

BGP and Remote Peering

BGP advanced topics

Wolfgang Tremmel
wolfgang.tremmel@de-cix.net



Where networks meet

www.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

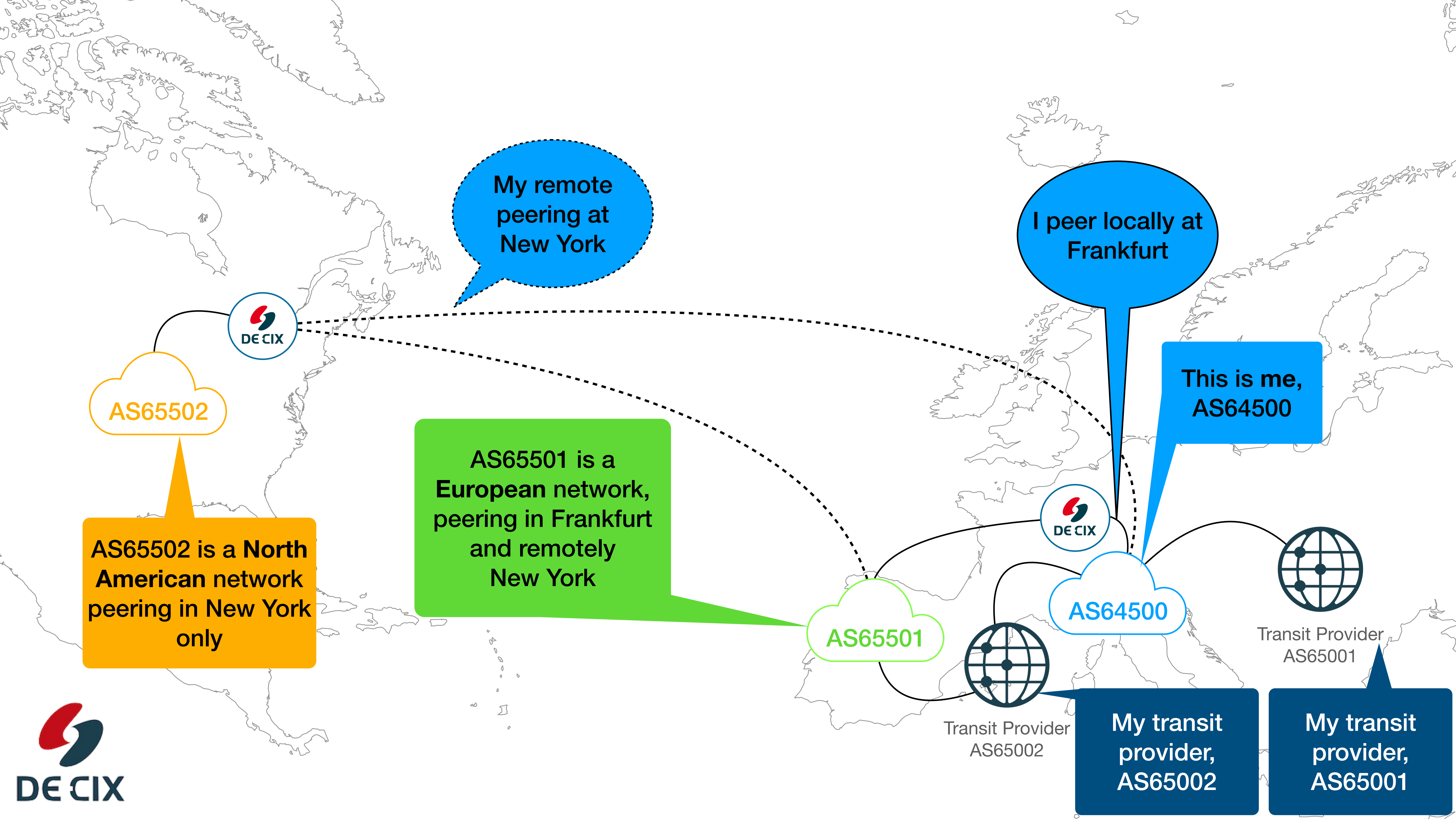


@wtremmel@hessen.social

Why remote peering?

DE-CIX offers remote peering at a number of sites

- You get peers you do not get at your home IXP
- You do not need your own connectivity to the remote IXP
 - DE-CIX delivers remote peering LANs simply via another VLAN ID
- Connectivity is monitored and maintained by DE-CIX
- You can book the bandwidth you need via the DE-CIX portal
- Let's have a look how remote peering can work for you...



