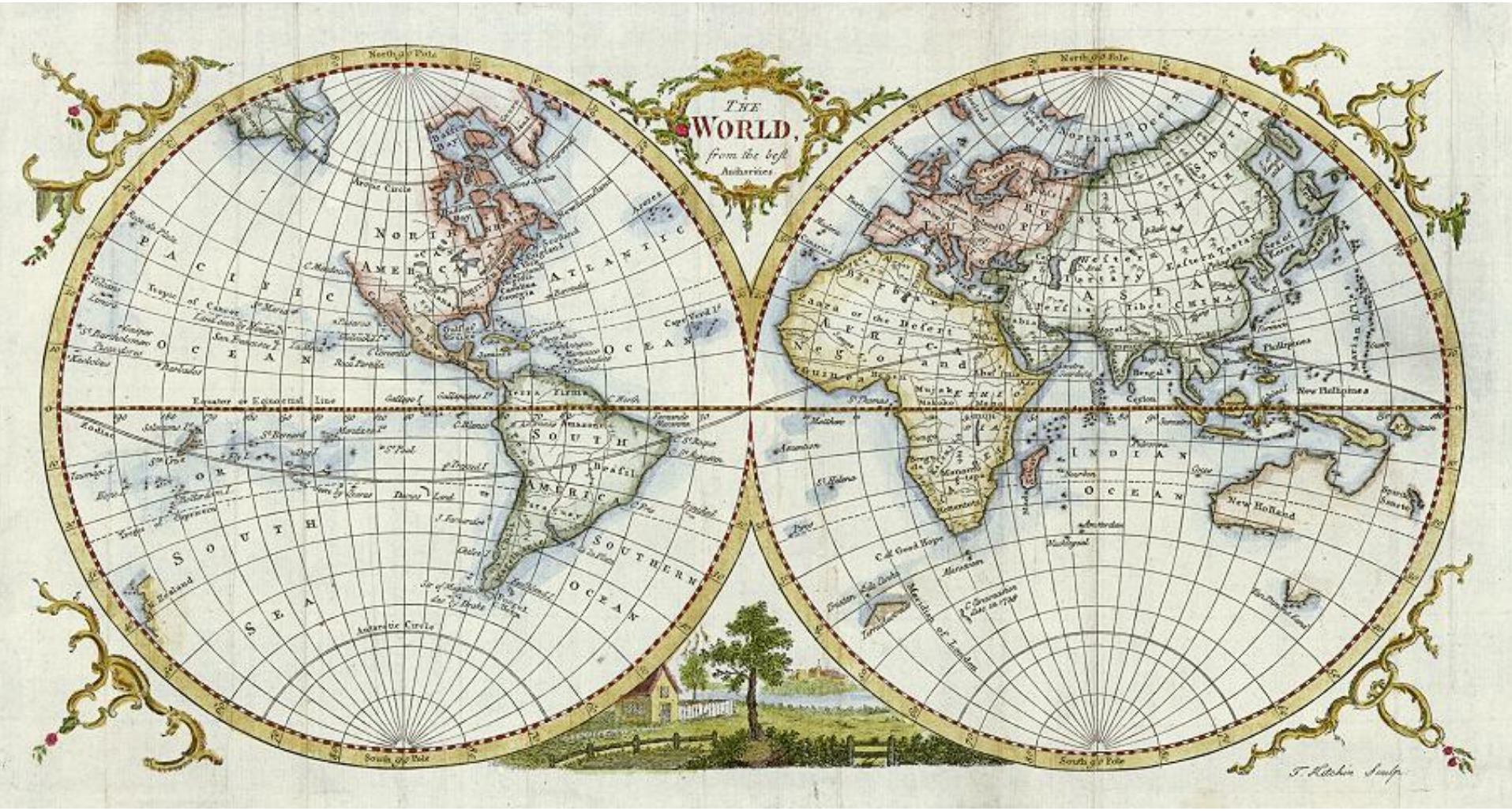
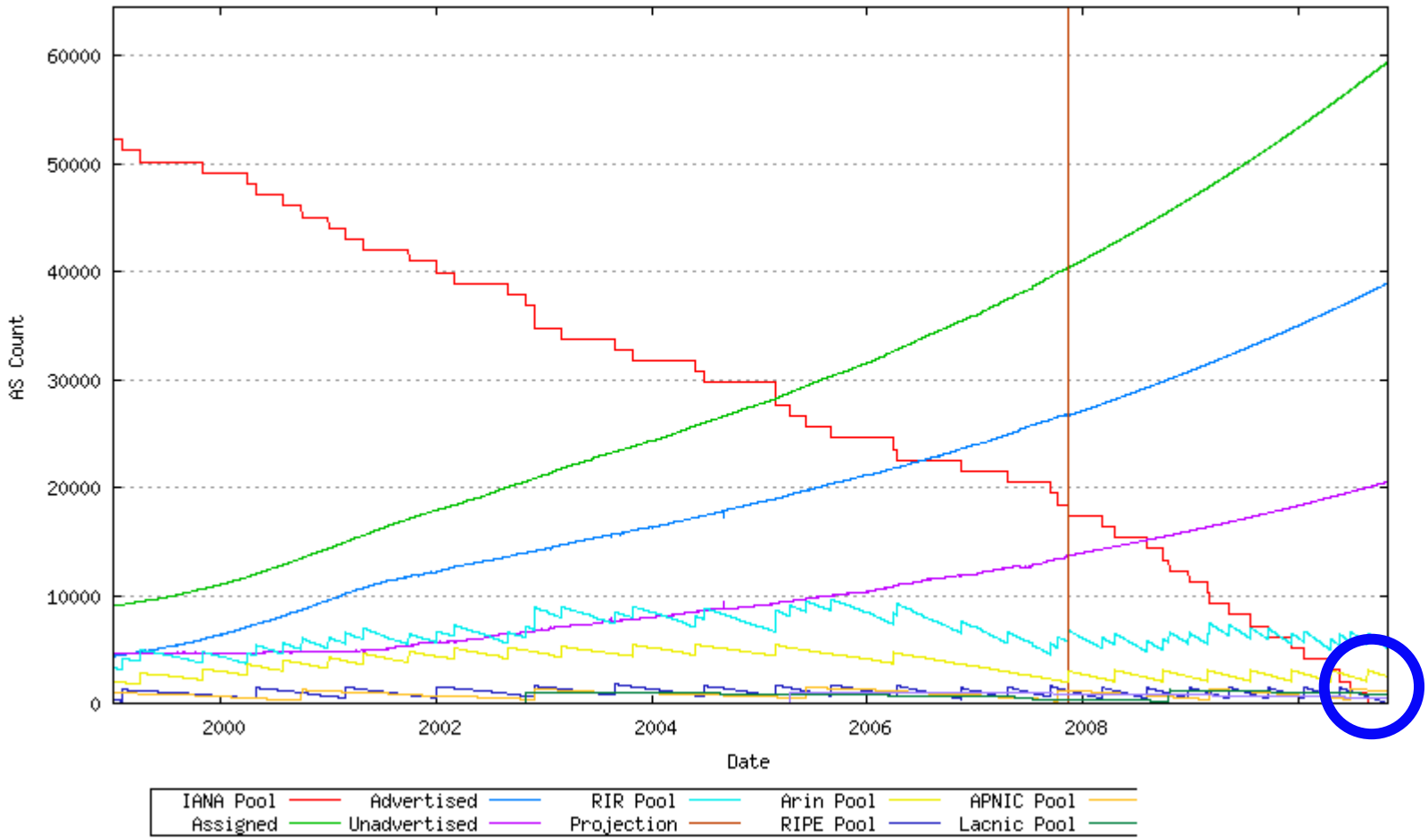
New World BGP



