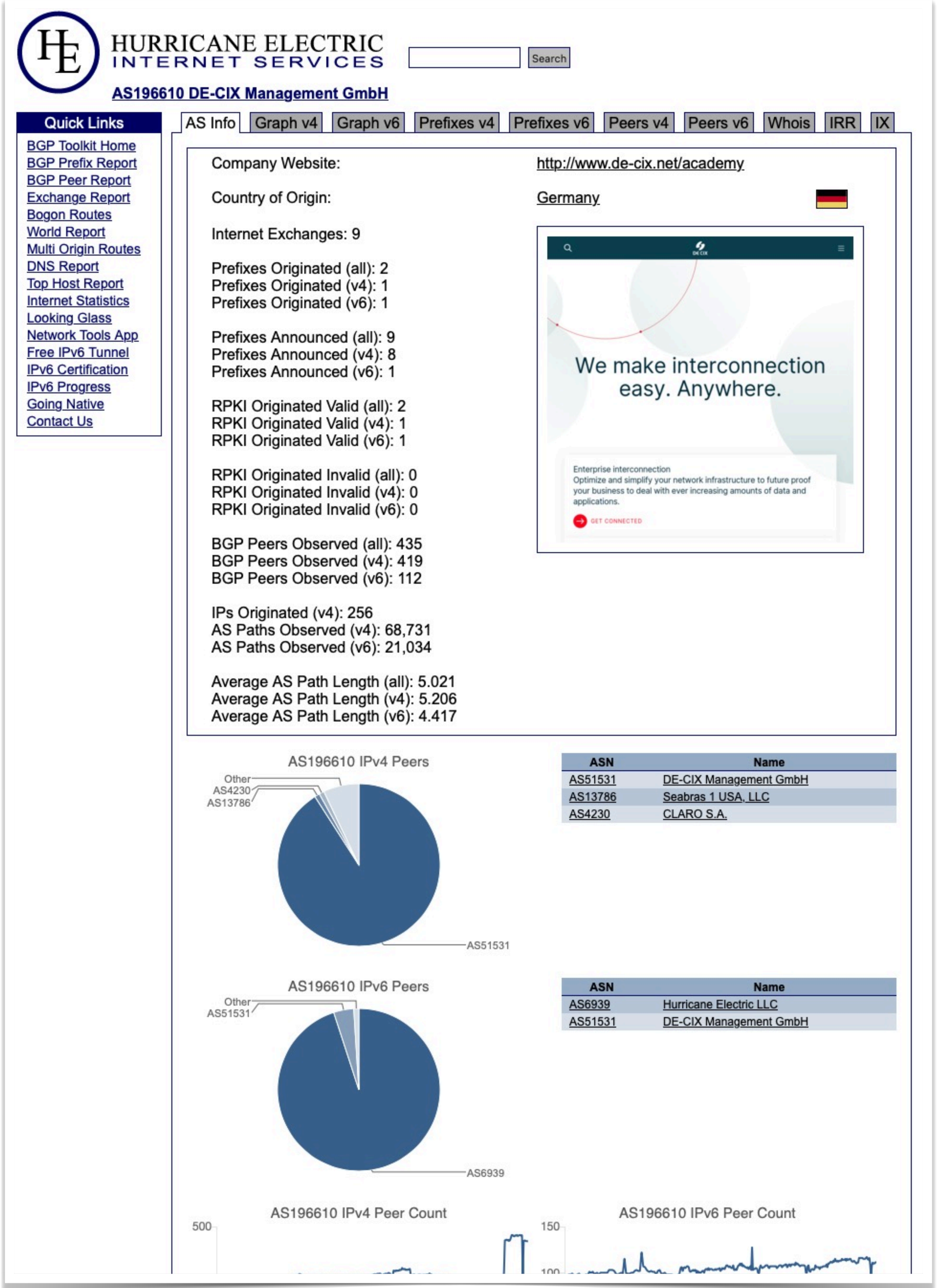
Tags:

Direct Feed

Public tools for BGP

bgp.he.net

- Operated by Hurricane Electric (he.net)
- Free, but shows only HEs point of view
- just check it out at https://bgp.he.net



AS196610 DE-CIX Management GmbH

Company Website:	http://www.de-cix.net/academy
Country of Origin:	Germany

Internet Exchanges: 9

Prefixes Originated (all): 2
Prefixes Originated (v4): 1
Prefixes Originated (v6): 1

Prefixes Announced (all): 9
Prefixes Announced (v4): 8
Prefixes Announced (v6): 1

RPKI Originated Valid (all): 2
RPKI Originated Valid (v4): 1
RPKI Originated Valid (v6): 1

RPKI Originated Invalid (all): 0
RPKI Originated Invalid (v4): 0
RPKI Originated Invalid (v6): 0

BGP Peers Observed (all): 435
BGP Peers Observed (v4): 419
BGP Peers Observed (v6): 112

IPs Originated (v4): 256
AS Paths Observed (v4): 68,731
AS Paths Observed (v6): 21,034

Average AS Path Length (all): 5.021
Average AS Path Length (v4): 5.206
Average AS Path Length (v6): 4.417

Enterprise interconnection
Optimize and simplify your network infrastructure to future proof your business to deal with ever increasing amounts of data and applications.

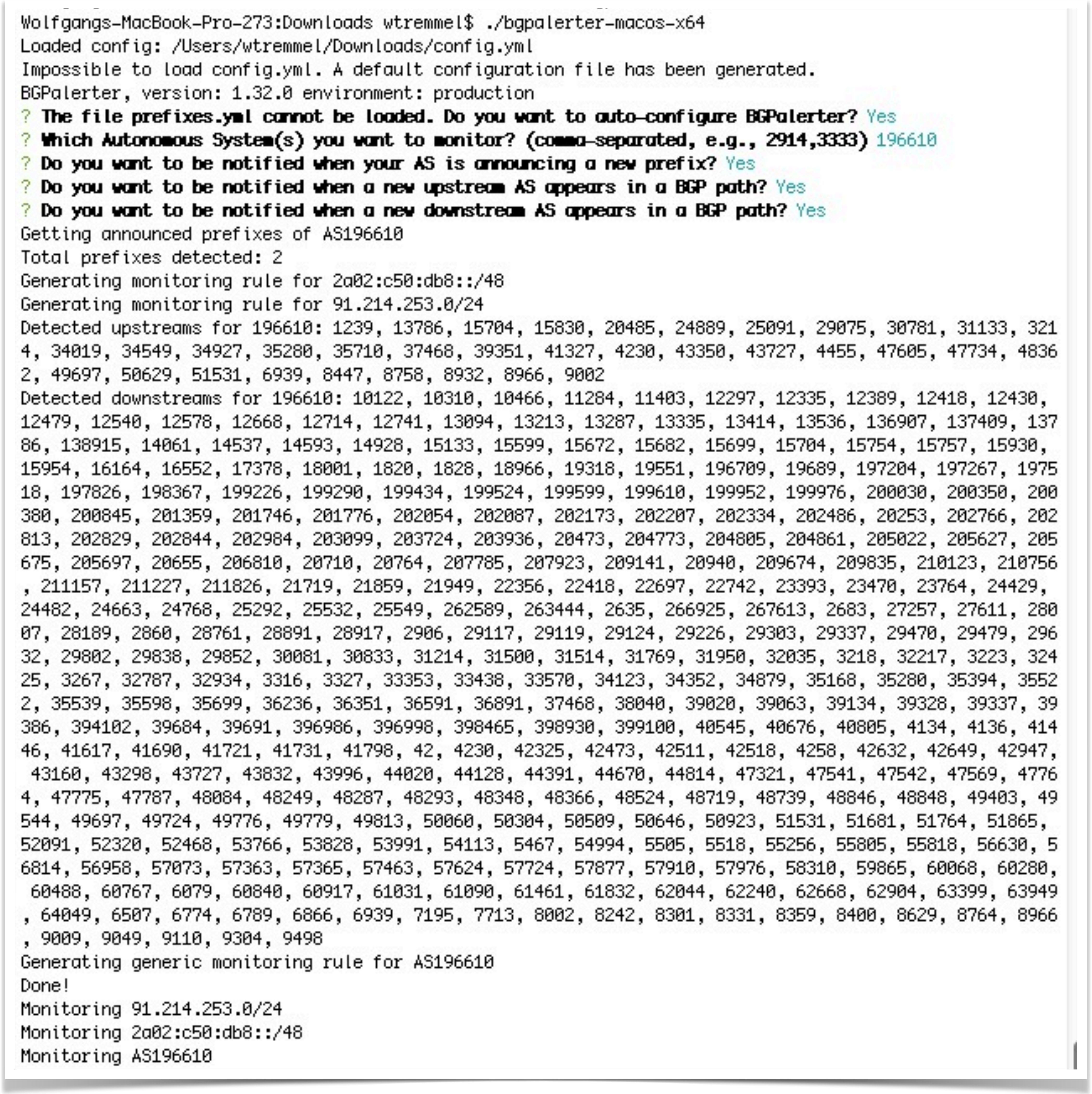
ASN	Name
AS51531	DE-CIX Management GmbH
AS13786	Seabras 1 USA, LLC
AS4230	CLARO S.A.

