BGP Best Path Selection

BGP for networks who peer: Part 5

Wolfgang Tremmel
wolfgang.tremmel@de-cix.net

BGP (new) Webinars Overview

→01 - Prefixes and AS numbers

→02 - BGP Introduction

→03a - Setting up iBGP

→03b - Setting up eBGP

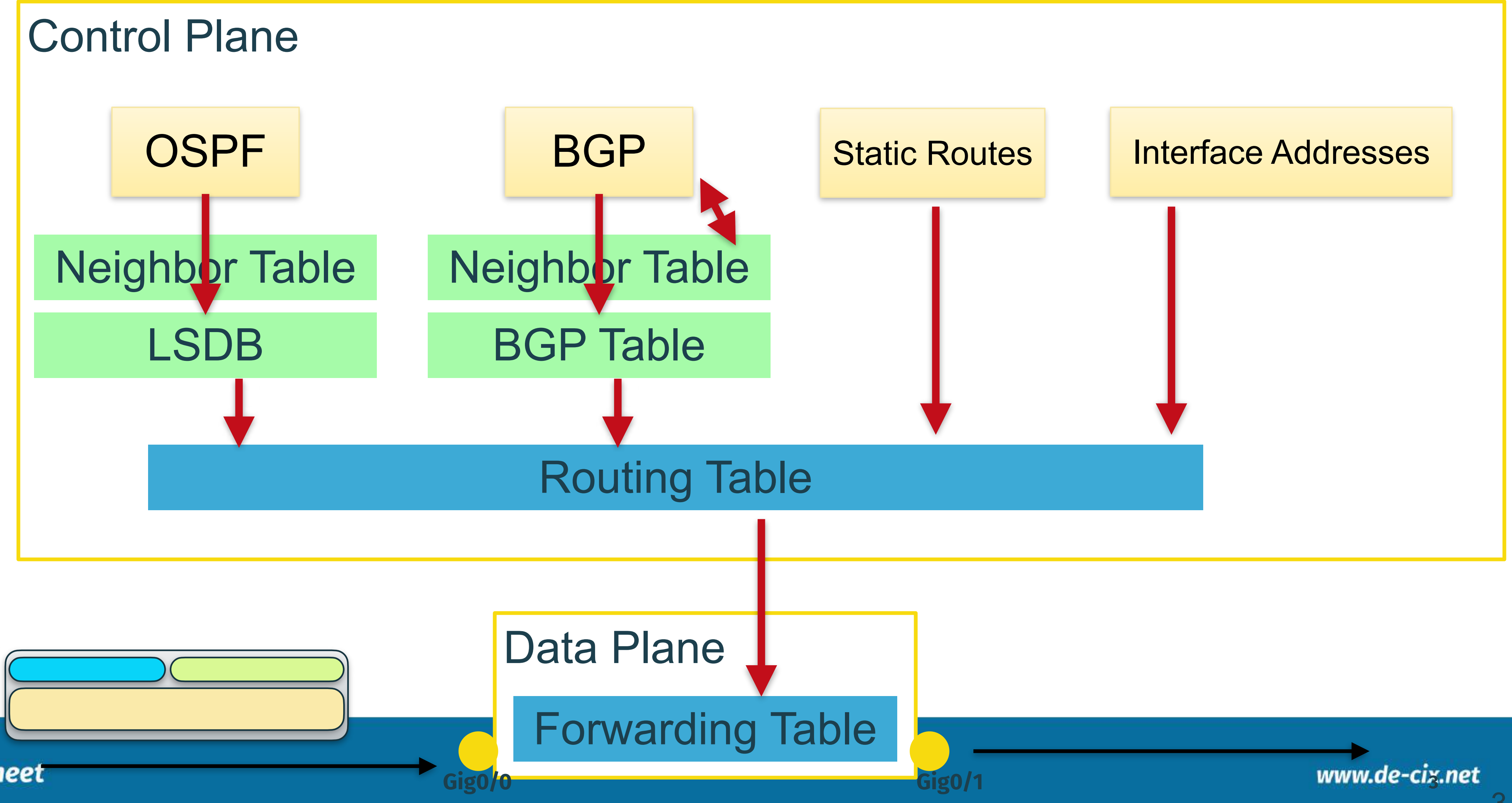
→04 - Becoming multi-homed

→05 - BGP Best Path Selection

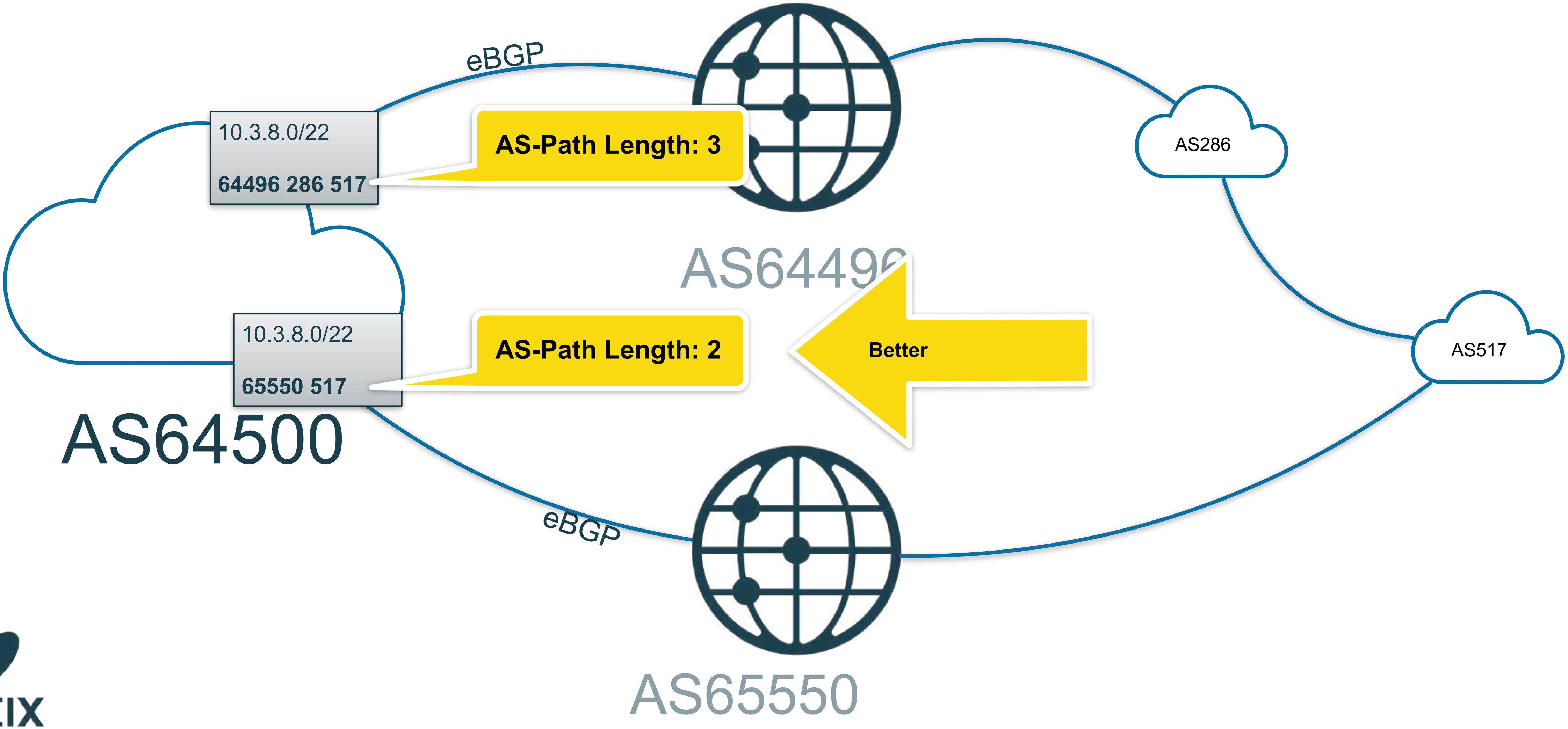
→06 - BGP Communities

DE CIX

How a router works



In part 4 we talked about path selection



10.3.8.0/22

64496 286 517

AS-Path Length: 3