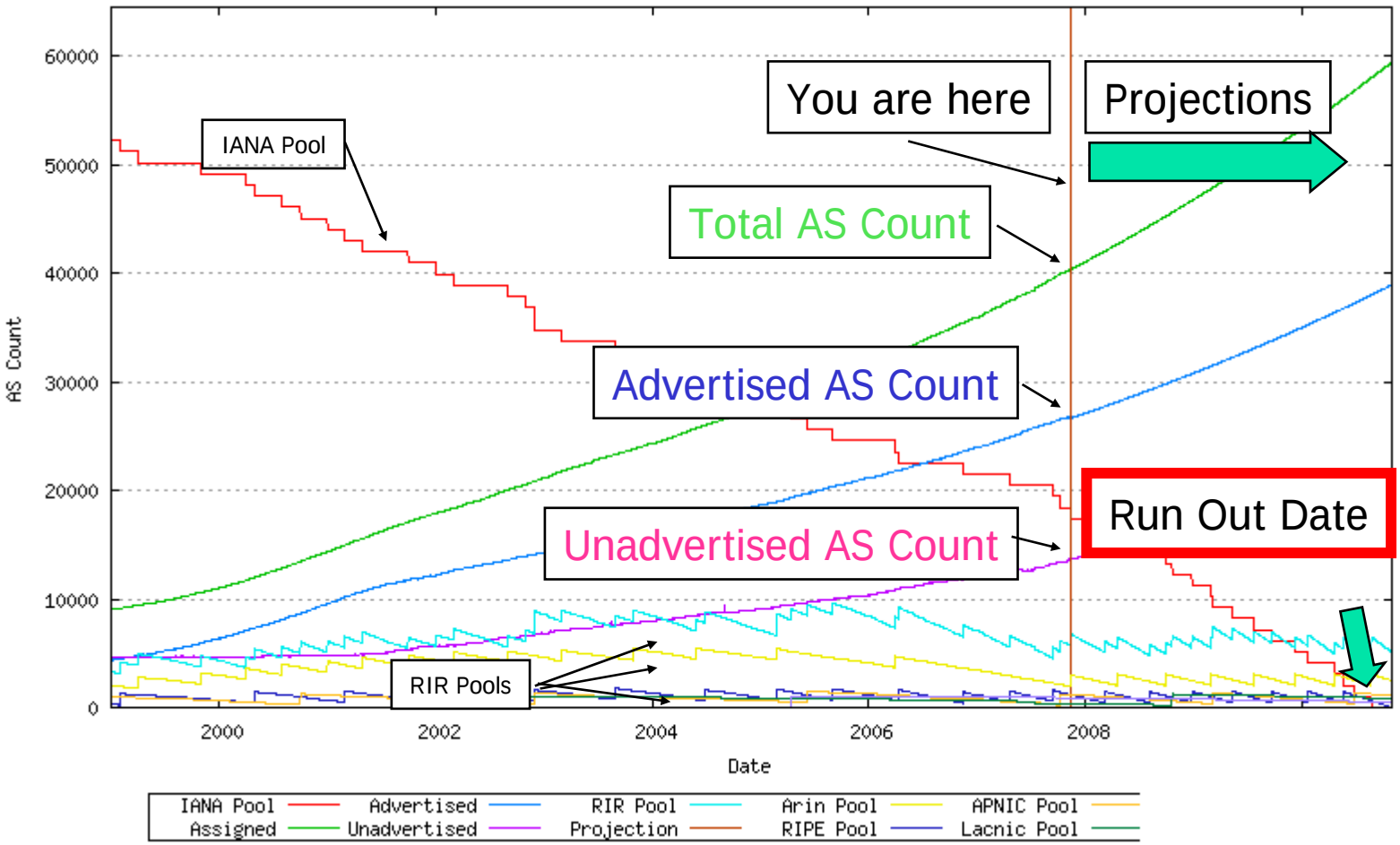
The view from the old BGP World

Geoff Huston
November 2007
APNIC

16-bit AS Number Consumption: The Big Picture

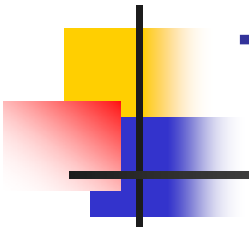


16-bit AS Number Consumption: The Big Picture Explained



16-bit AS Number Exhaustion

- We were running into exhaustion of the 16-bit AS Number pool
  - Estimated exhaustion time: 1200 UTC 1 November 2010
  - See http://www.potaroo.net/tools/asns



The 32-bit ASN Spec

RFC4893

Objectives of the 32-bit ASN design:

- **Change as little as possible in the BGP spec**
- **Be ‘backward compatible’ with 16-bit AS BGP implementations**
  - Attempt to negotiate 32-bit capability when opening a BGP session
  - Automatically adjust behaviour when peering with 16-bit BGP peers and assume a 16-bit “persona” with 16-bit peers
  - Use 32-bit “persona” only with 32-bit peers
- **Preserve “basic” AS semantics in BGP when peering with 16-bit AS BGP peers**
  - Preserve BGP’s loop detection properties
  - Preserve AS Path length metric properties
- **No ‘flag day’ transition**
  - Allow 16-bit BGP implementations to continue to operate indefinitely in a mixed 16 / 32-bit AS bgp world with complete address prefix reachability
  - Allow for piecemeal deployment of 32-bit BGP implementations

What's changed?

- BGP Update messages in the 16-bit world:
  - May contain "lies" in parts of the AS Path
  - May be larger in size due to tunnelled additional information
- But prefix reachability information is still communicated between 16-bit and 32-bit BGP "realms"

RIRs and 32-bit AS Numbers

- From 1 January 2007 the RIRs are allocating 32-bit AS numbers (upon specific request)
- From 1 January 2009 the RIRs will be allocating 32-bit AS numbers by default (leaving some 16-bit AS numbers available upon specific request)

What does this imply?

If you are a 16-bit AS

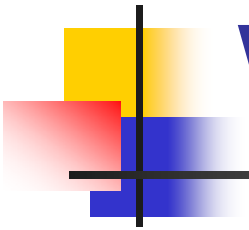
as most (all) of you are today

and you don't want to upgrade all your instances of BGP today

something you probably want to avoid (or at least defer!)

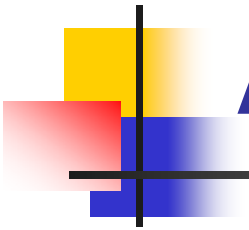
then you don't have to do anything at all!

NOTHING changes!



Well, **almost** nothing!