My remote peering at New York

I peer locally at Frankfurt

This is me, AS64500

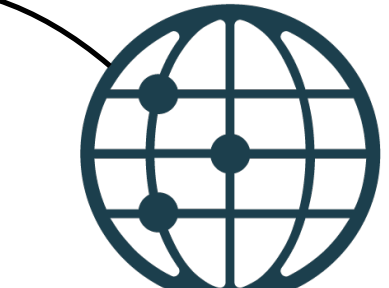
AS65502

AS65502 is a North American network peering in New York only

AS65501 is a European network, peering in Frankfurt and remotely New York

AS65501

AS64500



Transit Provider AS65001



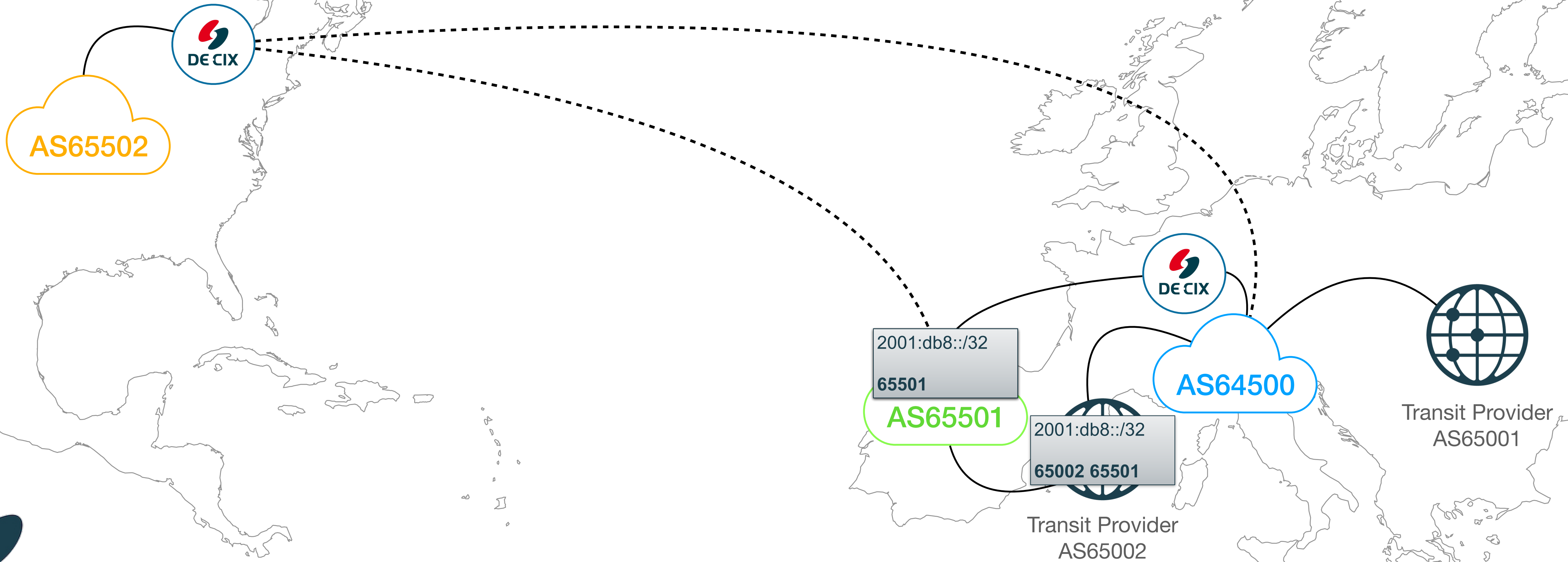
Transit Provider AS65002

My transit provider, AS65002

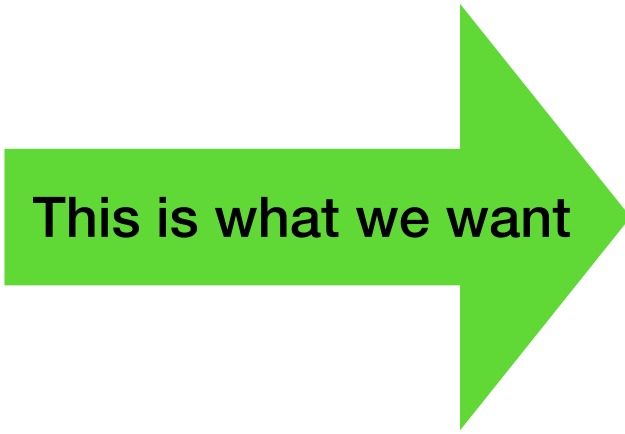
My transit provider, AS65001



BGP Announcements



BGP Announcements



Via Frankfurt Peering:

2001:db8::/32
65501

Best?

Via New York Peering:

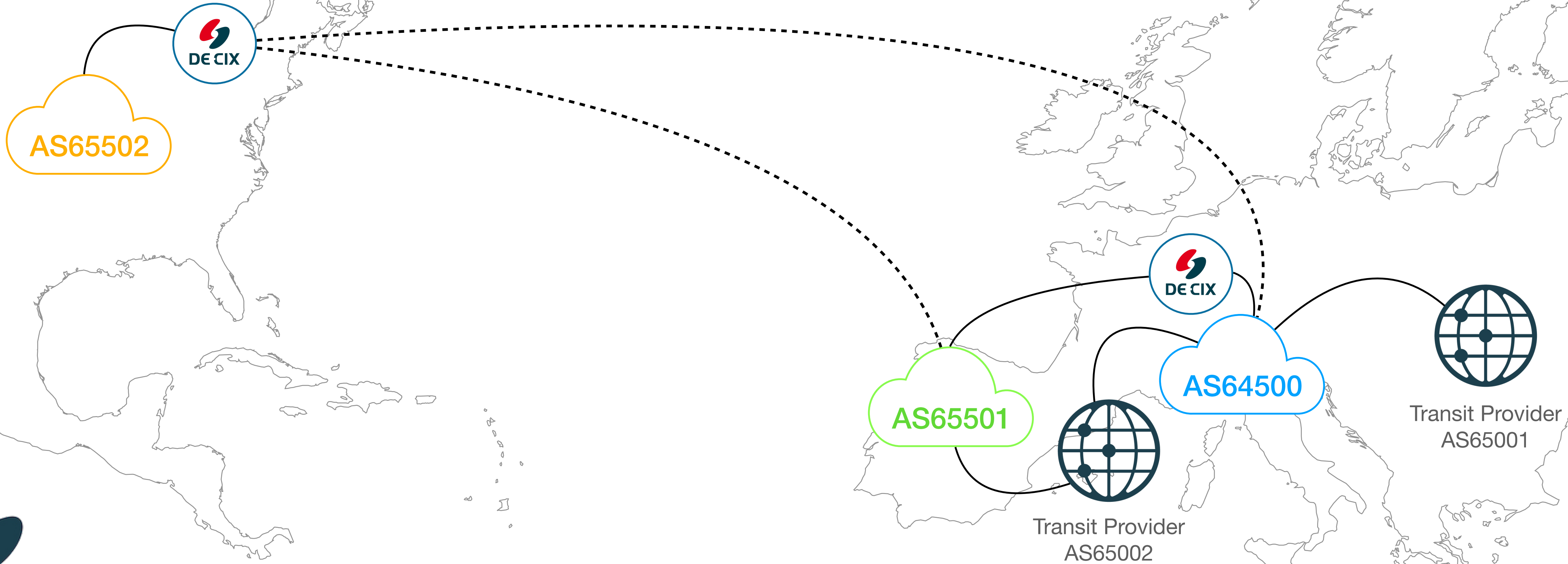
~~2001:db8::/32
65501~~

Best?

Via European Transit:

2001:db8::/32
65002 65501

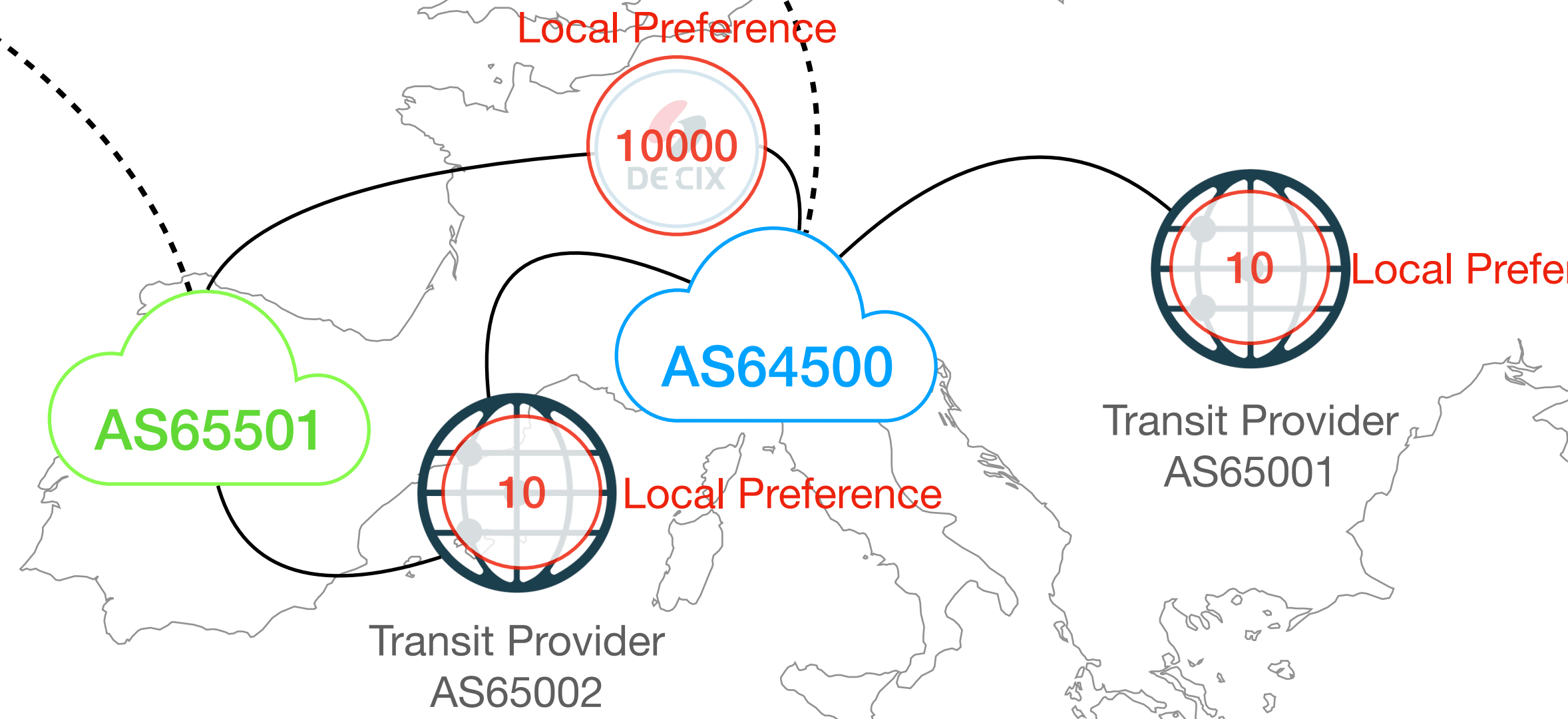
Longer AS Path



BGP Announcements

Local Preference

Via Frankfurt Peering:	2001:db8::/32 65501	10000	1
Via New York Peering:	2001:db8::/32 65501	5000	2
Via European Transit:	2001:db8::/32 65002 65501	10	3



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



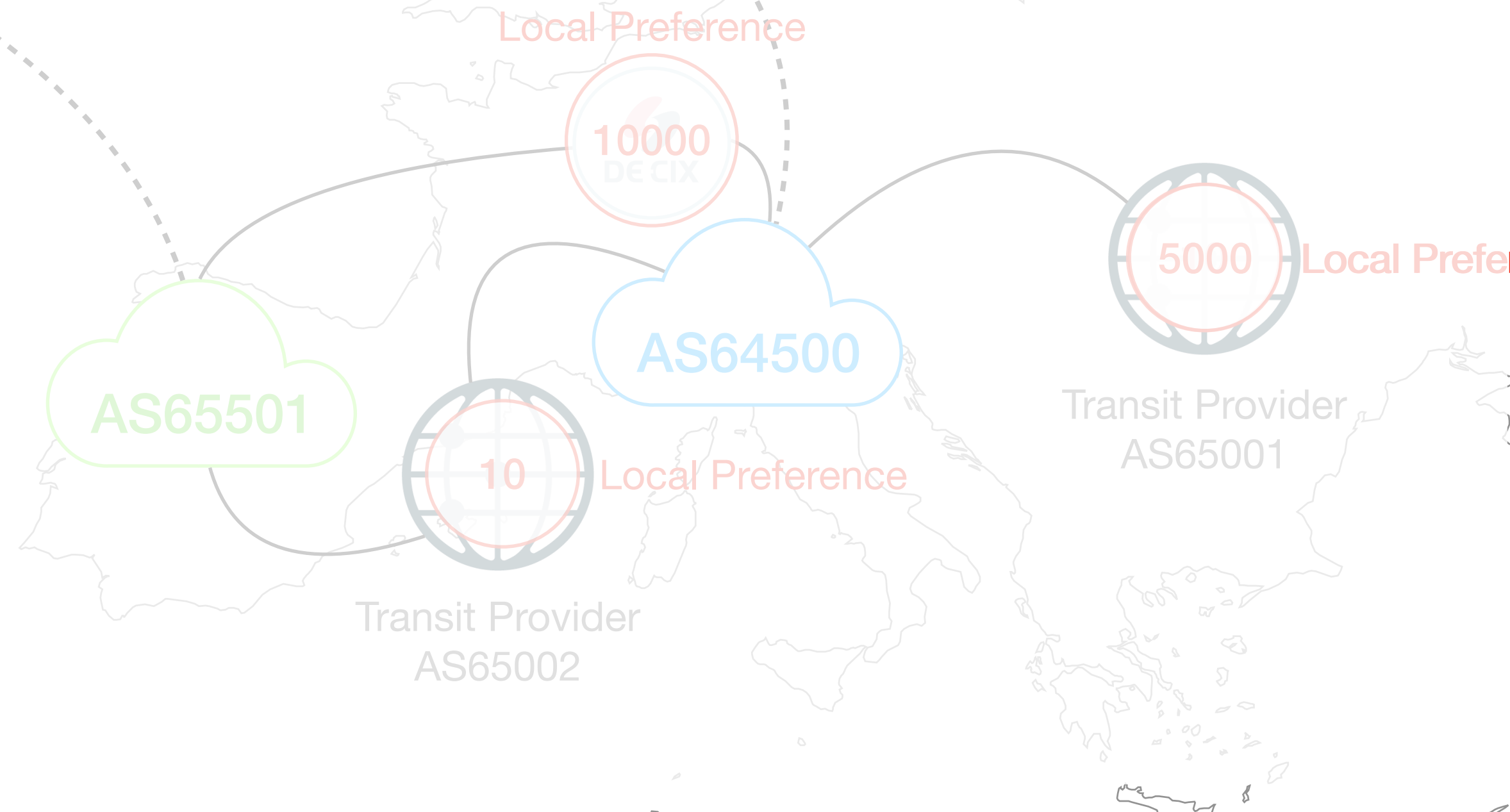
BGP Announcements

Path	ASes	Local Preference	AS Path Length
Via Frankfurt Peering:	2001:db8::/32 65501	10000	1
Via New York Peering:	2001:db8::/32 65501	10	never used
Via European Transit:	2001:db8::/32 65002 65501	5000	2



This does not work

Order	Criterion	Rule
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



Configuring for remote peering

Local Preference alone does not solve this

- Using only local preference for all remote peers does not work
- You want:
 - Prefer local peers
 - Do **not** send traffic to local peers via a remote IXP
 - Use your transit provider if your local peering goes away
- How to solve this in a scalable way?



