md brasil

TI&TELECOM

BGP Filtering with RouterOS

External Connectivity Strategies for Multi- Homed ISP’s, connected to an IXP Environment and providing transit services

European MUM – 2013 - Zagreb / Croatia

Wardner Maia

Copyright Notice
(Aviso sobre direitos autorais)

English:

This material is an effort intended to improve the level of knowledge of professionals that work with Mikrotik RouterOS and should be used solely for self-study purposes.

Digital copies and/or any printed material contained in this presentation or derived from it are property of MD Brasil TI & Telecom and cannot be used for any kind of training, presentation or workshop, even non-commercial ones.

Reproduction of any part or picture requires previous written authorization of MD Brasil. For information about how to obtain such authorization, please contact mdbrasil@mdbrasil.com.br.

Portuguese:

Este material é um esforço que visa aprimorar o grau de conhecimento de profissionais que trabalham com Mikrotik RouterOS e deve ser usado apenas com objetivos de auto estudo.

Cópias digitais e/ou materiais impressos com conteúdo desta apresentação ou dela derivados são de propriedade a MD Brasil TI & Telecom a não podem ser usados para qualquer tipo de treinamento, apresentação ou seminário, mesmo os de finalidades não comerciais.

A reprodução de qualquer parte ou figura requer prévia autorização por escrito da MD Brasil. Para informações em como obter esta autorização, por favor contate mdbrasil@mdbrasil.com.br.

Introduction

Wardner Maia

Electronic and Telecommunications Engineer;

Internet Service Provider since 1995;

Radio Frequency Trainings since 2002;

Certified Mikrotik Trainer since 2007;

MD Brasil IT & Telecom CTO

Member of the board of directors of LACNIC

Introduction

MD Brasil IT & Telecom

Internet Access Provider in São Paulo state - Brazil;

Telecom equipment manufacturer and integrator;

Mikrotik Training Partner since 2007;

Mikrotik distributor;

Consulting services worldwide;

http://www.mdbrasil.com.br http://mikrotikbrasil.com.br

Objectives and Target Audience

Objectives:

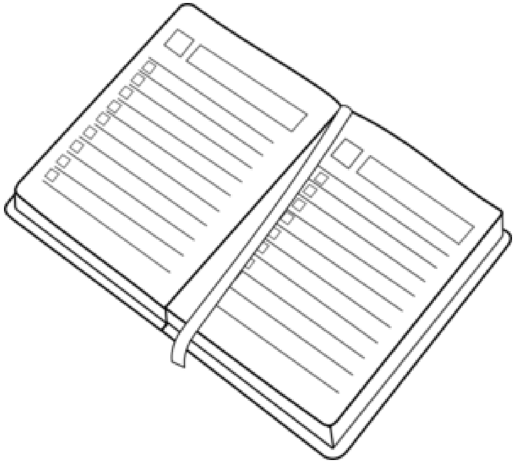
To understand BGP filtering techniques to be applied to a multi connected network and intended to implement external routing policies, providing traffic balance, security and reliability.

Target Audience:

ISP’s and Telecom operators running or intending to run BGP with Mikrotik RouterOS.

Agenda

1) BGP essentials and basics of BGP filtering;



2) Case Studies:

2.1) Overview

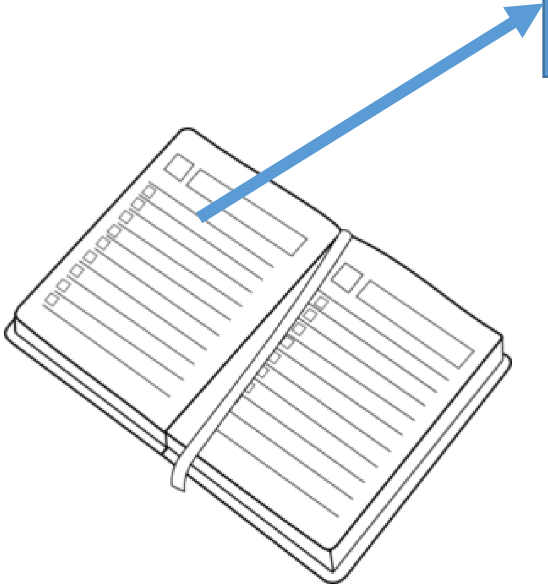
2.2) Single-Homed Provider

2.3) Single-Homed + IXP

2.4) Multi-Homed + IXP

2.5) Multi-Homed + IXP + Providing transit services

Agenda



1) BGP essentials and basics of BGP filtering;

2) Case Studies:

  2.1) Overview

  2.2) Single-Homed Provider

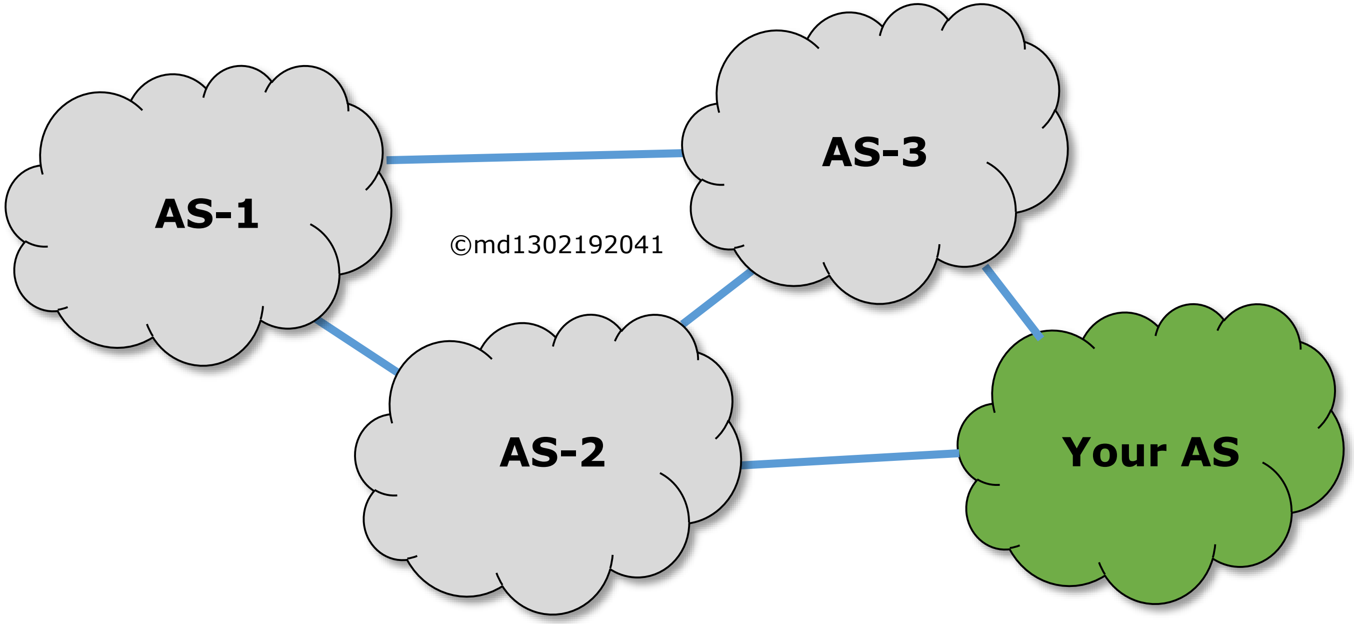
  2.3) Single-Homed + IXP

  2.4) Multi-Homed + IXP

  2.5) Multi-Homed + IXP + Providing transit services

Autonomous Systems and the Internet

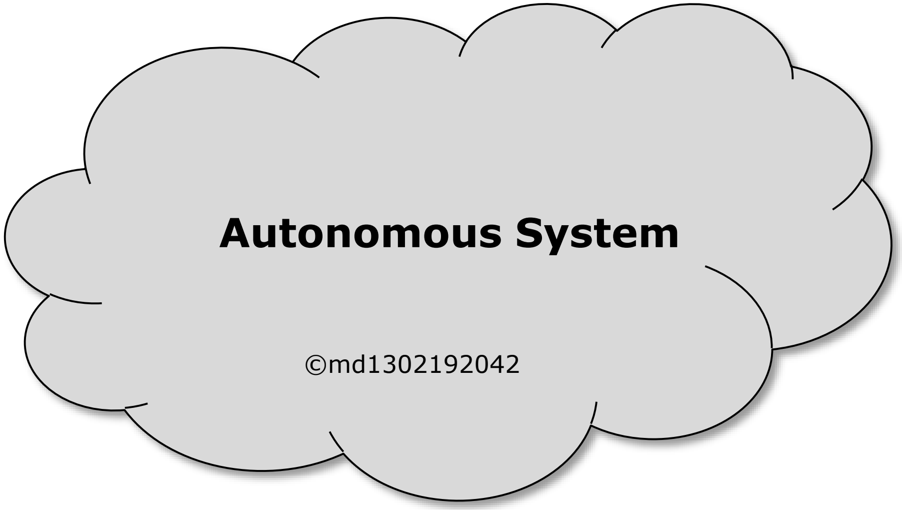
The Internet is composed of lots of interconnected networks, each one under an independent technical administration. Such networks are called an “Autonomous System”.



Autonomous System

One definition for an AS can be:

> “An Autonomous System (AS) is a group of IP networks run by one or more network operators with a single, clearly defined routing policy.”



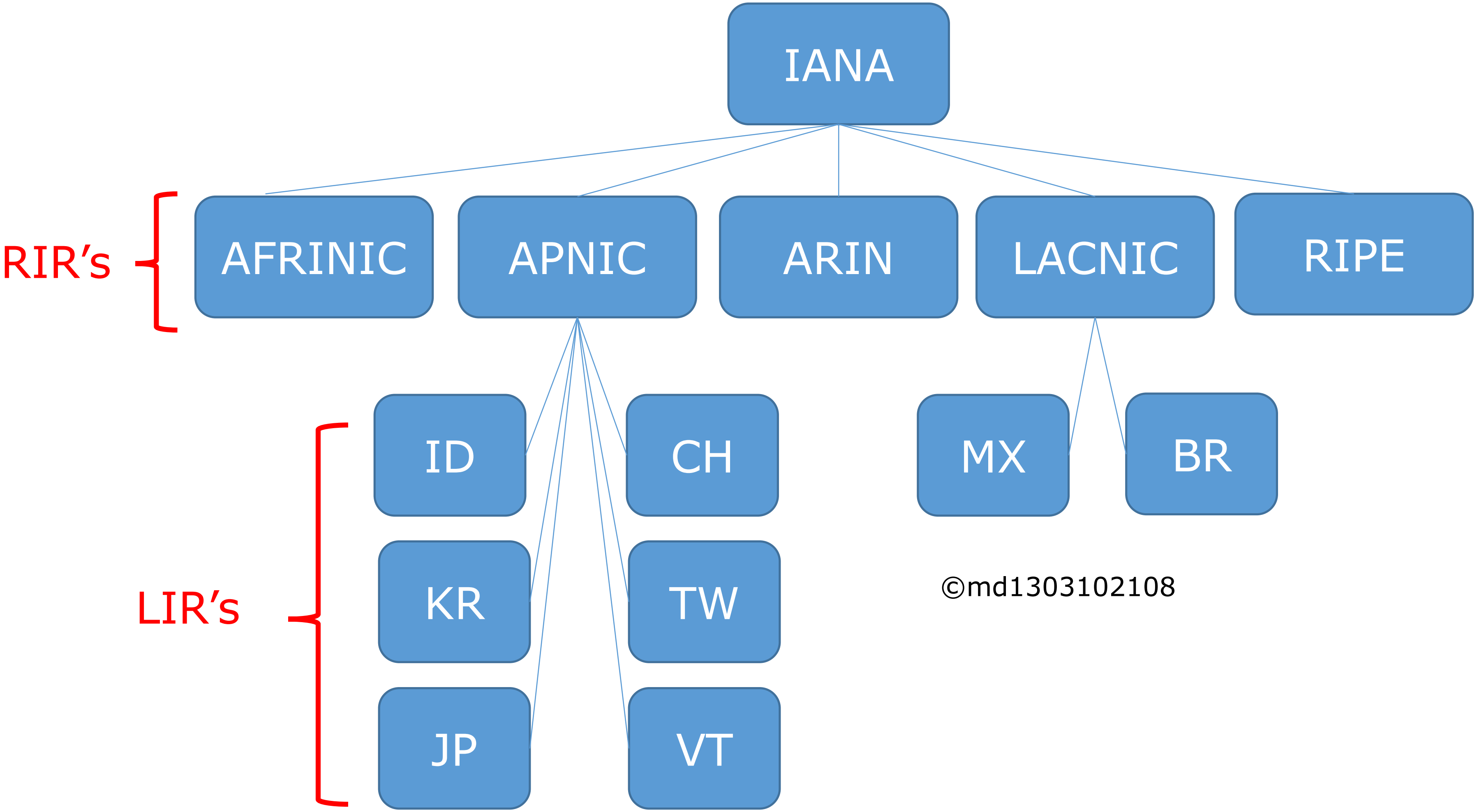
Autonomous System

©md1302192042

In practice you could become an AS with a administrative process, requesting numeration resources from a RIR (Regional Internet Registry)

For Europe: RIPE NCC

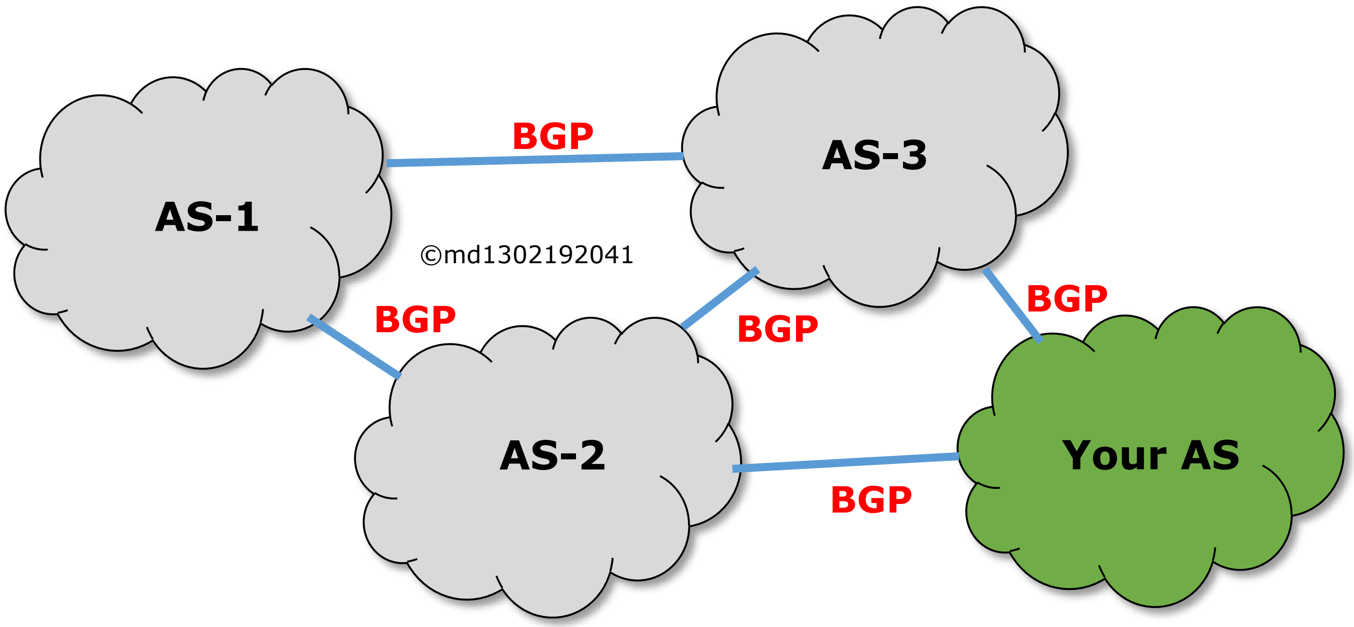
Internet Numbering Resources



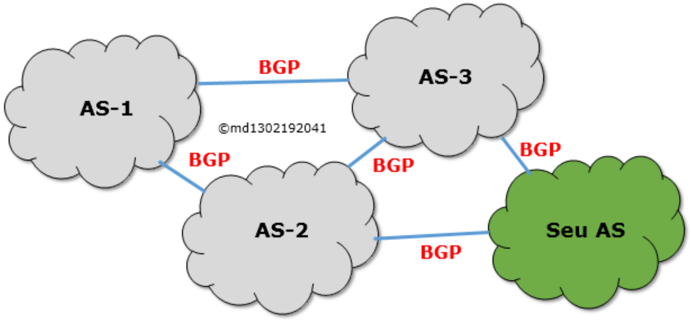
©md1303102108

Autonomous Systems, Internet and BGP Protocol

BGP protocol is the “language” that AS’s talk each other, exchanging routing information and making all destinations reachable.



BGP Protocol



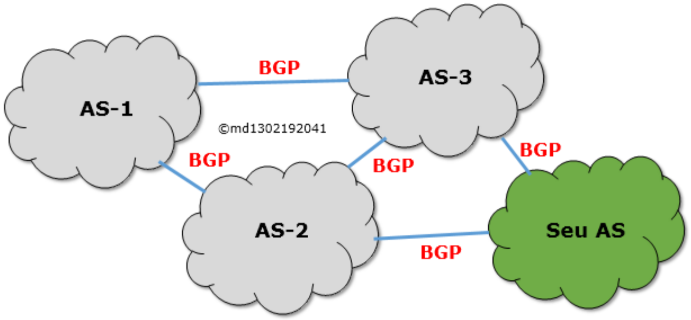
To deal with all Internet traffic, BGP should:

→ be a scalable protocol capable to handle with a huge amount of network prefixes always growing;

→ have robustness and reliability;

→ provide tools to in some way to influence on external traffic not under the direct control of the administrator.

BGP protocol



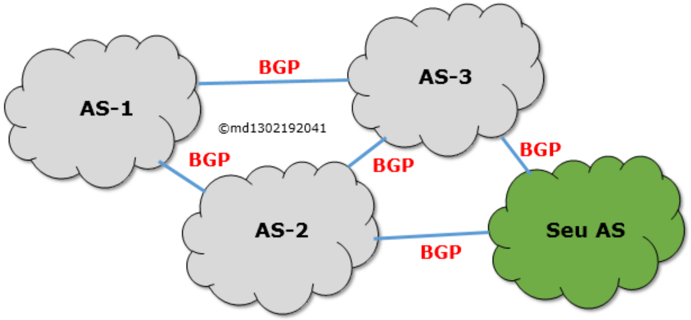
BGP Characteristics:

→ Can be considered a “vector distance” protocol, where each AS represents a single routing hop;

→ No matter how big is the network BGP doesn’t care about internal topology but only how can reach the networks.

→ Current BGP version is BGPv4 according to RFC-1771

BGP Protocol



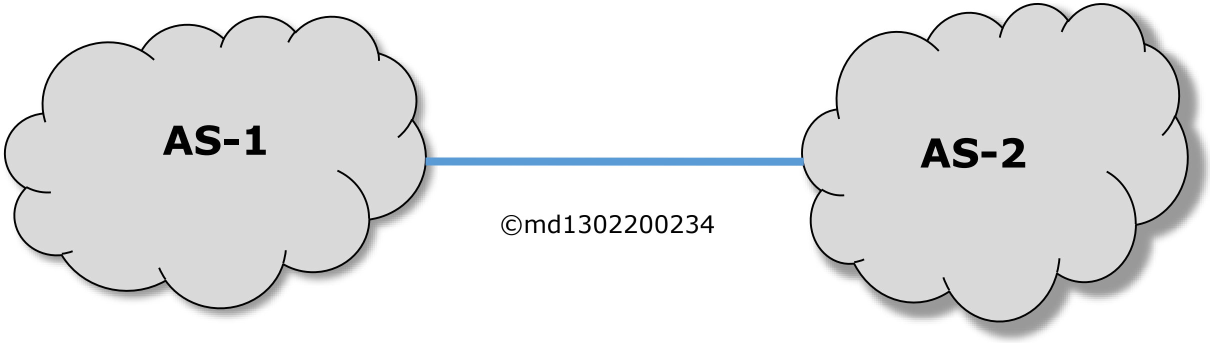
Basic principles:

BGP works exchanging routing information about reachability of networks with NLRI (Network Layer Reachability Information) messages;

NLRI messages have one or more network **prefixes** and **attributes** associated with them;

To ensure data integrity, information are transported over a TCP connection (port 179).

Running BGP



→ Both administrators configure the BGP peering;

→ A TCP session to port 179 is established and over it the BGP session;

→ Both sides exchange routing information until total convergence;

→ After this only information about new and withdrawn routes are excehanged.

BGP Messages

OPEN

First message sent after TCP connection establishment and confirmed with a KEEPALIVE;

KEEPALIVE

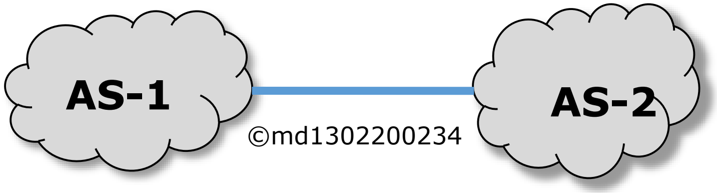
Messages exchanged in intervals of 60 seconds to check peer state;

UPDATE

Information about network prefixes;

NOTIFICATION

Sent when an error occurs;



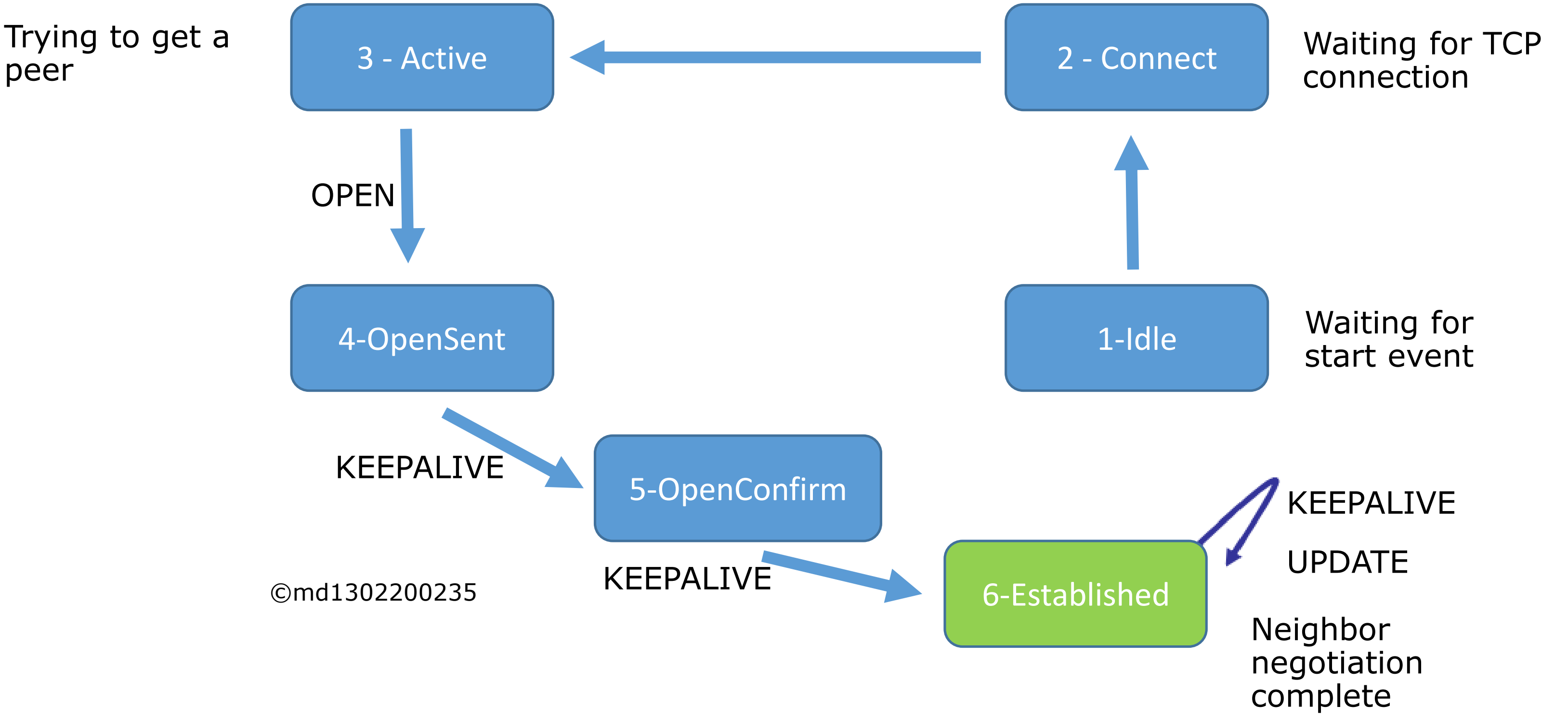
©md1302200234

Optional message:

ROUTE REFRESH

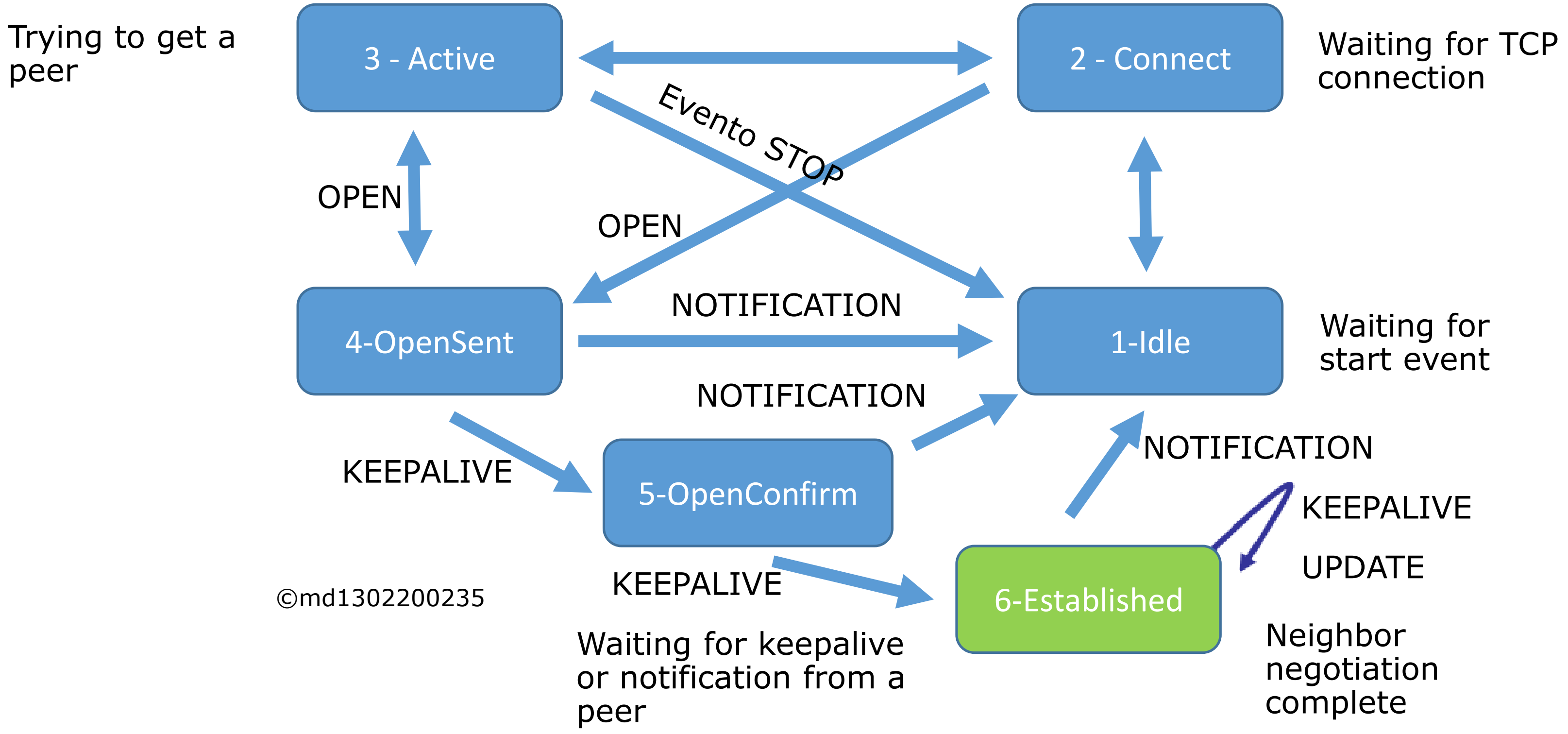
Ask the neighbor to send the routes again.