ASN	Name
AS6939	Hurricane Electric LLC
AS51531	DE-CIX Management GmbH

Public tools for BGP

BGP Alerter

- Open source tool running locally
- Using data from public datasets
  - like ris.ripe.net
- Get the source or a precompiled binary from https://github.com/nttgin/BGPalerter

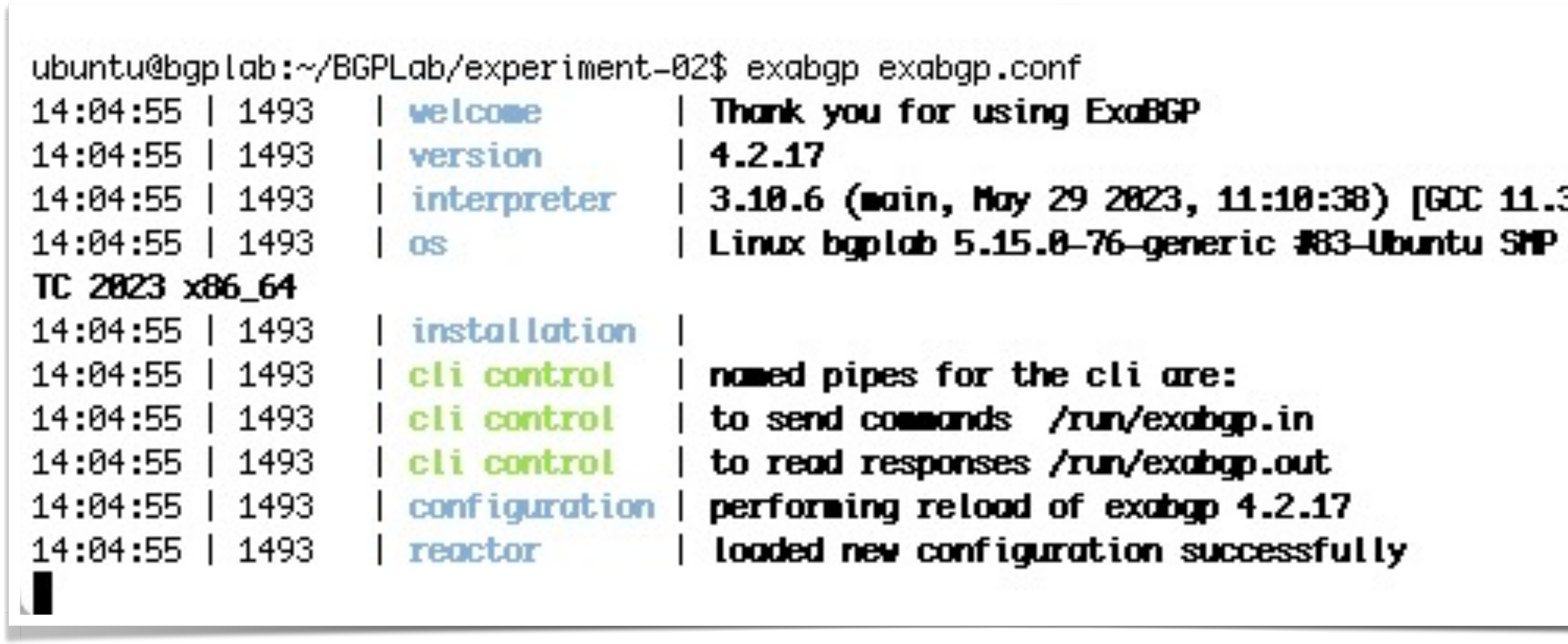


```
Wolfgangs-MacBook-Pro-273:Downloads wtremmel$ ./bgpalerter-macos-x64
Loaded config: /Users/wtremmel/Downloads/config.yml
Impossible to load config.yml. A default configuration file has been generated.
BGPalerter, version: 1.32.0 environment: production
? The file prefixes.yml cannot be loaded. Do you want to auto-configure BGPalerter? Yes
? Which Autonomous System(s) you want to monitor? (comma-separated, e.g., 2914,3333) 196610
? Do you want to be notified when your AS is announcing a new prefix? Yes
? Do you want to be notified when a new upstream AS appears in a BGP path? Yes
? Do you want to be notified when a new downstream AS appears in a BGP path? Yes
Getting announced prefixes of AS196610
Total prefixes detected: 2
Generating monitoring rule for 2a02:c50:db8::/48
Generating monitoring rule for 91.214.253.0/24
Detected upstreams for 196610: 1239, 13786, 15704, 15830, 20485, 24889, 25091, 29075, 30781, 31133, 321
4, 34019, 34549, 34927, 35280, 35710, 37468, 39351, 41327, 4230, 43350, 43727, 4455, 47605, 47734, 4836
2, 49697, 50629, 51531, 6939, 8447, 8758, 8932, 8966, 9002
Detected downstreams for 196610: 10122, 10310, 10466, 11284, 11403, 12297, 12335, 12389, 12418, 12430,
12479, 12540, 12578, 12668, 12714, 12741, 13094, 13213, 13287, 13335, 13414, 13536, 136907, 137409, 137
86, 138915, 14061, 14537, 14593, 14928, 15133, 15599, 15672, 15682, 15699, 15704, 15754, 15757, 15930,
15954, 16164, 16552, 17378, 18001, 1820, 1828, 18966, 19318, 19551, 196709, 19689, 197204, 197267, 1975
18, 197826, 198367, 199226, 199290, 199434, 199524, 199599, 199610, 199952, 199976, 200030, 200350, 200
380, 200845, 201359, 201746, 201776, 202054, 202087, 202173, 202207, 202334, 202486, 20253, 202766, 202
813, 202829, 202844, 202984, 203099, 203724, 203936, 20473, 204773, 204805, 204861, 205022, 205627, 205
675, 205697, 20655, 206810, 20710, 20764, 207785, 207923, 209141, 20940, 209674, 209835, 210123, 210756
, 211157, 211227, 211826, 21719, 21859, 21949, 22356, 22418, 22697, 22742, 23393, 23470, 23764, 24429,
24482, 24663, 24768, 25292, 25532, 25549, 262589, 263444, 2635, 266925, 267613, 2683, 27257, 27611, 280
07, 28189, 2860, 28761, 28891, 28917, 2906, 29117, 29119, 29124, 29226, 29303, 29337, 29470, 29479, 296
32, 29802, 29838, 29852, 30081, 30833, 31214, 31500, 31514, 31769, 31950, 32035, 3218, 32217, 3223, 324
25, 3267, 32787, 32934, 3316, 3327, 33353, 33438, 33570, 34123, 34352, 34879, 35168, 35280, 35394, 3552
2, 35539, 35598, 35699, 36236, 36351, 36591, 36891, 37468, 38040, 39020, 39063, 39134, 39328, 39337, 39
386, 394102, 39684, 39691, 396986, 396998, 398465, 398930, 399100, 40545, 40676, 40805, 4134, 4136, 414
46, 41617, 41690, 41721, 41731, 41798, 42, 4230, 42325, 42473, 42511, 42518, 4258, 42632, 42649, 42947,
 43160, 43298, 43727, 43832, 43996, 44020, 44128, 44391, 44670, 44814, 47321, 47541, 47542, 47569, 4776
4, 47775, 47787, 48084, 48249, 48287, 48293, 48348, 48366, 48524, 48719, 48739, 48846, 48848, 49403, 49
544, 49697, 49724, 49776, 49779, 49813, 50060, 50304, 50509, 50646, 50923, 51531, 51681, 51764, 51865,
52091, 52320, 52468, 53766, 53828, 53991, 54113, 5467, 54994, 5505, 5518, 55256, 55805, 55818, 56630, 5
6814, 56958, 57073, 57363, 57365, 57463, 57624, 57724, 57877, 57910, 57976, 58310, 59865, 60068, 60280,
 60488, 60767, 6079, 60840, 60917, 61031, 61090, 61461, 61832, 62044, 62240, 62668, 62904, 63399, 63949
, 64049, 6507, 6774, 6789, 6866, 6939, 7195, 7713, 8002, 8242, 8301, 8331, 8359, 8400, 8629, 8764, 8966
, 9009, 9049, 9110, 9304, 9498
Generating generic monitoring rule for AS196610
Done!
Monitoring 91.214.253.0/24
Monitoring 2a02:c50:db8::/48
Monitoring AS196610
```

Public tools for BGP

ExaBGP

- Open source tool to "talk" BGP
- Use cases:
  - for testing or even in production
  - announce prefixes
  - with any attributes you want
  - https://github.com/Exa-Networks/exabgp



```
ubuntu@bgplab:~/BGPLab/experiment-02$ exabgp exabgp.conf
14:04:55 | 1493    | welcome       | Thank you for using ExaBGP
14:04:55 | 1493    | version       | 4.2.17
14:04:55 | 1493    | interpreter   | 3.10.6 (main, May 29 2023, 11:10:38) [GCC 11.3
14:04:55 | 1493    | os            | Linux bgplab 5.15.0-76-generic #83-Ubuntu SMP
TC 2023 x86_64
14:04:55 | 1493    | installation  |
14:04:55 | 1493    | cli control   | named pipes for the cli are:
14:04:55 | 1493    | cli control   | to send commands  /run/exabgp.in
14:04:55 | 1493    | cli control   | to read responses /run/exabgp.out
14:04:55 | 1493    | configuration | performing reload of exabgp 4.2.17
14:04:55 | 1493    | reactor       | loaded new configuration successfully
▌
```

Public tools for BGP

DE-CIX Academy BGP lab

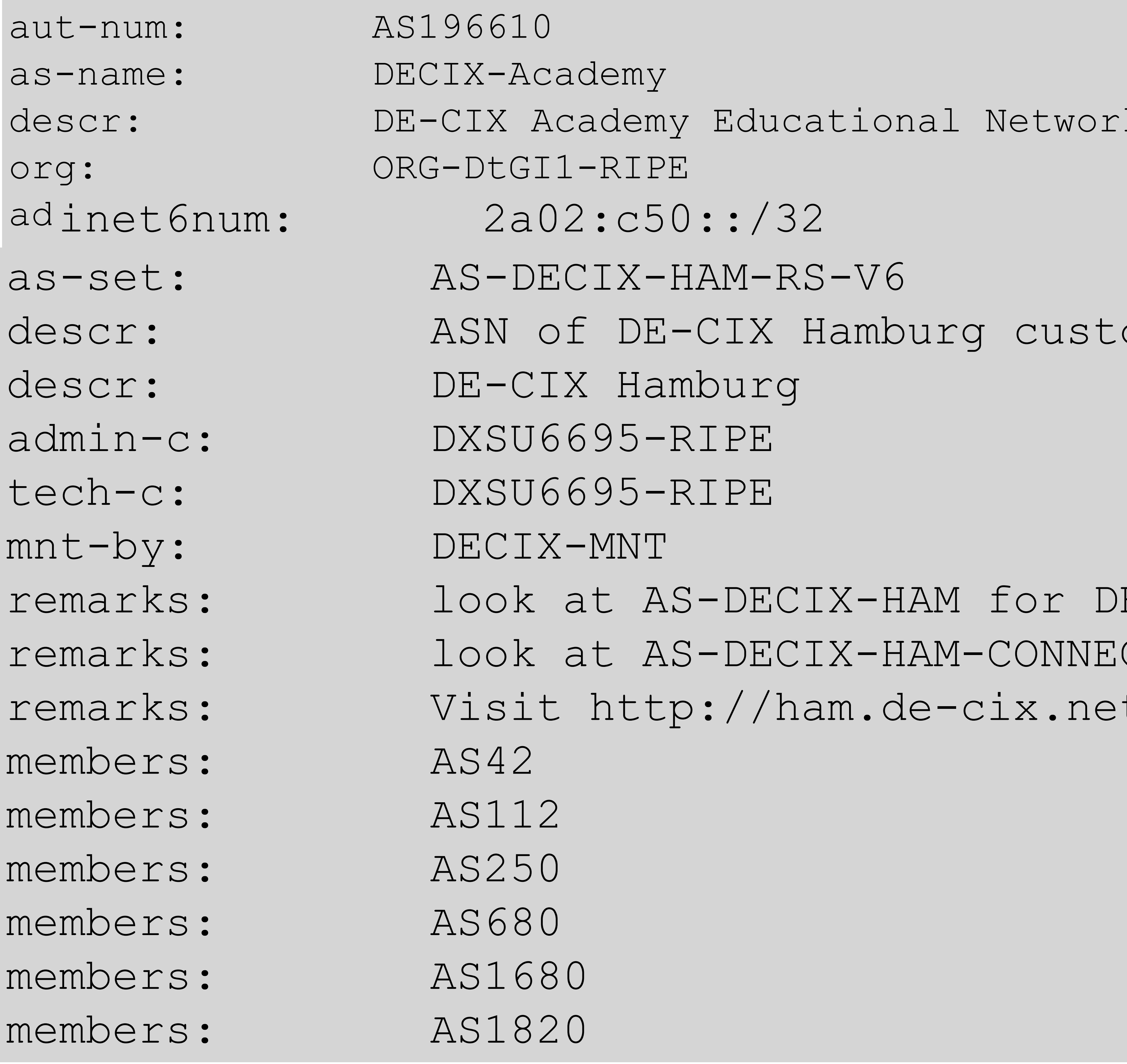
- For teaching a BGP seminar
- Based on FRRouting
- Runs (multiple) routers in Docker containers
- Just needs a linux server as host
- Get it at https://gitlab.com/de-cix-public/team-academy/bgp/BGPLab