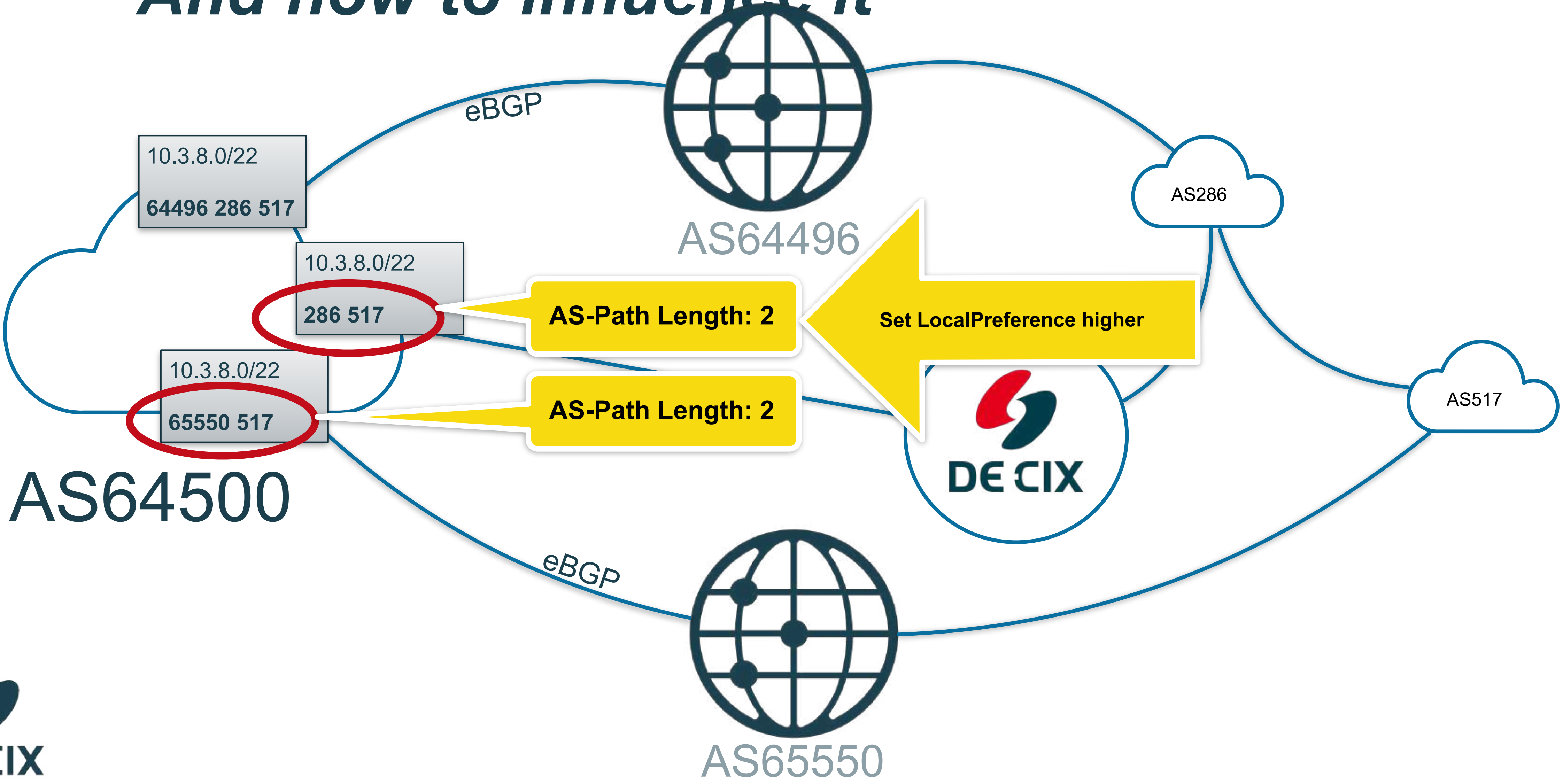
10.3.8.0/22

65550 517

AS-Path Length: 2

Better

And how to influence it

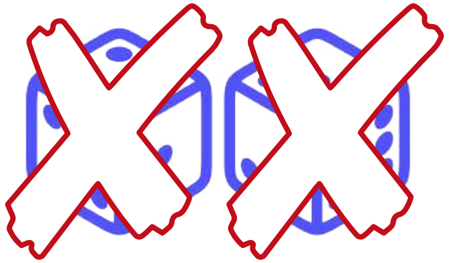


BGP Route Selection Algorithm

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

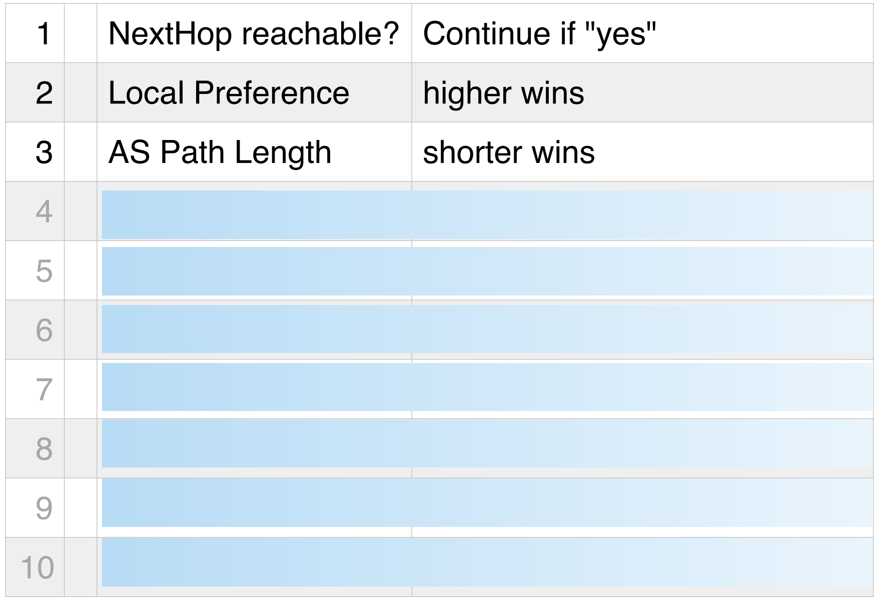
BGP Route Selection Algorithm: Motivation

- → 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- → Some are really outdated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 we will focus on the most important ones
- → But all will be covered.