BGP states



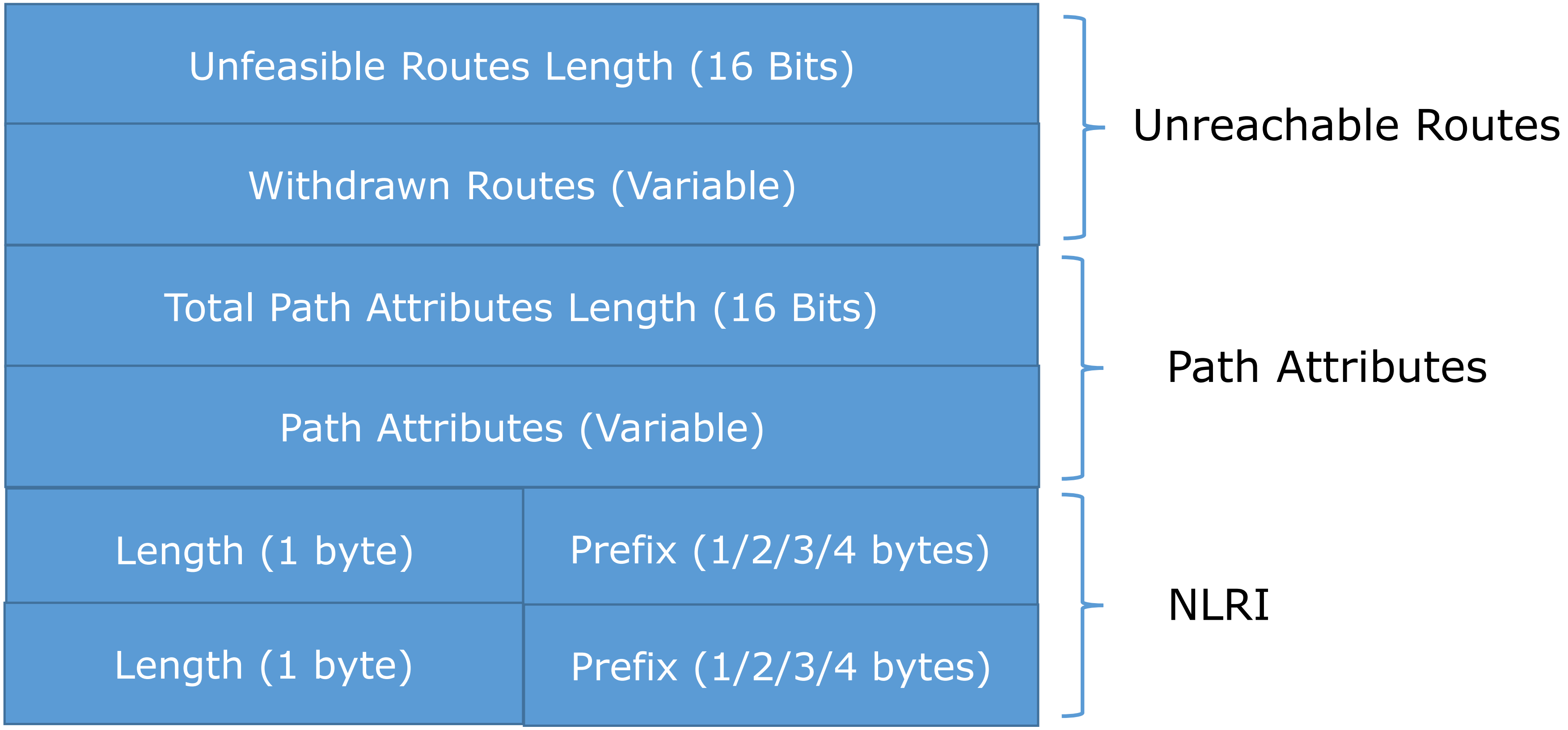
©md1302200235

BGP states



©md1302200235

UPDATE message



Unfeasible Routes Length (16 Bits)

Withdrawn Routes (Variable)

Unreachable Routes

Total Path Attributes Length (16 Bits)

Path Attributes (Variable)

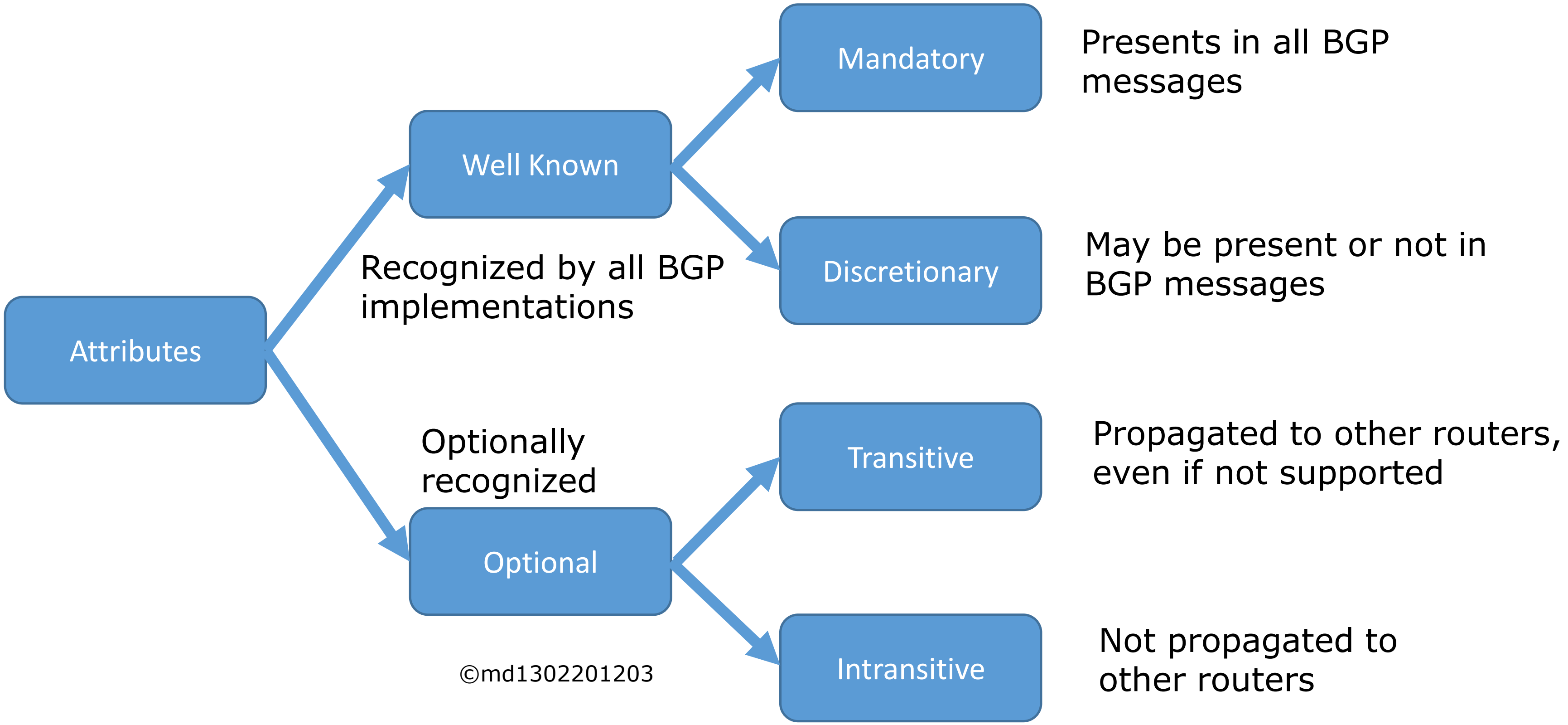
Path Attributes

Length (1 byte) | Prefix (1/2/3/4 bytes)

Length (1 byte) | Prefix (1/2/3/4 bytes)

NLRI

Attributes Types



Attributes

Well Known

Recognized by all BGP implementations

Mandatory

Presents in all BGP messages

Discretionary

May be present or not in BGP messages

Optionally recognized

Optional

Transitive

Propagated to other routers, even if not supported

Intransitive

Not propagated to other routers

©md1302201203

AS-Path, Next-Hop, Communities and Local Preference

AS-Path:

AS sequence through which a network is reachable;

Next-Hop:

IP address of the next hop router

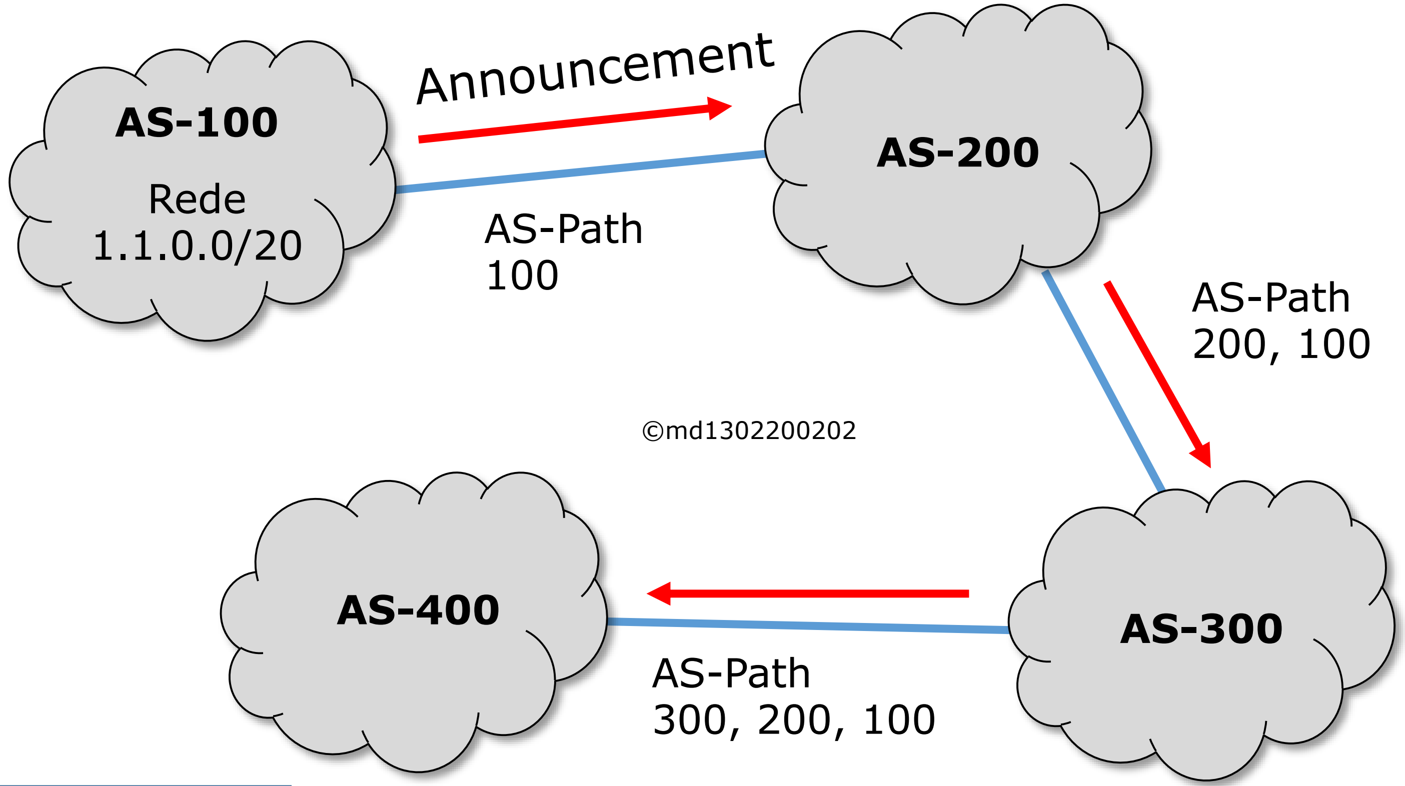
Community:

Numeric value that can be attached to a prefix with some specific purpose;

Local Preference:

Attribute used to choose a preferred outbound path inside an AS;

Understanding the AS-Path



Announcement

AS-100

Rede 1.1.0.0/20

AS-Path 100

AS-200

AS-Path 200, 100

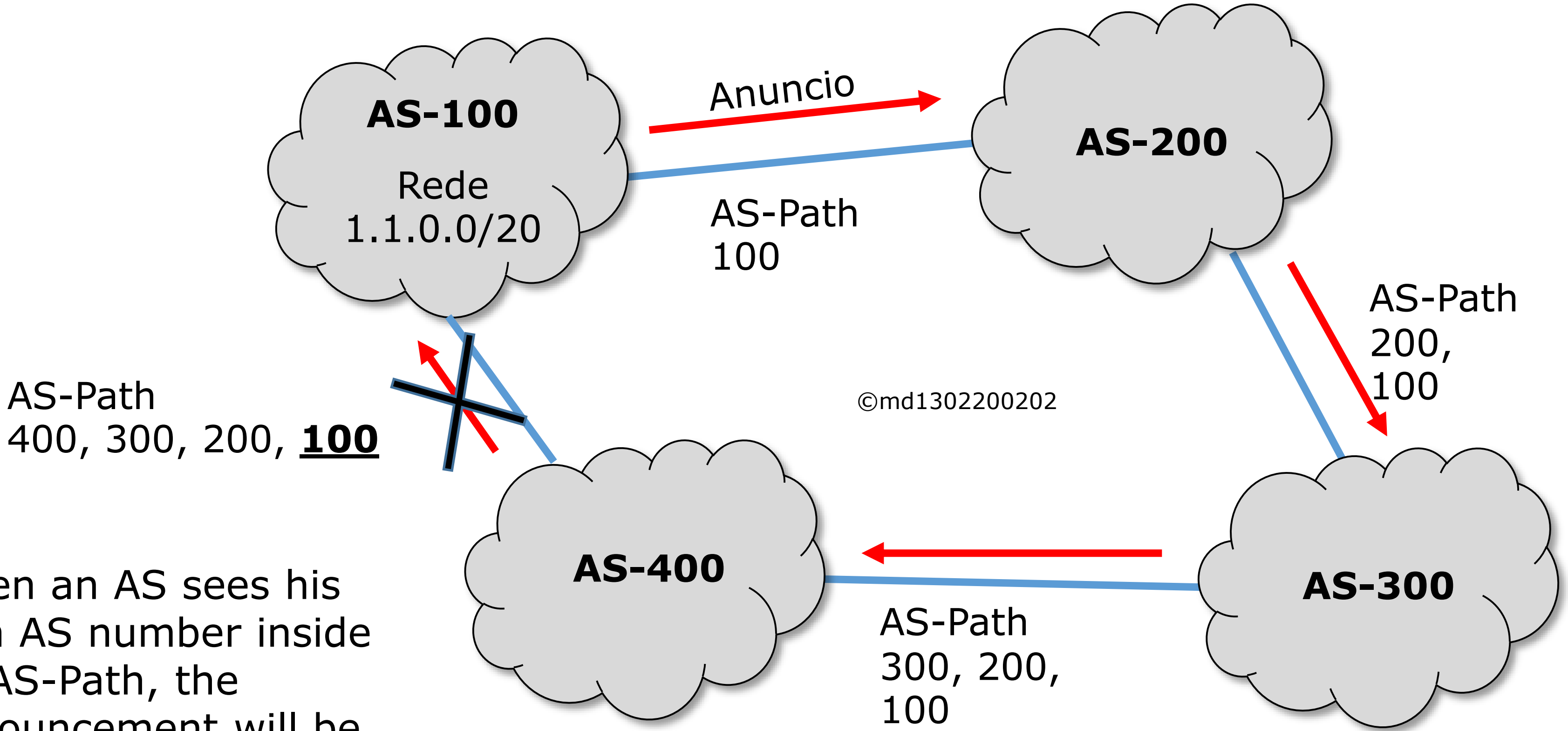
AS-300

AS-Path 300, 200, 100

AS-400

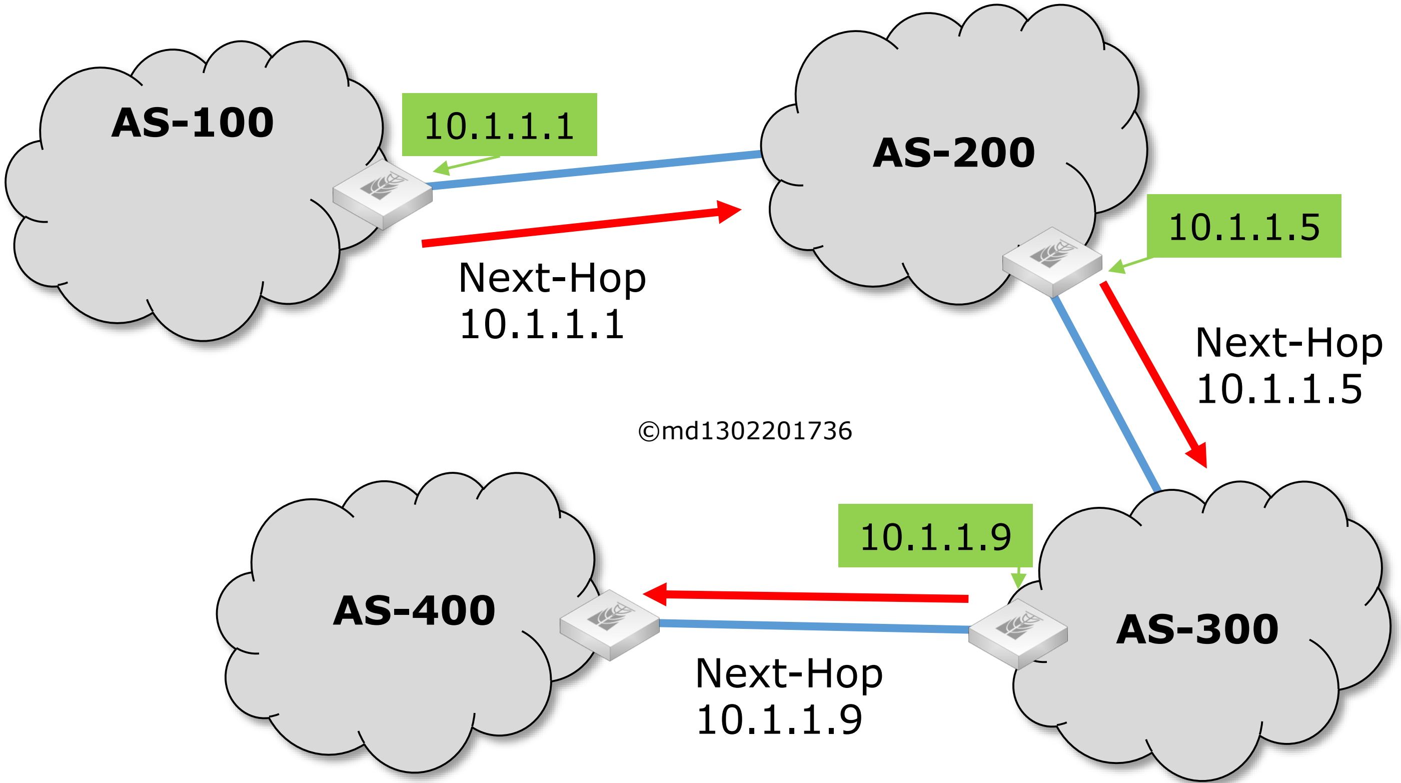
NETWORK	AS-path
1.1.0.0/20	300,200,100

Looping Prevention



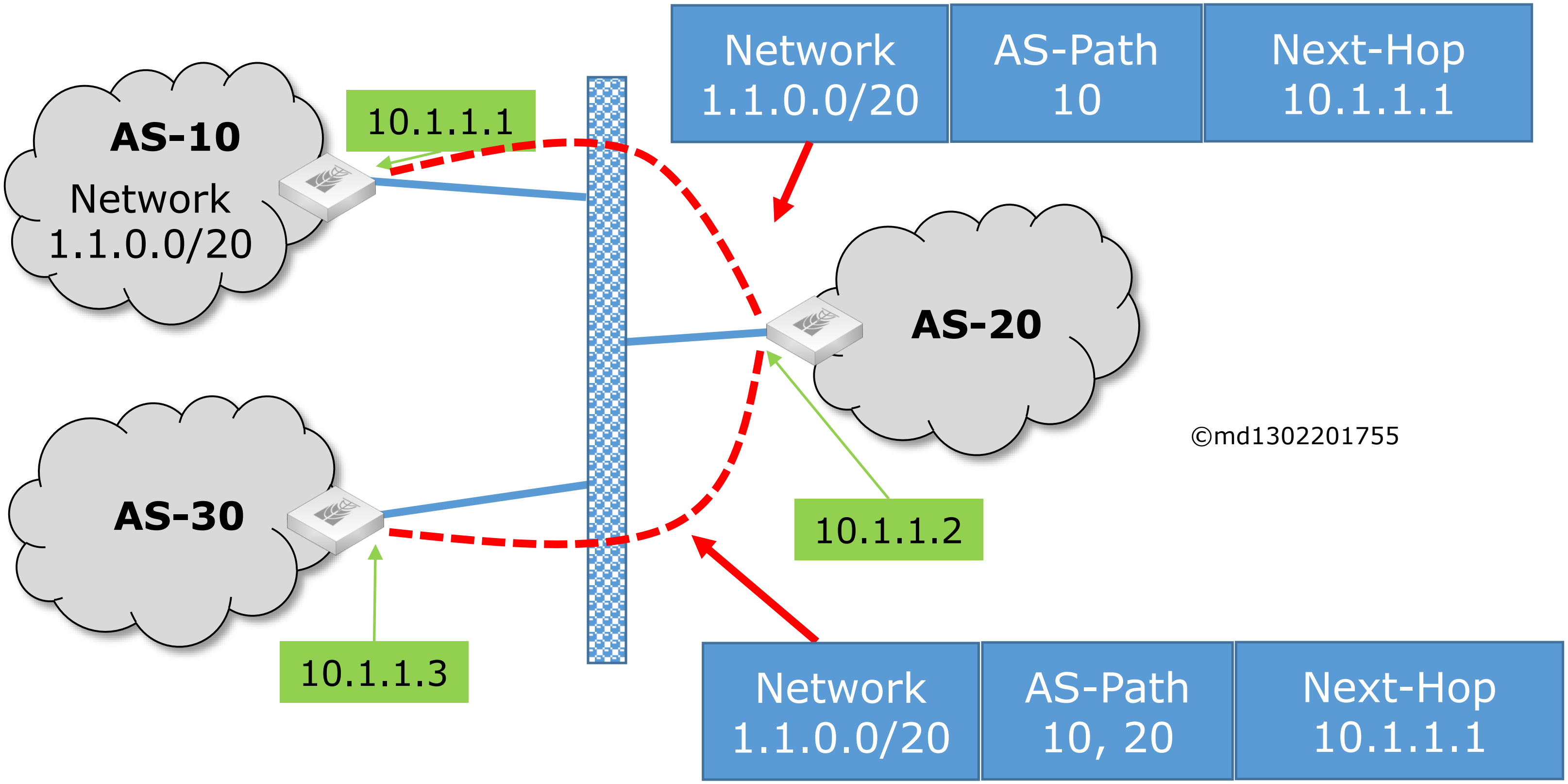
When an AS sees his own AS number inside an AS-Path, the announcement will be discarded.

Understanding Next-Hop



NETWORK	AS-Path	Next-Hop
1.1.0.0/20	300,200,100	10.1.1.9

Next-Hop on an shared network (e.g. IXP)



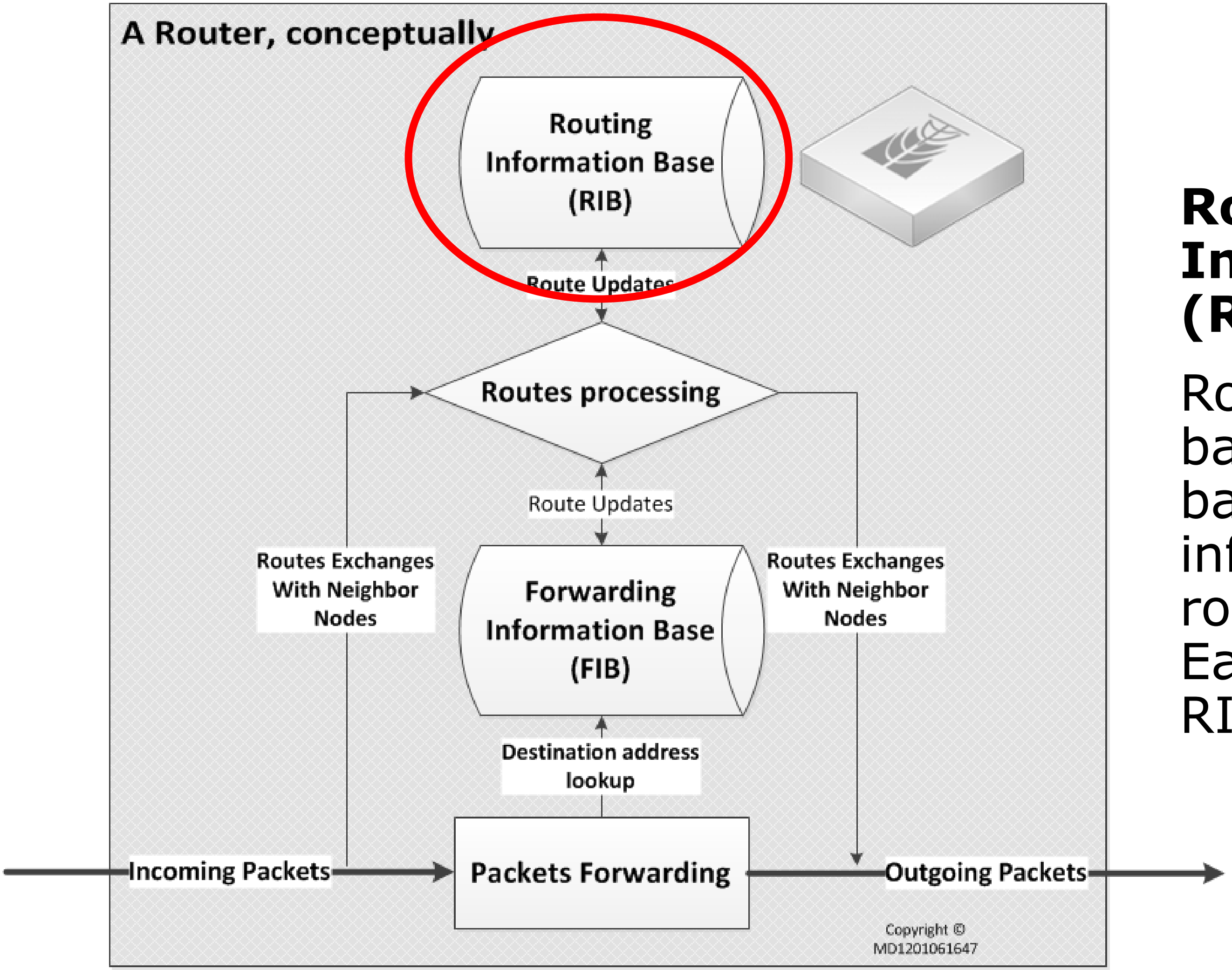
Network	AS-Path	Next-Hop
1.1.0.0/20	10	10.1.1.1

Network	AS-Path	Next-Hop
1.1.0.0/20	10, 20	10.1.1.1

To optimize packet forwarding, in a shared subnet, next hop will be kept.

How BGP decides about the best route?