Managing BGP relationships

What is the RIPE database?

Documenting our resources

- A public resource database
- It documents:
  - AS numbers, their owners and their use
  - IP resources, their owners and their use
  - AS-sets, lists of ASes
- To access it, you can use the "whois" command
- Or go to the RIPE database website

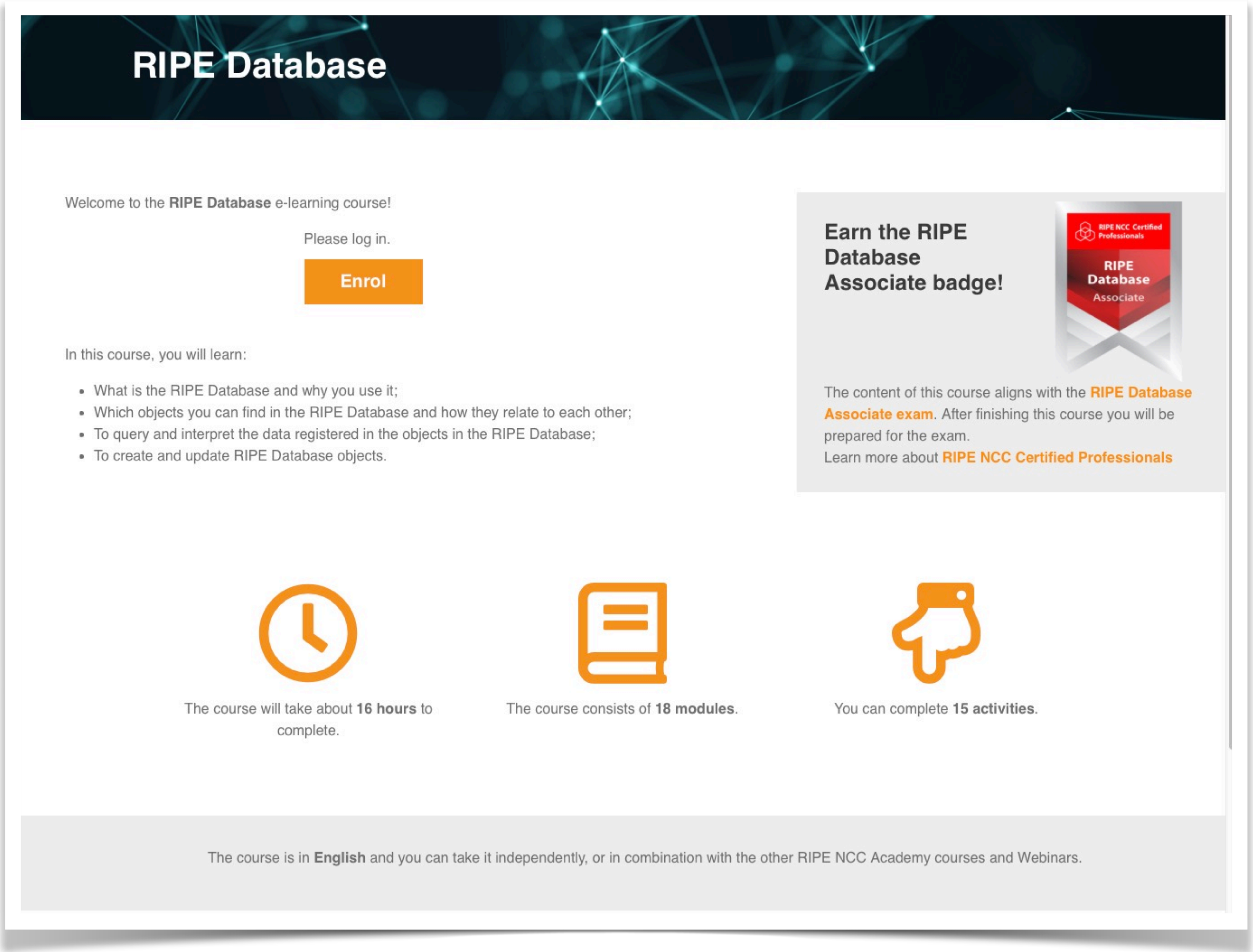


```
aut-num:        AS196610
as-name:        DECIX-Academy
descr:          DE-CIX Academy Educational Networ
org:            ORG-DtGI1-RIPE
adinet6num:         2a02:c50::/32
as-set:           AS-DECIX-HAM-RS-V6
descr:            ASN of DE-CIX Hamburg custo
descr:            DE-CIX Hamburg
admin-c:          DXSU6695-RIPE
tech-c:           DXSU6695-RIPE
mnt-by:           DECIX-MNT
remarks:          look at AS-DECIX-HAM for DE
remarks:          look at AS-DECIX-HAM-CONNE
remarks:          Visit http://ham.de-cix.ne
members:          AS42
members:          AS112
members:          AS250
members:          AS680
members:          AS1680
members:          AS1820
```

More Information?

RIPE Database Training

- The training is free
- The training is online
- Just go the ripe.net website



RIPE Database

Welcome to the **RIPE Database** e-learning course!

Please log in.

Enrol

In this course, you will learn:

- What is the RIPE Database and why you use it;
- Which objects you can find in the RIPE Database and how they relate to each other;
- To query and interpret the data registered in the objects in the RIPE Database;
- To create and update RIPE Database objects.

Earn the RIPE Database Associate badge!

The content of this course aligns with the **RIPE Database Associate exam**. After finishing this course you will be prepared for the exam.
Learn more about **RIPE NCC Certified Professionals**

The course will take about **16 hours** to complete.

The course consists of **18 modules**.

You can complete **15 activities**.

The course is in **English** and you can take it independently, or in combination with the other RIPE NCC Academy courses and Webinars.