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
  - → EGP - route is received from EGP
  - → incomplete - redistributed from another protocol
- → ***This rule is not really important***

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			

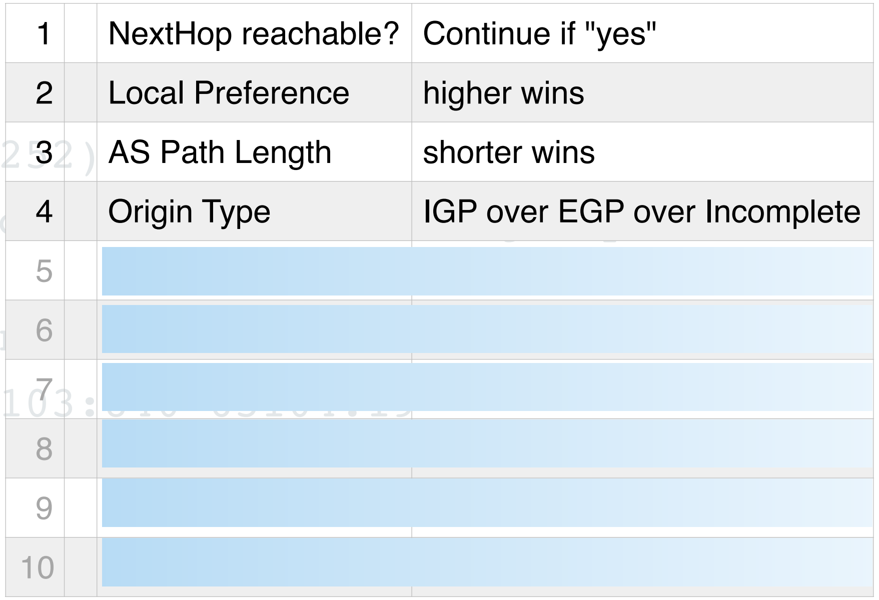
BGP Route Selection: Origin Type Examples

```
show ip bgp

Origin codes: i - IGP, e - EGP, ? - incomplete
* i1.0.4.0/22    206.130.10.8    634    200      0 6939 i
* i1.0.137.0/24   80.81.194.12 5000    200      0 9318 23969 ?
```