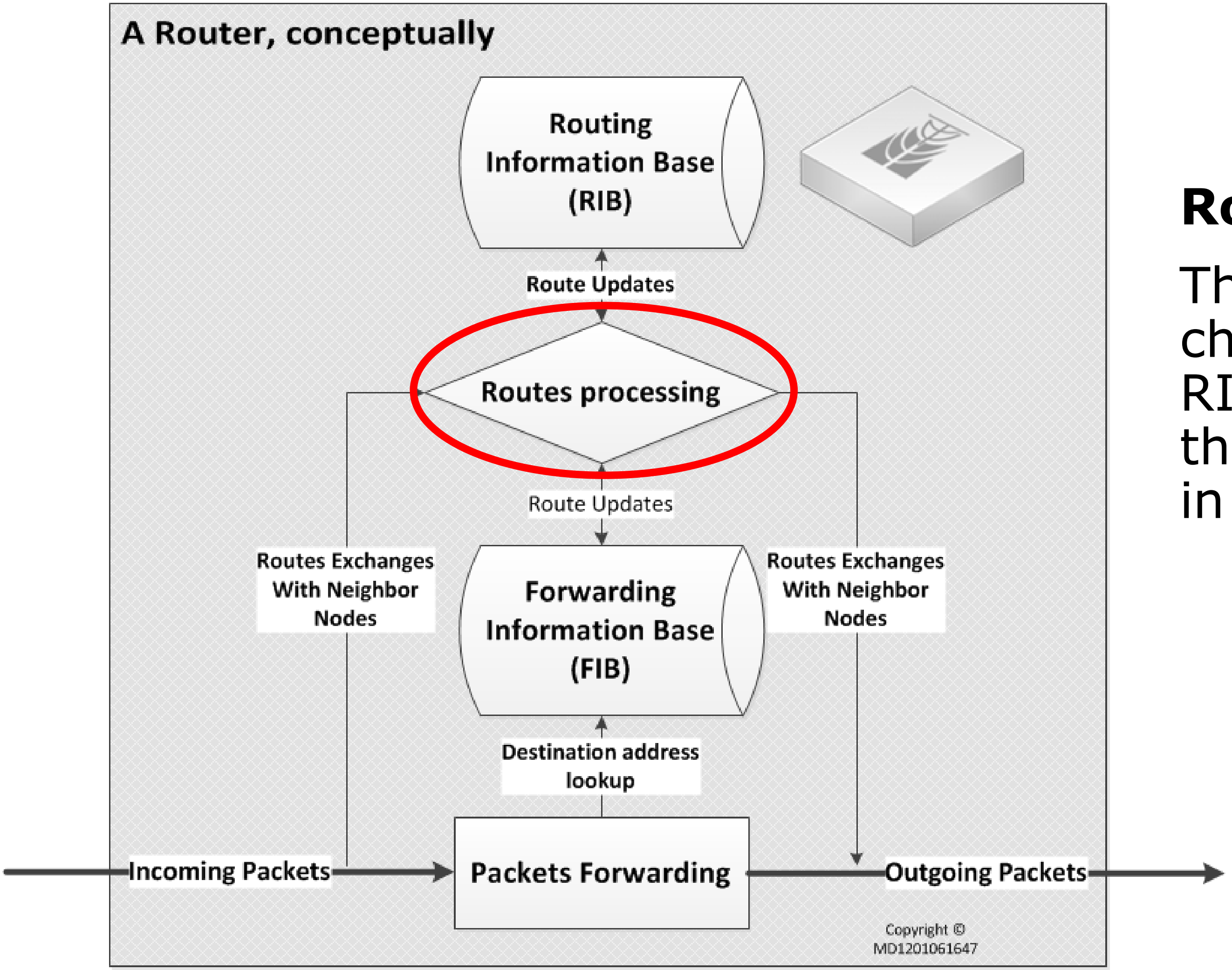
Routing Essentials


A Router, conceptually
Routing Information Base (RIB)
Route Updates
Routes processing
Route Updates
Routes Exchanges With Neighbor Nodes
Forwarding Information Base (FIB)
Routes Exchanges With Neighbor Nodes
Destination address lookup
Incoming Packets
Packets Forwarding
Outgoing Packets
Copyright © MD1201061647

Routing Information Base (RIB)

Routing Information base is the data base where all information about IP routes are stored. Each protocol has its RIB

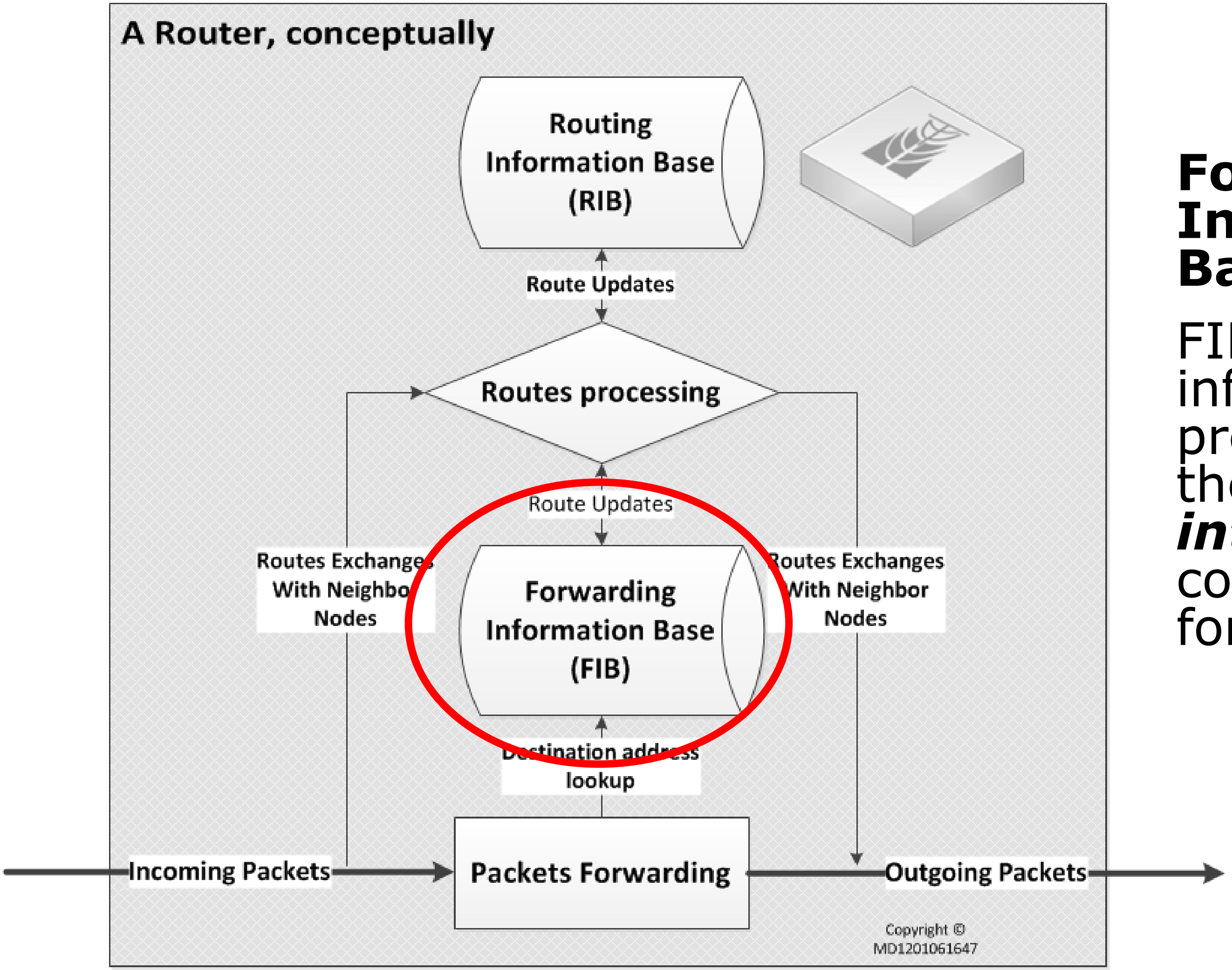
Routing Essentials


A Router, conceptually
Routing Information Base (RIB)
Route Updates
Routes processing
Route Updates
Routes Exchanges With Neighbor Nodes
Forwarding Information Base (FIB)
Routes Exchanges With Neighbor Nodes
Destination address lookup
Incoming Packets
Packets Forwarding
Outgoing Packets
Copyright © MD1201061647

Routes Processing

This process will choose among the RIB routes, the ones that will be installed in the FIB

Routing Essentials

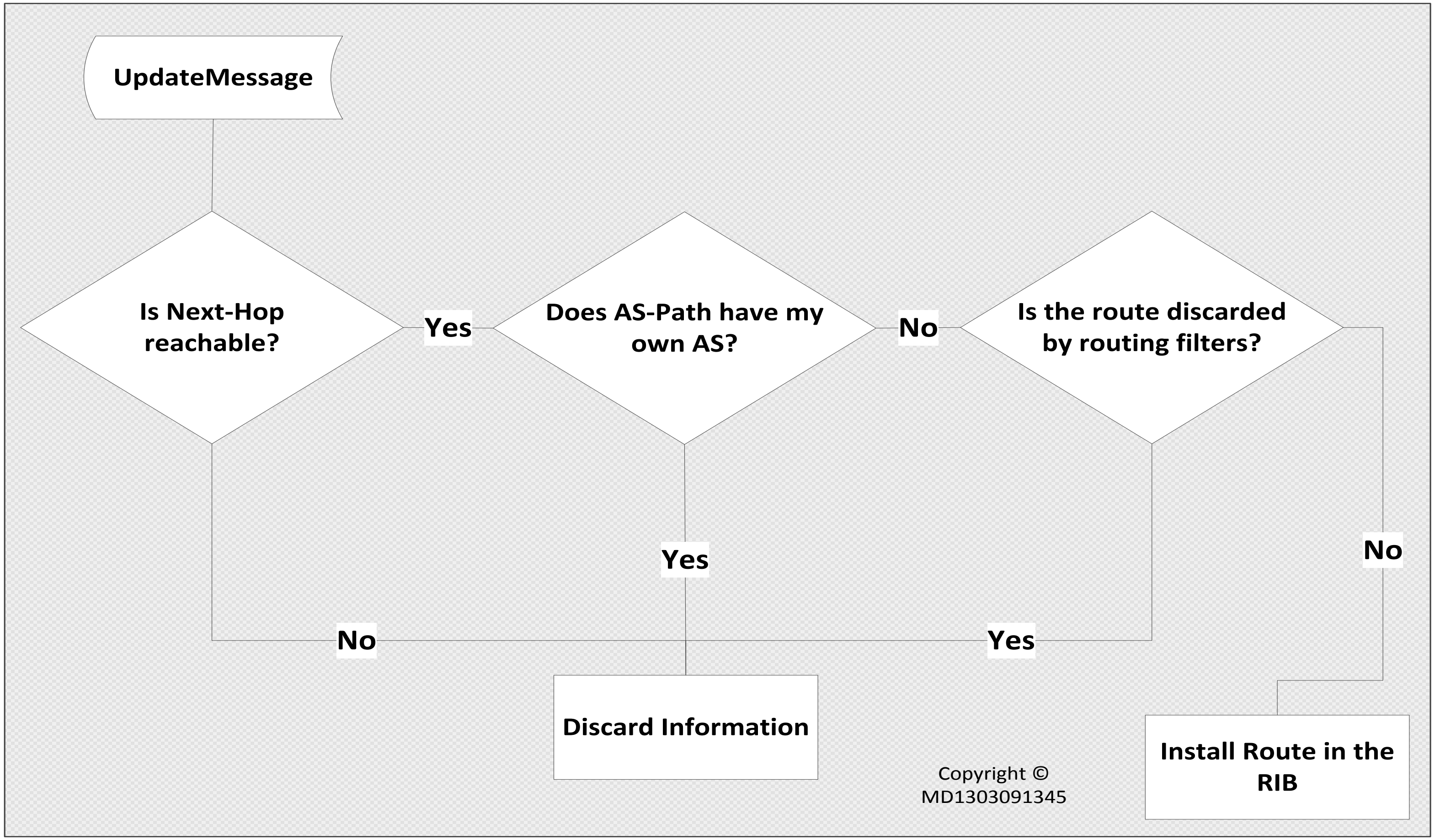

A Router, conceptually
Routing Information Base (RIB)
Route Updates
Routes processing
Route Updates
Routes Exchanges With Neighbor Nodes
Forwarding Information Base (FIB)
Routes Exchanges With Neighbor Nodes
Destination address lookup
Incoming Packets
Packets Forwarding
Outgoing Packets
Copyright © MD1201061647

Forwarding Information Base (FIB)

FIB contains information of prefixes related to the ***network interfaces*** that could be used to forward packets.

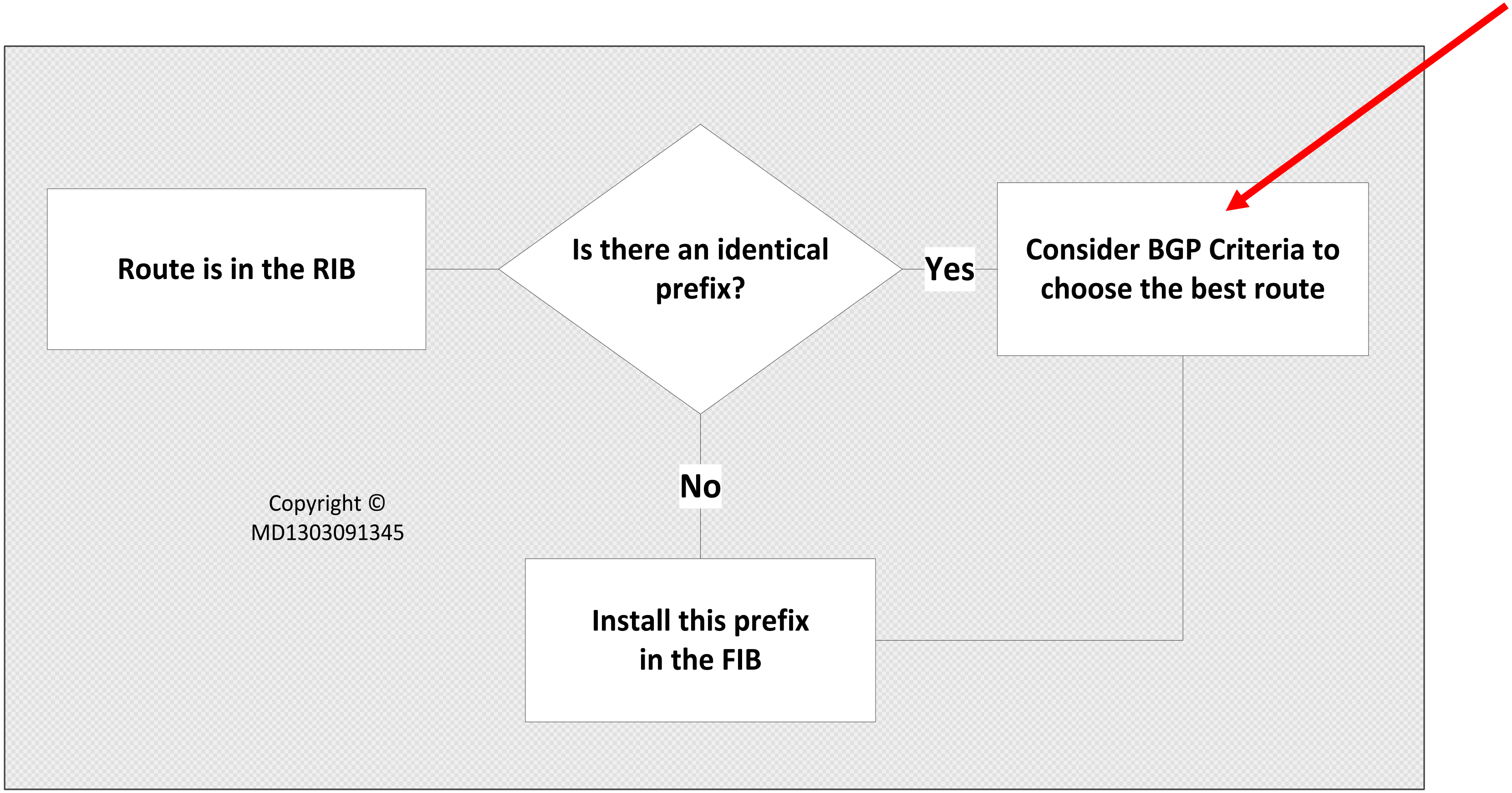
How BGP decides about the best routes

When receiving a BGP update message:



How BGP decides about the best routes

If the route is the first one in the RIB, it will be chosen. Otherwise, BGP decision criteria will be considered for selection


Route is in the RIB
Is there an identical prefix?
Yes
Consider BGP Criteria to choose the best route
No
Install this prefix in the FIB
Copyright © MD1303091345

BGP criteria for decision

BGP will compare identical prefixes in the following order:

1) Prefers the path with highest **WEIGHT** (default = 0);
2) Prefers path with highest **LOCAL-PREFERENCE** (default = 100);
3) Prefers path with the shortest **AS-Path**;
4) Prefers the path locally originated via aggregate or BGP network announce;
5) Prefers the path with lowest **ORIGIN** (igp < egp < incomplete);
6) Prefers the path with the lowest **MED** (default = 0);
7) Prefers the path learned by eBGP over the ones by iBGP;
8) Prefers the path received from the router with lower Router ID;
9) Prefers the path with shortest route reflection cluster list (default = 0);
10) Prefers the path that comes from the lowest neighbor address.

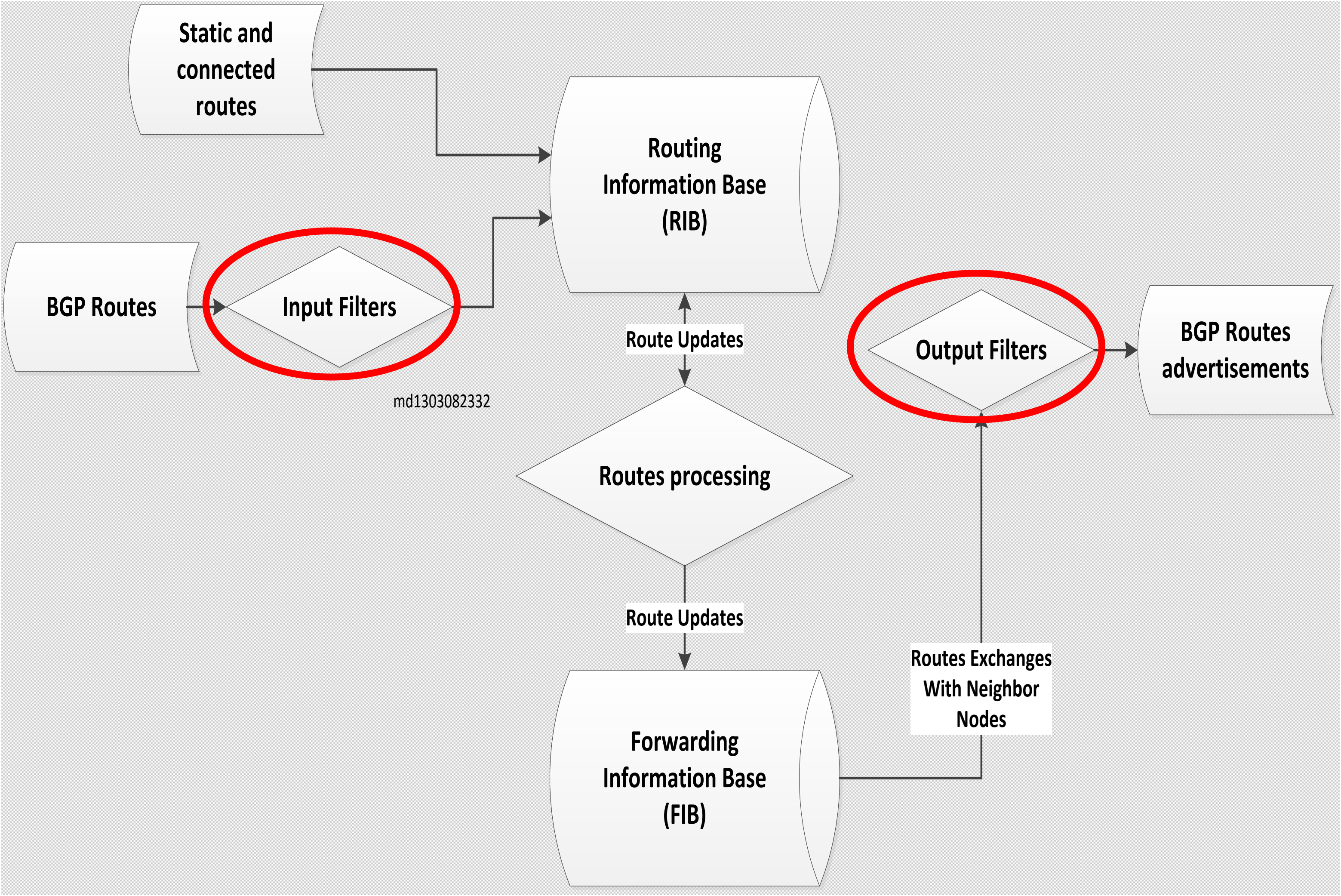
Routing Filters

The way to influence BGP decision is by configuring routing filters.

Filtering **incoming** routes will change, how we see the external world, thus influencing how we **send** traffic;

Filtering **outgoing** routes will change how the world see us, thus influencing how we **receive** traffic.

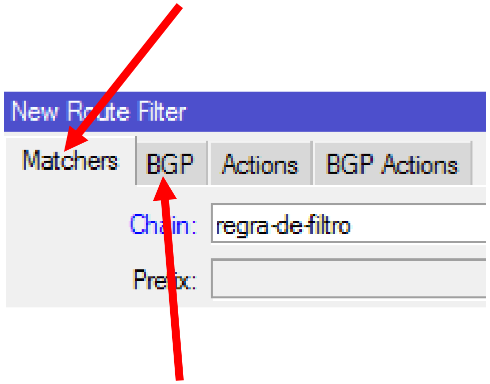
BGP Filter placement


Static and connected routes
Routing Information Base (RIB)
BGP Routes
Input Filters
Route Updates
Output Filters
BGP Routes advertisements
md1303082332
Routes processing
Route Updates
Routes Exchanges With Neighbor Nodes
Forwarding Information Base (FIB)

Understanding Routing Filters “Semantics” in RouterOS

Matchers

Matchers by the prefix itself, prefix-length, protocol, routing marks, etc.



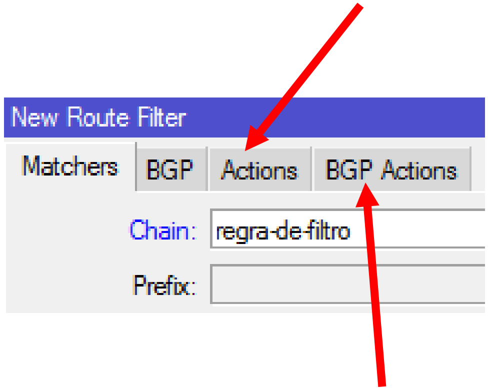
Chain: regra-de-filtro

Prefix:

Matchers by BGP attributes inside the UPDATE message.

Actions

Actions to be done, like accept, discard etc.



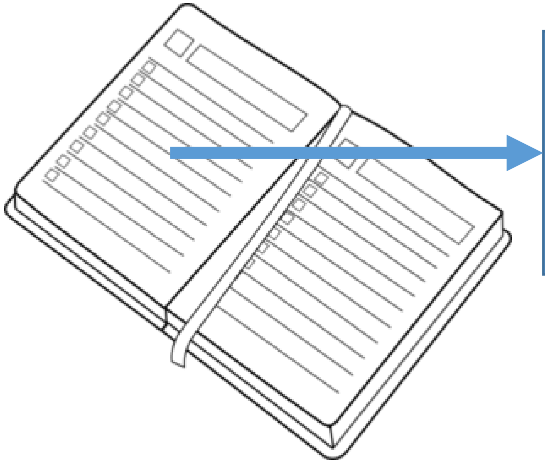
Chain: regra-de-filtro

Prefix:

Actions intended to modify BGP attributes on a specific route.

Agenda

1) BGP essentials and basics of BGP filtering; ✓



2) Case Studies:

2.1) Overview

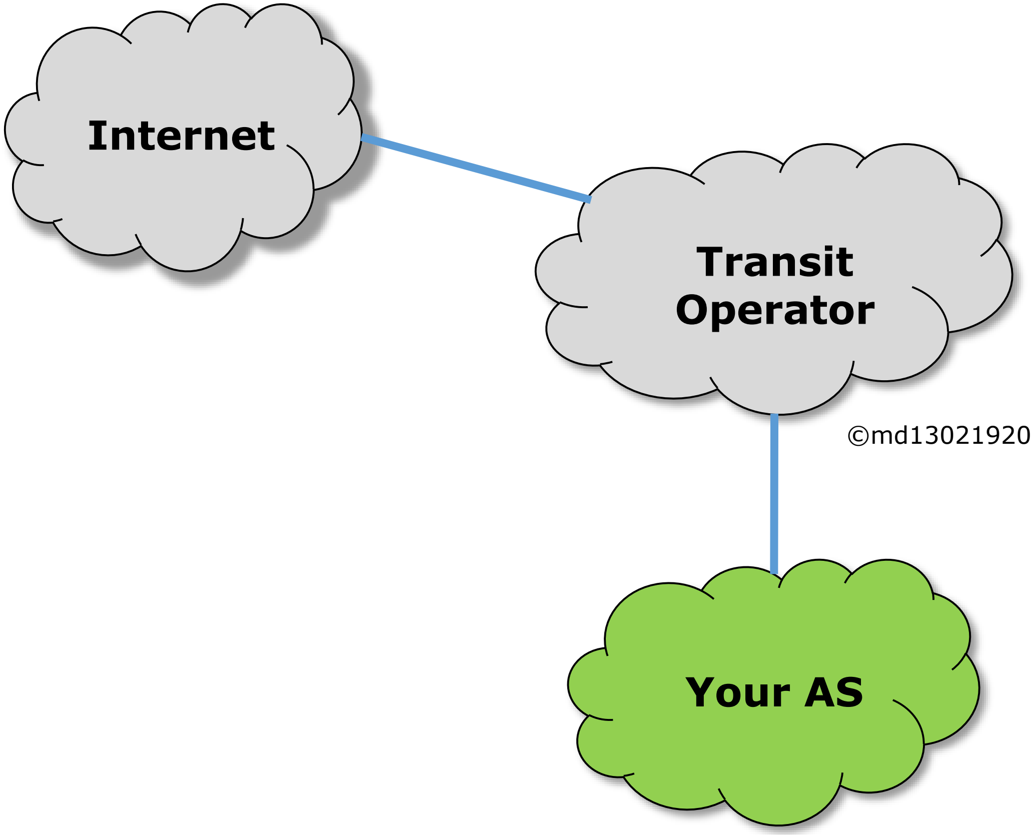
2.2) Single-Homed Provider

2.3) Single-Homed + IXP

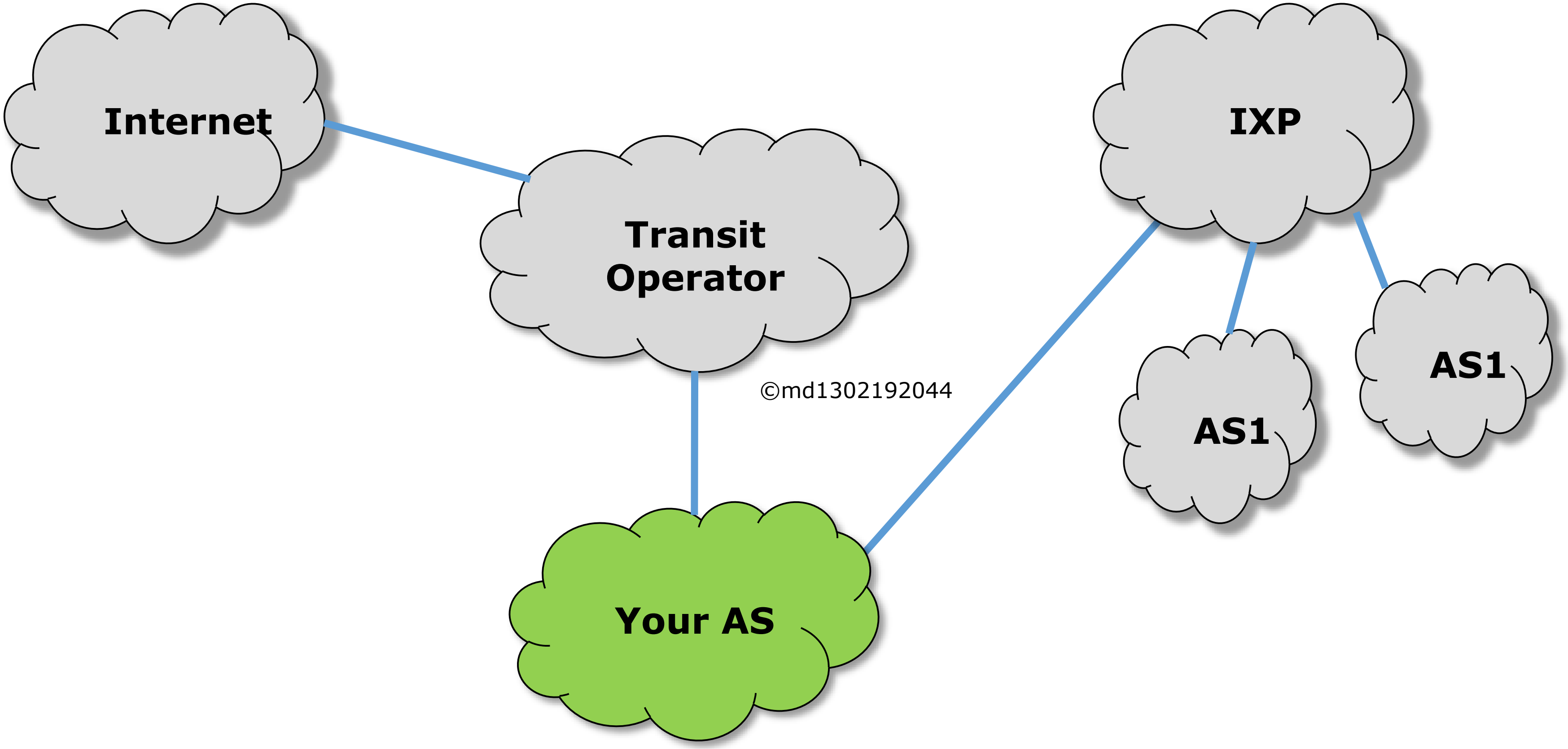
2.4) Multi-Homed + IXP

2.5) Multi-Homed + IXP + Providing transit services

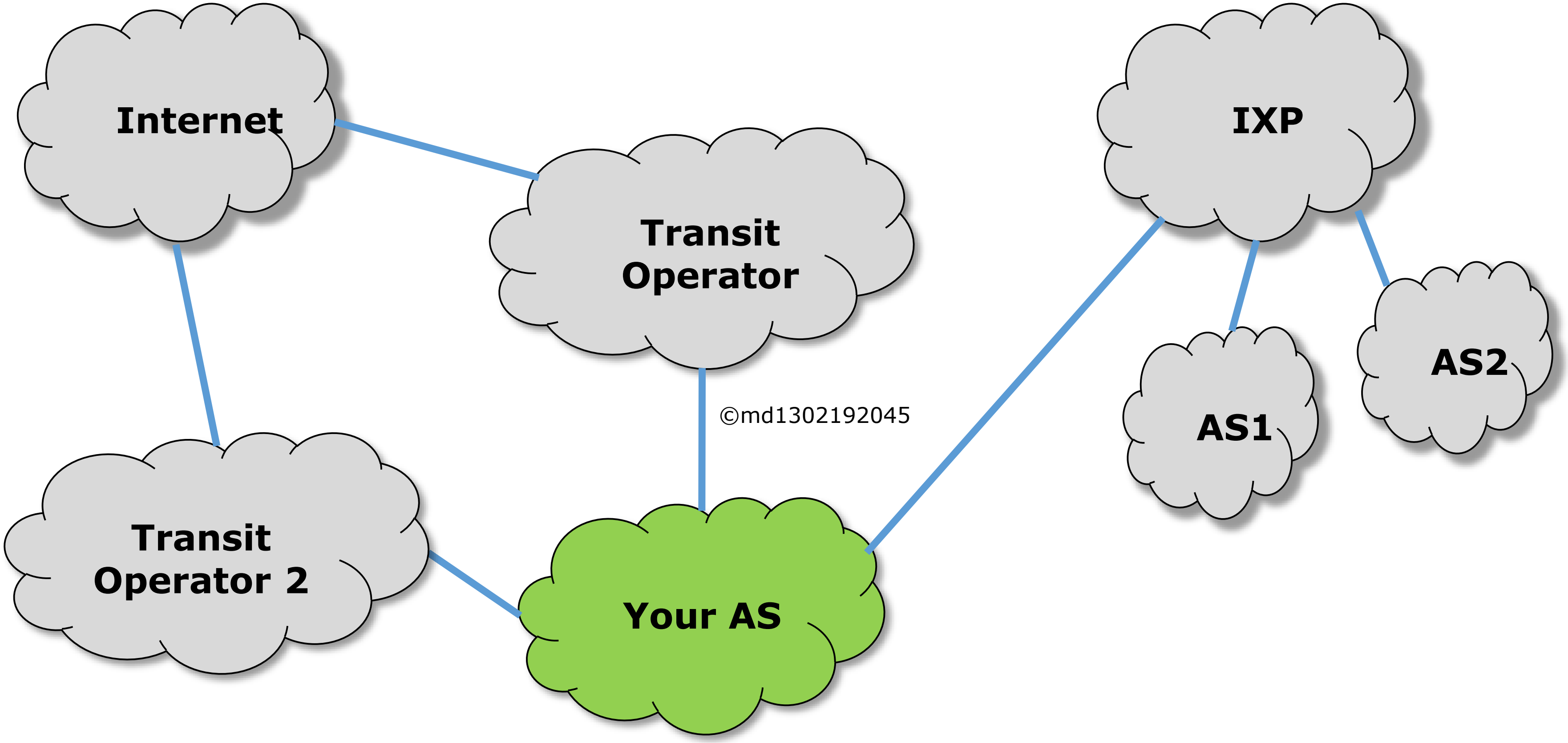
Scenario I
Single Homed ISP



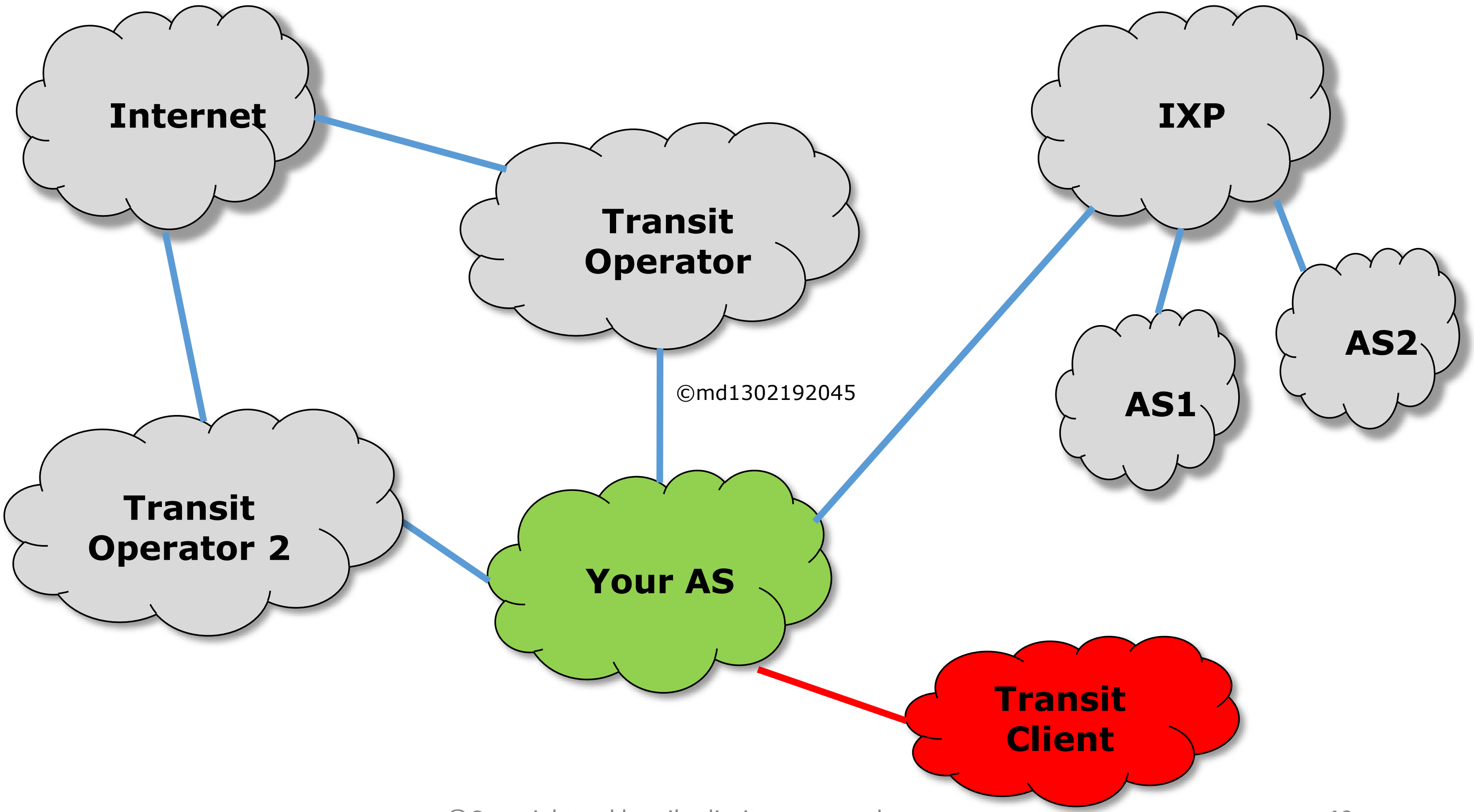
Scenario II
Single-Homed + IXP



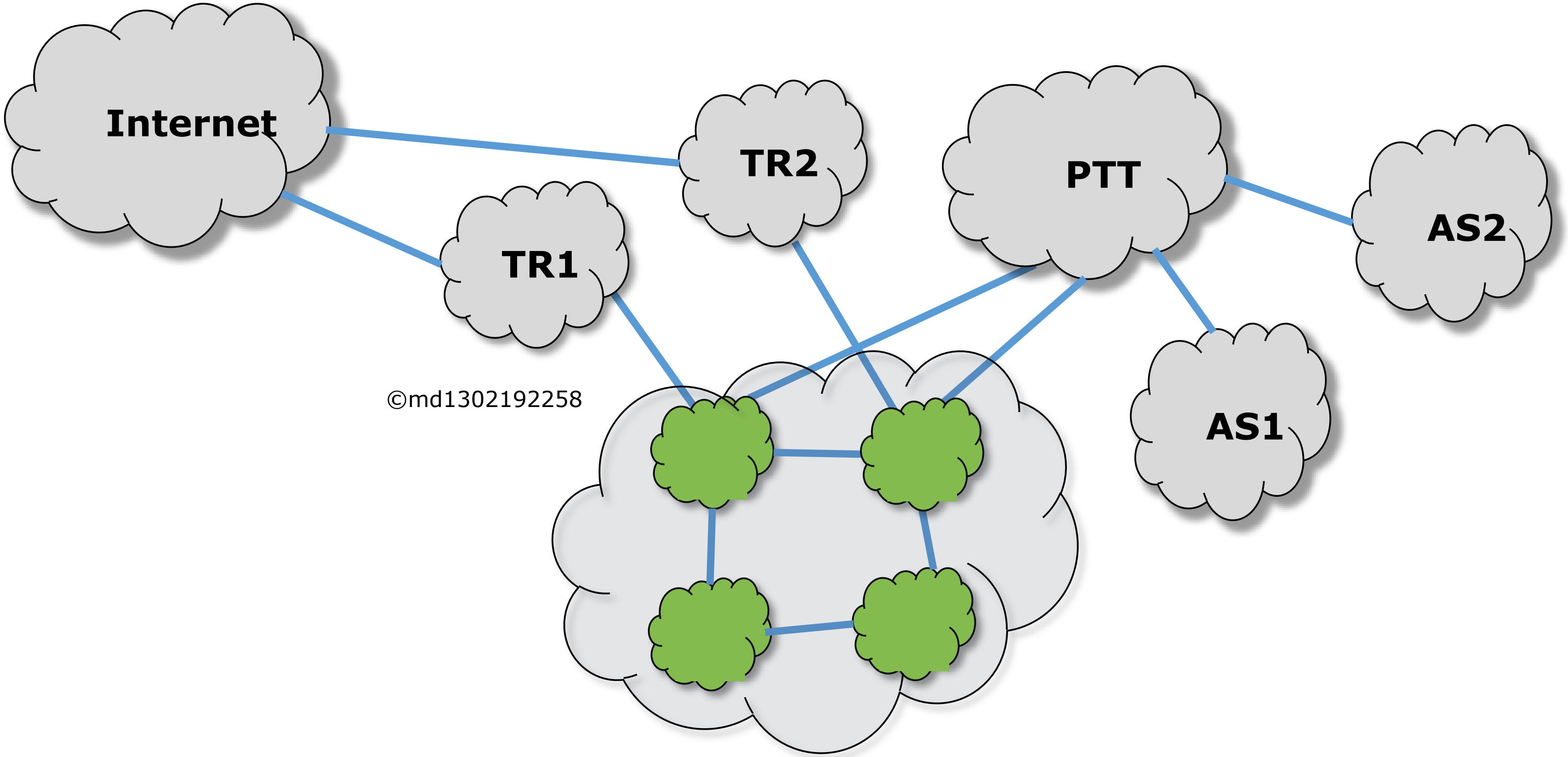
Scenario III
Dual-Homed + IXP



Scenario IV
Dual-Homed + IXP
Providing Transit services



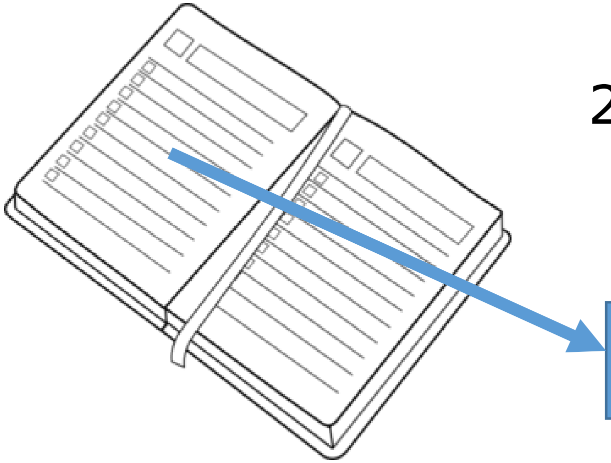
Scenario V – Multi-Homed + IXP + iBGP + Confederation



Agenda

1) BGP essentials and basics of BGP filtering; ✓

2) Case Studies: ✓



2.1) Overview ✓

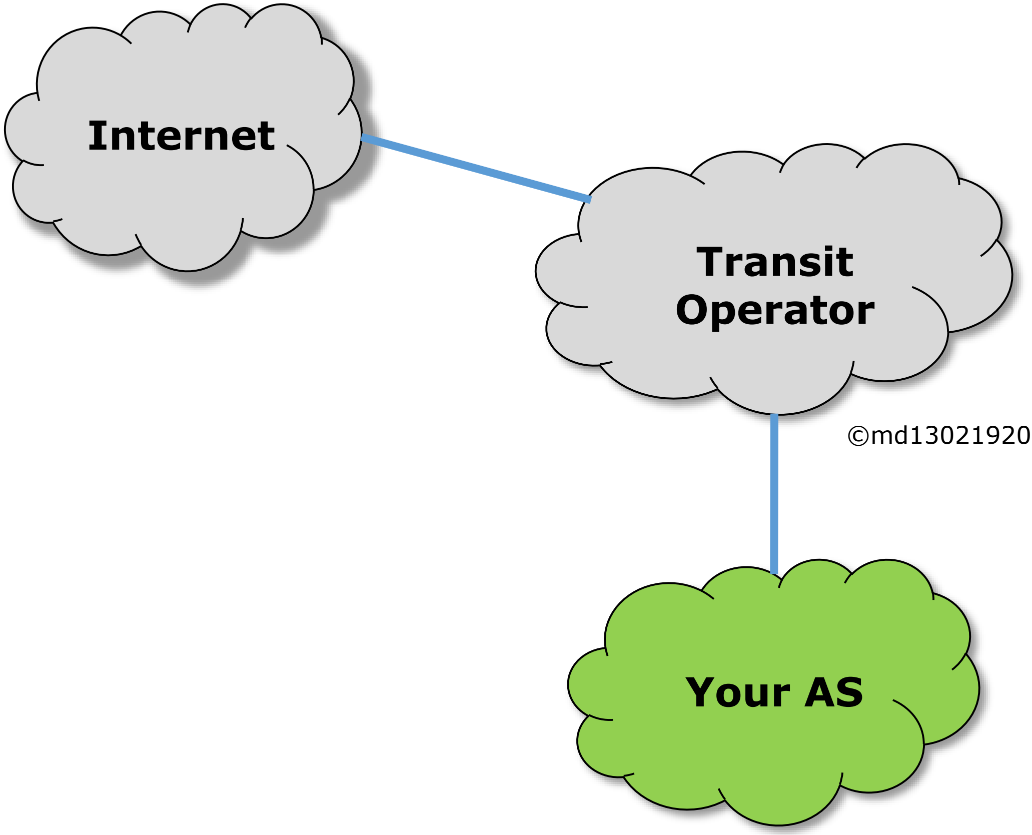
2.2) Single-Homed Provider

2.3) Single-Homed + IXP

2.4) Multi-Homed + IXP

2.5) Multi-Homed + IXP + Providing transit services

Scenario I
Single-Homed ISP



©md1302192043

Before BGP Session

You should sign an agreement with your transit provider to define some policies for you BGP session, like:

→ If you want Full or Partial Routing;

→ Which prefixes you intend to announce;

→ If you want a default Route;

→ MD5 password;

→ If the session should be established with a loopback interface;

etc.

Simulating the Scenario

For the purpose of this presentation, we are going to assume that:

→ Our transit provider is sending us a Full routing table;

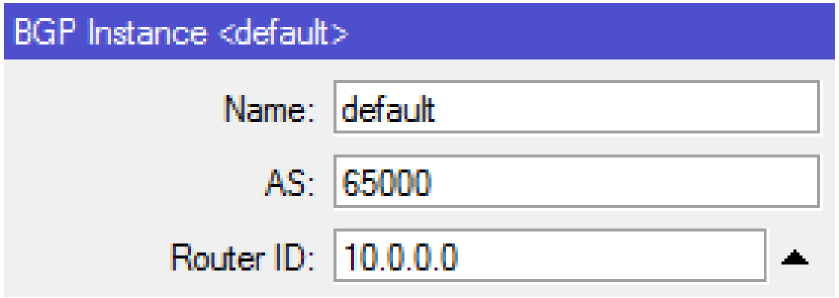
→ We're announcing the prefix 11.11.0.0/20;

→ Our peer will be established with a direct connected interface*

→ Our Transit Provider does not offer native IPv6 transit.

* Not a good practice. Please see work about routing security: http://mum.mikrotik.com/presentations/HU11/maia.pdf

BGP Configuration

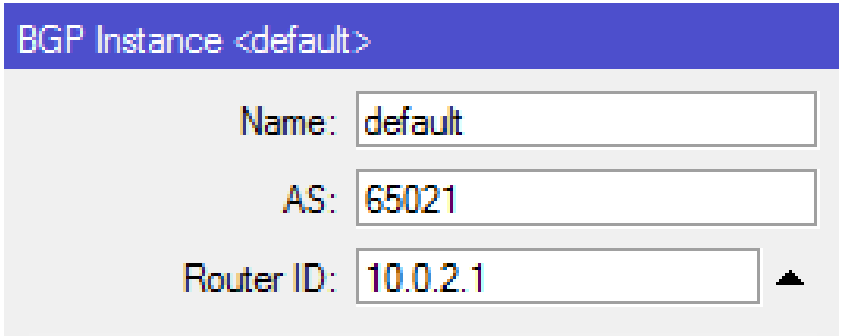


BGP Instance <default>

Name: default

AS: 65000

Router ID: 10.0.0.0

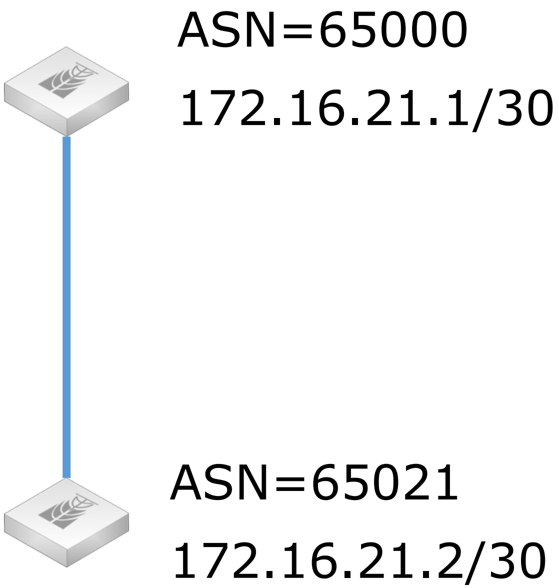


BGP Instance <default>

Name: default

AS: 65021

Router ID: 10.0.2.1

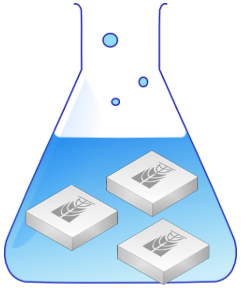


ASN=65000

172.16.21.1/30

ASN=65021

172.16.21.2/30



Minimal Configuration: AS Number and Peer

Router ID: Optional (but recommended)

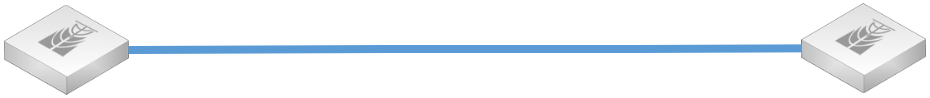
BGP Configuration

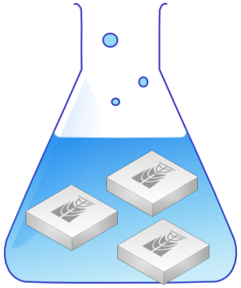
ASN=65000

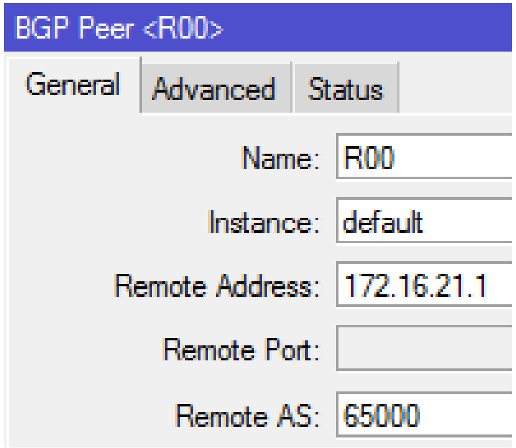
172.16.21.1/30

ASN=65021

172.16.21.2/30

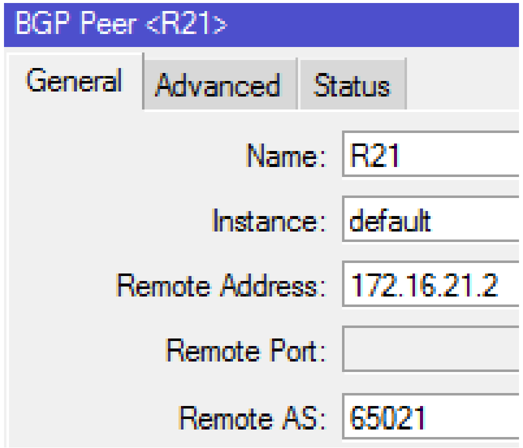






BGP Peer <R00>

Name:	R00
Instance:	default
Remote Address:	172.16.21.1
Remote Port:	
Remote AS:	65000



BGP Peer <R21>

Name:	R21
Instance:	default
Remote Address:	172.16.21.2
Remote Port:	
Remote AS:	65021

Minimal configuration for peer: Remote IP and Remote AS

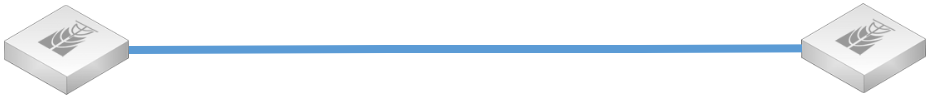
BGP Configuration

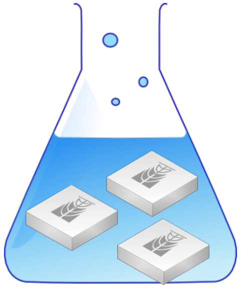
ASN=65000

172.16.21.1/30

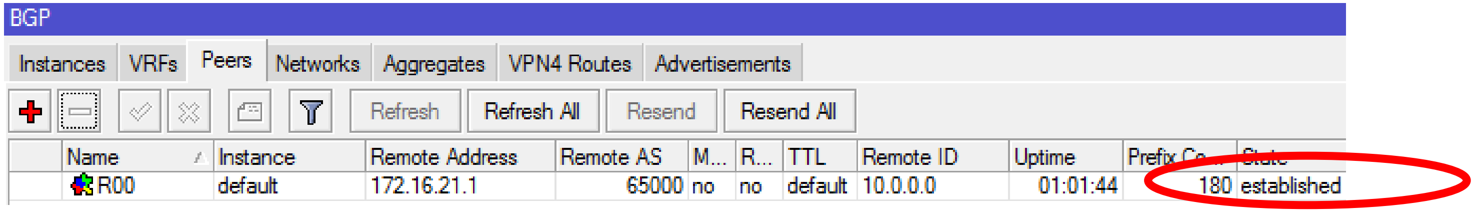
ASN=65021

172.16.21.2/30



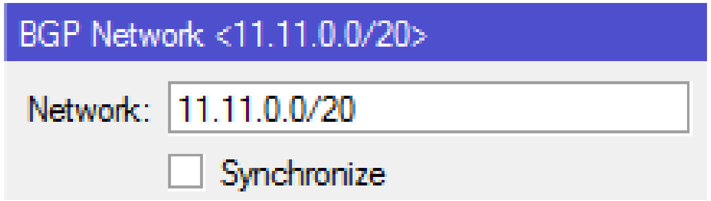


Checking results



Name	Instance	Remote Address	Remote AS	M...	R...	TTL	Remote ID	Uptime	Prefix Co...	State
R00	default	172.16.21.1	65000	no	no	default	10.0.0.0	01:01:44	180	established

Advertising the network

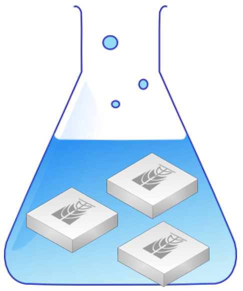


BGP Network <11.11.0.0/20>

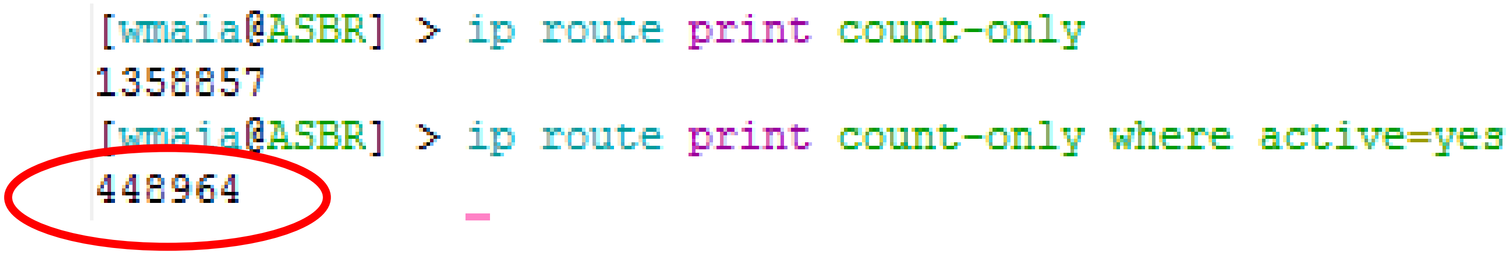
Network: 11.11.0.0/20

□ Synchronize

Testing the Results



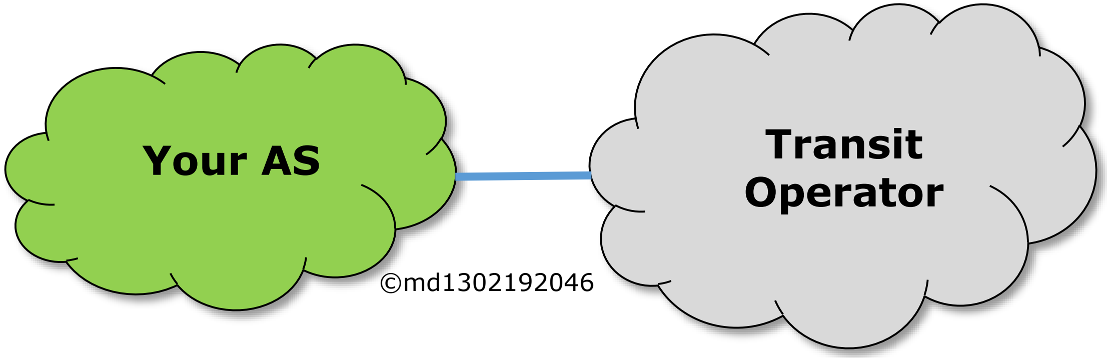
Supposing you ask for a Full routing, by this time you can look on your routing table and see ~400k network prefixes.



```
[wmaia@ASBR] > ip route print count-only
1358857
[wmaia@ASBR] > ip route print count-only where active=yes
448964
```

Do we need this bunch of prefixes?