AS Path Semantics in BGP



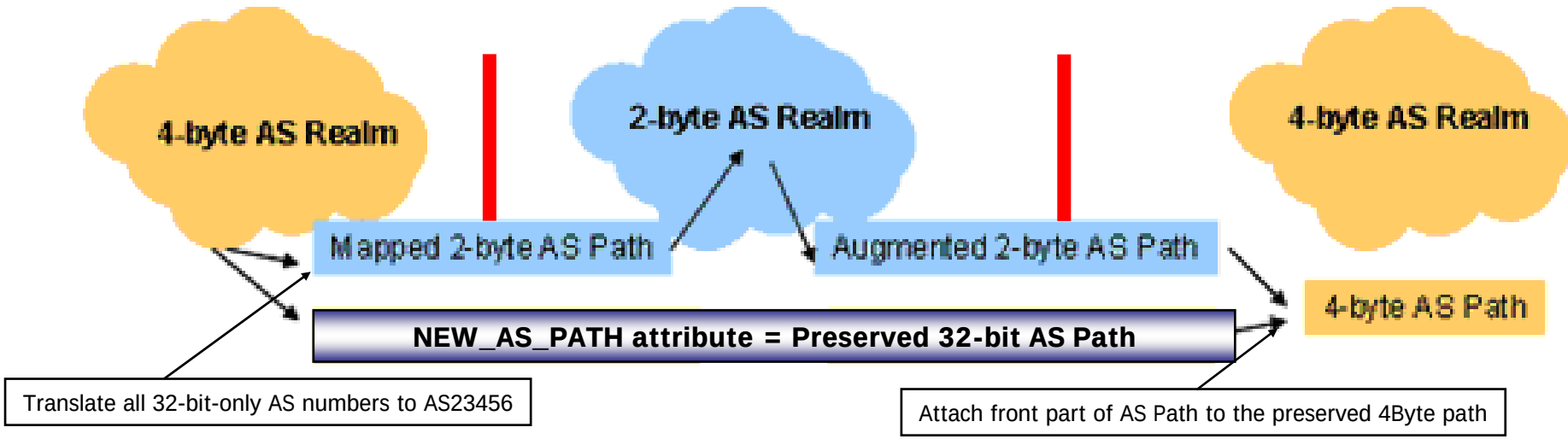
- It's a *path metric* where the length of the AS Path is used as in path selection

- It's a *loop detector* where the presence of your own AS in a PATH is an indicator of a distance-vector "I'm-going-to-loop-to-infinity-unless-you-stop-me" loop

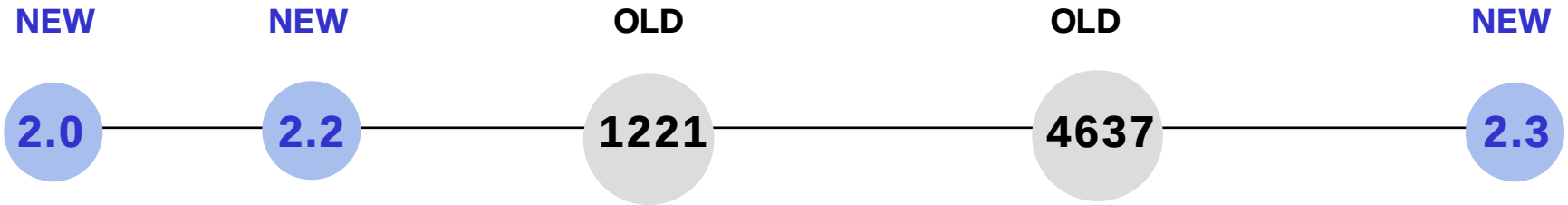
  You don't have to have an **entirely** accurate AS Path – but at a minimum you do have to have path-metric and loop-detecting properties for BGP to function correctly

32-bit AS Transition

- Think about this space as a set of NEW / OLD boundaries
- Define the NEW / OLD and the OLD / NEW transitions
- Preserve all BGP information at the transition interfaces
  - **Translate** 32-bit AS Path information into a 16-bit representation
  - **Tunnel** 32-bit AS Path information through 16-bit AS domain as an update attribute