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: Origin Type Examples

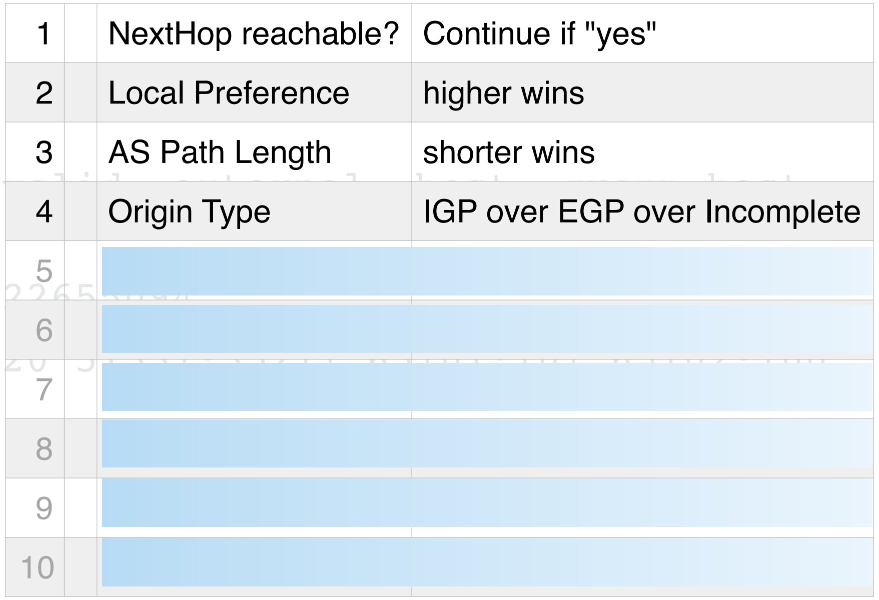
```
show ip bgp 1.0.4.0/22

Path #22: Received by speaker 0
  Advertised to update-groups (with more than one peer):
    0.10 0.11
  Advertised to peers (in unique update groups):
    46.31.120.208
  6939 4826 38803 56203
    206.130.10.8 from 206.130.10.252 (206.130.10.252)
      Origin IGP, metric 634, localpref 200, vali
import-candidate, import suspect
      Received Path ID 0, Local Path ID 1, versio
      Community: 51531:35214 65101:0 65102:200 65103:
      Origin-AS validity: not-found
```



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			

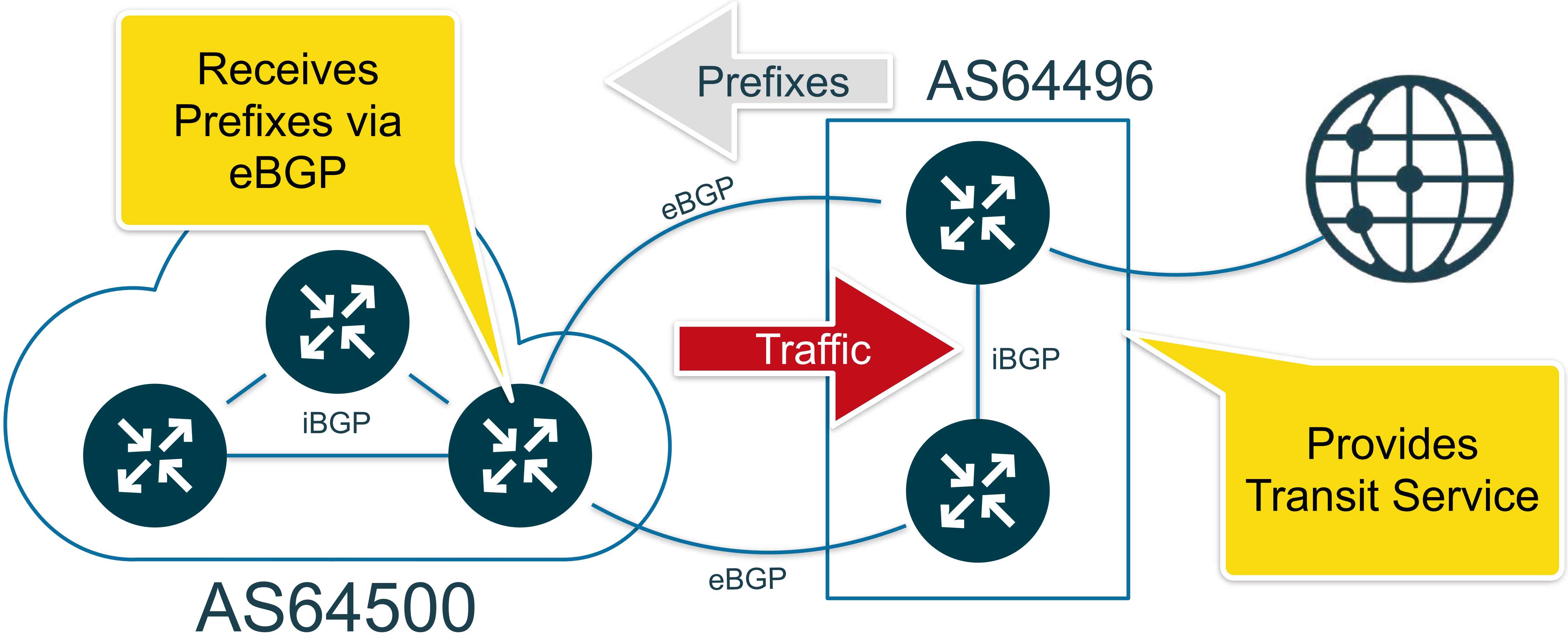
BGP Route Selection: Origin Type Examples

```
show ip bgp 1.0.137.0/24

Path #6: Received by speaker 0
  Advertised to update-groups (with more than one peer):
    0.10 0.11
  Advertised to peers (in unique update groups):
    46.31.120.208
  9318 38040 23969
    80.81.194.12 from 80.81.192.157 (80.81.192.157)
      Origin incomplete, metric 5000, localpref 200,
import-candidate, import suspect
      Received Path ID 0, Local Path ID 1, version 332265
      Community: 9318:120 9318:8300 9318:8330 9318:9020
65103:276 65104:150
      Origin-AS validity: not-found
```