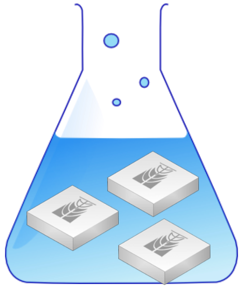
Prefixes Control



Your AS

Transit Operator

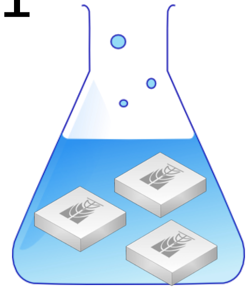
©md1302192046

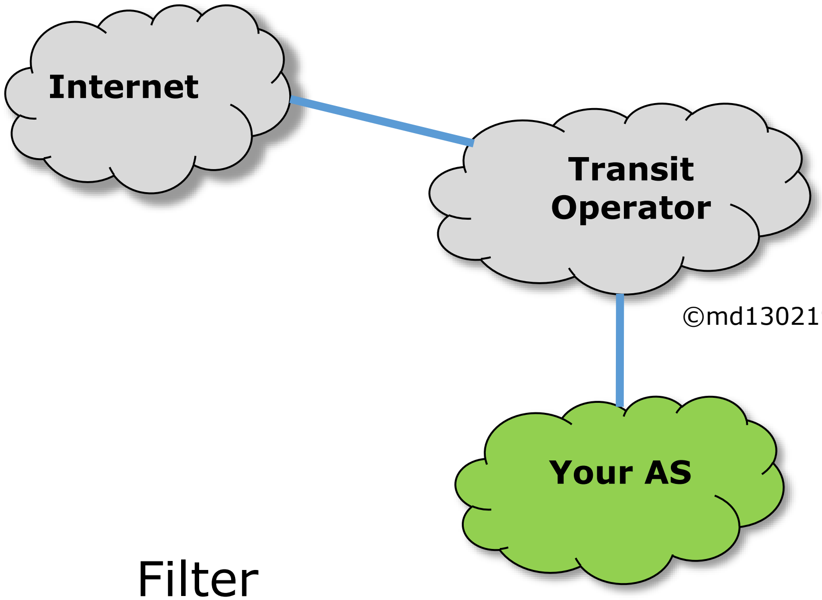


By default, nothing is filtered.

Routing filters allow the control of ingress and egress announcements.

BGP Filtering for Scenario 1





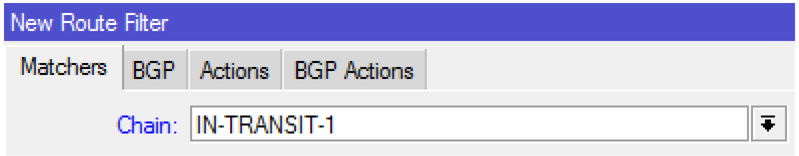
©md1302192043

To spare resources, you can:

→ Discard all routes received

→ Setup a static default route

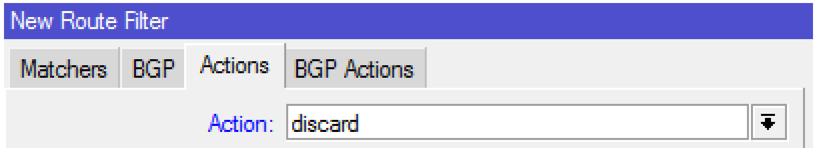
Filter



New Route Filter

Matchers | BGP | Actions | BGP Actions

Chain: IN-TRANSIT-1

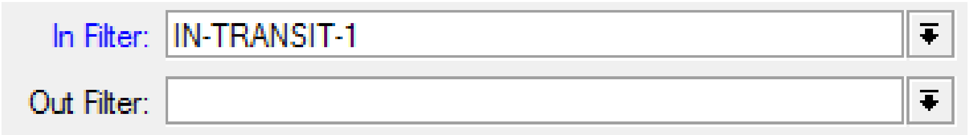


New Route Filter

Matchers | BGP | Actions | BGP Actions

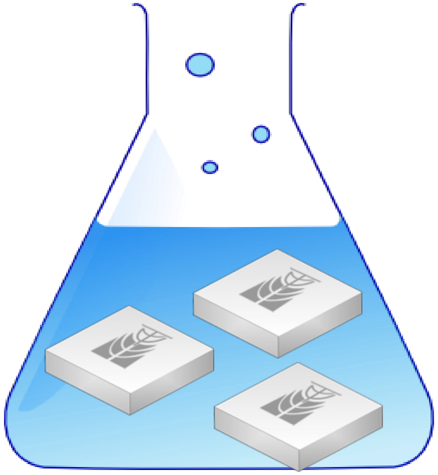
Action: discard

Peer



In Filter: IN-TRANSIT-1

Out Filter:



Break for hands on!

Discarding all routes and configuring a default one
(“Internet” → 99.99.0.1)

Anything else to do with a Single-Homed ISP?

QUIZZ

If we have a default route, should we do anything else?

Having a default route, all packets to any destination will be forwarded, **including** that ones destined to bogons networks.

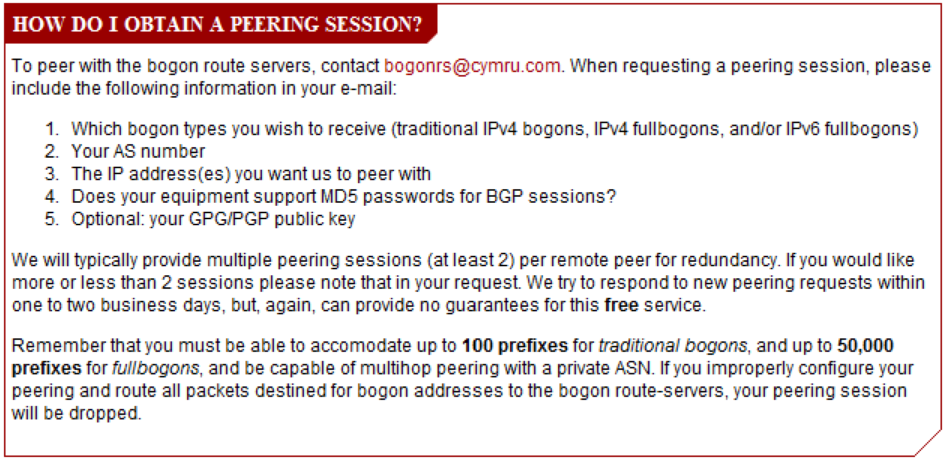
Bogons prefixes are valid ones, but not allocated to any provider or final consumer (they remain “in stock” of the RIR’s0

It is a good practice to deal with BOGONS prefixes!

Bogons treatment

To get automatic information about bogons prefixes, we'll establish a BGP session with Cymru Team http://www.team-cymru.org/

TEAM CYMRU
COMMUNITY
SERVICES



HOW DO I OBTAIN A PEERING SESSION?

To peer with the bogon route servers, contact bogonrs@cymru.com. When requesting a peering session, please include the following information in your e-mail:

1. Which bogon types you wish to receive (traditional IPv4 bogons, IPv4 fullbogons, and/or IPv6 fullbogons)
2. Your AS number
3. The IP address(es) you want us to peer with
4. Does your equipment support MD5 passwords for BGP sessions?
5. Optional: your GPG/PGP public key

We will typically provide multiple peering sessions (at least 2) per remote peer for redundancy. If you would like more or less than 2 sessions please note that in your request. We try to respond to new peering requests within one to two business days, but, again, can provide no guarantees for this **free** service.

Remember that you must be able to accomodate up to **100 prefixes** for *traditional bogons*, and up to **50,000 prefixes** for *fullbogons*, and be capable of multihop peering with a private ASN. If you improperly configure your peering and route all packets destined for bogon addresses to the bogon route-servers, your peering session will be dropped.

Cymru will send bogons prefixes via BGP with the **COMMUNITY** attribute **65332:888**

Understanding Communities

A Community is a 32 bit number you can attach to a route with the purpose to signalize something to other AS’s. A community can be viewed like a “flag” in the route.

There are well known communities, like no-export, no-advertise etc. but any AS can set this own set of communities. The usual format of a community is to split the 32 bit in 2 numbers: AS number:some_number

Communities are widely used to implement routing policies, like:

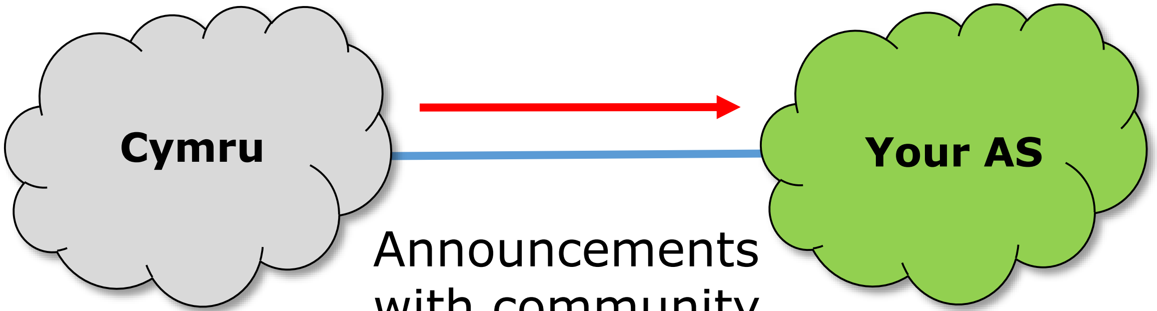
→ Allowing a remote AS to set some Local Preference when sending the announcements;

→ Putting a IP address or network in black hole.

Etc.

Understanding Communities

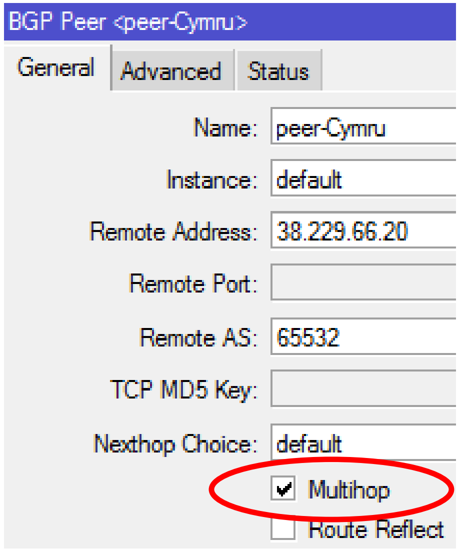
In our case, we know that Cymru will send bogons prefixes with the community 65332:888 and then we'll set up an ingress filter seeing in de BGP attributes if such "flag" is present.



Announcements with community 65332:888

Peering with Cymru

Note that peering with Cymru is a Multihop session



BGP Peer <peer-Cymru>

Name: peer-Cymru

Instance: default

Remote Address: 38.229.66.20

Remote Port:

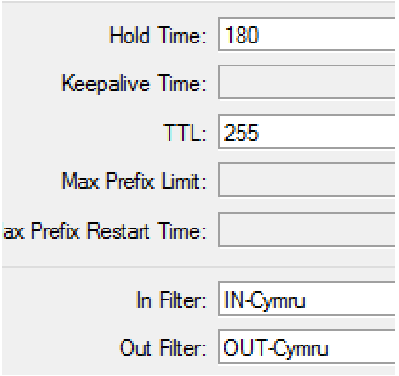
Remote AS: 65532

TCP MD5 Key:

Nexthop Choice: default

☑ Multihop

□ Route Reflect



Hold Time: 180

Keepalive Time:

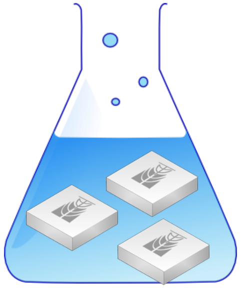
TTL: 255

Max Prefix Limit:

ax Prefix Restart Time:

In Filter: IN-Cymru

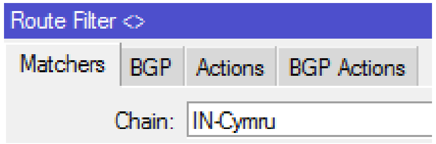
Out Filter: OUT-Cymru



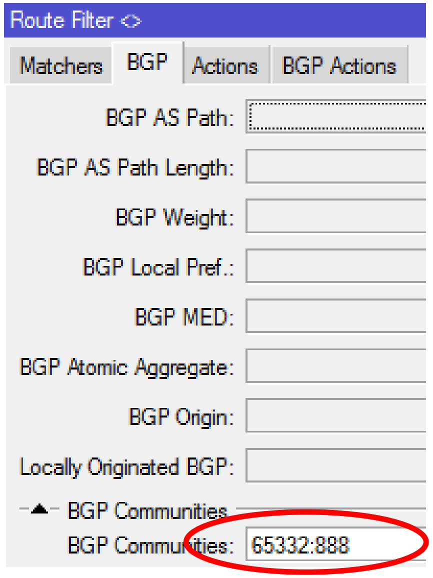
md brasil
TI&TELECOM

Filtering Routes with Cymru

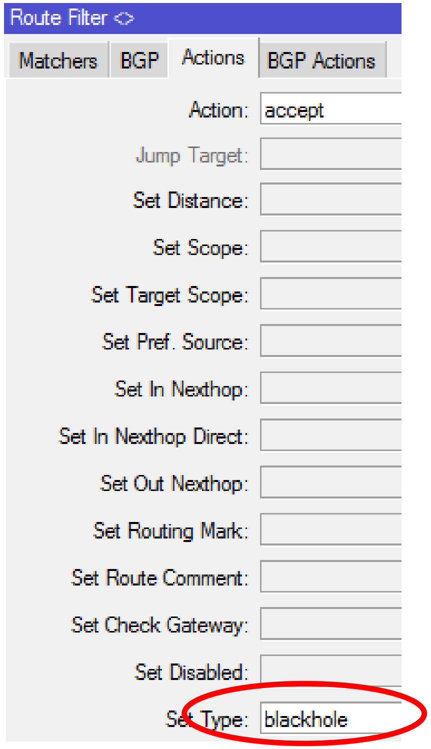
Accepting Cymru routes and setting them as blackhole



Chain: IN-Cymru

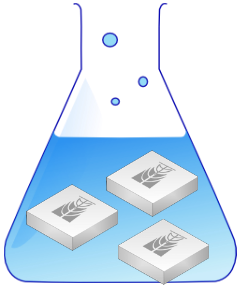


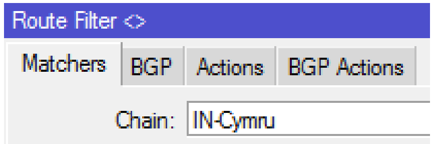
BGP AS Path:
BGP AS Path Length:
BGP Weight:
BGP Local Pref.:
BGP MED:
BGP Atomic Aggregate:
BGP Origin:
Locally Originated BGP:
▲ BGP Communities
BGP Communities: 65332:888



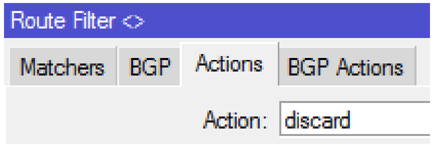
Action: accept
Jump Target:
Set Distance:
Set Scope:
Set Target Scope:
Set Pref. Source:
Set In Nexthop:
Set In Nexthop Direct:
Set Out Nexthop:
Set Routing Mark:
Set Route Comment:
Set Check Gateway:
Set Disabled:
Set Type: blackhole

Avoiding other routes IN and OUT

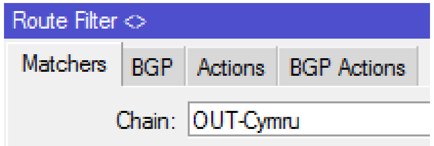




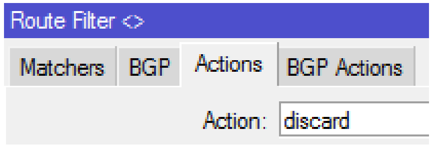
Chain: IN-Cymru



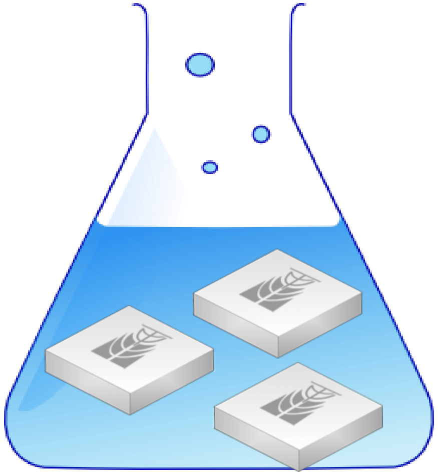
Action: discard



Chain: OUT-Cymru



Action: discard



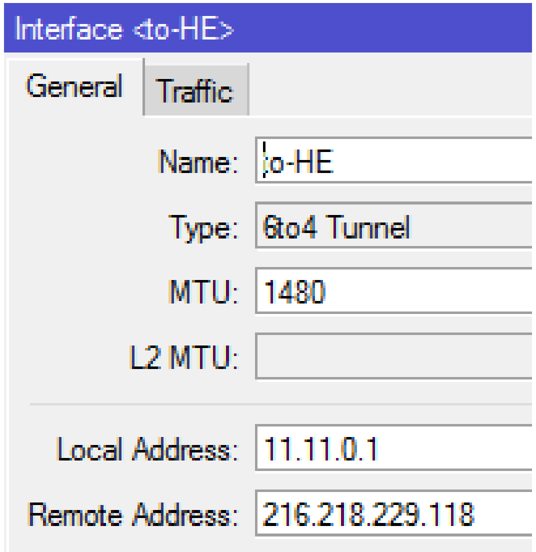
Break for hands on!

Establishing a peering to Cymru and putting routes in blackhole

What about IPv6?

Supposing our transit provider doesn’t supply native IPv6 connectivity, and we want to use this protocol, we can, via a Tunnel Broker, to be IPv6 worldwide connected.

Tunnel configuration



Interface <to-HE>

Name: to-HE

Type: 6to4 Tunnel

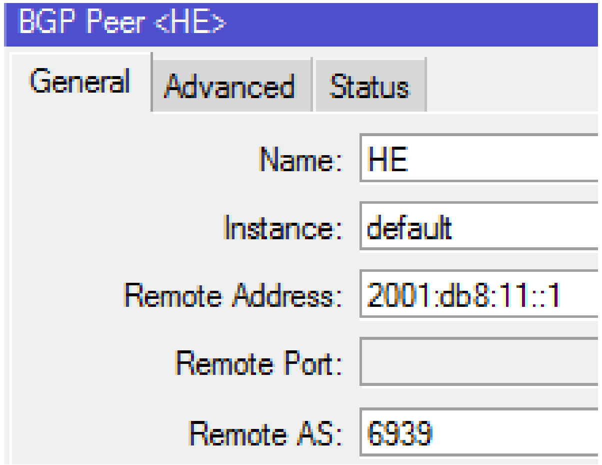
MTU: 1480

L2 MTU:

Local Address: 11.11.0.1

Remote Address: 216.218.229.118

BGP configuration



BGP Peer <HE>

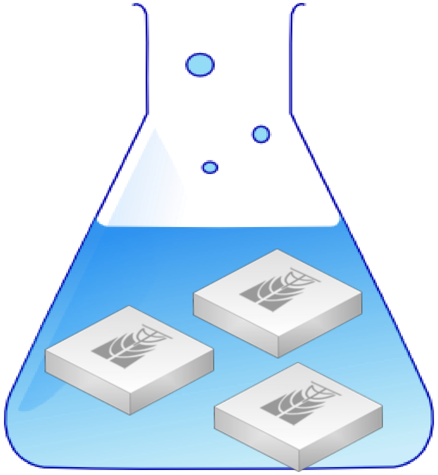
Name: HE

Instance: default

Remote Address: 2001:db8:11::1

Remote Port:

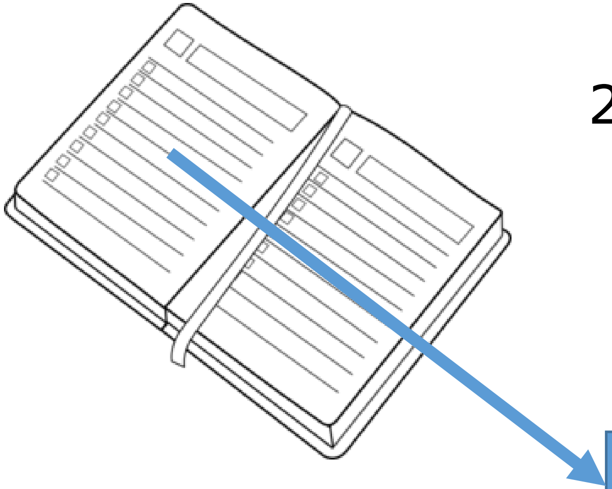
Remote AS: 6939



Break for hands on!

Establishing a IPv6 tunnel and receiving the routes
("Internet" → 2001:a::1)

Agenda



1) BGP essentials and basics of BGP filtering; ✓

2) Case Studies: ✓

2.1) Overview ✓

2.2) Single-Homed Provider ✓

2.3) Single-Homed + IXP

2.4) Multi-Homed + IXP

2.5) Multi-Homed + IXP + Providing transit services

What is an IXP?

IXP – Internet Exchange Point

(Or **NAP** – Network Exchange Point or **MAE** – Metropolitan Area Exchange)

Network solution whose purpose is to facilitate direct connections between Autonomous Systems, promoting the exchange of Internet traffic.

An IXP optimizes AS interconnection, allowing:

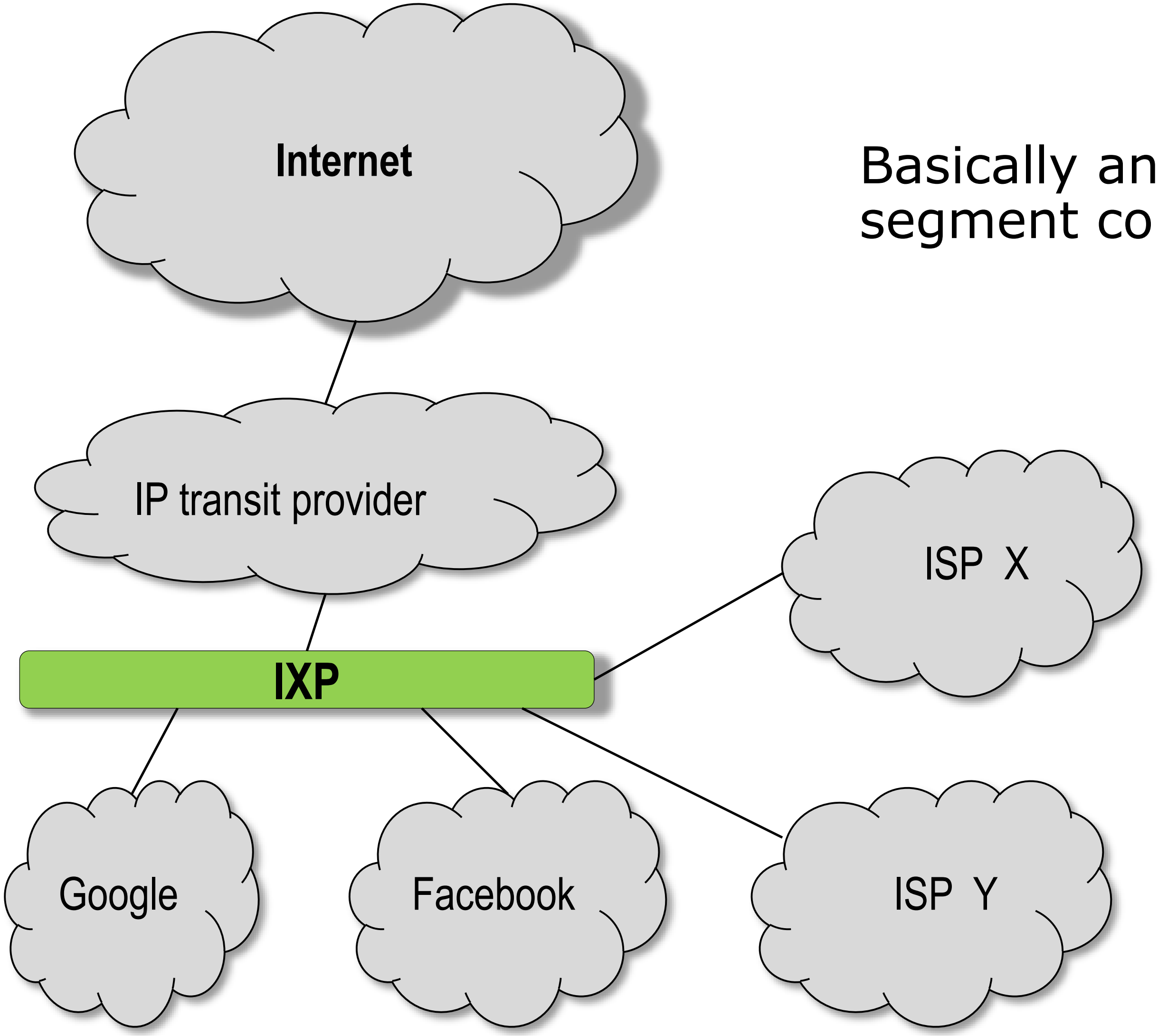
Better quality (low latency);

Avoid intermediates;

Lowering of costs (with a MLPA);

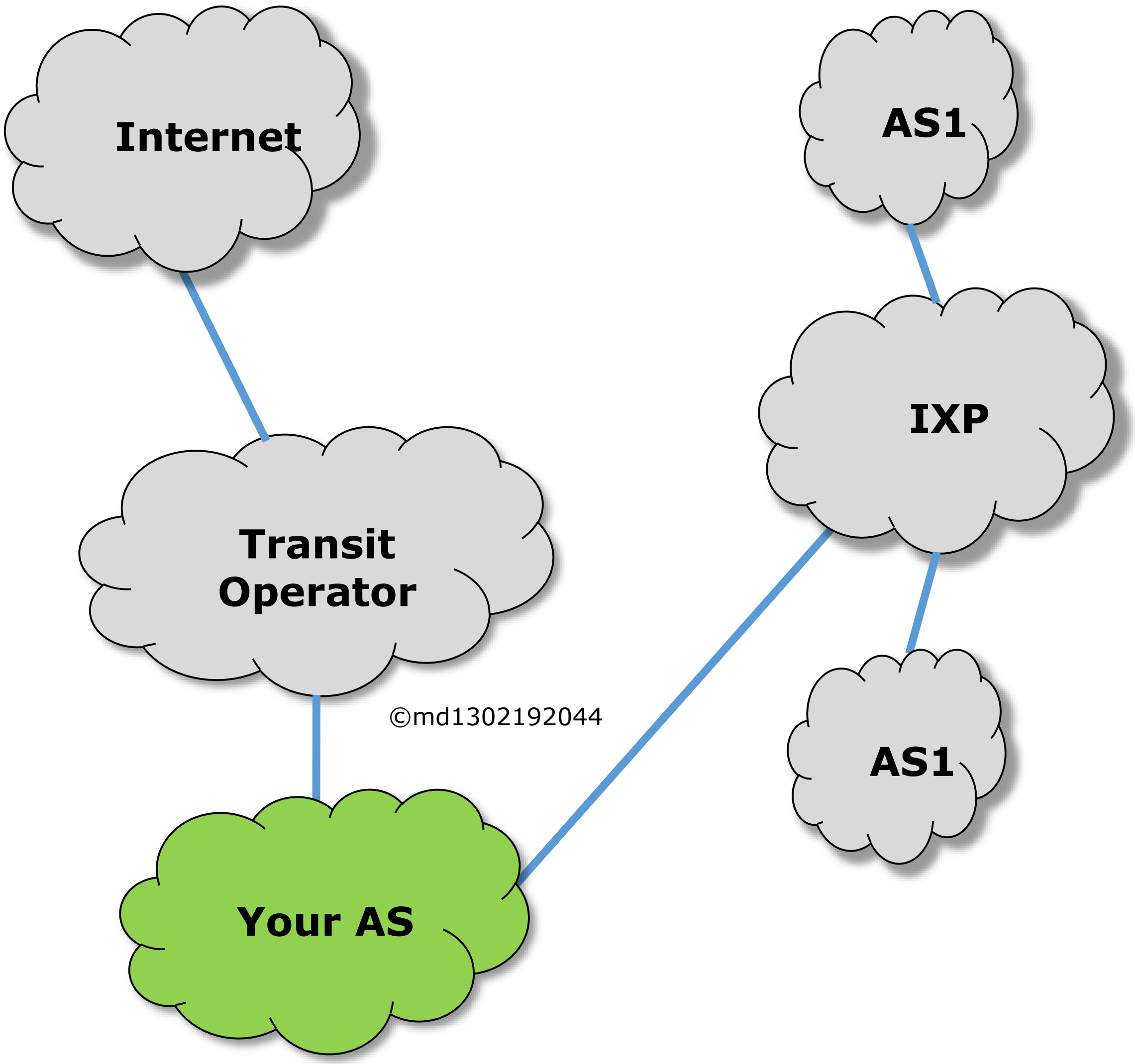
Better organization of regional networks.