Configuring for remote peering

Local Preference alone does not solve this

- Idea: Use MED and "bgp bestpath med always" (or similar)
 - Reminder: MED is by default only compared for prefixes with same next-hop-AS
 - Because MED (by default) is to signal your neighbor AS which path to prefer
 - If you enable MED always compare, you **must** set the MED on your side for **all** received prefixes and **all** peers.
 - Lets see how that would look like



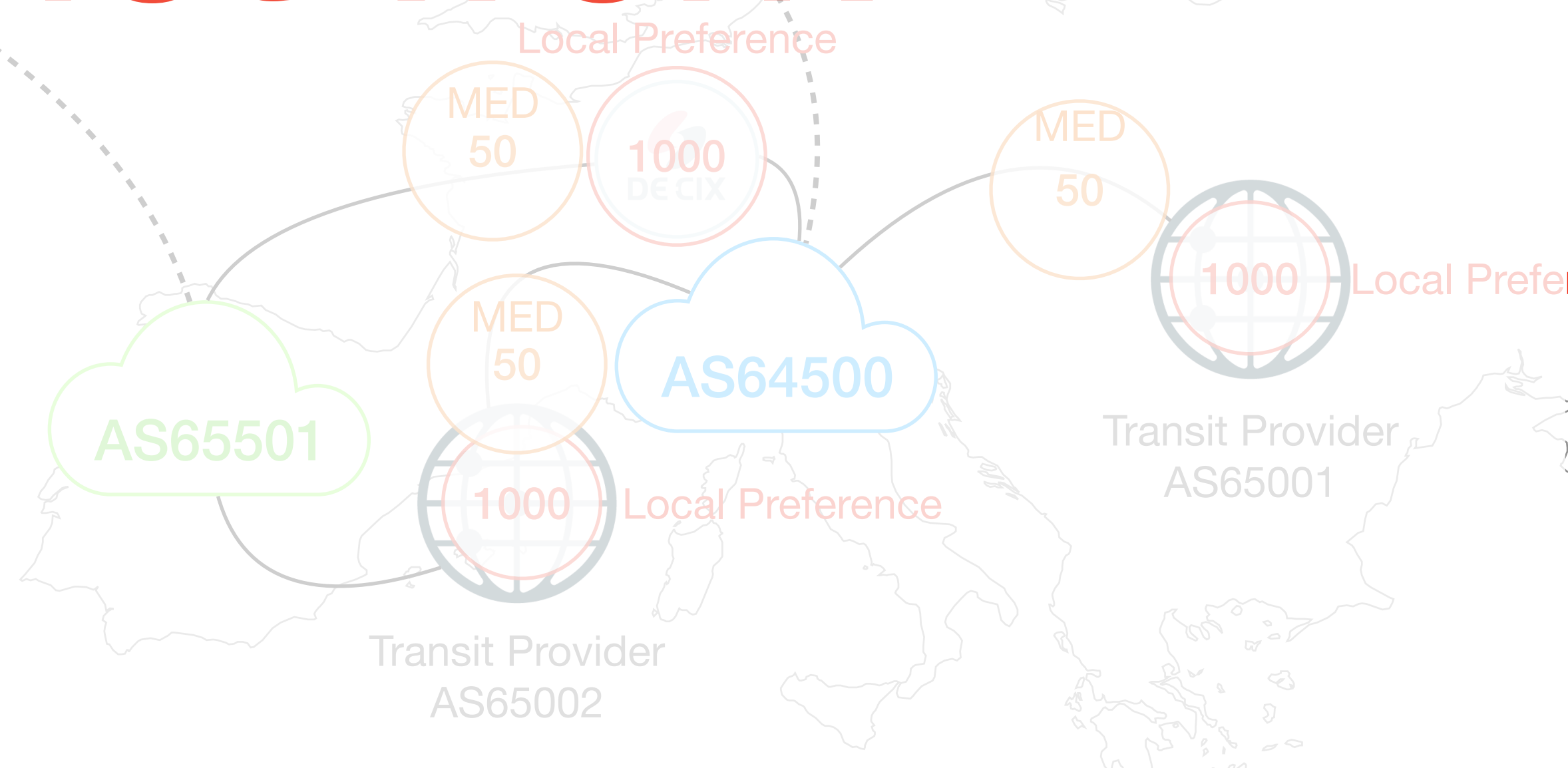


Local Preference MED

Via Frankfurt Peering:	2001:db8::/32 65501	1000	50	1
Via New York Peering:	2001:db8::/32 65501	1000	5000	2
Via European Transit:	2001:db8::/32 65002 65501	1000	50	3



This does not work



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



What now?