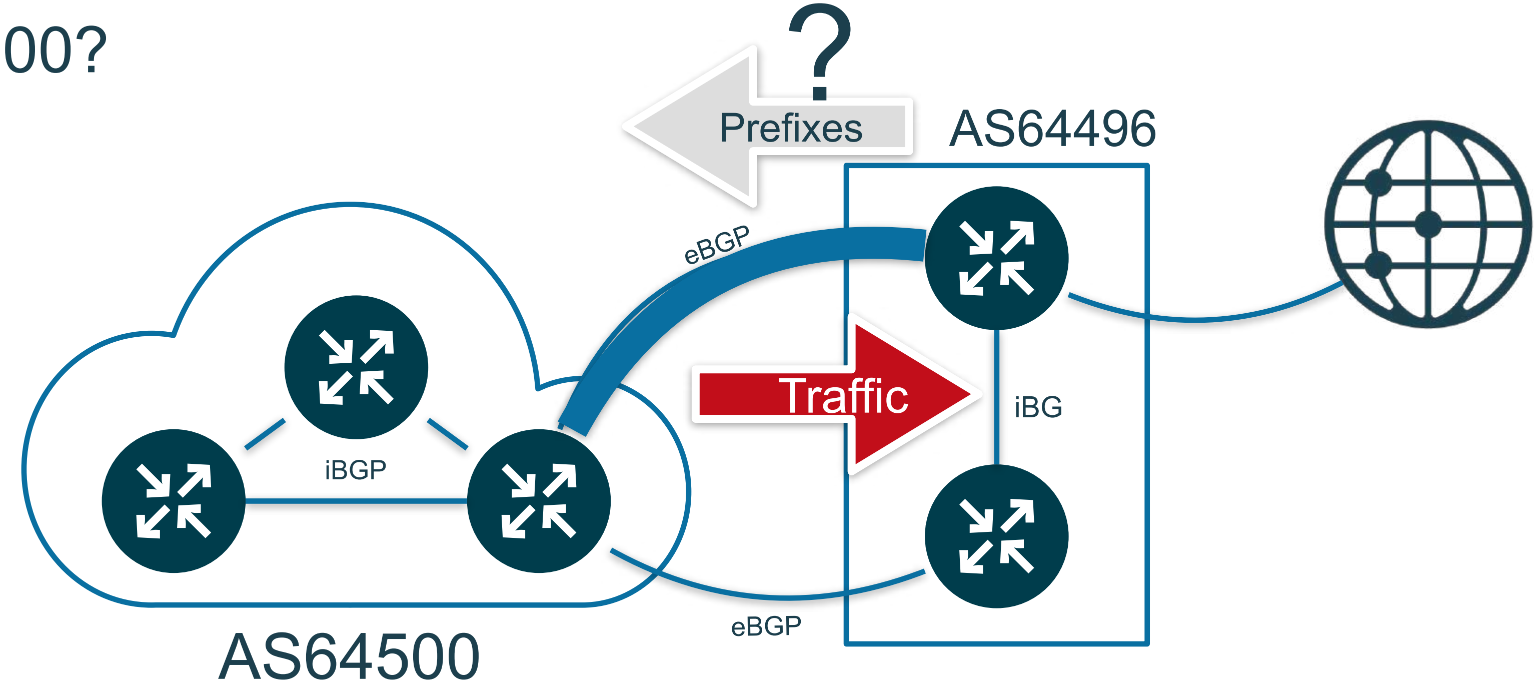
1		NextHop reachable?	Continue if "yes"
2		Local Preference	higher wins
3		AS Path Length	shorter wins
4		Origin Type	IGP over EGP over Incomplete
5			
6			
7			
8			
9			
10			