Internet Exchange Point



Basically an IXP is a Layer2 segment connecting AS's

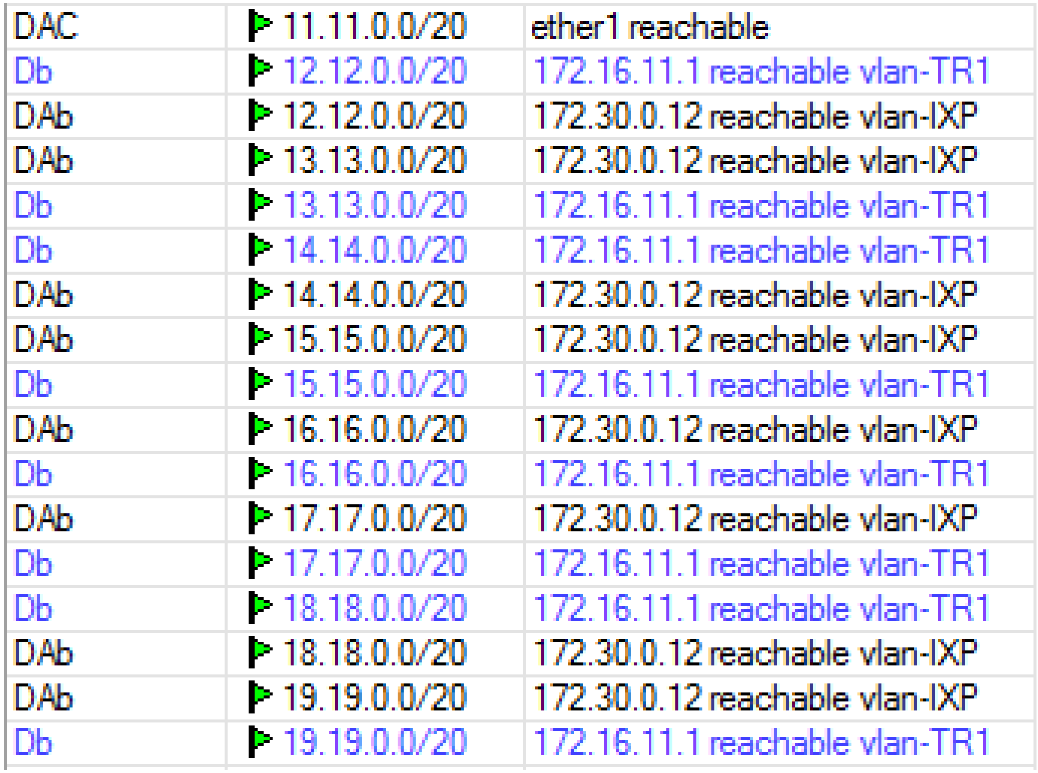
Scenario II
Single Homed + IXP



Scenario II – without filtering

12.12.0.0/20, 13.13.0.0/20, …, 19.19.0.0/20 are networks announced to IXP.

Note that without any filtering the IXP has "won" the election



DAC	11.11.0.0/20	ether1 reachable
Db	12.12.0.0/20	172.16.11.1 reachable vlan-TR1
DAb	12.12.0.0/20	172.30.0.12 reachable vlan-IXP
DAb	13.13.0.0/20	172.30.0.12 reachable vlan-IXP
Db	13.13.0.0/20	172.16.11.1 reachable vlan-TR1
Db	14.14.0.0/20	172.16.11.1 reachable vlan-TR1
DAb	14.14.0.0/20	172.30.0.12 reachable vlan-IXP
DAb	15.15.0.0/20	172.30.0.12 reachable vlan-IXP
Db	15.15.0.0/20	172.16.11.1 reachable vlan-TR1
DAb	16.16.0.0/20	172.30.0.12 reachable vlan-IXP
Db	16.16.0.0/20	172.16.11.1 reachable vlan-TR1
DAb	17.17.0.0/20	172.30.0.12 reachable vlan-IXP
Db	17.17.0.0/20	172.16.11.1 reachable vlan-TR1
Db	18.18.0.0/20	172.16.11.1 reachable vlan-TR1
DAb	18.18.0.0/20	172.30.0.12 reachable vlan-IXP
DAb	19.19.0.0/20	172.30.0.12 reachable vlan-IXP
Db	19.19.0.0/20	172.16.11.1 reachable vlan-TR1

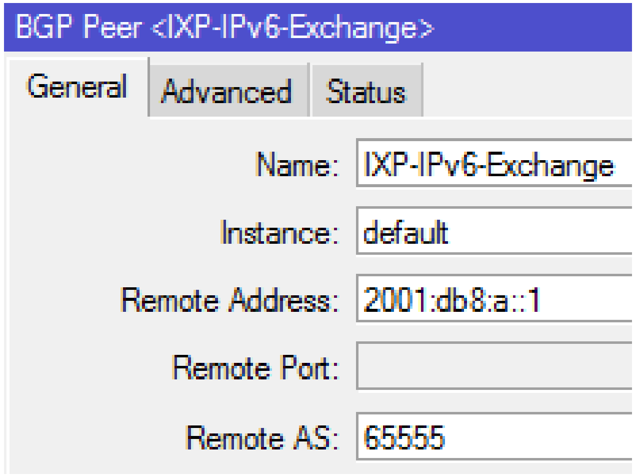
Note that we also have 2 destinations to the same network.

Why?

What about IPv6?

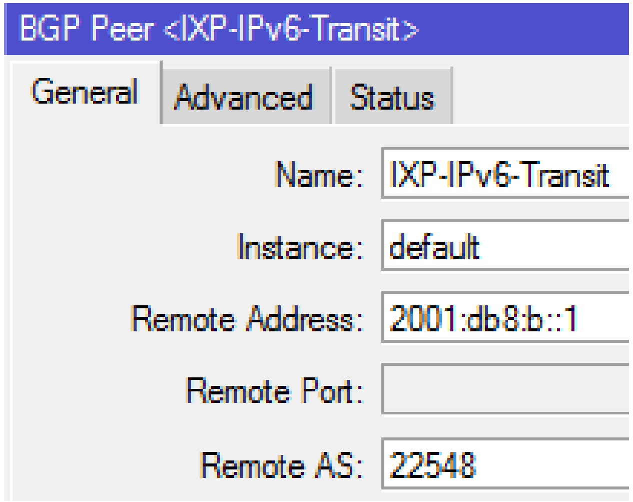
In our IXP we have native IPv6 transit to the Internet and we will use this as preferred path to IPv6 world keeping the tunnel to HE as a backup.

IPv6 exchange peering

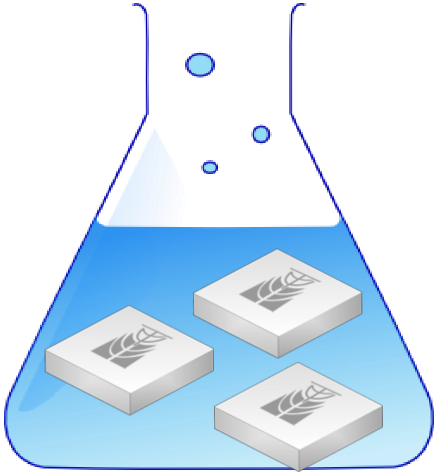

BGP Peer <IXP-IPv6-Exchange>

Name: IXP-IPv6-Exchange
Instance: default
Remote Address: 2001:db8:a::1
Remote Port:
Remote AS: 65555

IPv6 transit peering


BGP Peer <IXP-IPv6-Transit>

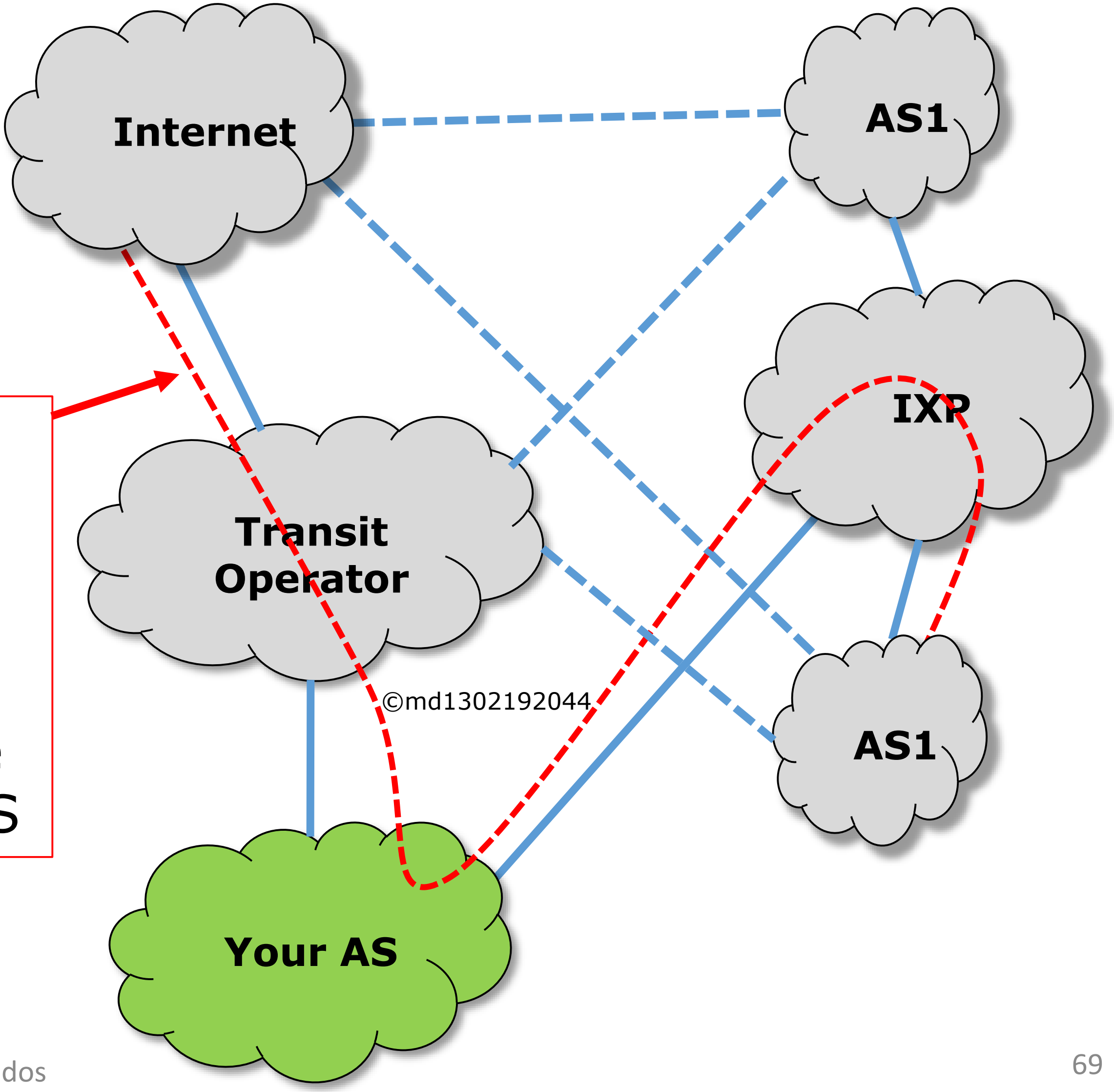
Name: IXP-IPv6-Transit
Instance: default
Remote Address: 2001:db8:b::1
Remote Port:
Remote AS: 22548



Break for hands on!

Establishing the peering with IXP for: IPv4 exchange, IPv6 exchange and IPv6 transit.

Scenario II
Single Homed + IXP

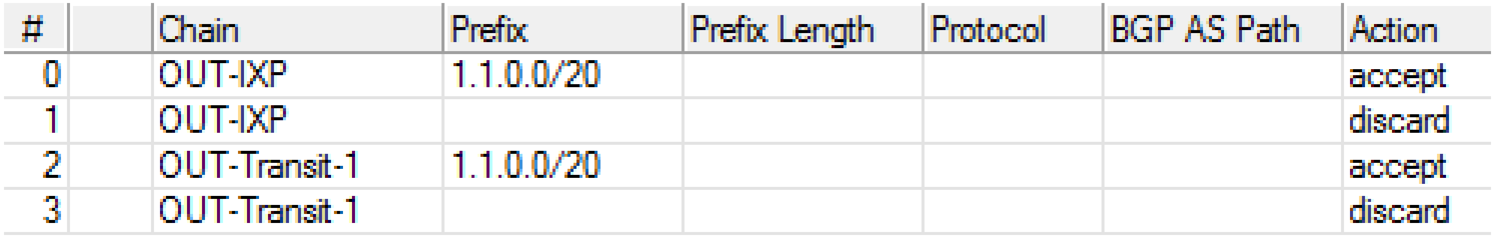


Transit Effect (undesired)

Without filtering AS-1 could decide that the best path to go to the Internet is via Your AS

Transit effect in Scenario II

To protect against undesirable “transit effect” your AS should advertise only its own prefixes.



#		Chain	Prefix	Prefix Length	Protocol	BGP AS Path	Action
0		OUT-IXP	1.1.0.0/20				accept
1		OUT-IXP					discard
2		OUT-Transit-1	1.1.0.0/20				accept
3		OUT-Transit-1					discard

Above filters applied to peers IXP and Transit-1 in out-filter channel

More Filtering – Ingress good Practices

Good practices for ingress filters for all peers are:

→ Discard receiving own prefix;

→ Discard private and reserved networks stated at RFC 5735;

→ Discard default route (we are assuming a Full Routing)

QUIZZ

Is necessary to discard routes that contain own AS number in the AS-Path?

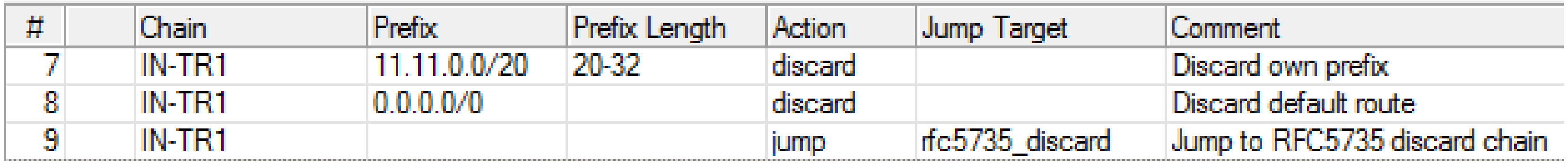
RFC 5735 – Summary Table 1/2

Address Block	Present Use	Reference
0.0.0.0/8	"This" Network	RFC 1122
10.0.0.0/8	Private-Use Networks	RFC 1918
127.0.0.0/8	Loopback	RFC 1122
169.254.0.0/16	Link Local	RFC 3927
172.16.0.0/12	Private-Use Networks	RFC 1918
192.0.0.0/24	IETF Protocol Assignments	RFC 5736
192.0.2.0/24	TEST-NET-1	RFC 5737
192.88.99.0/24	6to4 Relay Anycast	RFC 3068

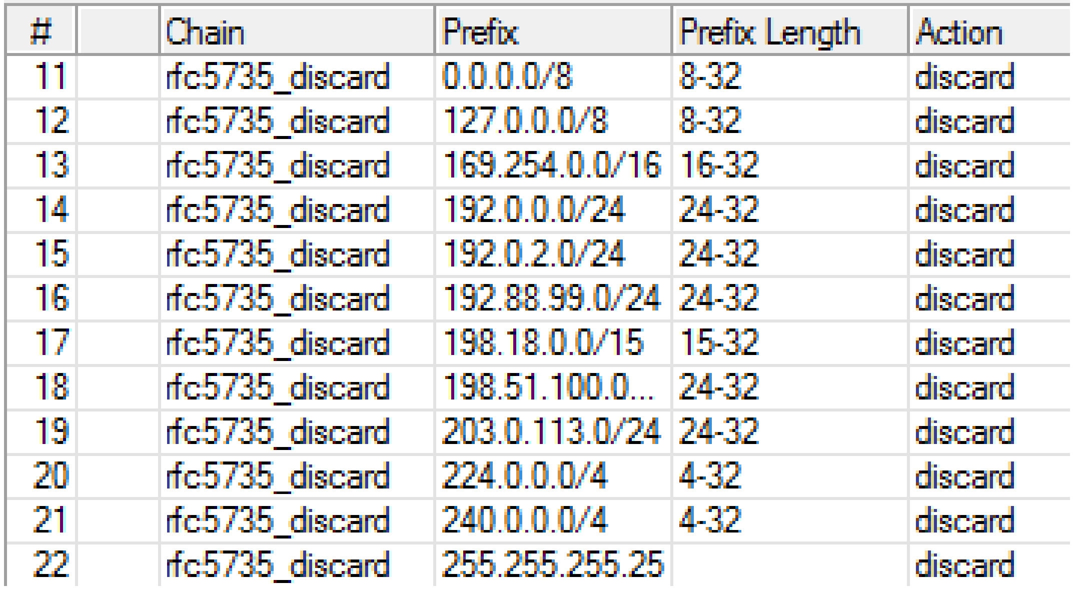
RFC 5735 – Summary Table 2/2

Address Block	Present Use	Reference
192.168.0.0/16	Private-Use Networks	RFC 1918
198.18.0.0/15	Device Benchmark Testing	RFC 2544
198.51.100.0/24	TEST-NET-2	RFC 5737
203.0.113.0/24	TEST-NET-3	RFC 5737
224.0.0.0/4	Multicast	RFC 3171
240.0.0.0/4	Reserved for Future Use	RFC 1112
255.255.255.255/32	Limited Broadcast	RFC 919 RFC 922

Ingress Filters for (almost) All Peers



#		Chain	Prefix	Prefix Length	Action	Jump Target	Comment
7		IN-TR1	11.11.0.0/20	20-32	discard		Discard own prefix
8		IN-TR1	0.0.0.0/0		discard		Discard default route
9		IN-TR1			jump	rfc5735_discard	Jump to RFC5735 discard chain

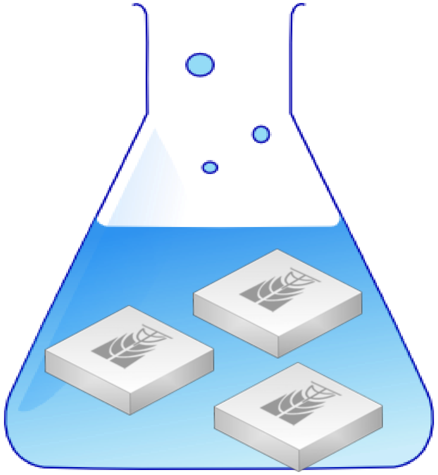


#		Chain	Prefix	Prefix Length	Action
11		rfc5735_discard	0.0.0.0/8	8-32	discard
12		rfc5735_discard	127.0.0.0/8	8-32	discard
13		rfc5735_discard	169.254.0.0/16	16-32	discard
14		rfc5735_discard	192.0.0.0/24	24-32	discard
15		rfc5735_discard	192.0.2.0/24	24-32	discard
16		rfc5735_discard	192.88.99.0/24	24-32	discard
17		rfc5735_discard	198.18.0.0/15	15-32	discard
18		rfc5735_discard	198.51.100.0...	24-32	discard
19		rfc5735_discard	203.0.113.0/24	24-32	discard
20		rfc5735_discard	224.0.0.0/4	4-32	discard
21		rfc5735_discard	240.0.0.0/4	4-32	discard
22		rfc5735_discard	255.255.255.25		discard

N.B: Private networks suppressed from this list because we're using them.

Hint:

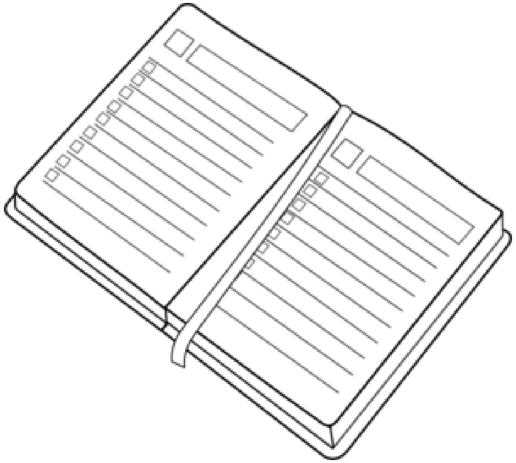
Action Jump can turn your filters more readable!



Break for hands on!

Enable protection filters for undesired transit effect and good practices ingress filters