Read the fine documentation

- What we want:
 - Receive/Accept no prefixes from remote peers at a remote IXP
 - Do not announce prefixes to remote peers at a remote IXP
 - At a remote IXP we only want peers **local to that IXP**
- Check the [route server guide](#) on the DE-CIX website!
 - BGP Communities are used to tag prefixes
 - You can filter for continent, country, metro area, edge device
- Be aware of the limitations - **this only works if the peer is connected via remote peering**
 - **NOT** if the peer uses its own circuit to reach the remote IXP



Filter received prefixes

Using BGP Communities

- Example: At DE-CIX New York you only want prefixes from the US
 - Use BGP Community `63034:1913:840`
 - "840" is the country code for USA according to [M49 standard](#)
- Example: At DE-CIX Frankfurt you only want local "Frankfurt" prefixes:
 - Use BGP Community `6695:1912:0`
 - "0" is our own numbering - check our [website](#) for all

AS-Number of
Route Server
New York

Country
based origin

USA

AS-Number of
Route Server
Frankfurt

Metro-Area based
origin

Frankfurt

**This is the solution for sending
traffic**

What about receiving traffic?

Receiving Traffic

(Not) announcing your prefixes to remote peers

- Peers only send you traffic if they receive your prefixes
- So you must instruct the route server not to **announce** your prefixes to remote peers
- This can be done by [BGP action communities](#)
 - There are many options
 - Useful for our case: "Shortcut" community *65200:65212*
 - = "Redistribute to local peers only"



Summary

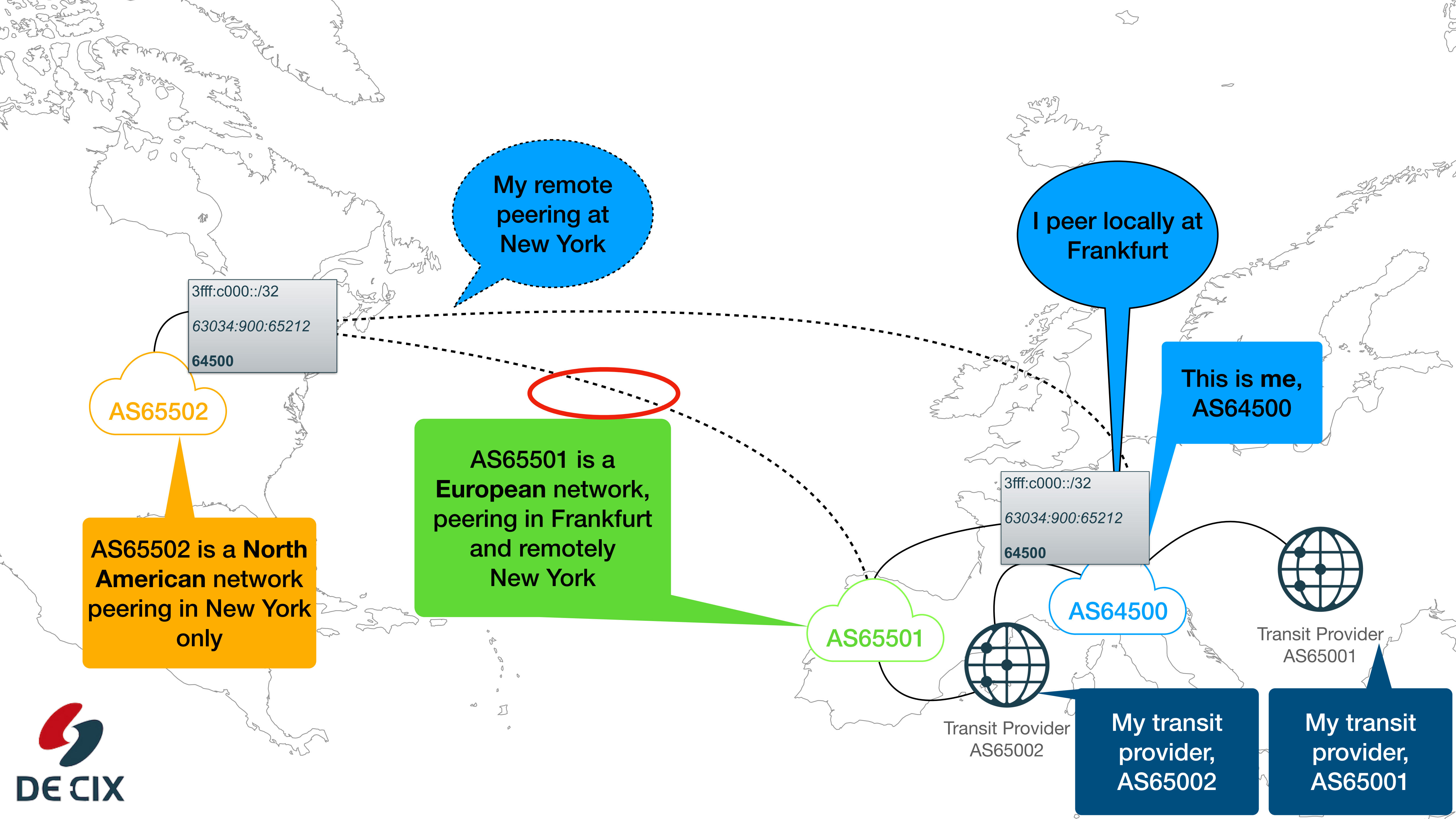
How to configure remote peering (for our New York example)