Consider the following network



Consider the following network

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

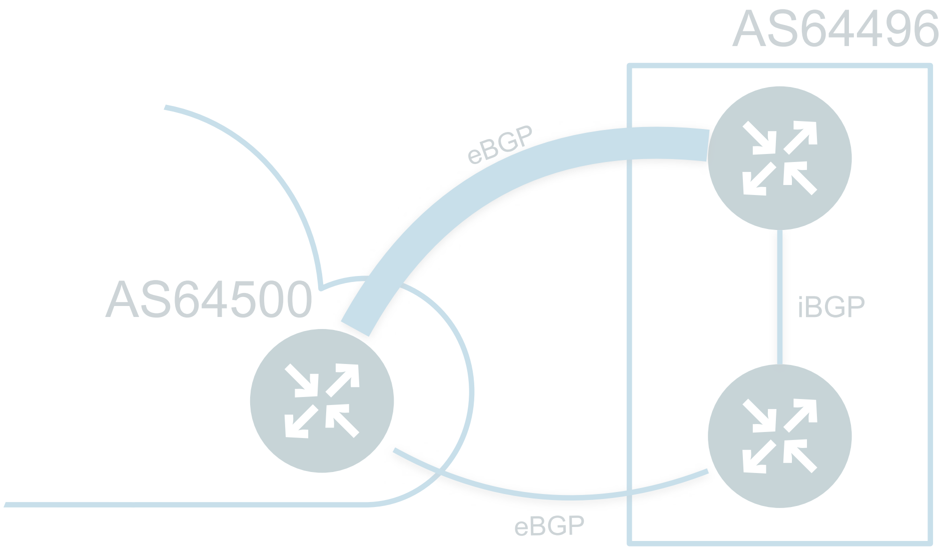


BGP Route Selection Algorithm:

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

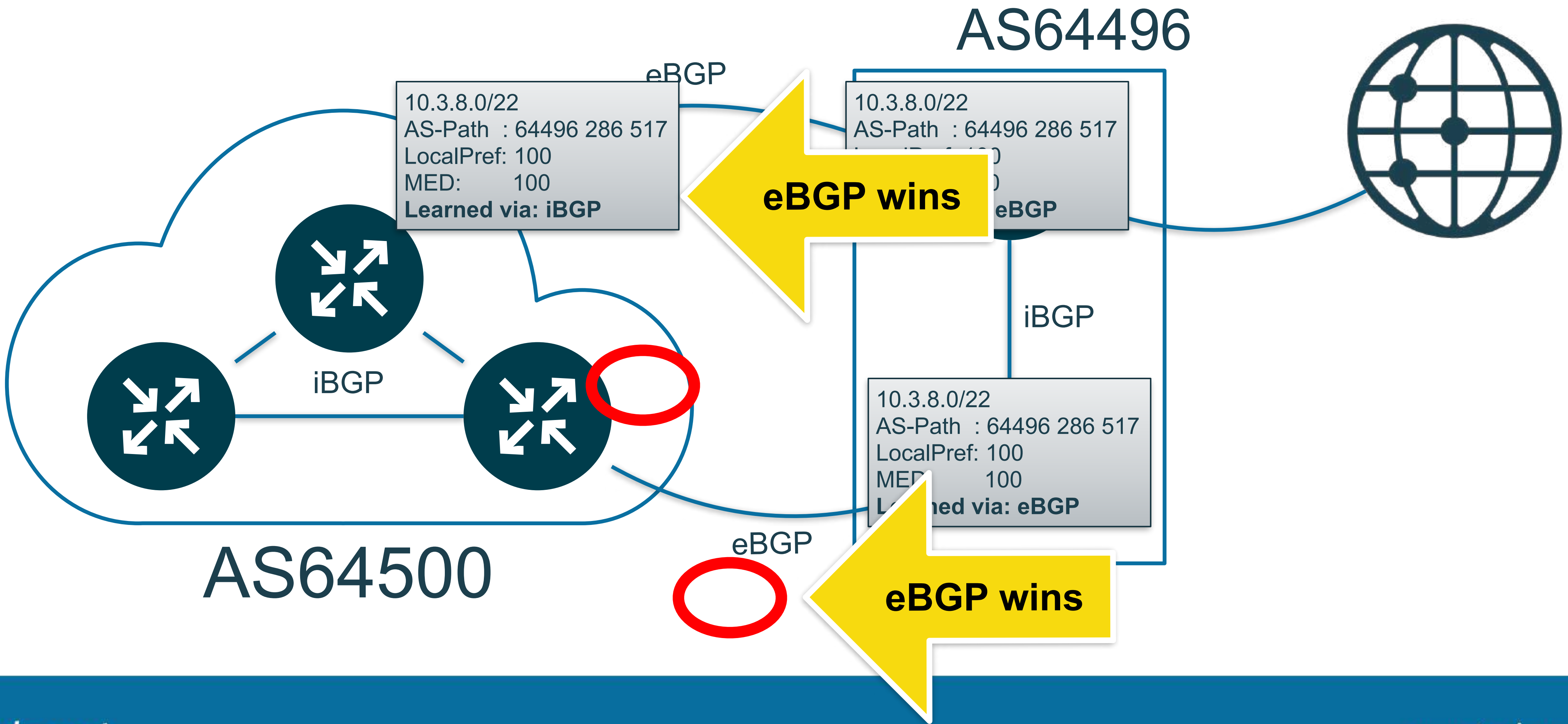


- → MED = **M**ulti-**E**xit **D**iscriminator
- → Only compared if next-hop AS is the same
- → 32bit value (0..4294967294)
- → Lower wins
- → Optional (does not have to be there)
- → A missing MED can be treated as "best" (=0, default) or "worst" (=4294967294)
- → Option "always-compare-med" **not recommended**!
- → 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