https://academy.ripe.net/enrol/index.php?id=9

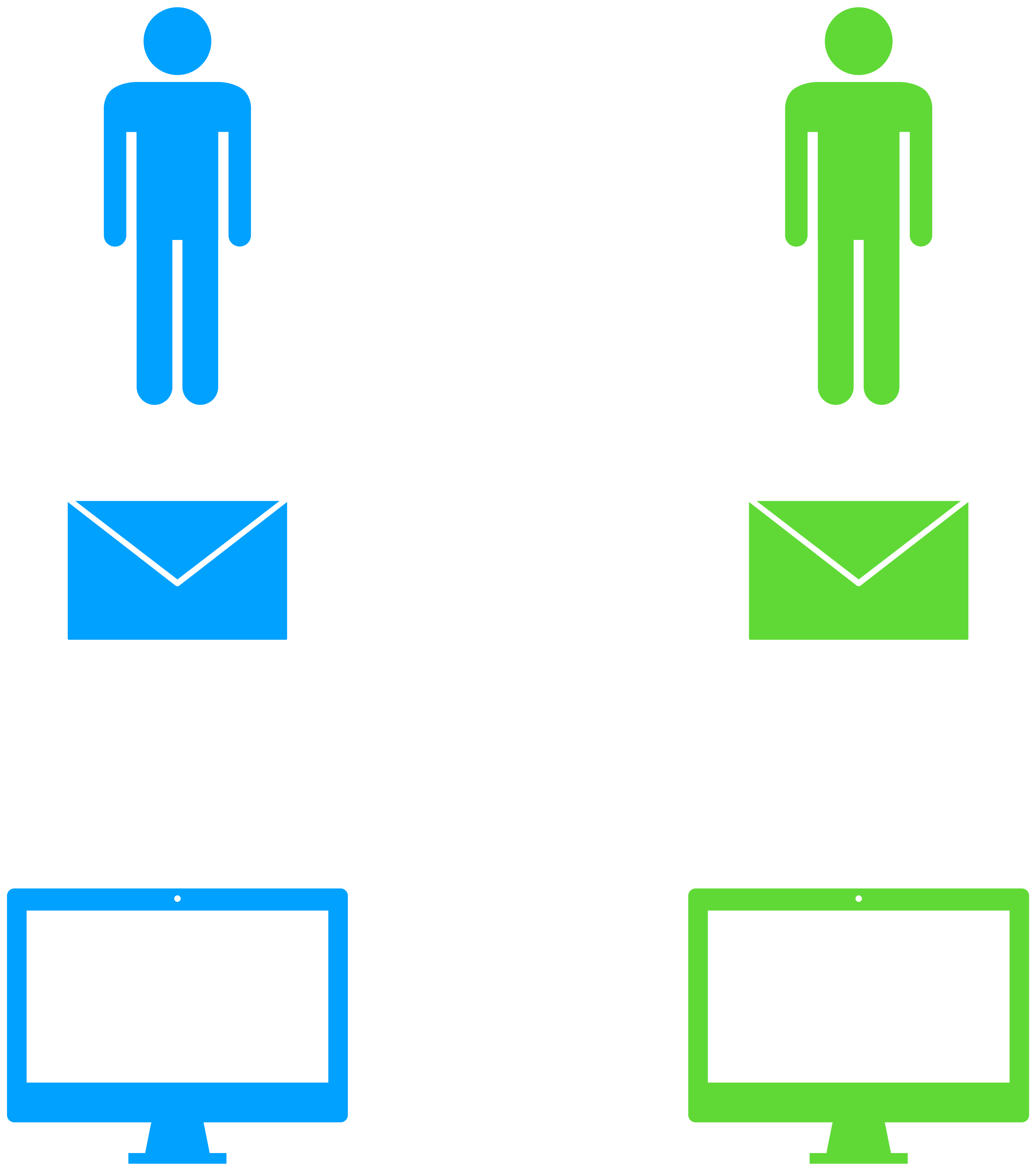
The lazy Network Manager

How to keep record of your peers

Setting up BGP sessions

Standard procedure

- Contact your neighbor
- Exchange a few emails
- Configure BGP

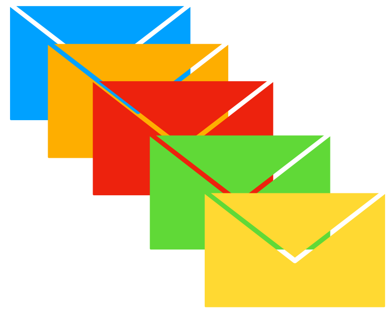


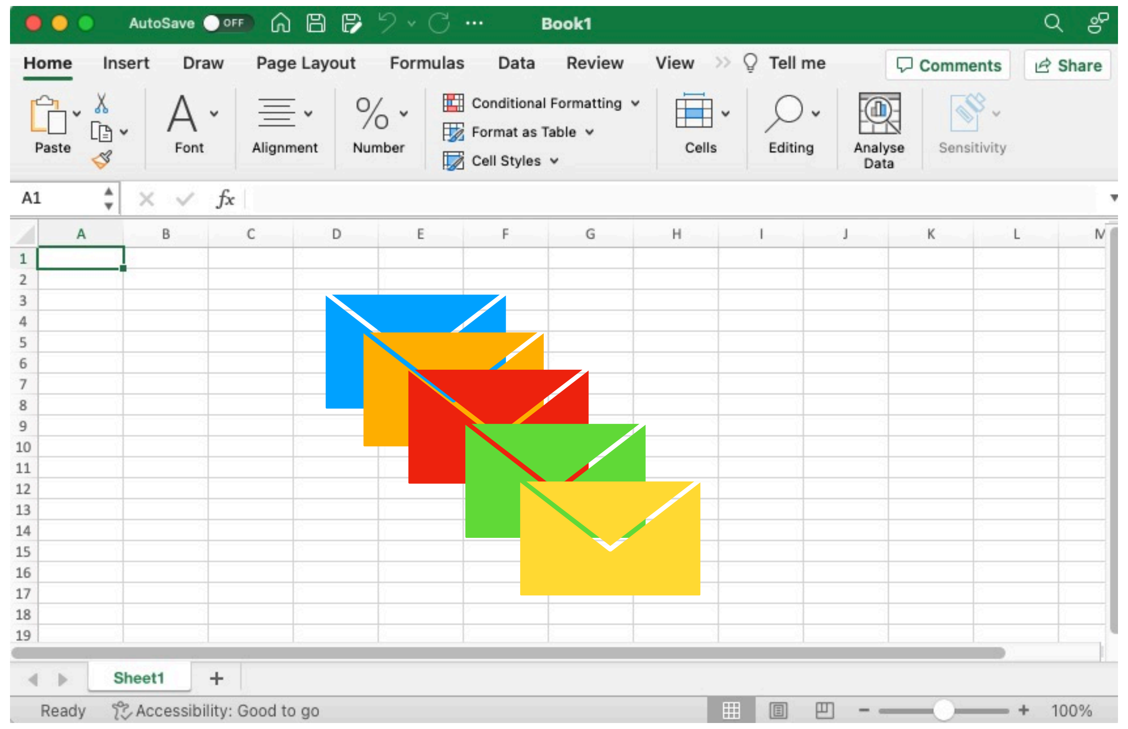
Years later…

You need to contact your neighbor

But where did I put the contact information

- I might have my original emails somewhere
- Or I put the contact information into an Excel sheet
- Or I configured it as a comment on my router
- Or….





But then you notice...

But then you notice...

Surprise, surprise...

- The contact you emailed with works no longer there
- The company name of your peer has changed
- The email address you have (peering@...) is no longer valid
- What now?

There is a solution

Why not have a common database?

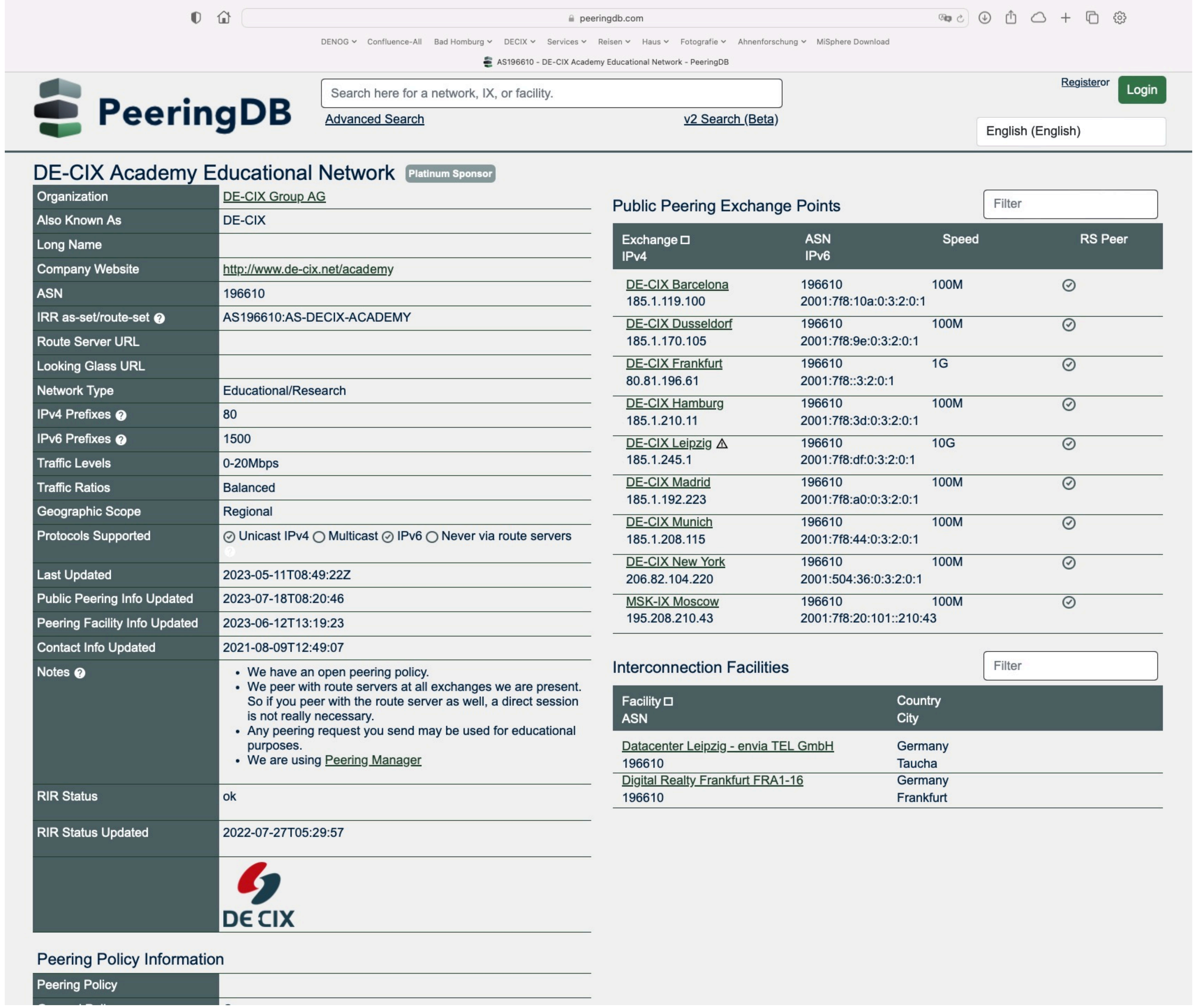
For networks who peer...

- Put contact information into a central database
- Make it accessible for all networks who peer
- Everybody maintains their own information (hopefully)
- If you need some information, simply look it up

PeeringDB

A database for networks who peer

- Free for users
- Financed by sponsoring
- Some public information
- Contact data is private
- Check it out at https://peeringdb.com



DE-CIX Academy Educational Network

Organization	DE-CIX Group AG
Also Known As	DE-CIX
Long Name	
Company Website	http://www.de-cix.net/academy
ASN	196610
IRR as-set/route-set ?	AS196610:AS-DECIX-ACADEMY
Route Server URL	
Looking Glass URL	
Network Type	Educational/Research
IPv4 Prefixes ?	80
IPv6 Prefixes ?	1500
Traffic Levels	0-20Mbps
Traffic Ratios	Balanced
Geographic Scope	Regional
Protocols Supported	✓ Unicast IPv4 ○ Multicast ✓ IPv6 ○ Never via route servers ?
Last Updated	2023-05-11T08:49:22Z
Public Peering Info Updated	2023-07-18T08:20:46
Peering Facility Info Updated	2023-06-12T13:19:23
Contact Info Updated	2021-08-09T12:49:07
Notes ?	• We have an open peering policy. • We peer with route servers at all exchanges we are present. So if you peer with the route server as well, a direct session is not really necessary. • Any peering request you send may be used for educational purposes. • We are using Peering Manager
RIR Status	ok
RIR Status Updated	2022-07-27T05:29:57
	DE CIX