Agenda



1) BGP essentials and basics of BGP filtering; ✓

2) Case Studies: ✓

   2.1) Overview ✓

   2.2) Single-Homed Provider ✓

   2.3) Single-Homed + IXP ✓

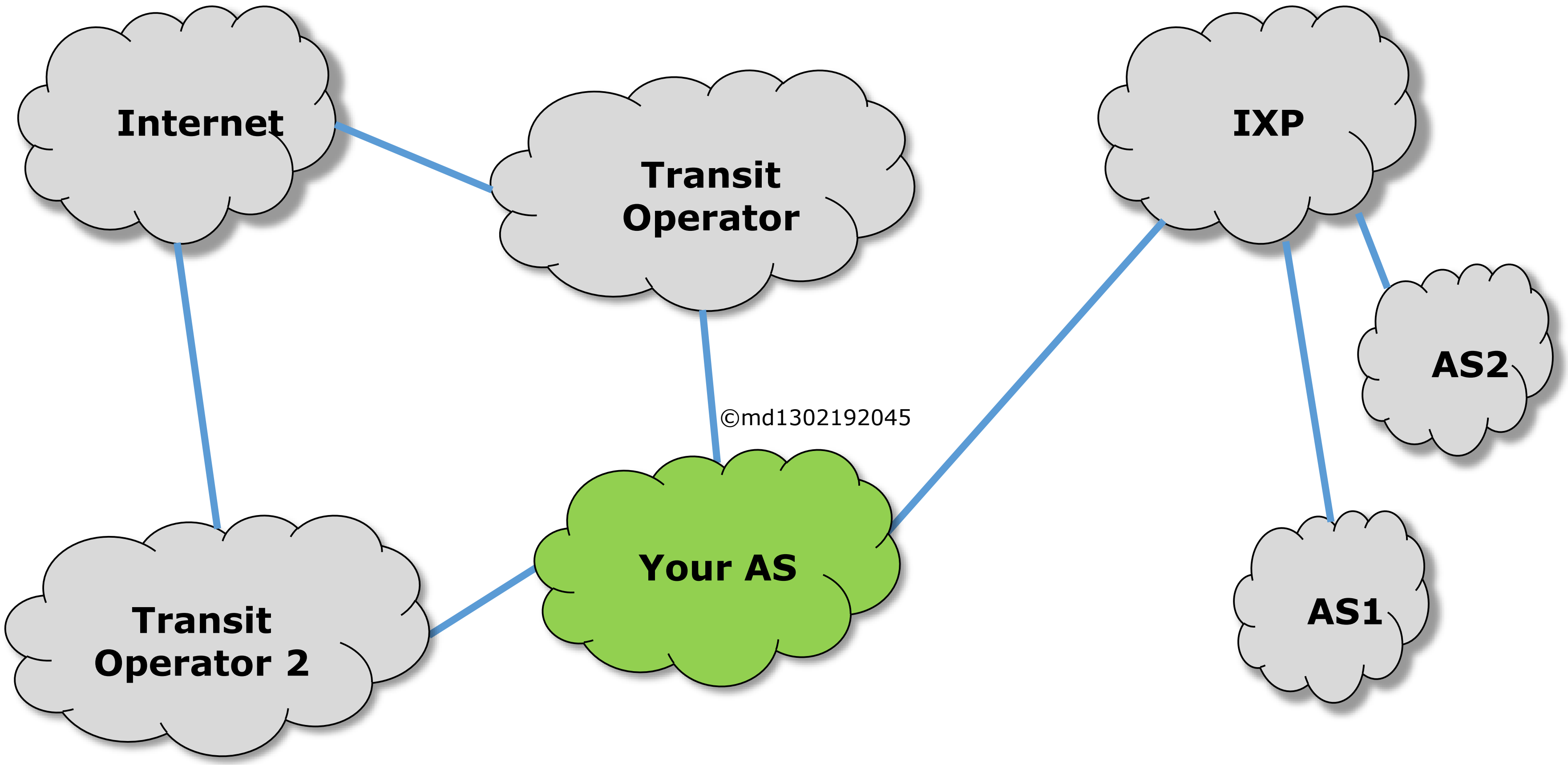
   2.4) Multi-Homed + IXP

   2.5) Multi-Homed + IXP + Providing transit services

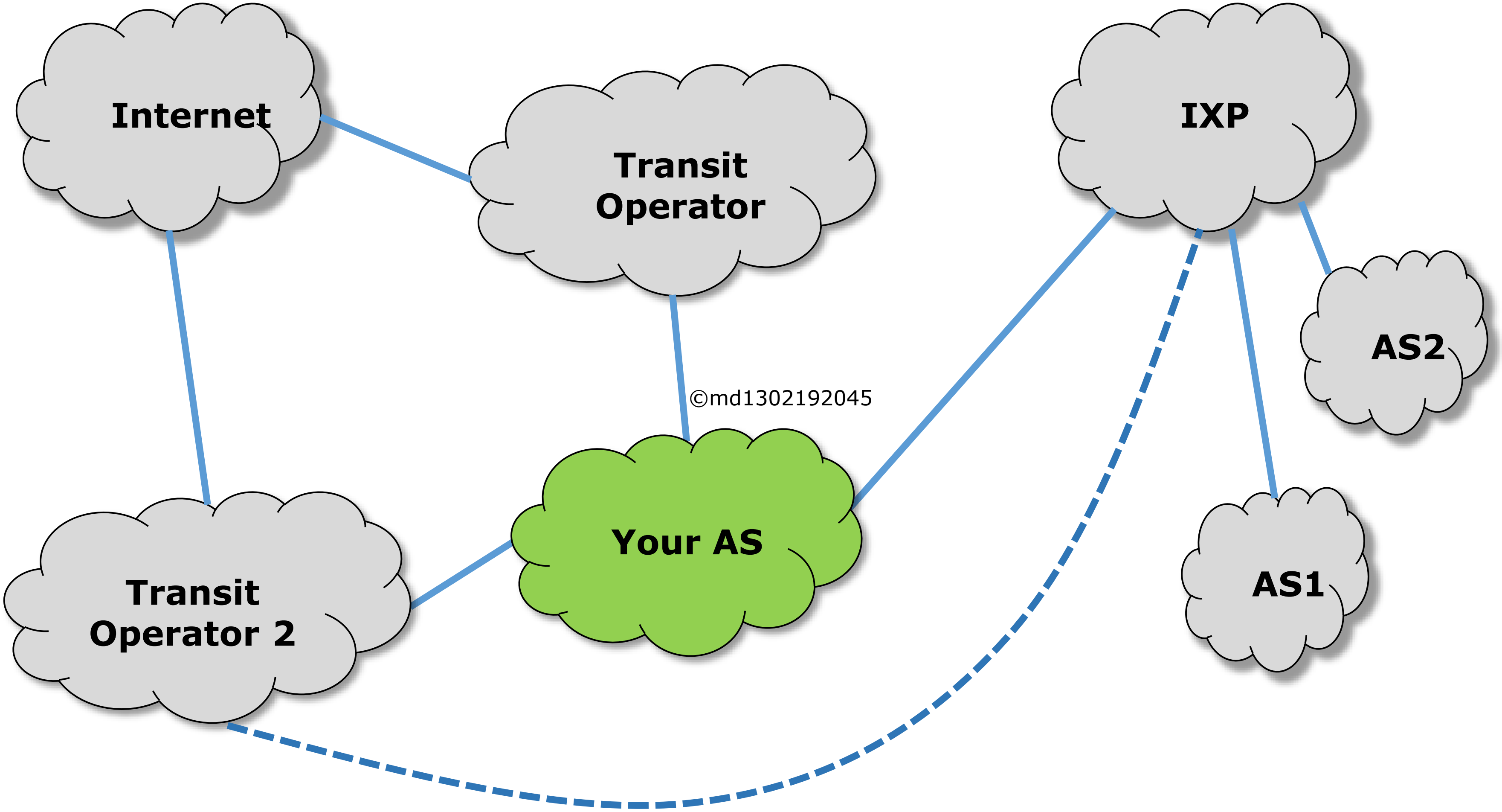
Scenario III

Multi-Homed + IXP

Scenario III
Multi-Homed + IXP

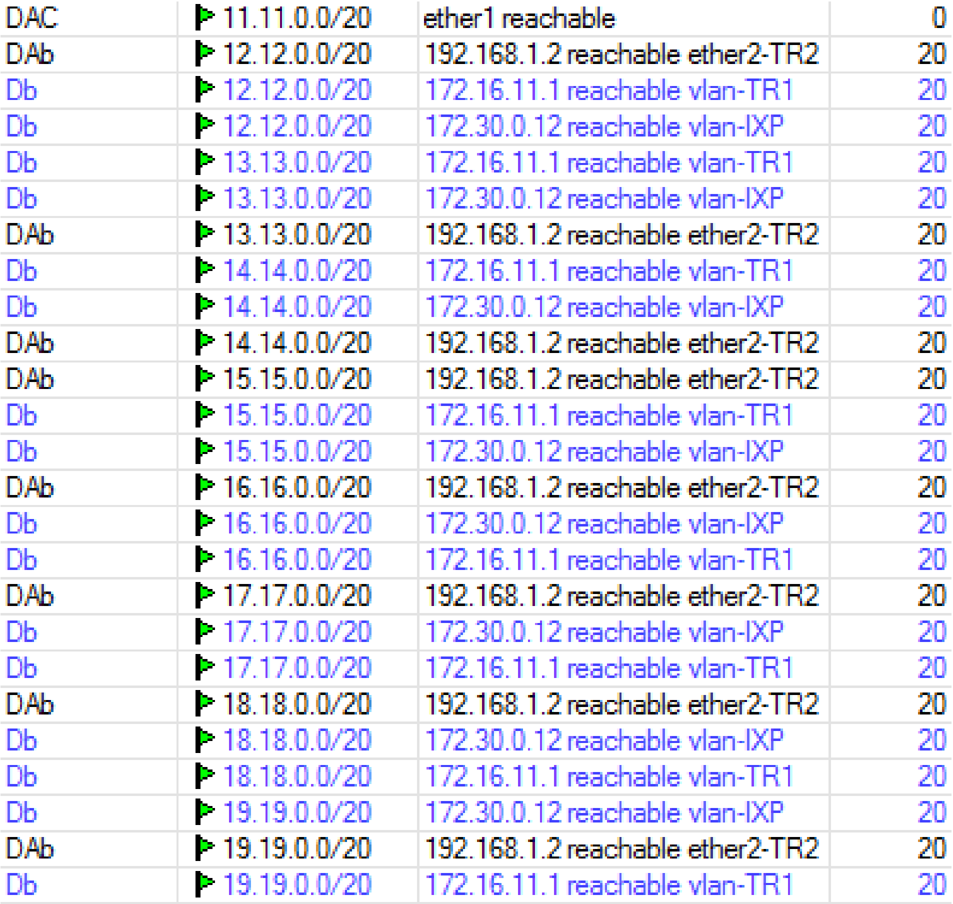


Scenario III
Multi-Homed + IXP



Scenario III without filtering

12.12.0.0/20, 13.13.0.0/20, …, 19.19.0.0/20 are networks belonging to TR-2 and announced to IXP and TR-1

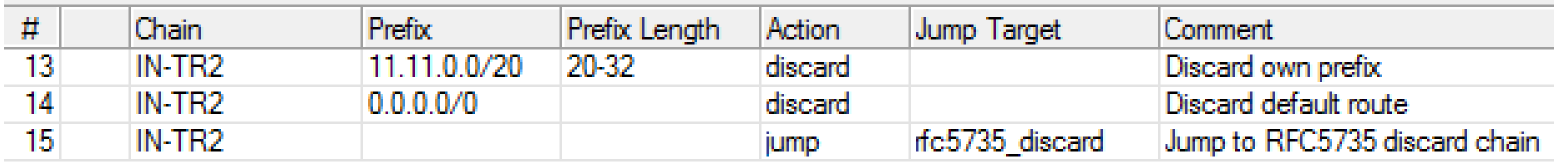


DAC	11.11.0.0/20	ether1 reachable	0
DAb	12.12.0.0/20	192.168.1.2 reachable ether2-TR2	20
Db	12.12.0.0/20	172.16.11.1 reachable vlan-TR1	20
Db	12.12.0.0/20	172.30.0.12 reachable vlan-IXP	20
Db	13.13.0.0/20	172.16.11.1 reachable vlan-TR1	20
Db	13.13.0.0/20	172.30.0.12 reachable vlan-IXP	20
DAb	13.13.0.0/20	192.168.1.2 reachable ether2-TR2	20
Db	14.14.0.0/20	172.16.11.1 reachable vlan-TR1	20
Db	14.14.0.0/20	172.30.0.12 reachable vlan-IXP	20
DAb	14.14.0.0/20	192.168.1.2 reachable ether2-TR2	20
DAb	15.15.0.0/20	192.168.1.2 reachable ether2-TR2	20
Db	15.15.0.0/20	172.16.11.1 reachable vlan-TR1	20
Db	15.15.0.0/20	172.30.0.12 reachable vlan-IXP	20
DAb	16.16.0.0/20	192.168.1.2 reachable ether2-TR2	20
Db	16.16.0.0/20	172.30.0.12 reachable vlan-IXP	20
Db	16.16.0.0/20	172.16.11.1 reachable vlan-TR1	20
DAb	17.17.0.0/20	192.168.1.2 reachable ether2-TR2	20
Db	17.17.0.0/20	172.30.0.12 reachable vlan-IXP	20
Db	17.17.0.0/20	172.16.11.1 reachable vlan-TR1	20
DAb	18.18.0.0/20	192.168.1.2 reachable ether2-TR2	20
Db	18.18.0.0/20	172.30.0.12 reachable vlan-IXP	20
Db	18.18.0.0/20	172.16.11.1 reachable vlan-TR1	20
Db	19.19.0.0/20	172.30.0.12 reachable vlan-IXP	20
DAb	19.19.0.0/20	192.168.1.2 reachable ether2-TR2	20
Db	19.19.0.0/20	172.16.11.1 reachable vlan-TR1	20

Note that we have a direct path and 2 other options

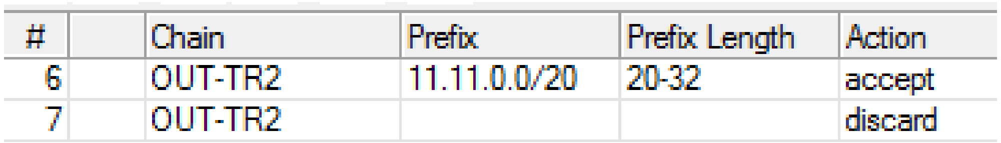
Filtering for Scenario III

1) Ingress Filters will be the same:



#		Chain	Prefix	Prefix Length	Action	Jump Target	Comment
13		IN-TR2	11.11.0.0/20	20-32	discard		Discard own prefix
14		IN-TR2	0.0.0.0/0		discard		Discard default route
15		IN-TR2			jump	rfc5735_discard	Jump to RFC5735 discard chain

2) Filters to avoid undesired traffic effect, as well



#		Chain	Prefix	Prefix Length	Action
6		OUT-TR2	11.11.0.0/20	20-32	accept
7		OUT-TR2			discard

What about filters to manipulate traffic?

Traffic Manipulation

Routing Filters

The way to influence BGP decision is by configuring routing filters.

Filtering **incoming** routes can change, how we see the external world, thus influencing how we **send** traffic;

Filtering **outgoing** routes can change how the world see us, thus influencing how we **receive** traffic.

Traffic Manipulation

How to check results?

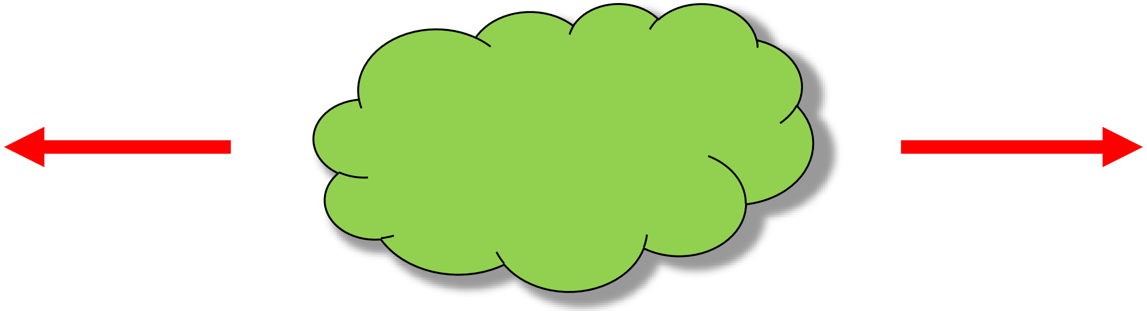
1) Tools that don't tell all the true:

Ping, traceroute, torch, bandwidth test…

2) Where should we see:

Results of our upload policy: **Our routing table**

Results of our download policy: **Our routes as seen by other AS's (looking glasses)**



Upload Control

©Copyright md brasil - direitos reservados

Upload Manipulation

To influence our upload, basically we can manipulate 2 attributes:

→**Weight**

→**Local-Preference**

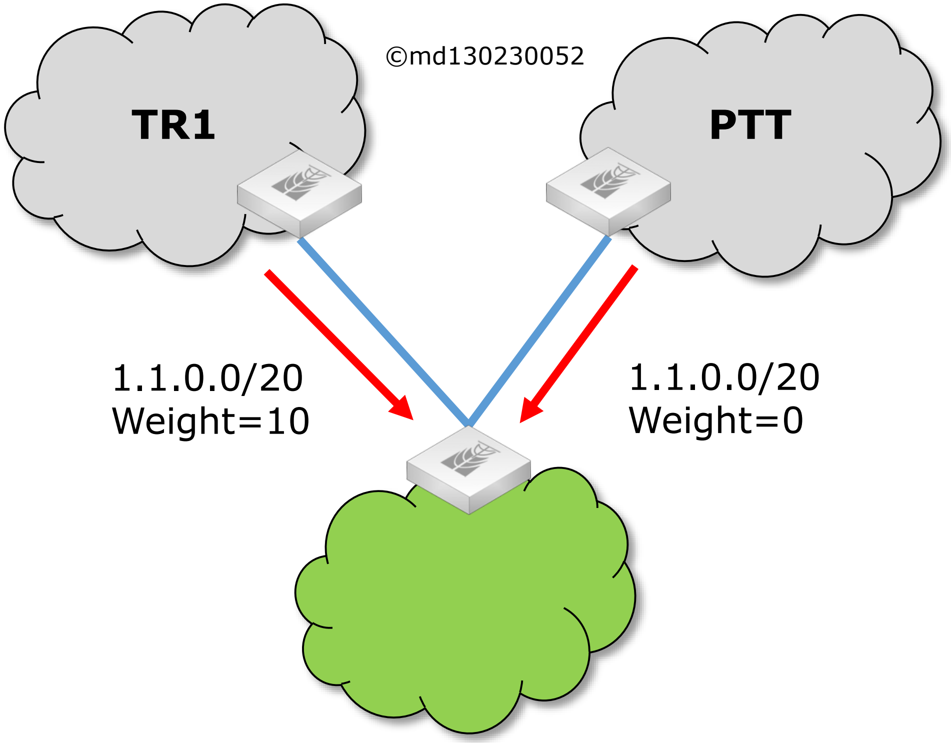
Both will cause the same effect if we have a single router.

Upload Manipulation Weight

Weight

Filters can set a “weight” to the route received from one peer. Routes with higher weight will be preferred (Default =0)

OBS: Although weight is usually treated as a BGP attribute, in fact is not, because it is not propagated inside the update messages.

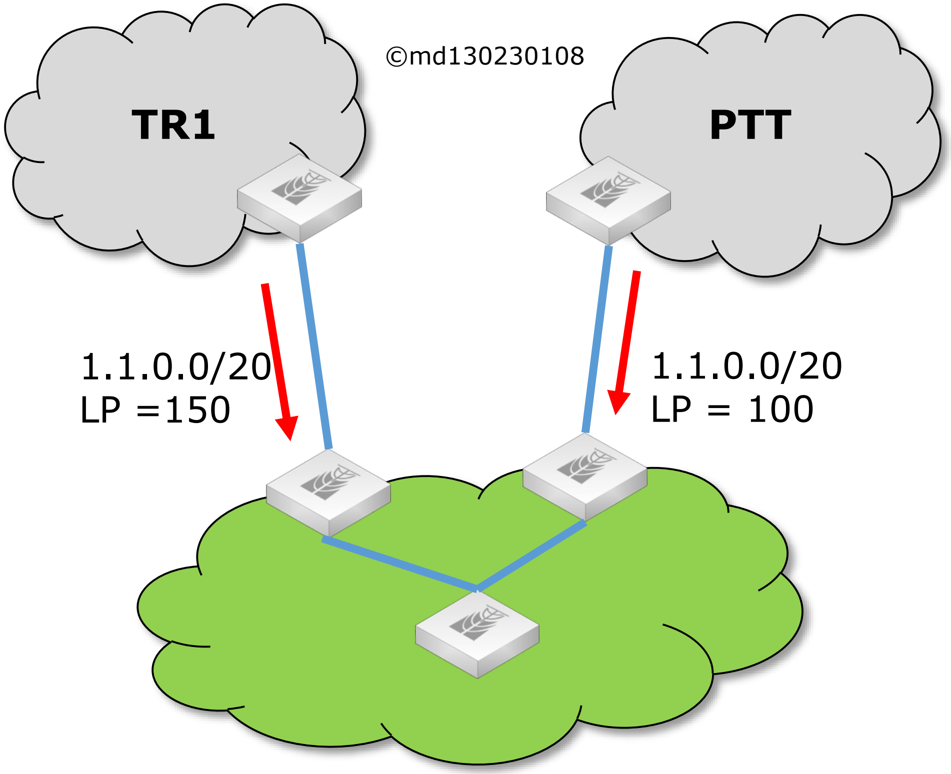


Upload Manipulation Local-Preference

Local-Preference

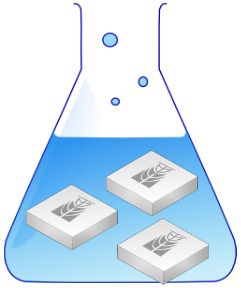
Filters can set a Local-Preference to the route(s) received from one peer. Routes with higher LP, will be preferred to send traffic. Default LP is 100.

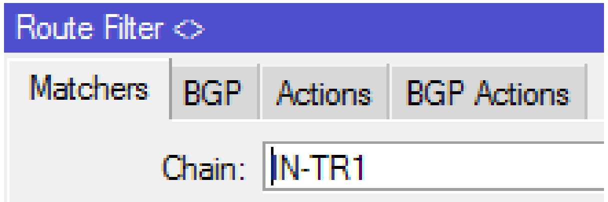
OBS: Local Preference is a real attribute that propagates inside the entire AS. Does not propagate to other AS's.



Upload Manipulation Weight or Local preference

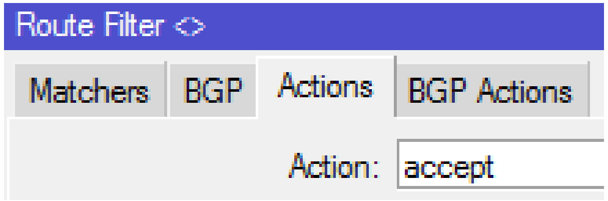
Natural upload preference is via TR2. Filter to set TR1 as the preferred path:





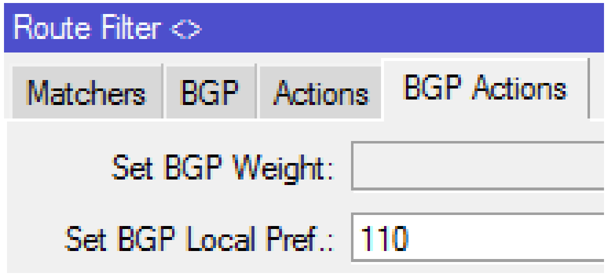
Route Filter <>

Chain: IN-TR1



Route Filter <>

Action: accept

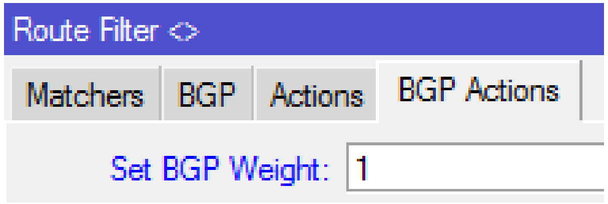


Route Filter <>

Set BGP Weight:

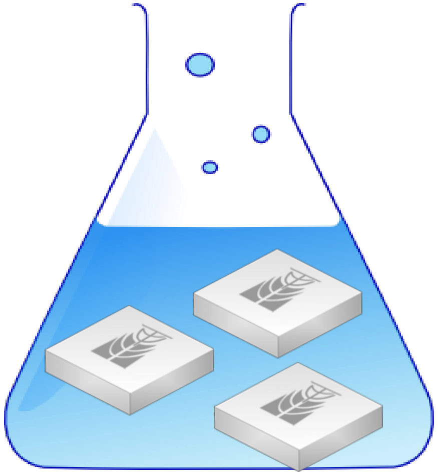
Set BGP Local Pref.: 110

or



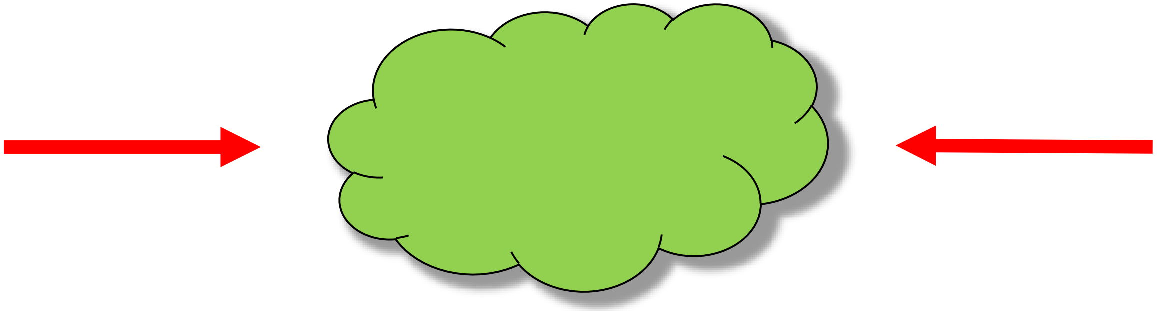
Route Filter <>

Set BGP Weight: 1



Break for hands on!

Enable Local Preference filter and show the effect on routing table



Download Control

Download Manipulation

Basically there are 3 ways to influence how downloads are received by our AS:

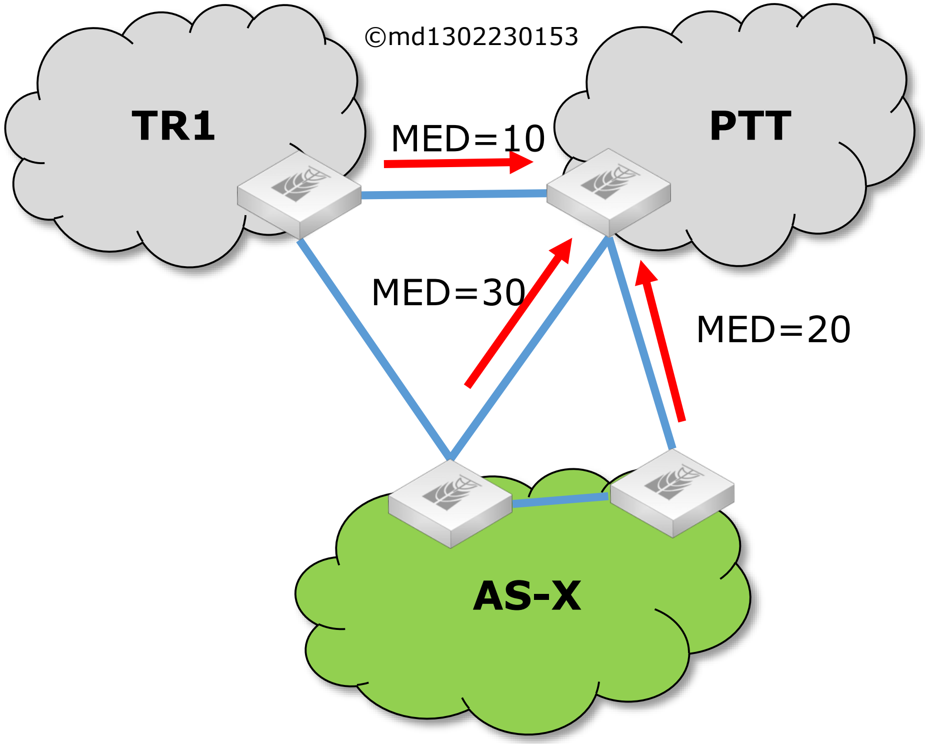
→ **Controlling network advertisements with longer or shorter prefixes;**

→ **Manipulating AS-Path attribute;**

→ **Manipulating MED attribute;**

Download Manipulation with MED attribute

With **MED** (Multi Exit Discriminator) one AS can inform a neighbor one, which is the preferred way to receive traffic. Lower MED will be used (default=0);



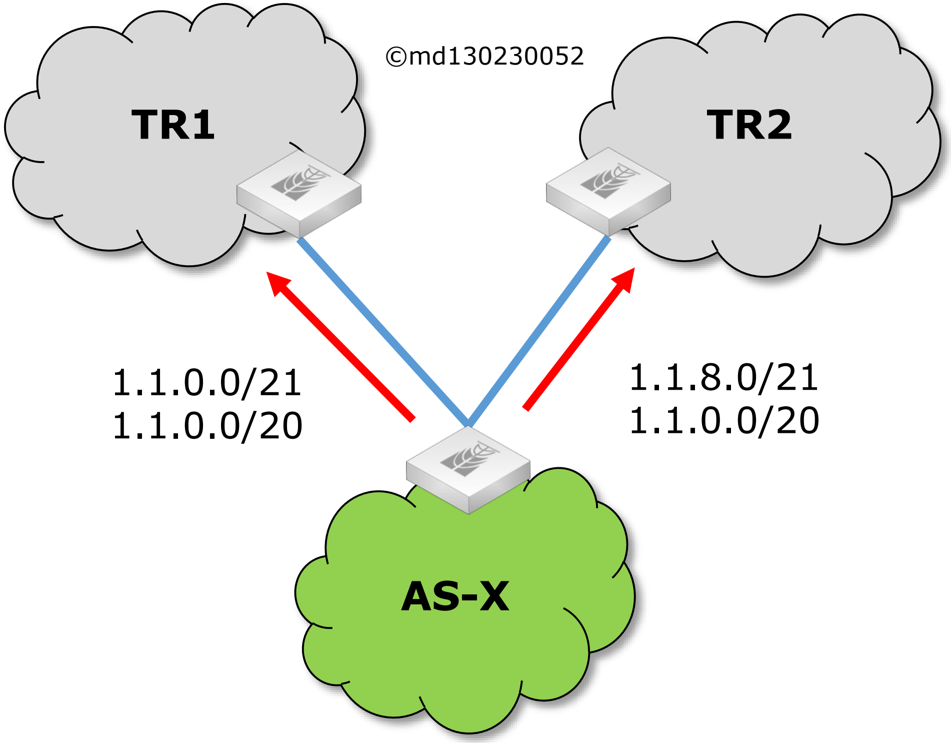
With RouterOS, MED will work only when there are **two ore more connections** between AS's.

NB: In a scenario like the picture, TR1 MED will be ignored

Download Manipulation by more specific announcements

e.g.:

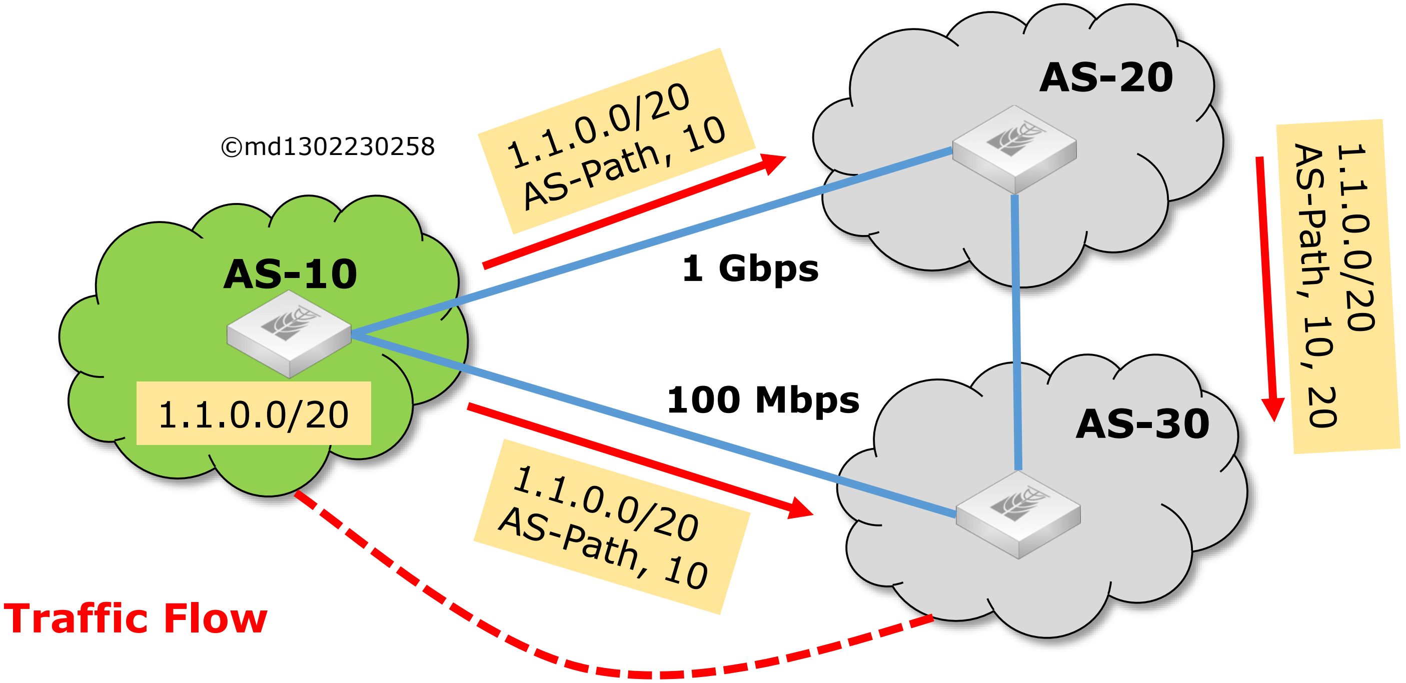
AS-x announces half of its addresses for each link and the whole IP range for both links. The goal is to "guarantee" the balance and redundancy.