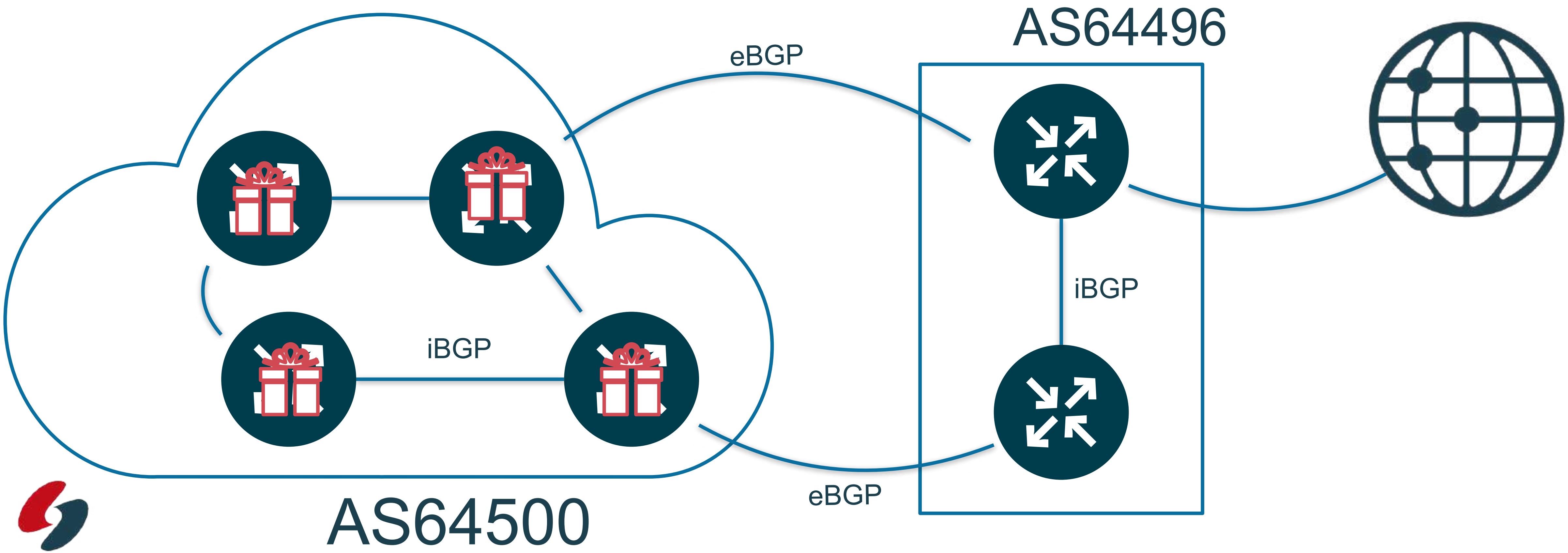


10.3.8.0/22
AS-Path : 64496 286 517
LocalPref: 100
MED: 100
Learned via: iBGP

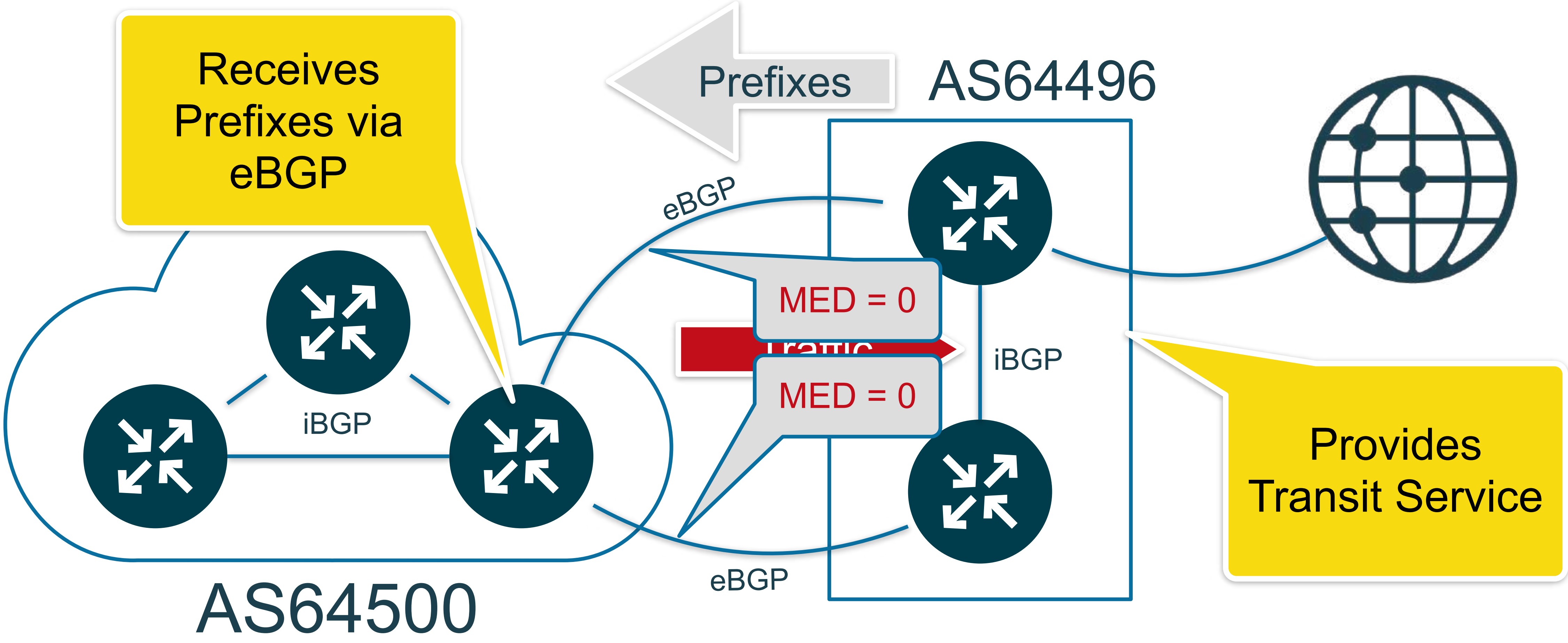
10.3.8.0/22
AS-Path : 64496 286 517
eBGP