Peering Policy Information

Peering Policy	

Public Peering Exchange Points

Exchange □ IPv4	ASN IPv6	Speed	RS Peer
DE-CIX Barcelona 185.1.119.100	196610 2001:7f8:10a:0:3:2:0:1	100M	✓
DE-CIX Dusseldorf 185.1.170.105	196610 2001:7f8:9e:0:3:2:0:1	100M	✓
DE-CIX Frankfurt 80.81.196.61	196610 2001:7f8::3:2:0:1	1G	✓
DE-CIX Hamburg 185.1.210.11	196610 2001:7f8:3d:0:3:2:0:1	100M	✓
DE-CIX Leipzig ⚠ 185.1.245.1	196610 2001:7f8:df:0:3:2:0:1	10G	✓
DE-CIX Madrid 185.1.192.223	196610 2001:7f8:a0:0:3:2:0:1	100M	✓
DE-CIX Munich 185.1.208.115	196610 2001:7f8:44:0:3:2:0:1	100M	✓
DE-CIX New York 206.82.104.220	196610 2001:504:36:0:3:2:0:1	100M	✓
MSK-IX Moscow 195.208.210.43	196610 2001:7f8:20:101::210:43	100M	✓

Interconnection Facilities

Facility □ ASN	Country City
Datacenter Leipzig - envia TEL GmbH 196610	Germany Taucha
Digital Realty Frankfurt FRA1-16 196610	Germany Frankfurt

BGP Communities

BGP Communities

- →A transitive, optional BGP attribute
  - →**Transitive**: Once attached, it stays until removed
  - →**Optional**: it does not have to be there
- →"BGP Communities are like a sticker on a suitcase"

Source: https://openclipart.org/detail/273877/suitcase-with-stickers-no-trademarks-remix

"Original" BGP Communities

→Definition:

"A community is a group of destinations which share some common property"

→Introduced in RFC1997 in year 1996

→A community is expressed by a 32Bit-Number

→High 16 bit are the AS defining the low 16 bits

→Notation: "6695:1000", "5669:32000"

→You can attach as many communities as you like (within reason)

→BGP max message size is 4096 Bytes

What are they useful for? Information!

198.51.100.0/24 80.81.192.15 from 80.81.192.15

Path: 1301 286 517

Origin IGP, metric 0, localpref 100, valid, external

Informational Communities

198.51.100.0/24 80.81.192.15 from 80.81.192.15

Path: 1301 286 517

Origin IGP, metric 0, localpref 100, valid, external

Received from: **Upstream**

Customer

Peering

Example: Encode geographical information

65010:1____

Example: "1" here means geographical community

You may encode the continent here (if you are global) like:

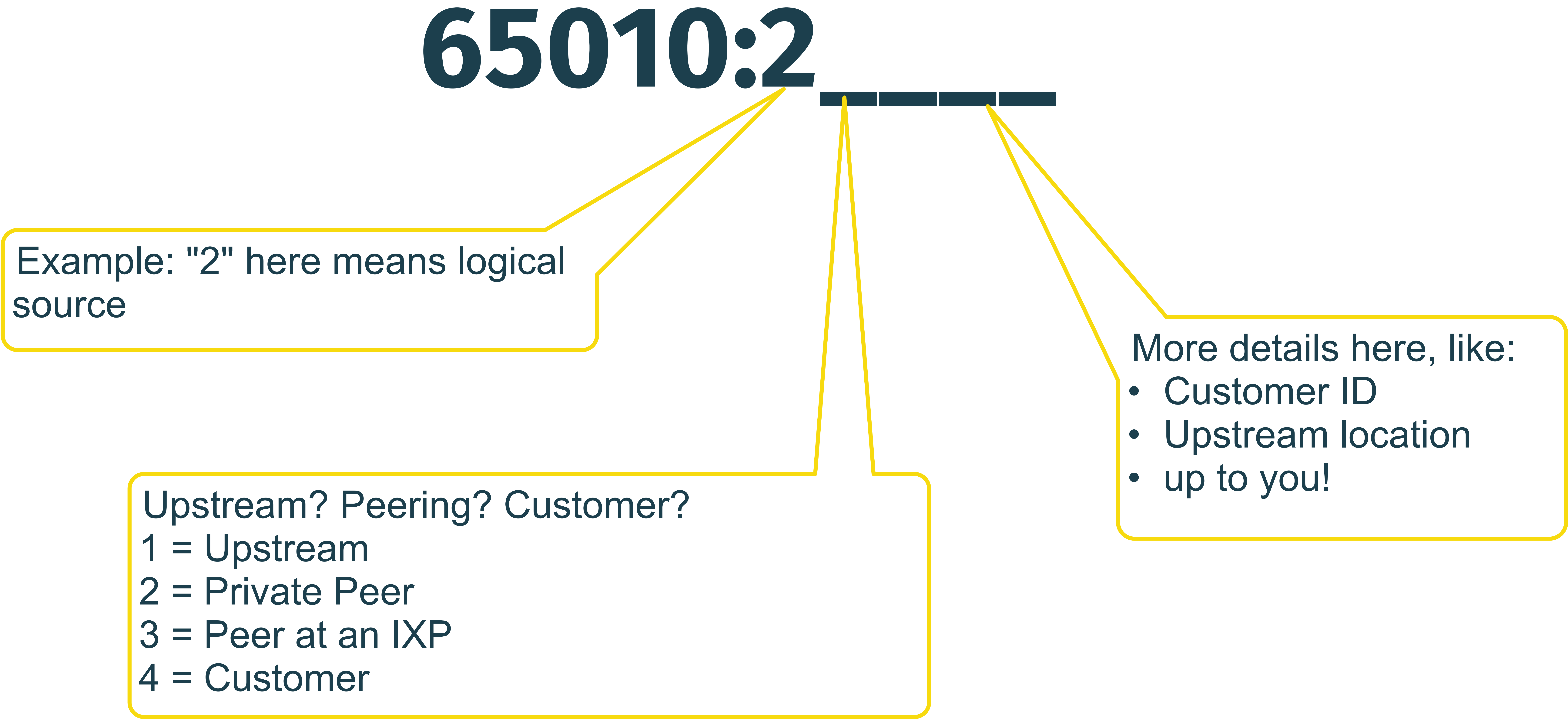
- 1 = Europe
- 2 = North America
- 3 = Asia …

ISO-Country-Codes here …

250 - France
276 - Germany
840 - USA

Just an Example!

Example: Encode logical information



65010:2____

Example: "2" here means logical source

Upstream? Peering? Customer?
1 = Upstream
2 = Private Peer
3 = Peer at an IXP
4 = Customer

More details here, like:

- Customer ID
- Upstream location
- up to you!

What are they useful for? Action!

198.51.100.0/24 80.81.192.15 from 80.81.192.15

Path: 65010

Origin IGP, metric 0, localpref 100, valid, external

Announce to Customers Peers Upstreams

DE CIX

Encoding up to you!

DE CIX

Action Communities: Encoding

→ Again - you only have two 16bit numbers … (with original BGP Communities)

→ Some Ideas …

- If you want your customers to send you "actions"
  - You really should have them put your AS number into the first 16bit number
  - You **must scrub** everything they should not send on incoming
- Possible actions:
  - (not) announce to upstream, peers, customers
  - fine granular announcement control (geographically, by IXP, …)
  - announce with longer AS path
  - change *local preference*
  - Blackhole

Action Communities: Well-Known

- A couple of communities are pre-defined by RFCs
- NO-EXPORT
  - Do not send the prefix to eBGP neighbours (other ASes)
- NO-ADVERTISE
  - Do not send the prefix to anyone (not even internal via iBGP)
- NO-PEER
  - Do not send to any peers
- BLACKHOLE
  - Sink all traffic to prefixes tagged with this community
  - Most commonly used with host routes
  - Implies NO-EXPORT

32Bit AS? No luck with original communities

65010:12345

→ Two 16-bit numbers

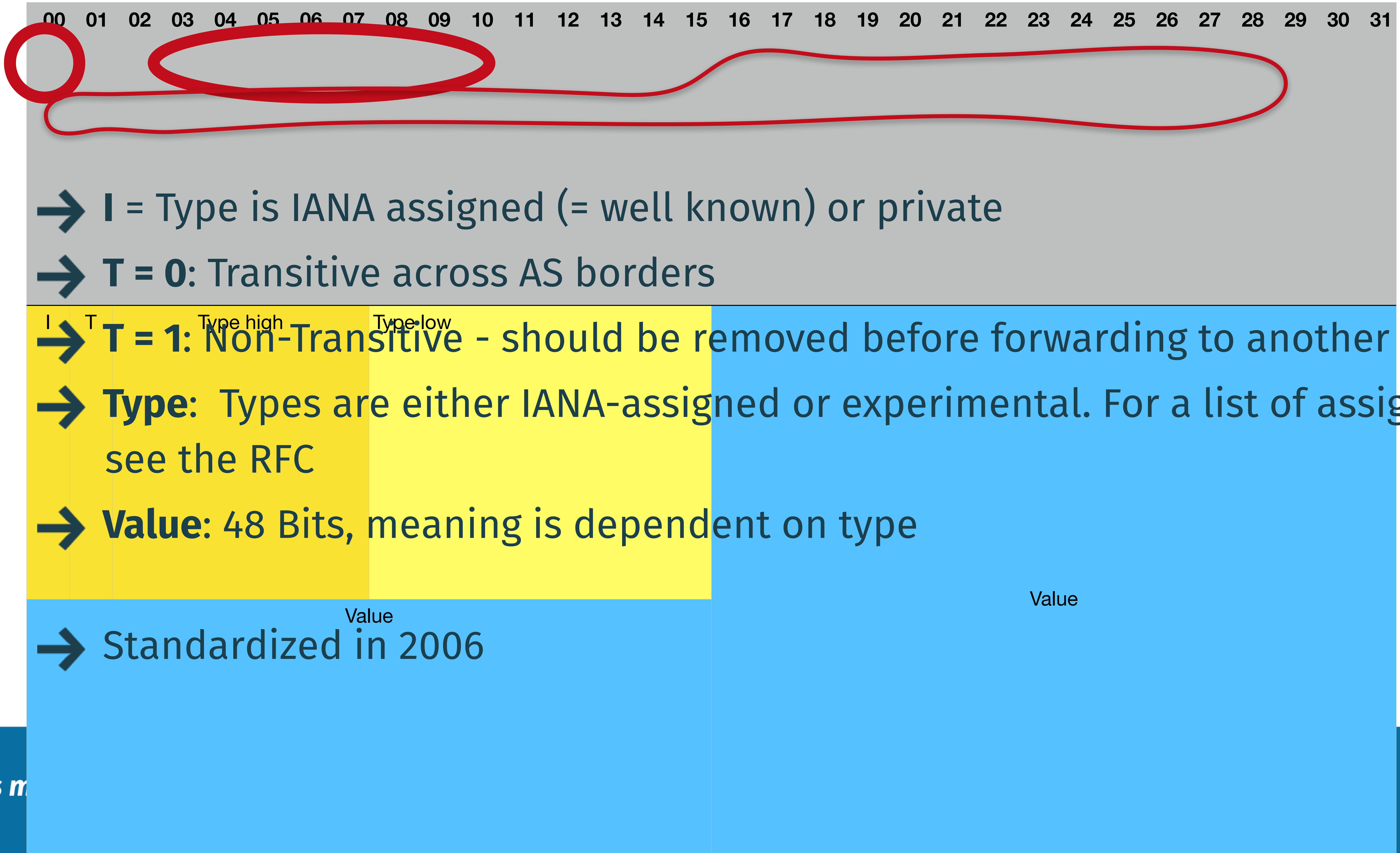
→ No way to encode a 32Bit AS number and something else …