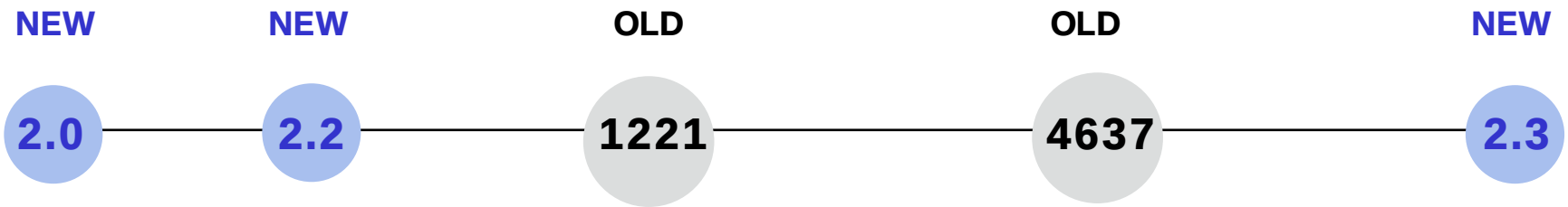
32-bit / 16-bit BGP Example



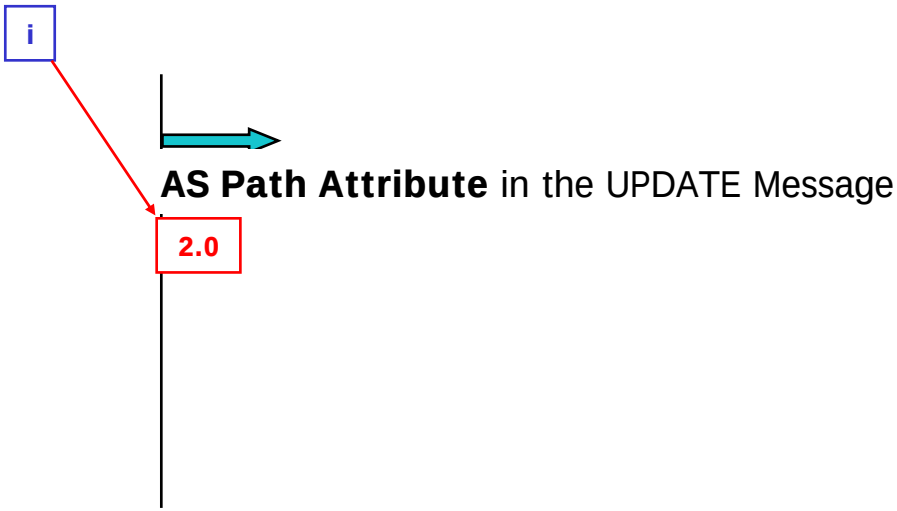
AS Path in the RIB

i

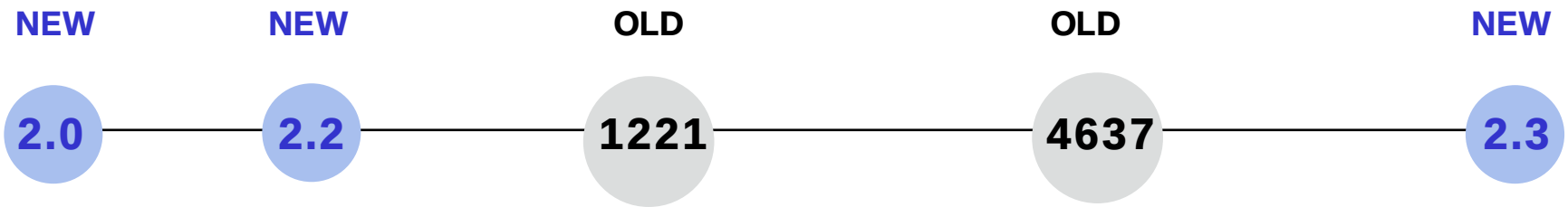
32-bit / 16-bit BGP Example



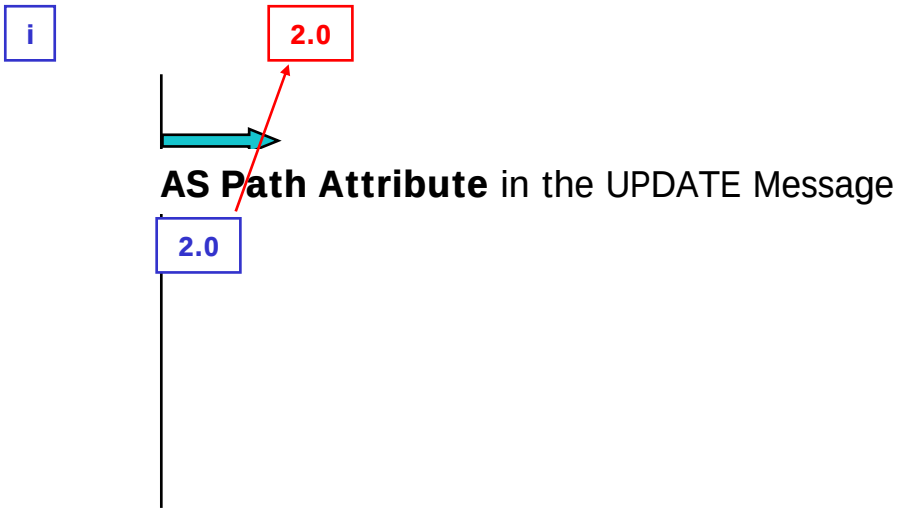