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

BGP Route Selection : nearest exit wins



Let's go back to our sample network



Receives Prefixes via eBGP

Provides Transit Service

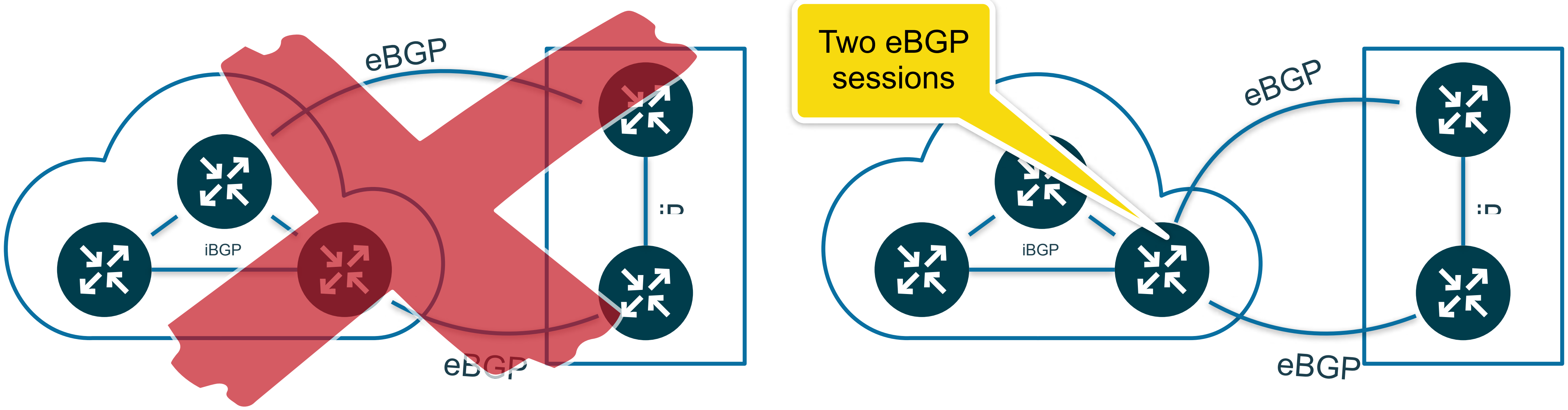
BGP Route Selection : Age / Stability

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		

DE CIX

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

Experiment: best path selection

Experiment 3

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

Links and further reading

- Definition of terms (all from RFC4271):
  - *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
- Best Path Selection process: Section 9.1
- BGP Route Selection Algorithm by vendor:
  - Cisco
  - Juniper
  - Mikrotik
  - Nokia
  - BIRD
  - Quagga

BGP Best Path Selection Algorithm

Bold items were covered in this webinar

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**

BGP Best Path Selection Algorithm

Local Preference is...

- → a 32bit integer value (0-4294967295)
- → Propagated via iBGP inside an Autonomous System
- → Usually set using rules when receiving prefixes
  - According to your routing policy
- → Typical values
  - 10000 (high value) for customer prefixes
  - 1000 (medium value) for prefixes received via peering
  - 100 (low value) for prefixes received via upstream
- → Rules to adjust local preference can be as complex as your router software allows it to be.

AS Path is...

- → an ordered list of AS numbers...
- → ...with the originator AS at the rightmost side
- → automatically built when prefixes are sent via eBGP
- → length of the path is used for selection (shorter wins)

BGP Best Path Selection Algorithm

Origin Type is...

→ a historic, but mandatory attribute

→ set by originator AS and forwarded unchanged

→ can have the values (in order of preference):

- IGP - prefix was originated via a network statement
- EGP - prefix was learned from Exterior Gateway Protocol (RFC904, historic)
- incomplete - prefix was learned by another protocol

Multi Exit Discriminator (MED) is...

→ a 32Bit value, lower wins

→ optional, if it is not there it's either treated as zero (best) or as 2^32-1 (worst)

→ non-transitive (set by an eBGP speaker and only sent to the next-hop AS)

→ usually set using rules when sending prefixes (according to the sender's routing policy)

→ only compared between eBGP speakers if next-hop AS is the same

Router ID is...

→ also called **BGP Identifier**

→ a 4 byte, unsigned integer (mostly it's the IPv4 loopback address of a router)

→ unique within one AS

→ set at startup and stays unchanged

→ the same for all BGP sessions

Neighbor IP is...

→ the last tie-breaker in the BGP Best Path Selection

→ the IP address of the eBGP speaker a prefix was learned from

DE CIX

https://de-cix.net/academy

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