- RFC4360 - Extended Communities

→ Extended Communities - Lots of new features

- In total 2*32Bits
- Introducing a "type" field
- Possible to encode 16Bit Type, 32Bit AS, 16Bit Data

Extended Communities



- **I** = Type is IANA assigned (= well known) or private
- **T = 0**: Transitive across AS borders
- **T = 1**: Non-Transitive - should be removed before forwarding to another AS
- **Type**: Types are either IANA-assigned or experimental. For a list of assigned types see the RFC
- **Value**: 48 Bits, meaning is dependent on type
- Standardized in 2006

Extended Communities and 32Bit ASes

00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
I	T	0x02						0x02 or 0x03								32Bit-AS Number															
32Bit-AS Number (continued)																Value															

- → You can encode a 32Bit AS-Number
  - → and a 16 Bit value

Extended Communities and 32Bit ASes

00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
I	T	0x02						0x02 or 0x03								16Bit-AS Number															
32Bit Value																															

- → You can encode a 32Bit AS-Number
  - → and a 16 Bit value
- → or a 16Bit AS-Number
  - → and a 32 Bit value

- → 32Bit AS and 32Bit Value?
  - → **not possible!**

Extended communities use cases

→ Notation:

- Similar to original communities: **RT:6500000:1234** or **RT:1234:6500000**

→ Disadvantages

- Only 48bits in total
- Only one 32Bit value is possible (and one 16Bit value)
- RT, RO and other types confusing to many operators

→ Conclusion

- Another community version was needed
- It took the IETF a while to realize that (11 years)

Introducing: Large Communities

<table>
<tr><th>00</th><th>01</th><th>02</th><th>03</th><th>04</th><th>05</th><th>06</th><th>07</th><th>08</th><th>09</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th><th>16</th><th>17</th><th>18</th><th>19</th><th>20</th><th>21</th><th>22</th><th>23</th><th>24</th><th>25</th><th>26</th><th>27</th><th>28</th><th>29</th><th>30</th><th>31</th></tr>
<tr><td colspan="32">Global Administrator (32Bit AS)</td></tr>
<tr><td colspan="32">Local Data Part 1 (Function)</td></tr>
<tr><td colspan="32">Local Data Part 2 (Parameter)</td></tr>
</table>

→ Very simple - three 32Bit values (finally something useful)

→ Global Administrator:

- An AS number (in 32Bit notation)
- Has defined meaning of two other fields
- May have published that meaning

→ Local Data

- Can be seen as "just two 32Bit numbers"
- Or as "Function" / "Parameter"

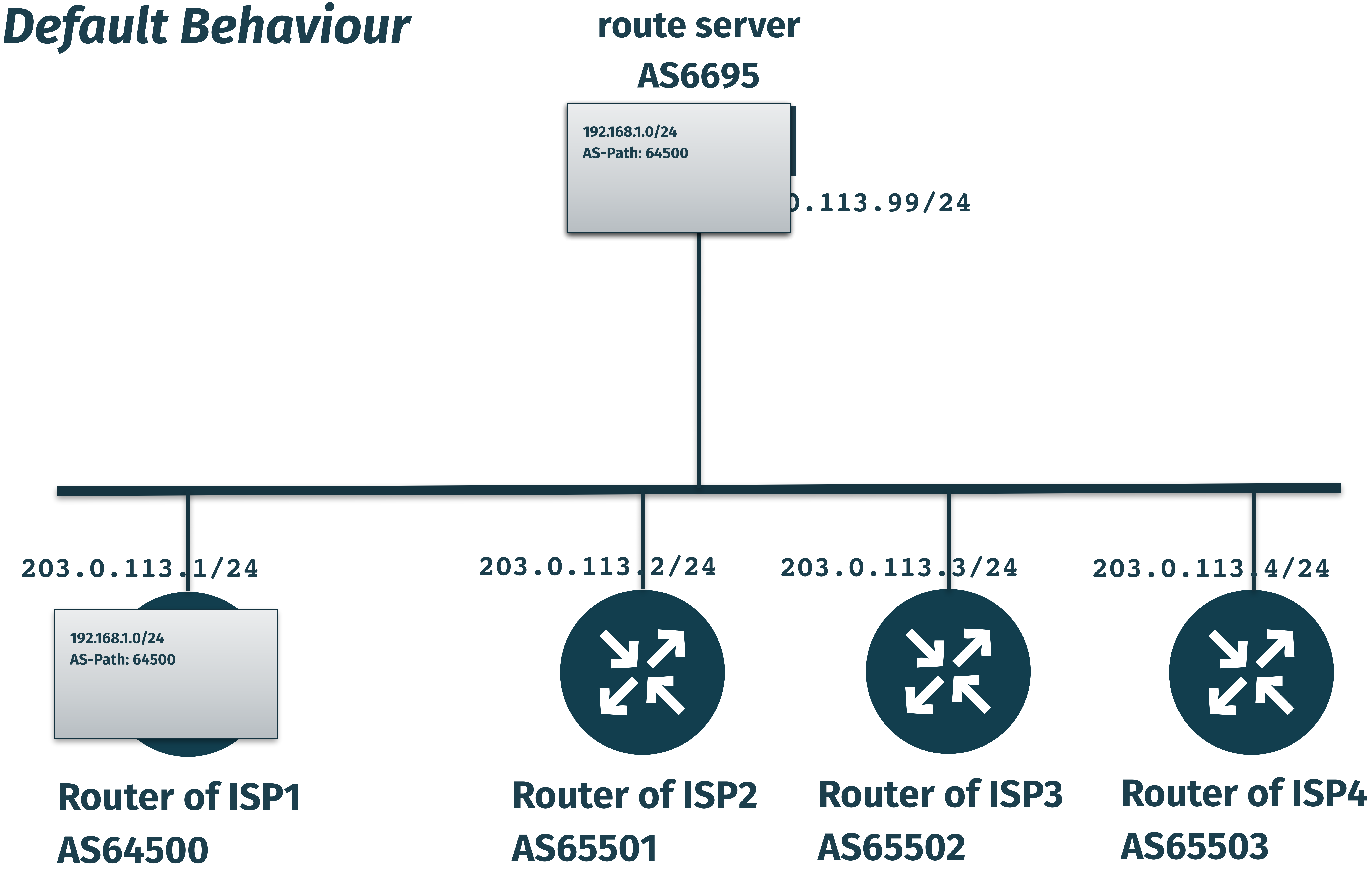
Large BGP Communities

<table>
<tr><td>00</td><td>01</td><td>02</td><td>03</td><td>04</td><td>05</td><td>06</td><td>07</td><td>08</td><td>09</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td><td>22</td><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td><td>29</td><td>30</td><td>31</td></tr>
<tr><td colspan="32">Global Administrator (32Bit AS)</td></tr>
<tr><td colspan="32">Local Data Part 1 (Function)</td></tr>
<tr><td colspan="32">Local Data Part 2 (Parameter)</td></tr>
</table>

- Notation:
  - Similar to Original Communities: ***196610:100:65000010***
- Defined in two RFCs:
  - RFC8092: BGP Large Communities Attribute
  - RFC8195: Use of BGP Large Communities
- A dedicated website exists: http://largebgpcommunities.net
  - Keeping track of Implementations, News etc.