AS Path in the RIB



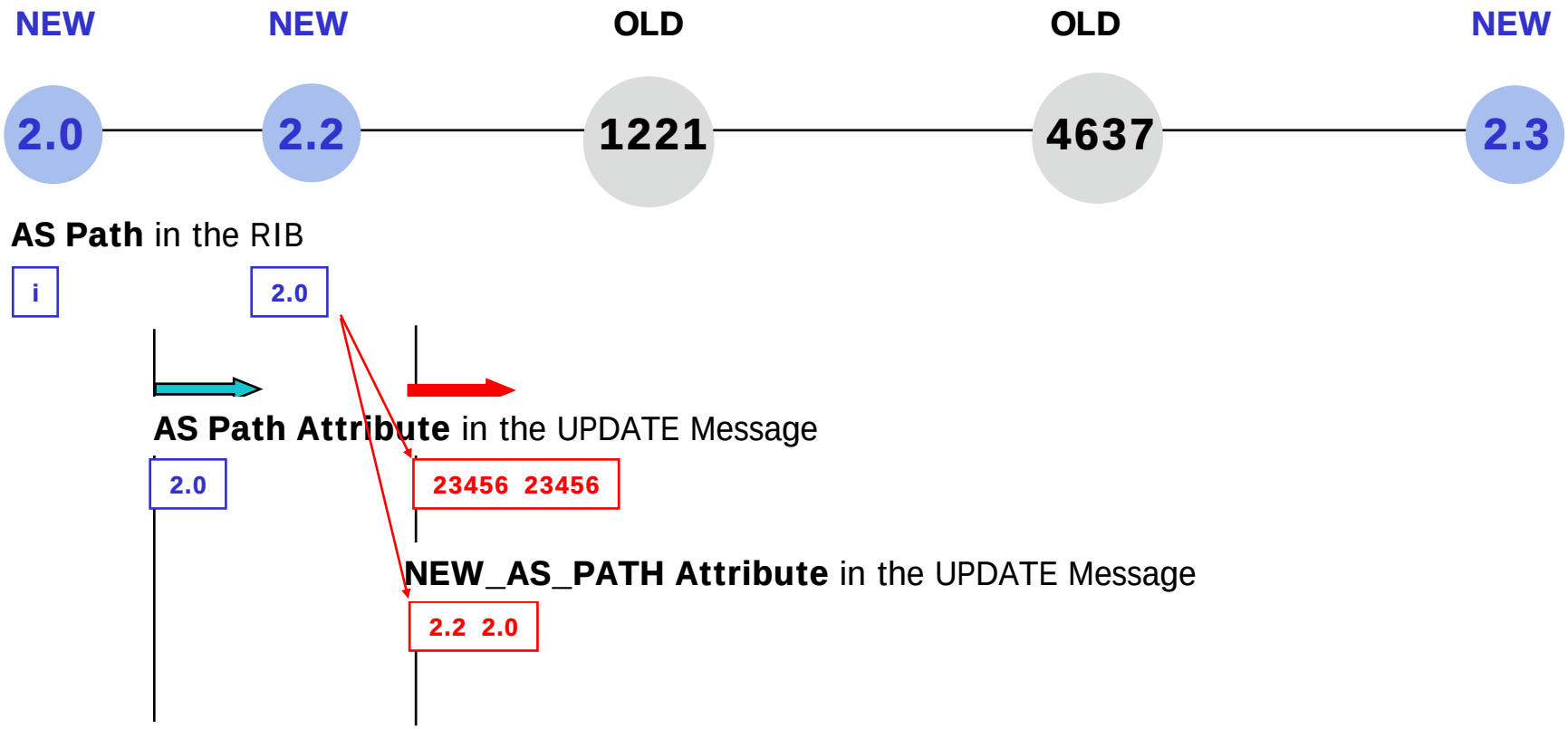
32-bit / 16-bit BGP Example



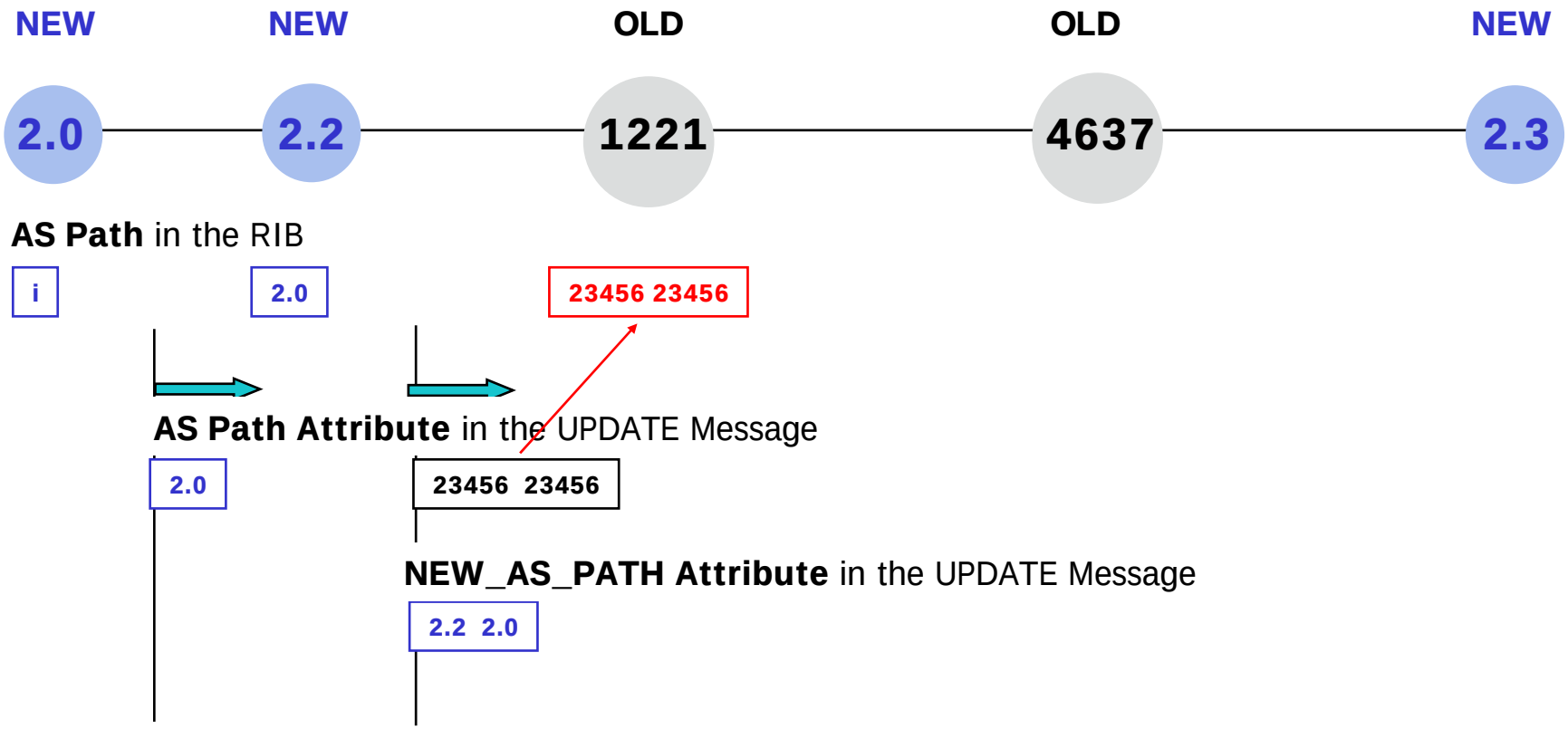
AS Path in the RIB



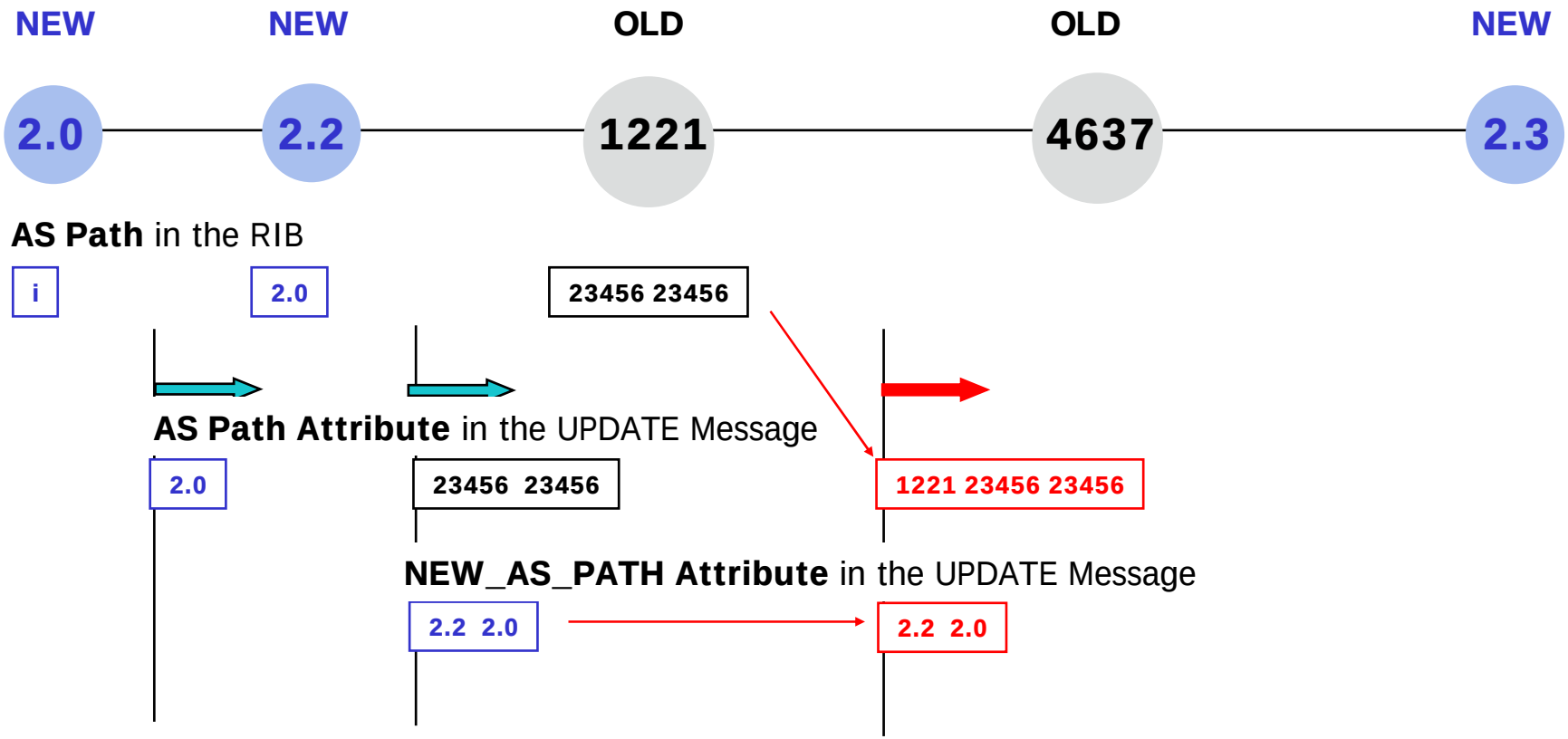
32-bit / 16-bit BGP Example