OBS: This policy will succeed only if the use of IP's are quite equilibrated.

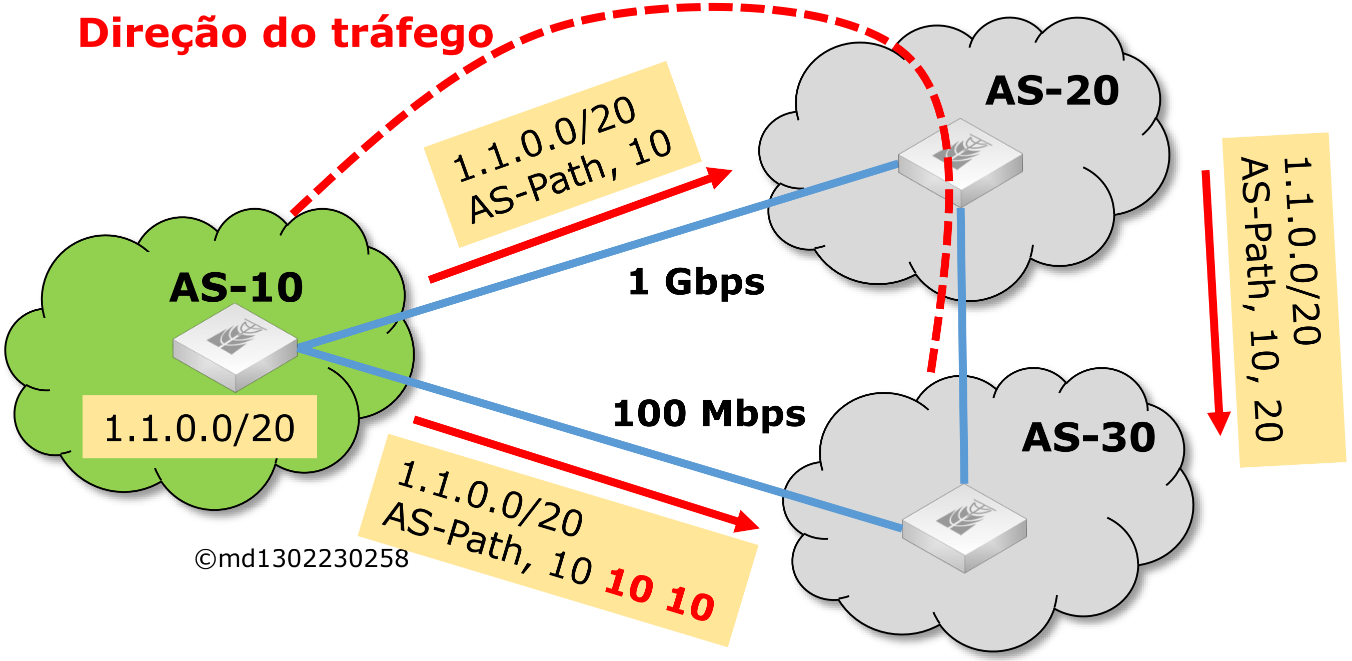
Download Manipulation
AS-Path prepend technique

Example: before prepending



Download Manipulation AS-Path prepend technique

Prepending 3 times self AS



©md1302230258

Download Manipulation

Comparing the methods:

MED:

Efficient, but limited when having 2 or more connections to the same AS;

More specific announcements:

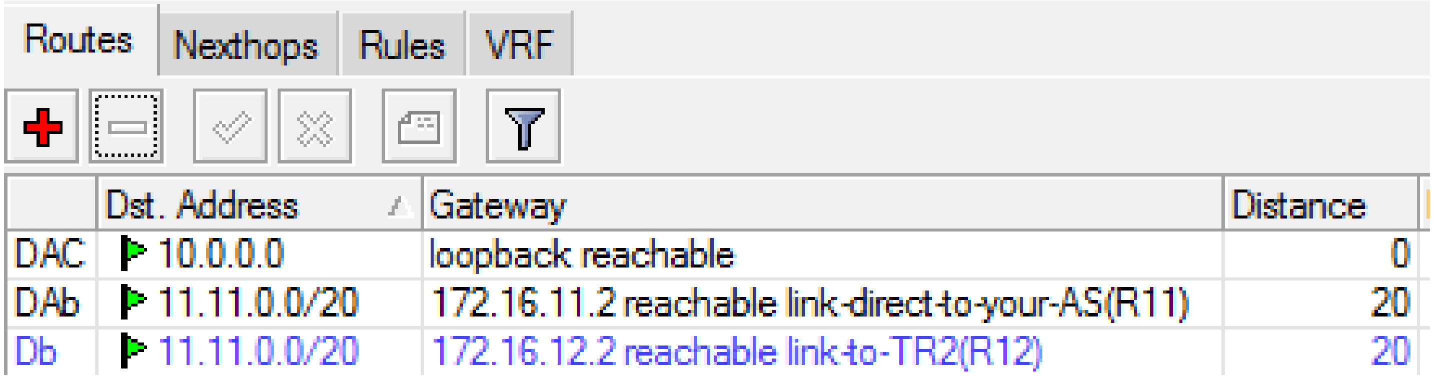
Aggressive resource. Will work regardless the number of hops. Can choose sub-optimal paths. Use in extreme cases.

AS-Path prepend:

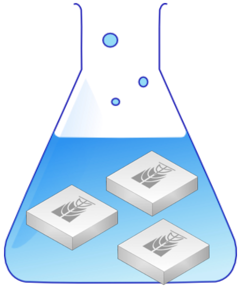
“Soft” resource. Also has limitations due to number of AS-Path’s and topology changes.

Download Manipulation with AS-Path prepend

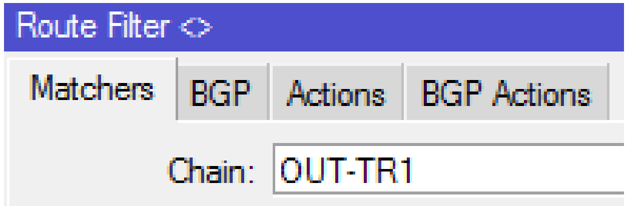
TR1 Routing table (looking glass) before AS-Path prepend



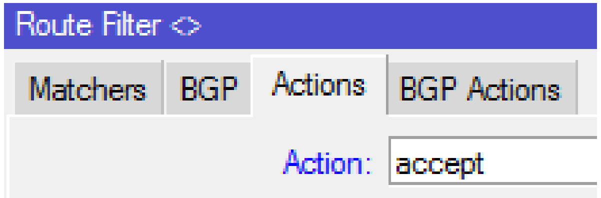
	Dst. Address	Gateway	Distance
DAC	10.0.0.0	loopback reachable	0
DAb	11.11.0.0/20	172.16.11.2 reachable link-direct-to-your-AS(R11)	20
Db	11.11.0.0/20	172.16.12.2 reachable link-to-TR2(R12)	20



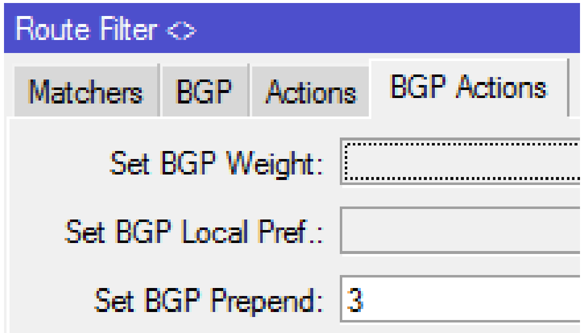
Filters:



Chain: OUT-TR1



Action: accept



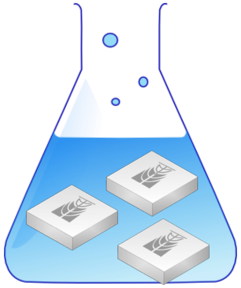
Set BGP Weight:

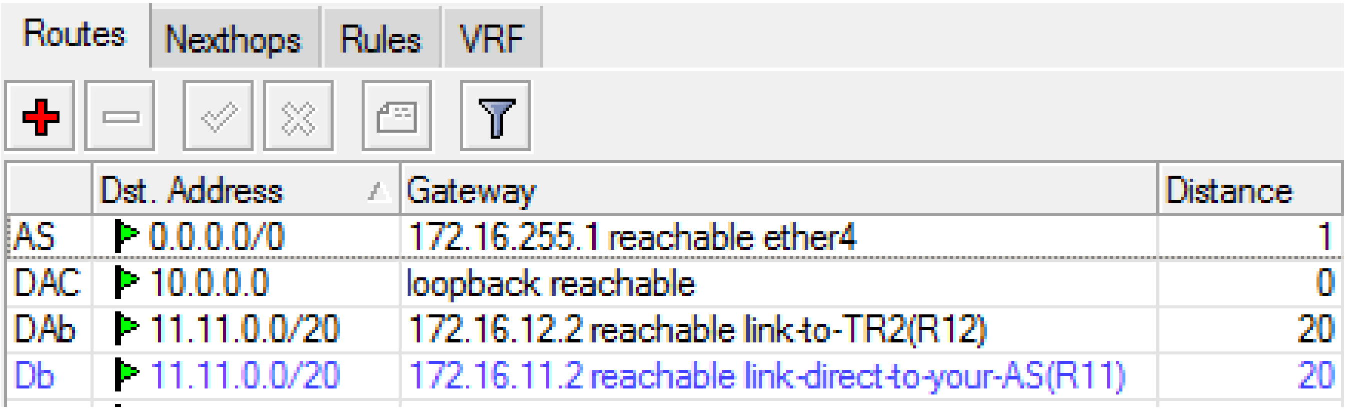
Set BGP Local Pref.:

Set BGP Prepend: 3

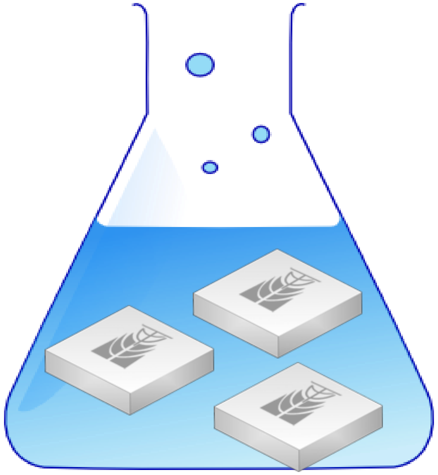
Download Manipulation with AS-Path prepend

TR1 Routing table (looking glass) after AS-Path prepend:





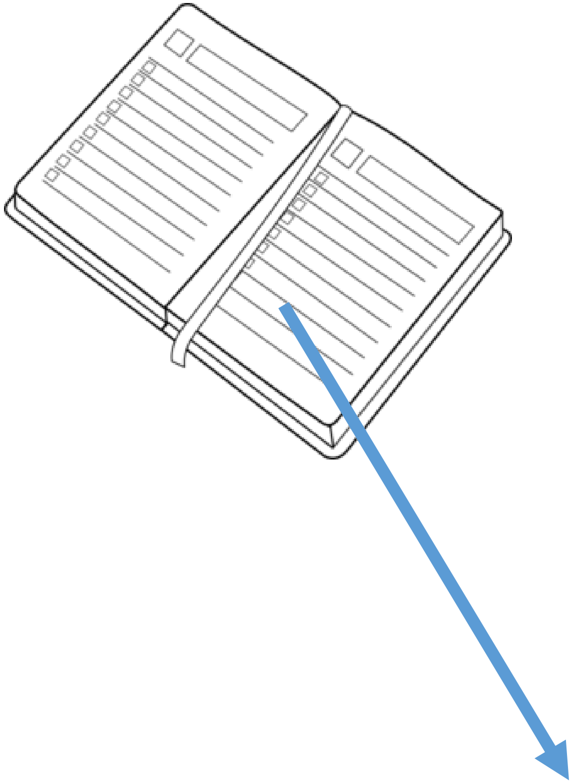
	Dst. Address	Gateway	Distance
AS	0.0.0.0/0	172.16.255.1 reachable ether4	1
DAC	10.0.0.0	loopback reachable	0
DAb	11.11.0.0/20	172.16.12.2 reachable link-to-TR2(R12)	20
Db	11.11.0.0/20	172.16.11.2 reachable link-direct-to-your-AS(R11)	20



Break for hands on!

Enable AS-Path prepend filter and show the results on the "looking glass"

Agenda

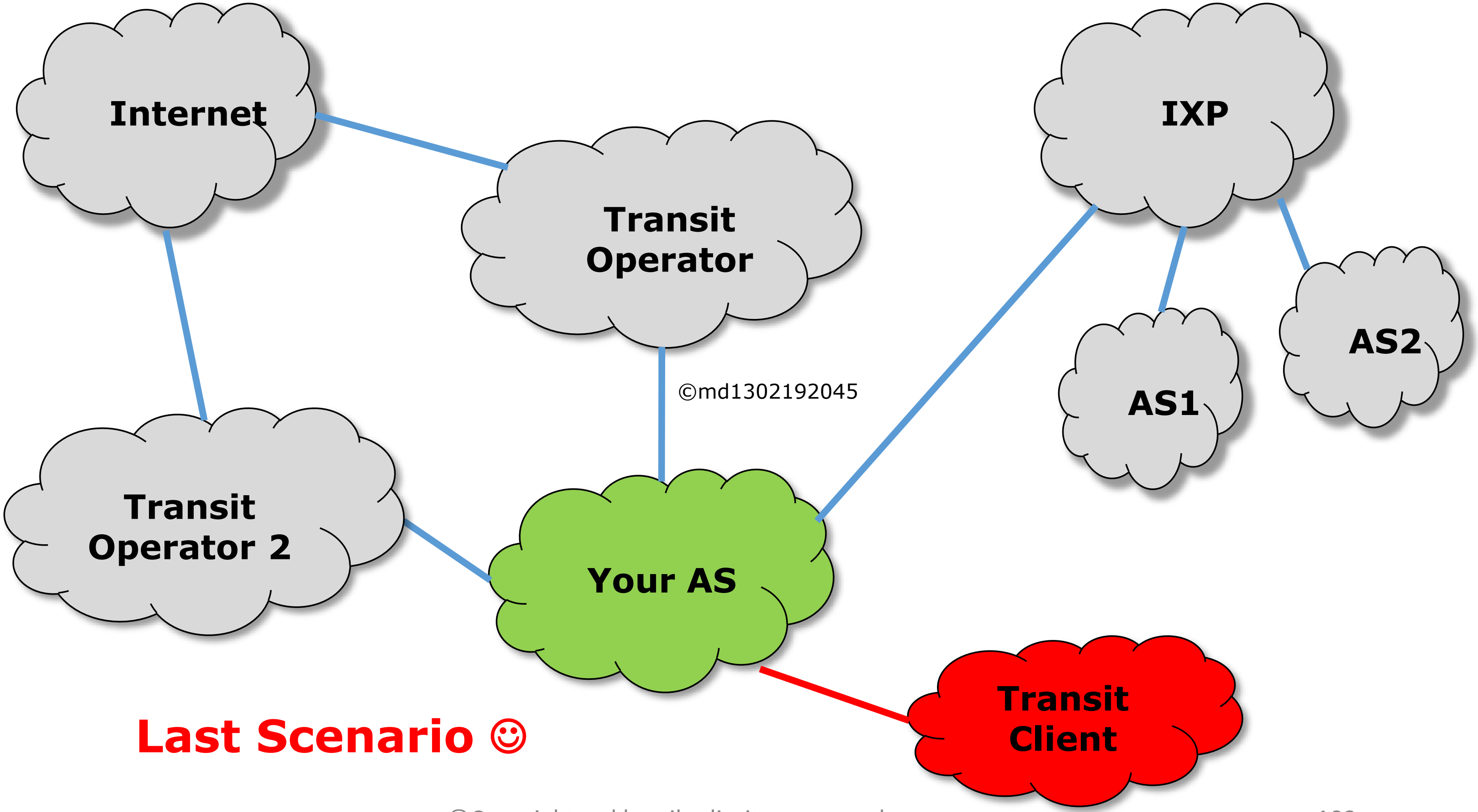


1) BGP essentials and basics of BGP filtering; ✓

2) Case Studies: ✓

- 2.1) Overview ✓
- 2.2) Single-Homed Provider ✓
- 2.3) Single-Homed + IXP ✓
- 2.4) Multi-Homed + IXP ✓
- 2.5) Multi-Homed + IXP + Providing transit services

Scenario IV
Dual Homed + IXP
providing transit services



Last Scenario ☺

Becoming a Transit AS

Supposing the agreement with our customer has the following statements:

→ He will announce prefix 200.0.0.0/20;

→ His AS number is 200 and we'll allow them to make any number of prepends;

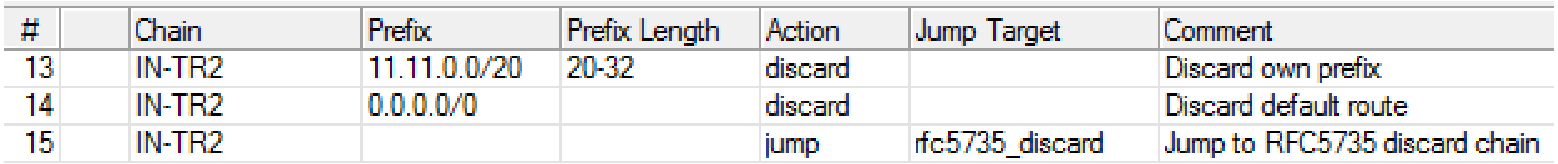
→ He is not transit to any other provider;

→ We'll offer him native IPv6 transit.

Filtering for Scenario IV

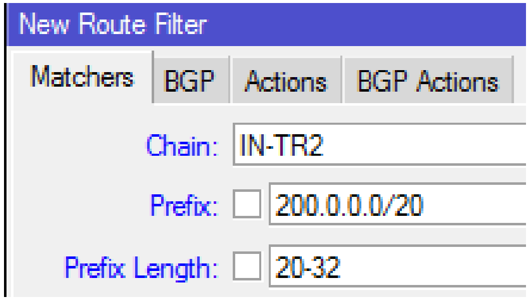
1) Ingress Filters:

The same for discarding default route and own prefix:



#		Chain	Prefix	Prefix Length	Action	Jump Target	Comment
13		IN-TR2	11.11.0.0/20	20-32	discard		Discard own prefix
14		IN-TR2	0.0.0.0/0		discard		Discard default route
15		IN-TR2			jump	rfc5735_discard	Jump to RFC5735 discard chain

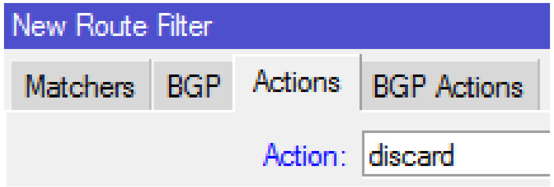
+ Discard receiving via external peers, our customer's prefixes (if we only want to communicate with him directly):



Chain: IN-TR2

Prefix: □ 200.0.0.0/20

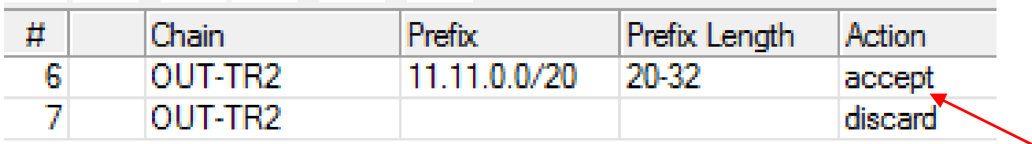
Prefix Length: □ 20-32



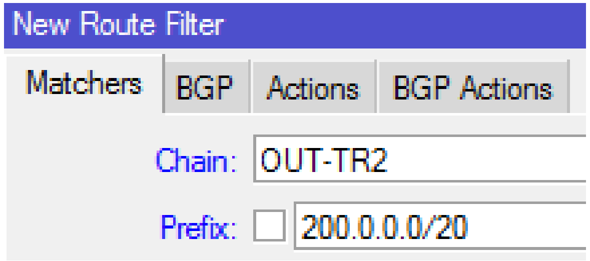
Action: discard

Filtering for Scenario IV

2) Filters to avoid undesired traffic, have to be modified to allow us sending the prefixes from our customer



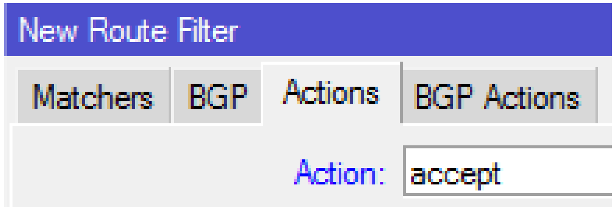
#		Chain	Prefix	Prefix Length	Action
6		OUT-TR2	11.11.0.0/20	20-32	accept
7		OUT-TR2			discard



New Route Filter

Chain: OUT-TR2

Prefix: □ 200.0.0.0/20



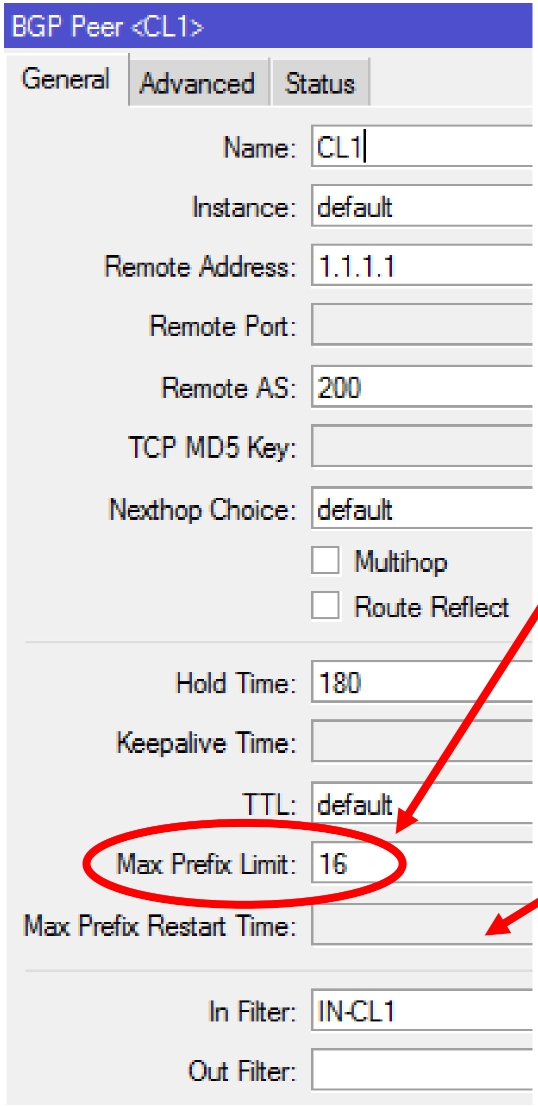
New Route Filter

Action: accept

Above filter should be done for each peer (TR1, TR2 and IXP) and placed before discard rule.

NB: We need also to notify external peers about the new prefix and we'll announce.

Filtering for Scenario IV Avoiding "garbage" from our Customer



BGP Peer <CL1>

General | Advanced | Status

Name: CL1

Instance: default

Remote Address: 1.1.1.1

Remote Port:

Remote AS: 200

TCP MD5 Key:

Nexthop Choice: default

□ Multihop

□ Route Reflect

Hold Time: 180

Keepalive Time:

TTL: default

Max Prefix Limit: 16

Max Prefix Restart Time:

In Filter: IN-CL1

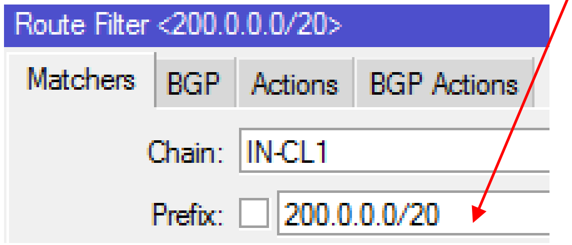
Out Filter:

Is possible to limit the number of prefixes received from peer.

Restart time will work in case of Prefix Limit has reached (BGP session is closed)

Filtering for Scenario IV
Avoiding “garbage” from our Customer

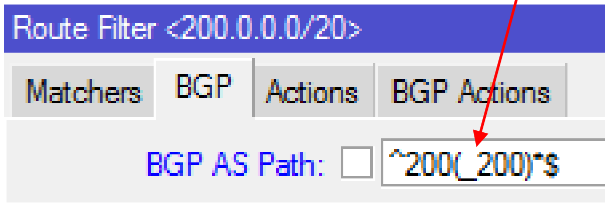
Accepting only his prefix and only his AS number (but allowing any number of prepends with regexp)



Route Filter <200.0.0.0/20>

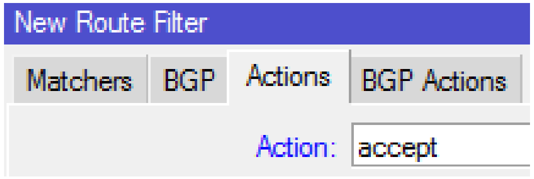
Chain: IN-CL1

Prefix: □ 200.0.0.0/20



Route Filter <200.0.0.0/20>

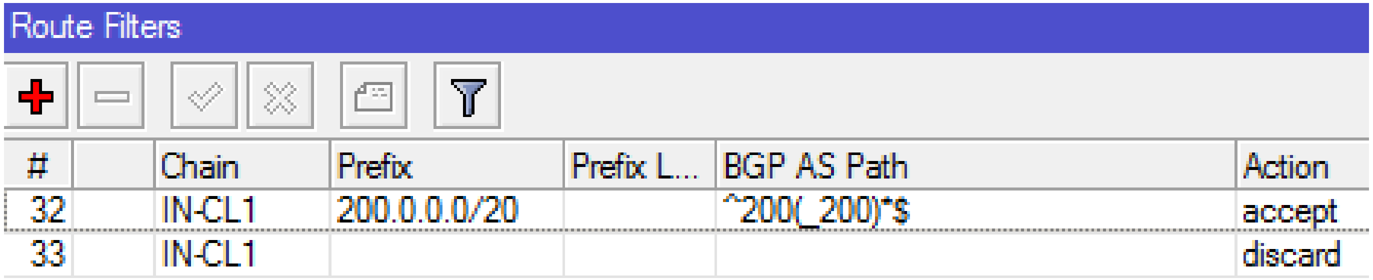
BGP AS Path: □ ^200(_200)*$



New Route Filter

Action: accept

Discarding all the rest

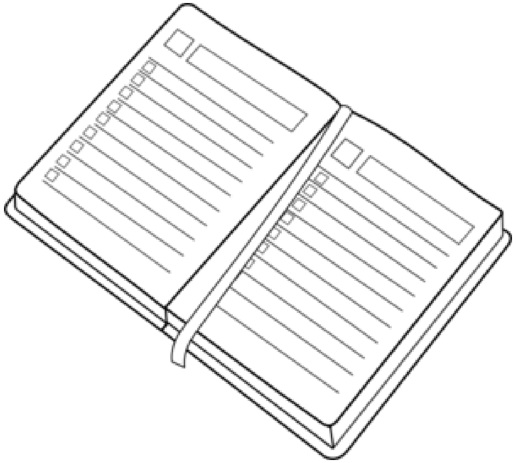


Route Filters

#		Chain	Prefix	Prefix L...	BGP AS Path	Action
32		IN-CL1	200.0.0.0/20		^200(_200)*$	accept
33		IN-CL1				discard

Agenda ✓

1) BGP essentials and basics of BGP filtering; ✓



2) Case Studies: ✓

2.1) Overview ✓

2.2) Single-Homed Provider ✓

2.3) Single-Homed + IXP ✓

2.4) Multi-Homed + IXP ✓

2.5) Multi-Homed + IXP + Providing transit services ✓

Final Considerations

Filtering techniques presented here are commonly used practices considering natural scenarios evolution for Small/Medium ISPs.

The purpose of this work is the orientation on how and where to use the filters with Mikrotik RouterOS and obviously they should be adapted for particular situations.

Some slides can have edition mistakes. So, if interested, ask for the export file of the router.

Thank you

Hvala!

Wardner Maia – maia@mdbrasil.com.br

Download this presentation

Soon, this presentation will be available for download at Mikrotik and MD Brasil Web sites.

www.mikrotikbrasil.com.br/artigos