- Use the DE-CIX Route Servers
- Announce your prefixes with BGP communities tagged *62500:65212* or *63034:900:65212* (= announce only locally)
- Only accept New York local prefixes by filtering for *65102:2000* or *63034:1912:2* (= New York prefixes)
- Lets have a look at the map.....

or use
63034:1913:840 for
"US prefixes only"

or use *63034:1914:19* for
"American (continent)
prefixes"





My remote peering at New York

I peer locally at Frankfurt

This is me, AS64500

AS65502

AS65502 is a North American network peering in New York only

AS65501 is a European network, peering in Frankfurt and remotely New York

AS65501

AS64500

Transit Provider AS65002

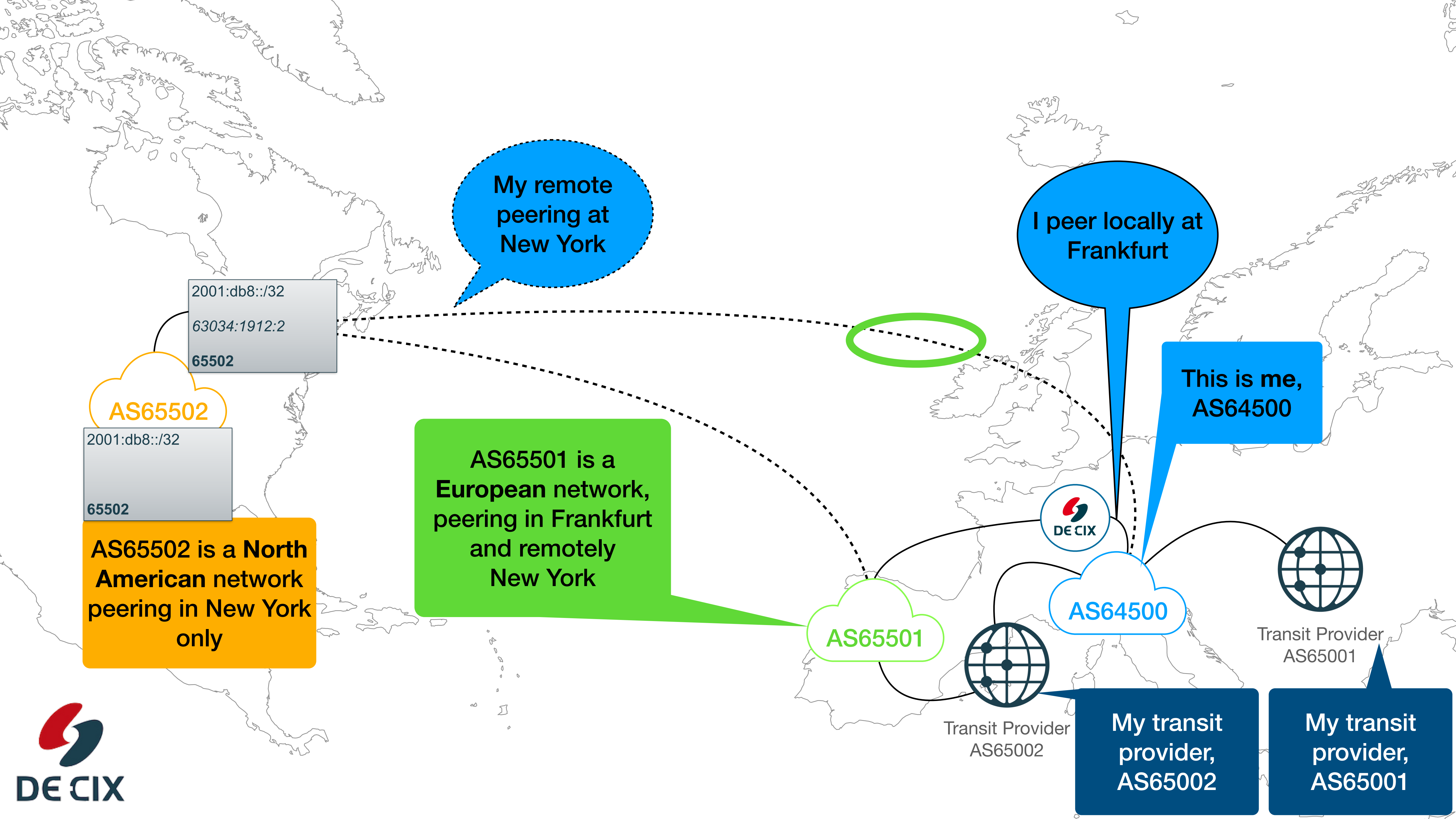
My transit provider, AS65002

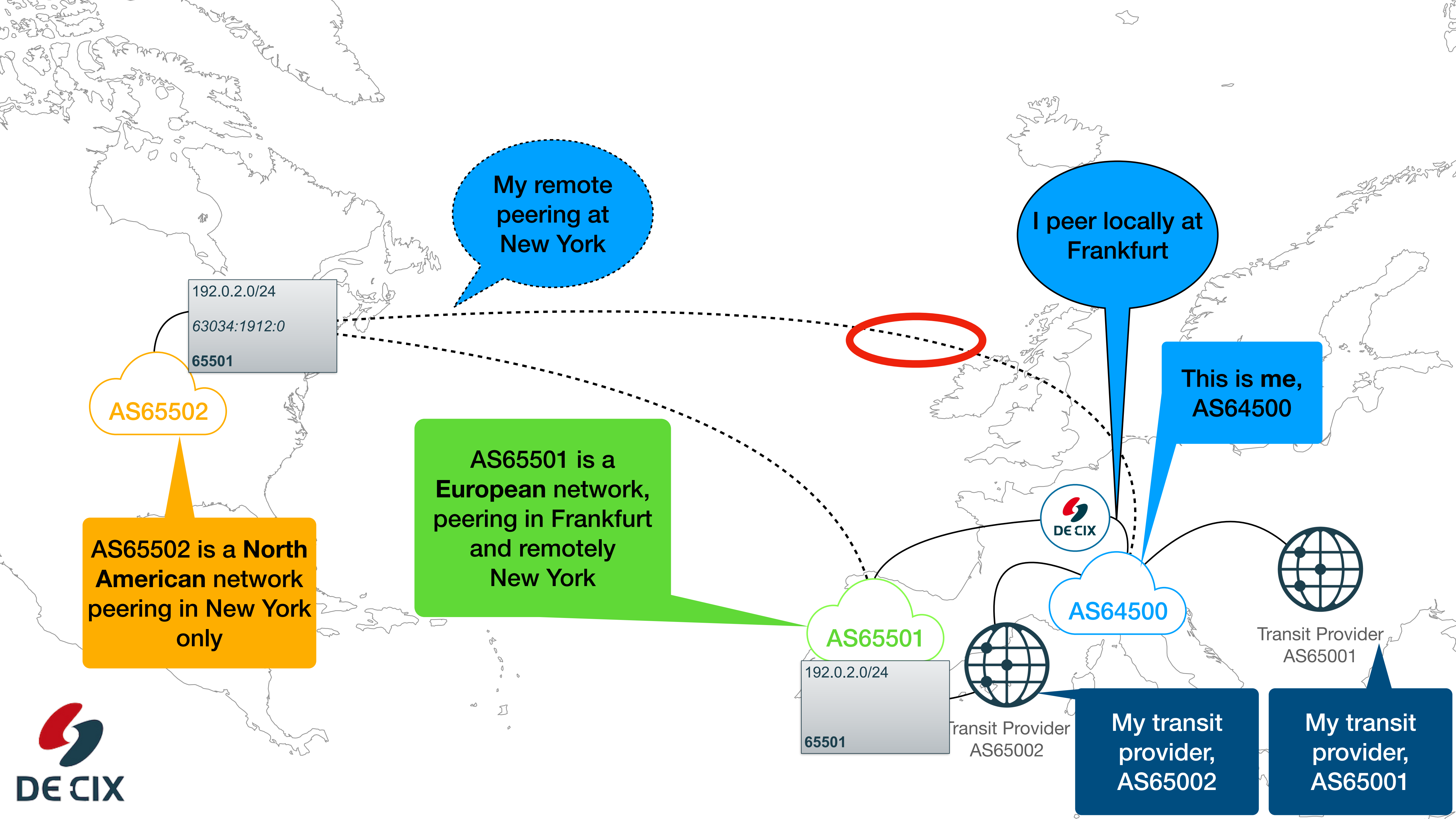
Transit Provider AS65001

My transit provider, AS65001

3fff:c000::/32
63034:900:65212
64500

3fff:c000::/32
63034:900:65212
64500





Is that all?

There is a bit more....

- What about direct peering?
 - Recommendation: ask your peering partners where they are located
 - Peer with them only at their (or your) home IXP
- What about "big ones" who are local at every IXP?
 - Recommendation: Peer with them only at your home IXP
 - And use your transit provider as backup.

Thank you!

academy@de-cix.net

Interested in more webinars? Please subscribe to our mailing list at <https://lists.de-cix.net/wws/subscribe/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>
 - "[Networking Basics](#)" Playlist

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

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



DE CIX

<https://de-cix.net/academy>

Interested in more webinars? Please subscribe to our mailing list at <https://lists.de-cix.net/www/subscribe/academy>



DE CIX

DE-CIX Management GmbH | Lindleystr. 12 | 60314 Frankfurt | Germany
Phone + 49 69 1730 902 0 | sales@de-cix.net | www.de-cix.net