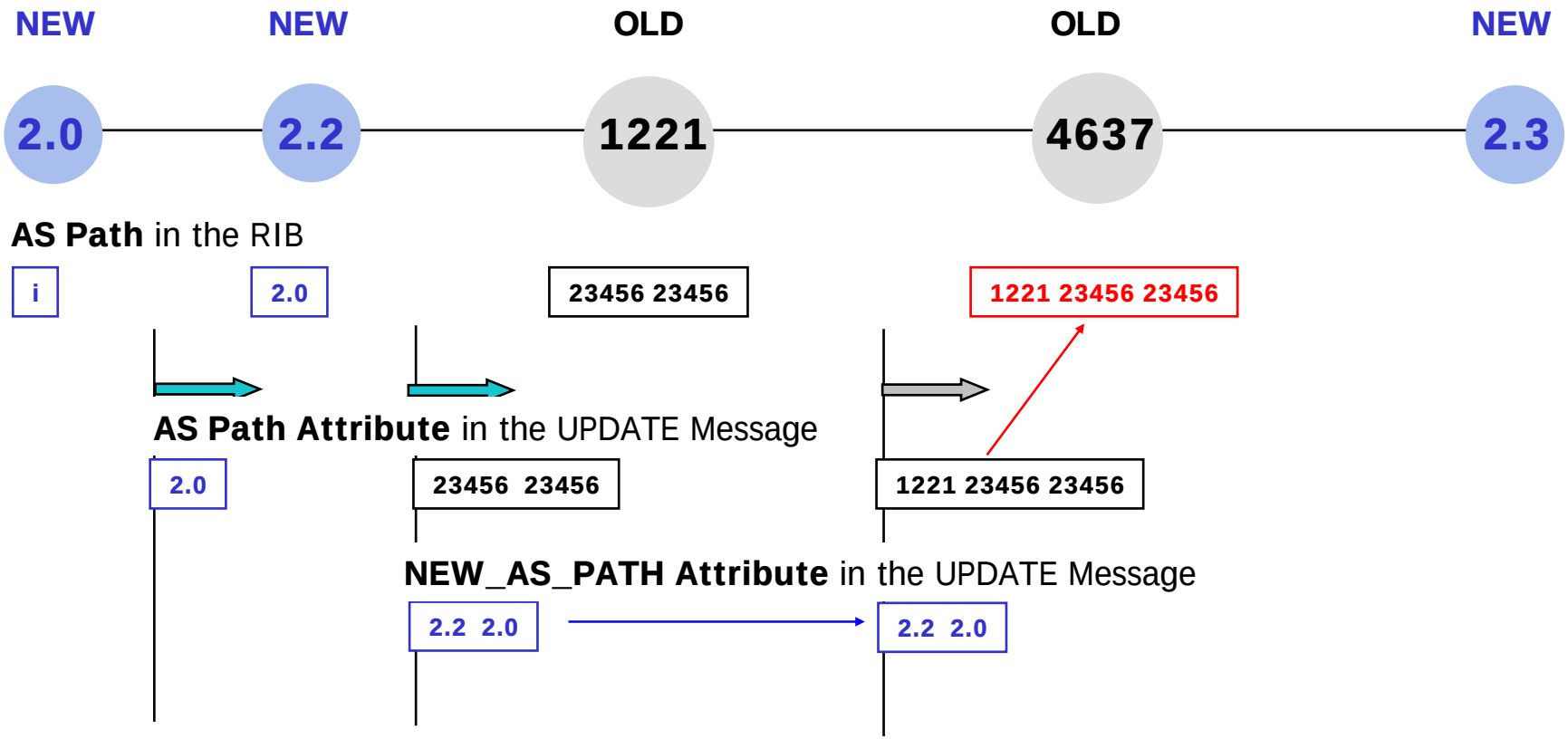
32-bit / 16-bit BGP Example



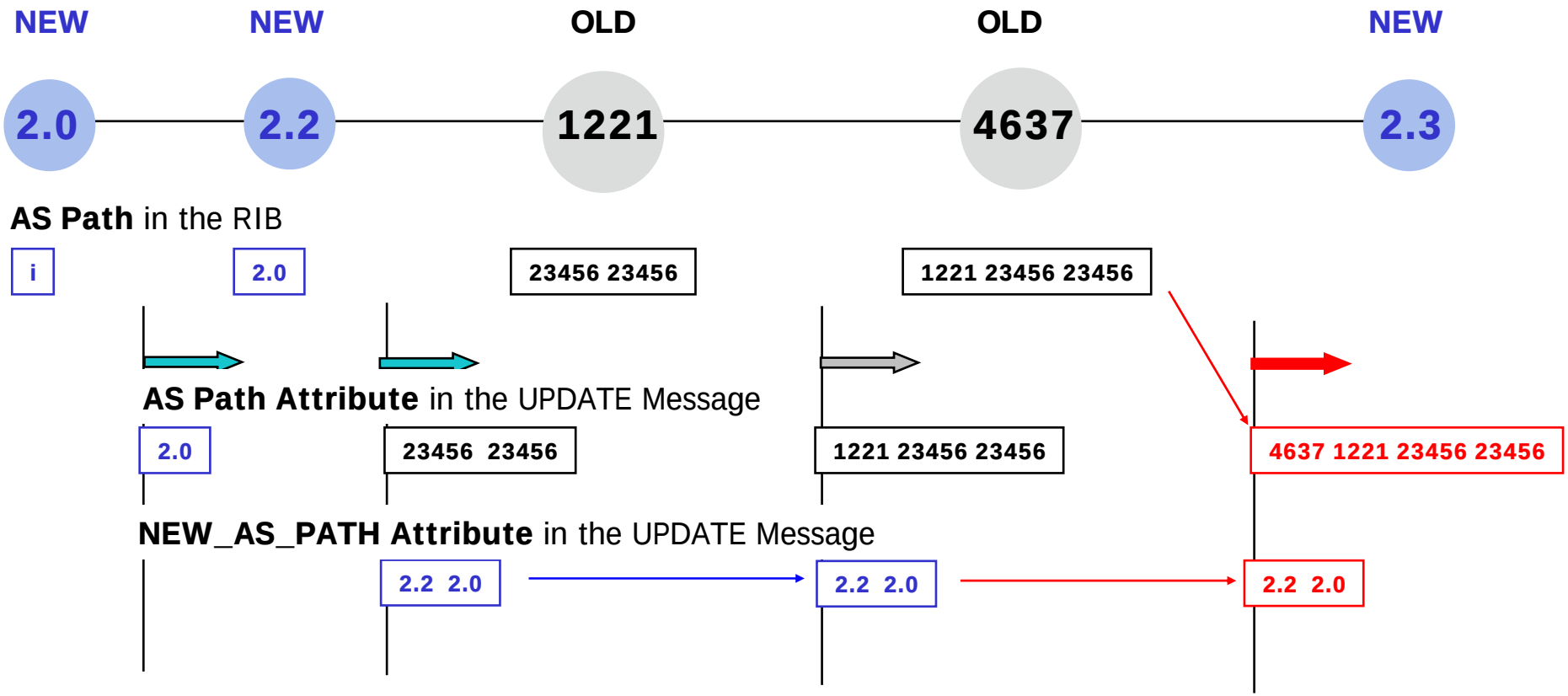
32-bit / 16-bit BGP Example



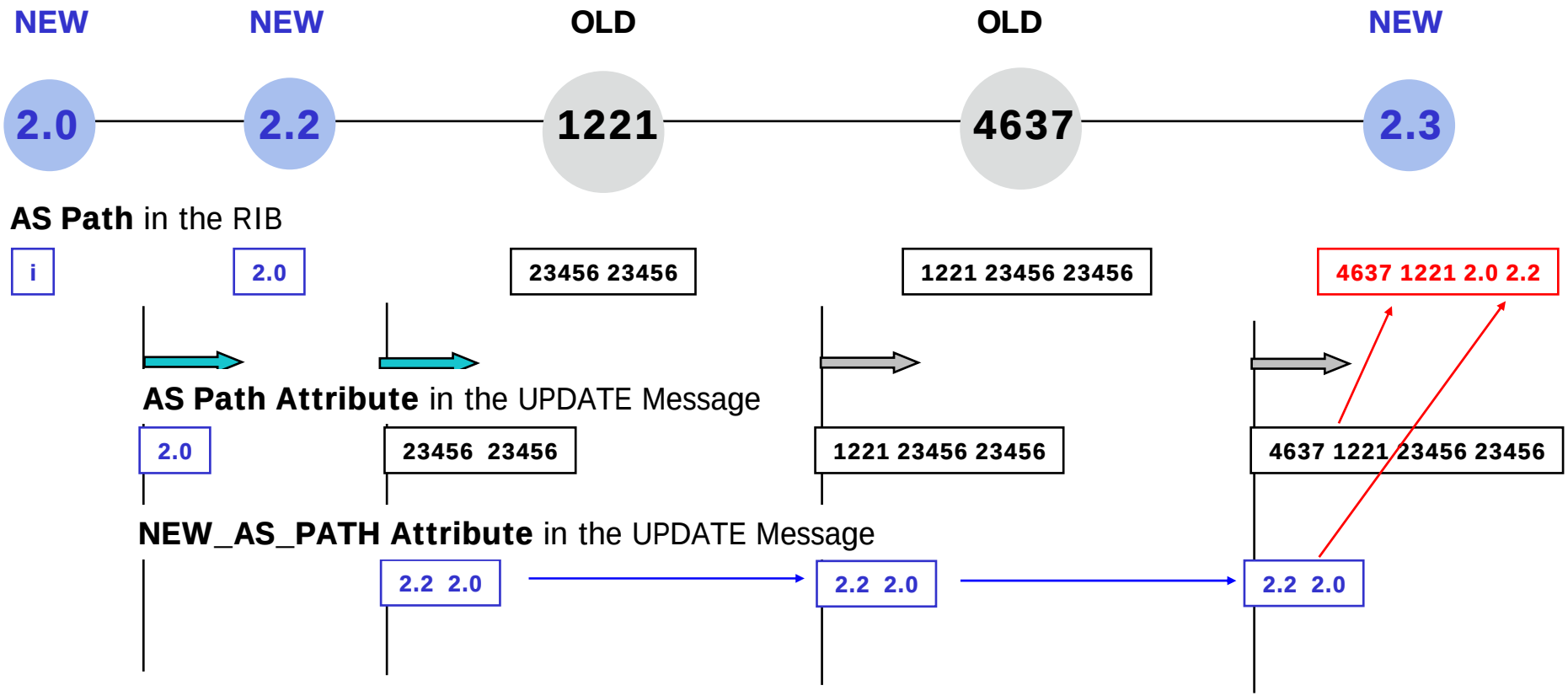
32-bit / 16-bit BGP Example