BGP Communities and the DE-CIX Route Servers

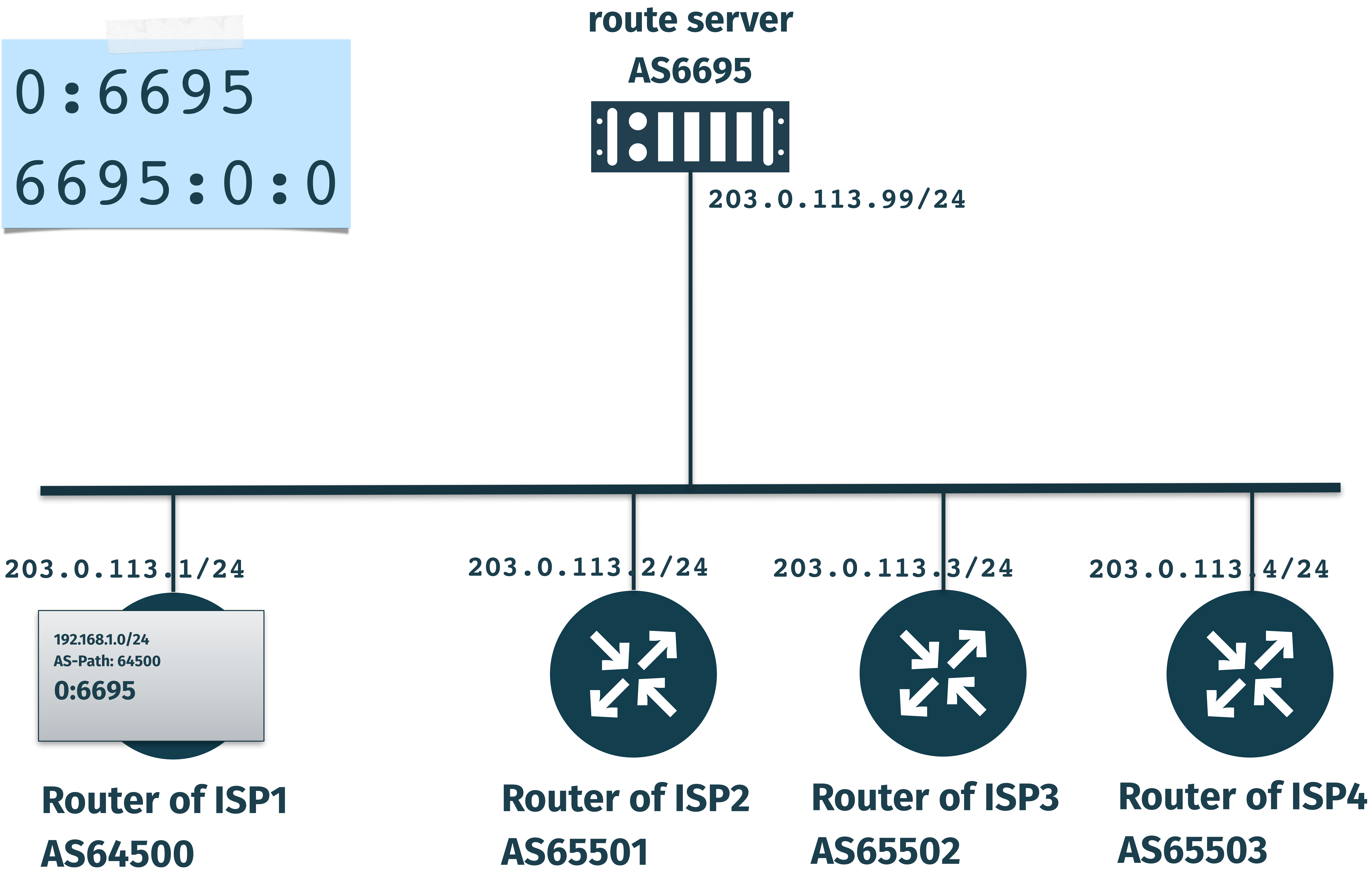
Default Behaviour



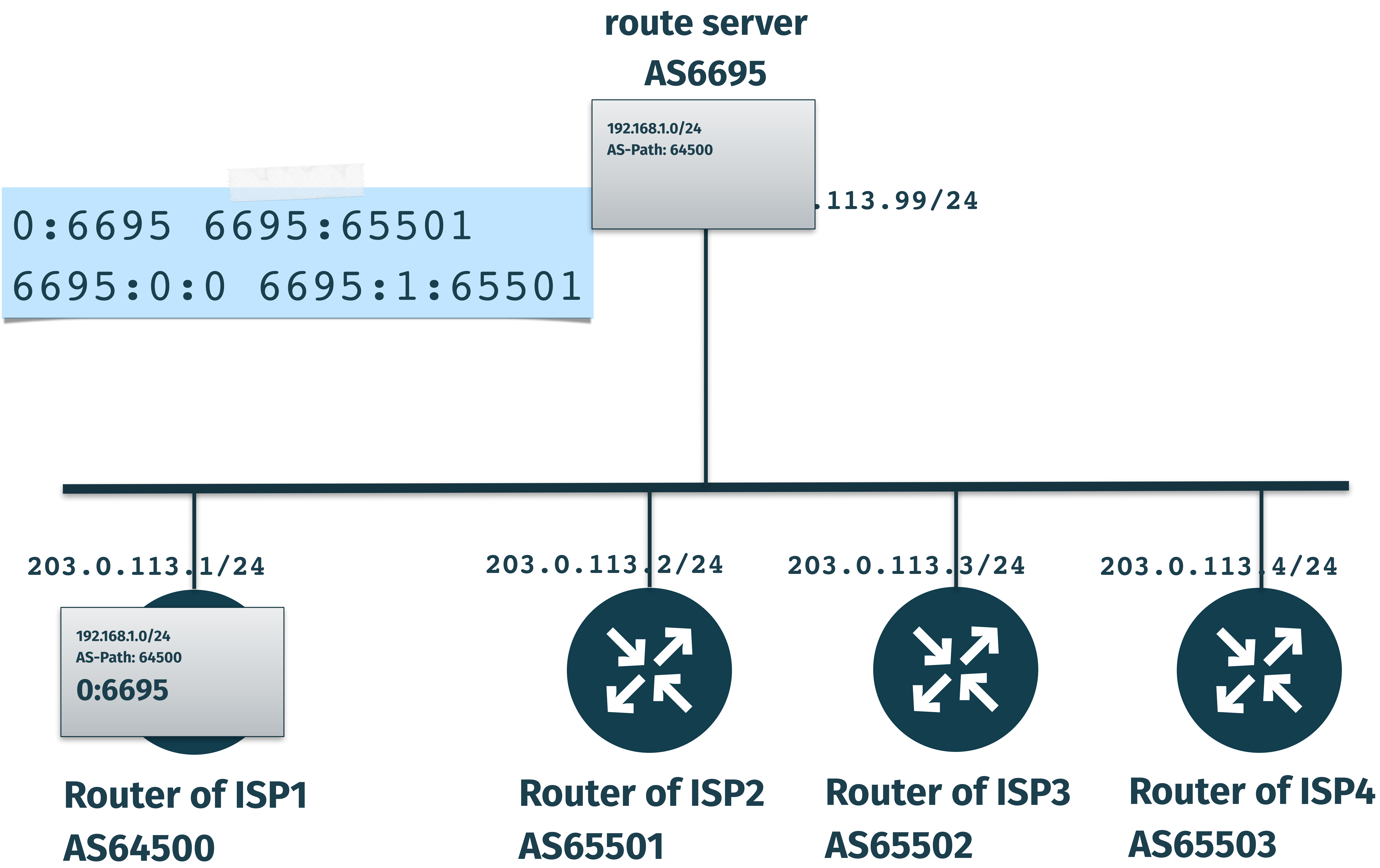
Do not announce to any AS

0:6695

6695:0:0



Do not announce to any AS, but announce to AS65501



route server
AS6695

192.168.1.0/24
AS-Path: 64500

.113.99/24

```
0:6695 6695:65501
6695:0:0 6695:1:65501
```

203.0.113.1/24

192.168.1.0/24
AS-Path: 64500
0:6695

203.0.113.2/24

203.0.113.3/24

203.0.113.4/24

Router of ISP1
AS64500

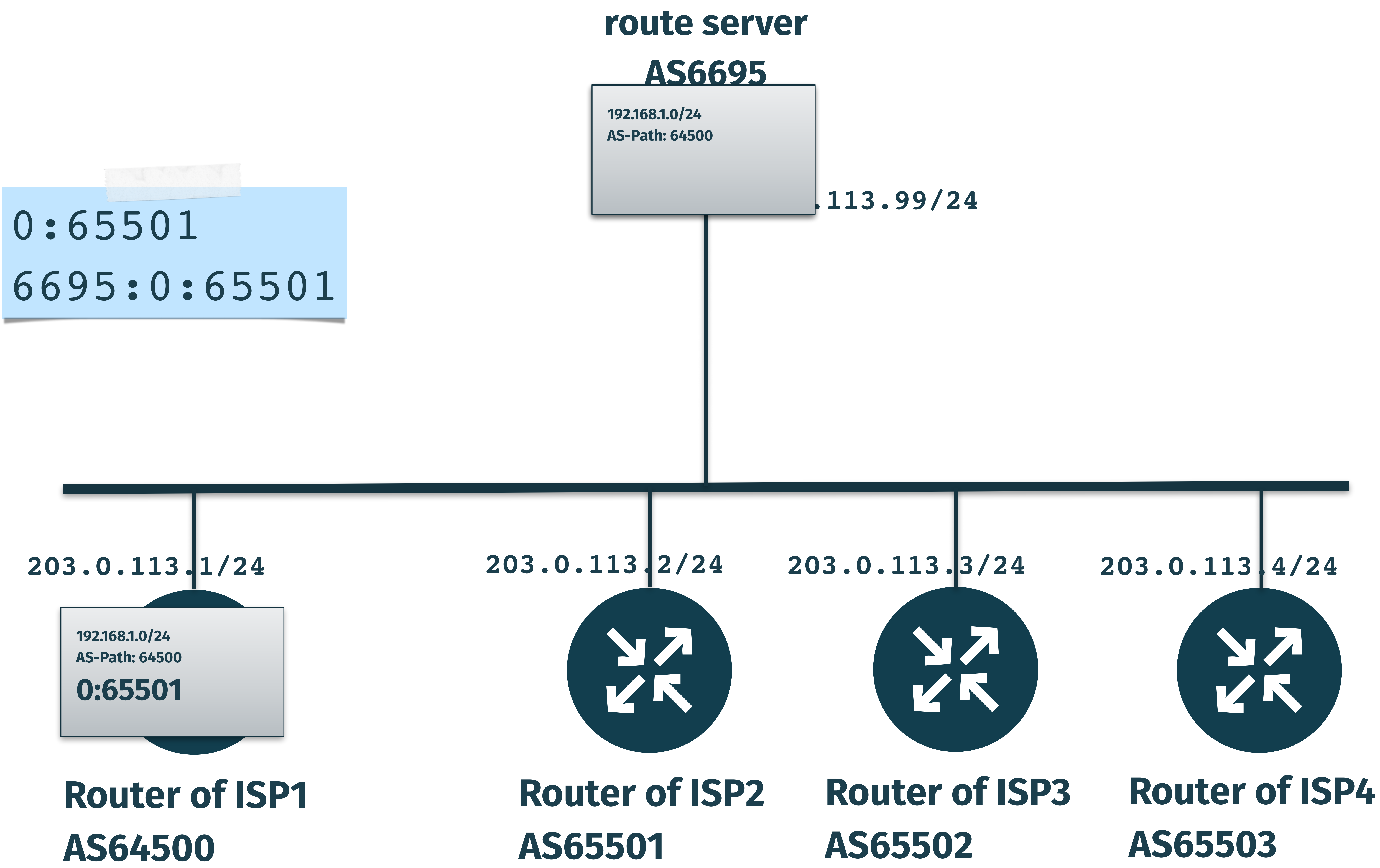
Router of ISP2
AS65501

Router of ISP3
AS65502

Router of ISP4
AS65503

Do not announce to AS65501

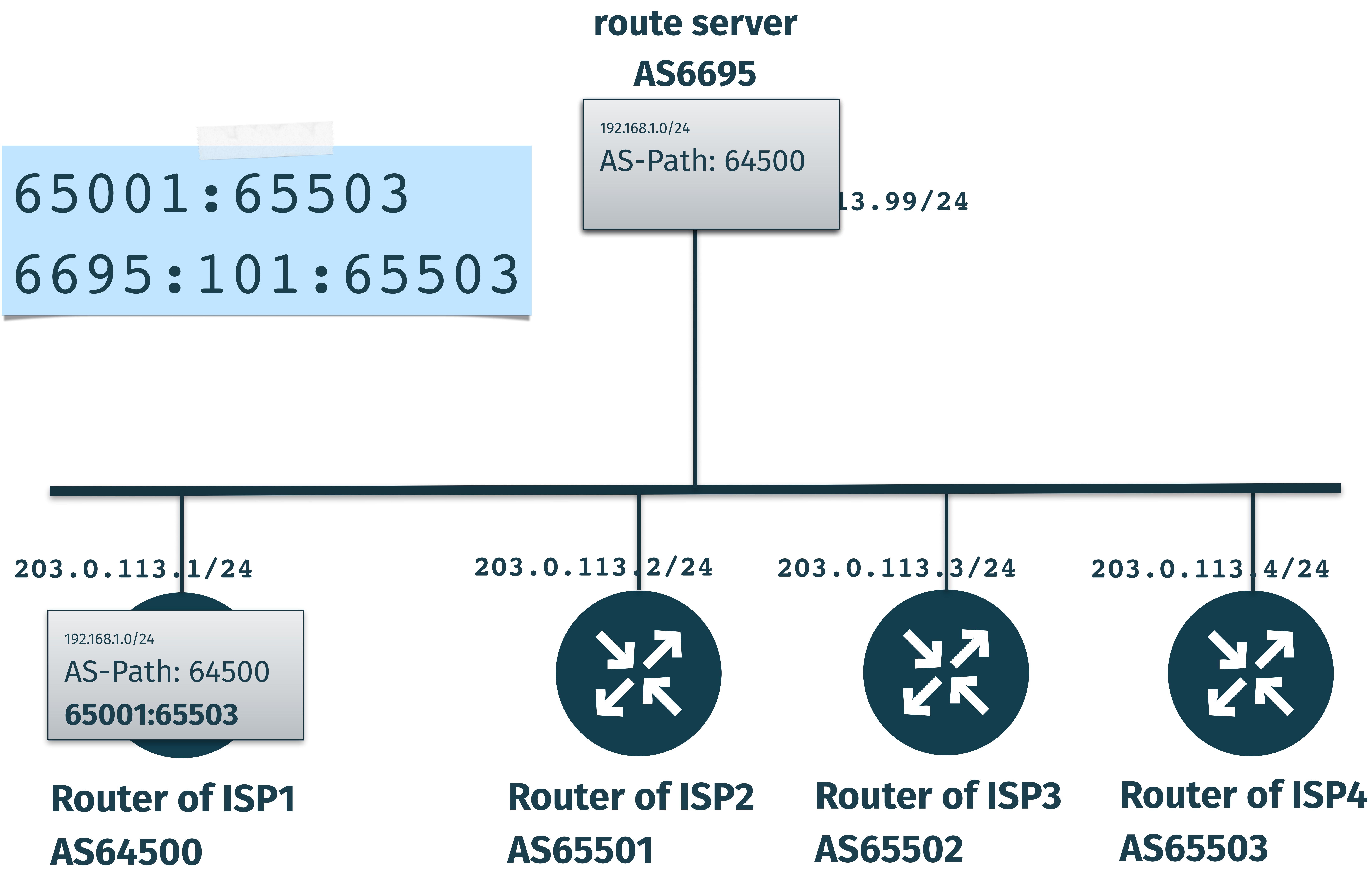
```
0:65501
6695:0:65501
```



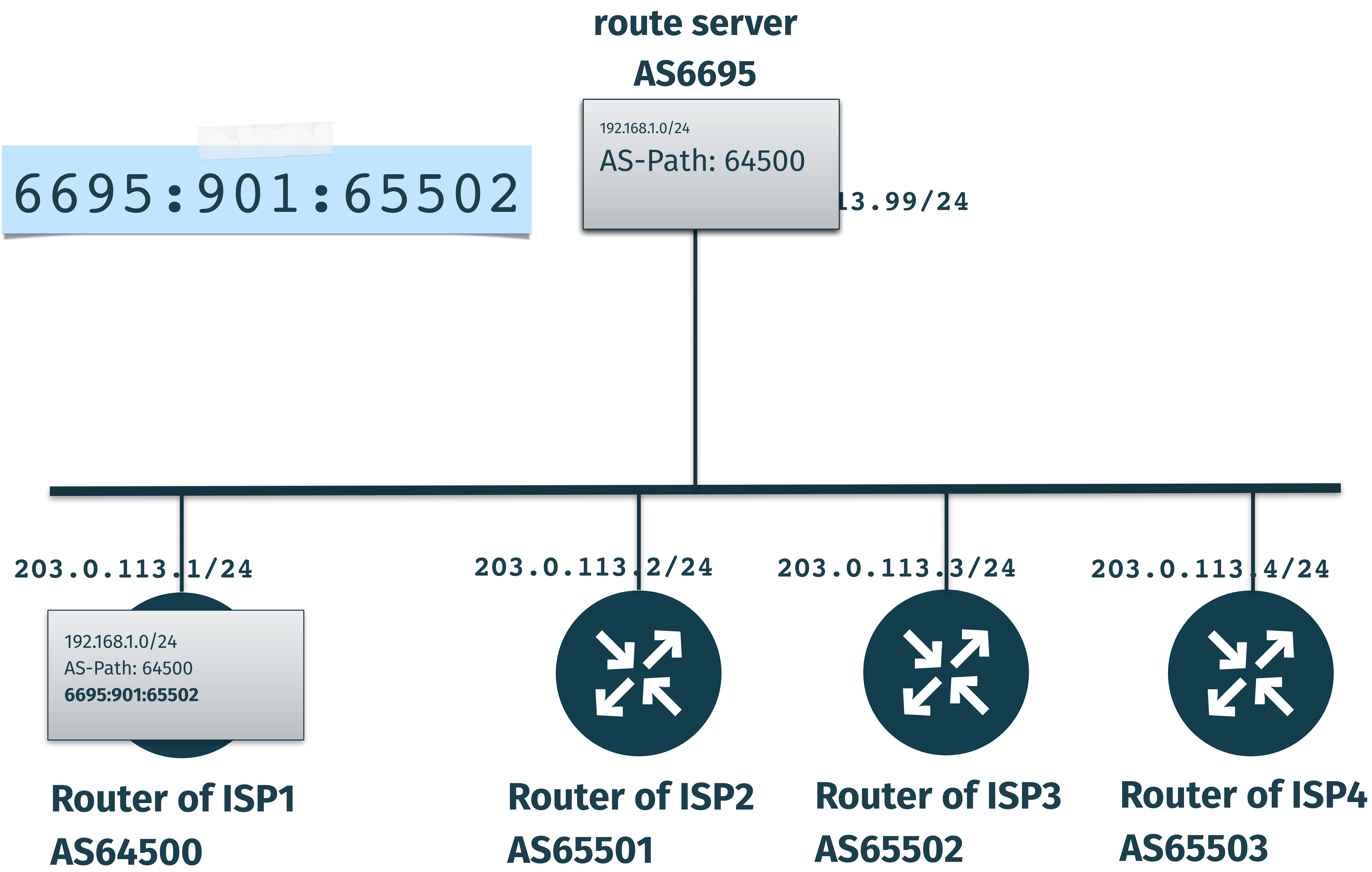
Prepend 1 time to AS65503

65001:65503

6695:101:65503



Add NO-EXPORT to AS65502



route server
AS6695

192.168.1.0/24
AS-Path: 64500

6695:901:65502

192.168.1.0/24
AS-Path: 64500
6695:901:65502

DE CIX

https://de-cix.net/academy

Links and further reading

DE-CIX Academy Resources

Lab and documentation

- DE-CIX Academy BGP Lab:
  https://gitlab.com/de-cix-public/team-academy/bgp/BGPLab
- Book: "BGP for networks who peer"
  https://github.com/wtremmel/BGP-for-networks-who-peer
- DE-CIX YouTube Channel: https://www.youtube.com/@DE-CIX

AS - Numbers

How to request an AS number

- Giving AS numbers to the RIRs: iana.org
- Requesting an AS number, links for:
  - ARIN
  - Lacnic
  - APNIC
  - RIPE NCC
  - Afrinic

BGP: Autonomous Systems

RFCs

- RFC1930: Guidelines for creation, selection, and registration of an Autonomous System (AS)

- RFC6793: BGP Support for Four-Octet Autonomous System (AS) Number Space

RFCs are Internet standards issued by the Internet Engineering Task Force (IETF)

Routing

Relevant RFCs

- RFC4632: Classless Inter-domain routing (CIDR)