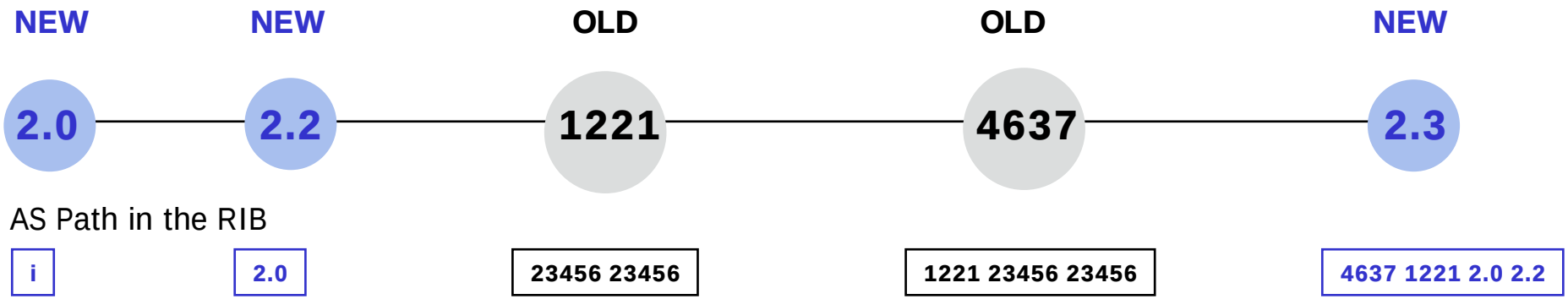
32-bit / 16-bit BGP Example



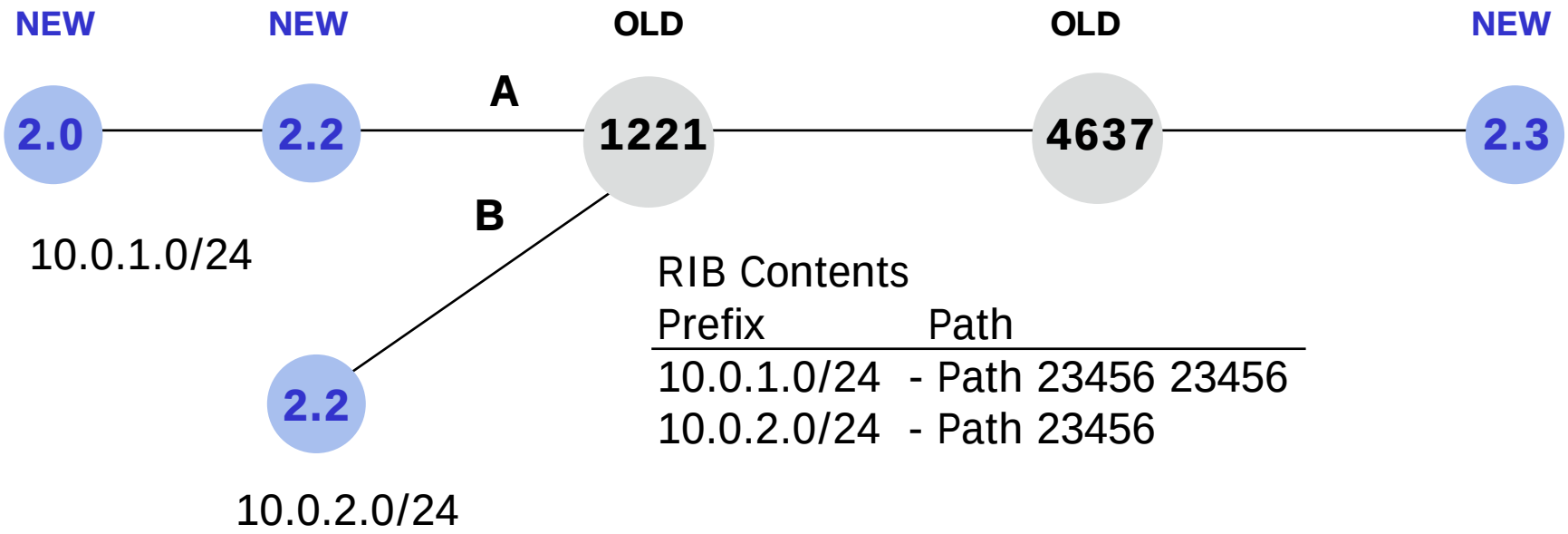
32-bit / 16-bit BGP Example



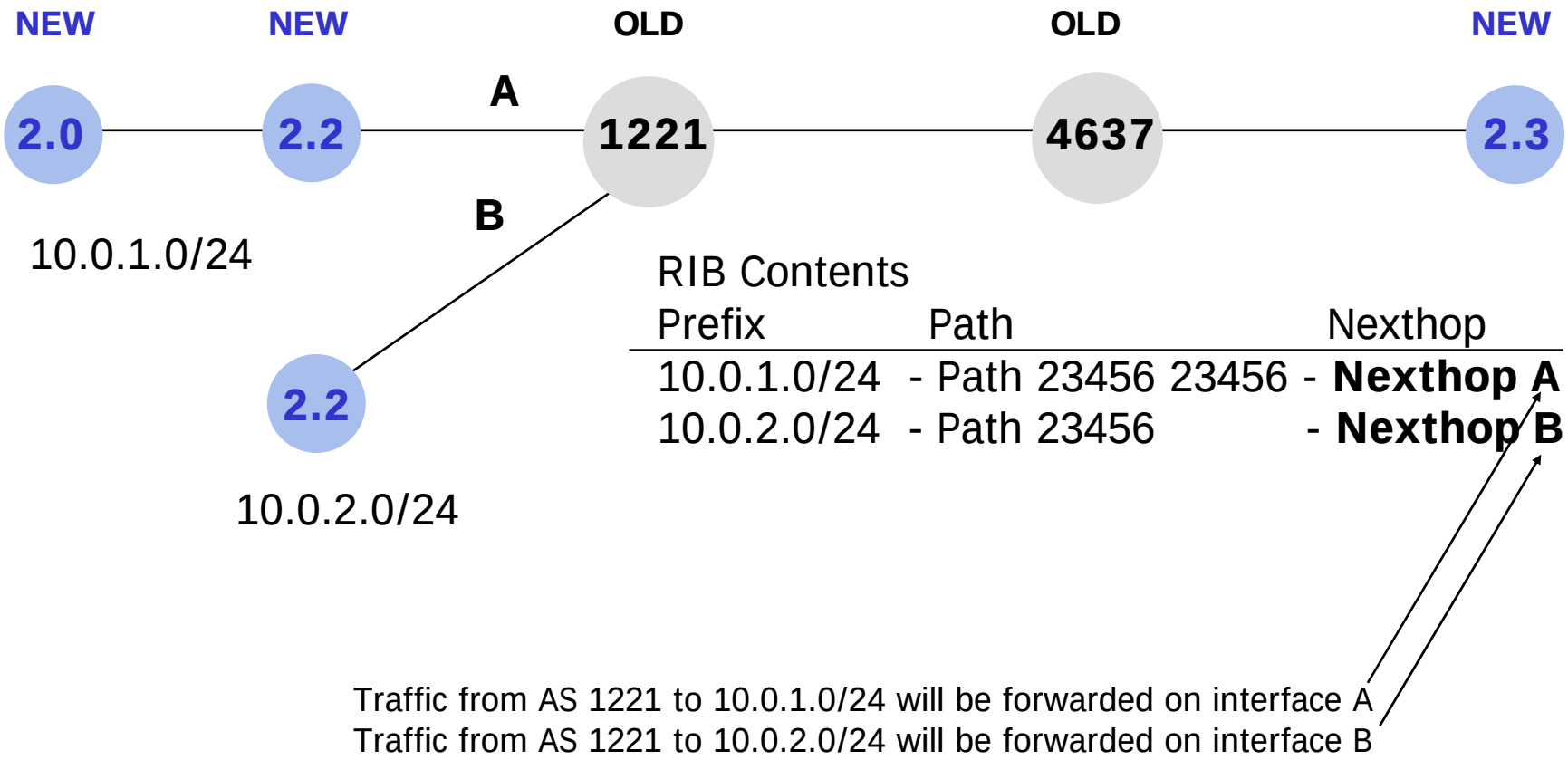
32-bit / 16-bit BGP Example



Can old-BGP get Confused?