RFCs are Internet standards issued by the Internet Engineering Task Force (IETF)

IPv6

Relevant RFCs

- RFC4291: IPv6 addressing architecture

RFCs are Internet standards issued by the Internet Engineering Task Force (IETF)

BGP - Best Path Selection

RFCs and Implementations

- RFC4271 - A Border Gateway Protocol 4 (BGP-4)
  - *Next Hop* is defined in Section 5.1.3
  - *AS Path* is defined in Section 5.1.2
  - *Local Preference*: Section 5.1.5
  - *Origin*: Section 5.1.1
  - *Multi Exit Discriminator (MED)*: Section 5.1.4
  - see 9.1 for the BGP best path selection algorithm
- BGP Best Path Selection by vendor
  - Cisco
  - Juniper
  - Mikrotik
  - Nokia
  - BIRD
  - FRRouting

1	NextHop reachable?	Continue if "yes"
2	Local Preference	higher wins
3	AS Path	shorter wins
4	**Origin Type**	**IGP over EGP over Incomplete**
5	**MED**	**lower wins**
6	**eBGP, iBGP**	**eBGP wins**
7	**Exit**	**nearest wins**
8	**Age of route**	**older wins**
9	**Router ID**	**lower wins**
10	**Neighbor IP**	**lower wins**

RFCs are Internet standards issued by the Internet Engineering Task Force (IETF)

BGP Attributes

Relevant RFCs

- BGP attribute types:
  - Registering new types: RFC2042
  - Published in BGP Parameters database at IANA

RFCs are Internet standards issued by the Internet Engineering Task Force (IETF)

BGP Security

Relevant RFCs

- RFC7454 - BGP Operations and Security
- Password protect BGP sessions
  - RFC2385 (obsolete) - Protection of BGP Sessions via the TCP MD5 Signature Option
  - RFC5925 - The TCP Authentication Option
- RFC5082 - The Generalized TTL Security Mechanism (GTSM)

RFCs are Internet standards issued by the Internet Engineering Task Force (IETF)

~~Relevant~~ RFCs

Historical (obsolete)

- RFC827: Exterior Gateway Architecture (EGP) (historical, obsolete)
-

RFCs are Internet standards issued by the Internet Engineering Task Force (IETF)