NO! BGP Nexthop is the key!



This is standard BGP behaviour – nothing changes here for BGP as it is used today

What changes with 32-bit ASs?

- If you are an “old” BGP speaker then what should you look out for?

NEW_AS_PATH Attribute

- BGP speakers in 16-bit AS domains should support **NEW_AS_PATH** as a transitive optional attribute in UPDATE messages
  - because that's where the 32-bit path is hiding
  - That's a "SHOULD" not a "MUST", by the way
  - Its better if you do, but nothing fatally breaks if you don't
    - Mixed 2 / 4 Byte loops will get detected in the 16-bit world as a fallback

Default BGP configurations will do the right thing here

NEW_AGGREGATOR Attribute

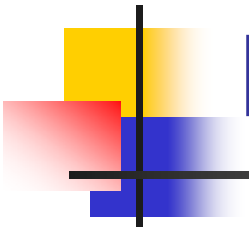
- BGP speakers in 16-bit AS domains should support **NEW_AGGREGATOR** as a transitive optional attribute in UPDATE messages
  - because that's where the 32-bit Aggregator AS is hiding
  - That's a "SHOULD" not a "MUST", by the way
  - Its better if you do, but nothing fatally breaks if you don't

  Default BGP configurations should do the right thing here

AS 23456

- **AS 23456** is going to appear in many 16-bit AS paths – both origin and transit

This is not an error – it's a 16-bit token holder for a 32-bit AS number



Netflow and Sflow

- Netflow analyzers may need to be reviewed
  - Netflow version 9 supports 32-bit AS numbers
    - But may not report the 32-bit ASN unless the netflow collector is a 32-bit BGP
    - Does your analyzer support 32-bit AS numbers?
  - Netflow version 8 and earlier are 16-bit AS constrained
    - Which implies that you'll be seeing AS 23456 more than you may want!
- Sflow
  - Appears to define a source and dest AS using a 32 bit field
  - So it *should* be ok!

BGP Communities

- If you want to explicitly signal to a 32-bit AS using communities in BGP then you will need to explicitly signal the 32-bit AS using **BGP Extended Communities**
  - Attempting to use AS 23456 in this context will have unintended consequences!

  See:
  - RFC 4630
  - draft-rekhter-as4octet-ext-community-01.txt

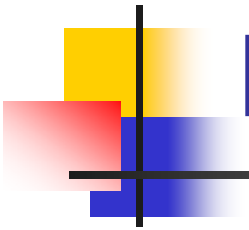
BGP Memory requirements

- BGP memory requirements will increase
  - 32-bit BGP speakers will need twice the memory used to hold AS paths[1]
  - 16-bit BGP speakers will need up to three times the memory used to hold AS paths plus NEW_AS_PATH extended community attribute[2]
    - 30,000 unique AS paths with an average length of 4 implies an additional memory requirement of 240Kb for 32-bit BGP and up to a further 480Kb for 16-bit BGP

1 - Not “twice the memory” but “twice the memory used for AS Path storage”
2 – Not “three times the memory”, but “three times the memory used for AS Path Storage”

BGP Bandwidth requirements

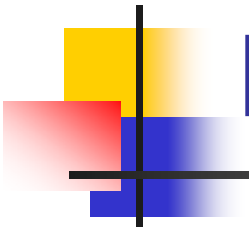
- BGP bandwidth requirements will increase (ever so slightly!)
  - 32-bit BGP speakers will need twice the size used to carry AS paths
  - 16-bit BGP speakers will need up to three times the size used to carry AS paths (factoring in the NEW_AS_PATH attribute)
    - The update will grow by an average of 20 bytes, assuming an average AS path length of 4



Performance

- 32-bit to 16-bit BGP session startup may be considerably slower
  - The 32-bit speaker will need to compress all the AS Paths into their 16-bit equivalent prior to generating updates

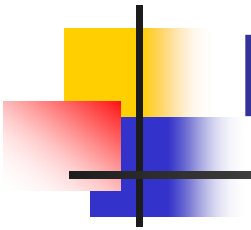
    (assuming that the 16-bit Paths for Update messages are generated on demand)
  - This may take some time to compute for some 30,000 distinct AS Paths (depending on the internal structure of the BGP implementation)



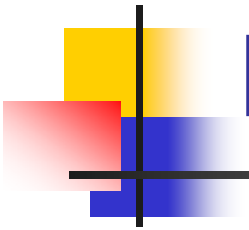
Performance

- BGP convergence times may increase in some cases
  - Any instance of 16-bit BGP world destruction of the tunnelled NEW_AS_PATH attribute implies extended times on loop detection in order to fully complete prefix withdrawal
  - Its not that the withdrawal will loop forever, its that the loop will take additional AS hops before it is detected in the 16-bit realm
  - The time to complete the withdrawal of a route may be extended

Proxy Aggregation



- If you proxy aggregate in the 16-bit world then make sure that the aggregate is strictly larger than the components
  - Or loop detection may be harder
    - As the AS Set object generated in the 16-bit word as a result of this proxy aggregation is not cleanly translatable into the 32-bit world, so 32-bit information is lost
- But proxy aggregation is not a common occurrence in today's BGP environment



Mixed environments

- No dynamic capability for 16/32-bit ASN mode shift
  - You cannot flick from “16-bit OLD” to “32-bit NEW” mode within an active BGP session
  - You need to clear the session and then perform a clean start to trigger the initial capability exchange

Transition within an AS

- In a complex iBGP AS that wants to transition to using a 32-bit “home” AS then you are going to have to think about the transition VERY carefully
  - You can undertake this transition one router at a time, but care and attention are required

Notation Confusion

- We have not (yet) converged to a uniform way of describing 32-bit AS Numbers:

  Numerics

  101, 65637

  Dotted Short Ints

  0.101, 1.101

  Dotted Short Ints+

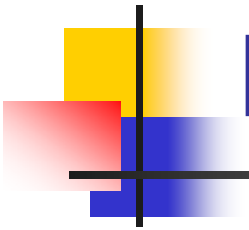
  101 , 1.101

See draft-michaelson-4byte-as-representation-02.txt

Operational Support Systems

What happens when you have a customer / transit / peer with a 32-bit AS Number?

- What's in the route registries and what your customers tell you about their AS and what's in your OSS and your routing system will differ:
  - E.g.: **AS 1.2** needs to be auto-translated into **AS 23456** in a number of places, including in your OSS
  - Your BGP routers may need to peer with **AS 23456**, transit across **AS 23456**, and have multiple customers on **AS 23456** at the same time, while also understanding that these refer to different external parties
  - Your OSS might get terminally confused!

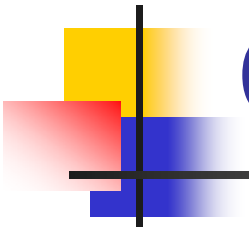


Related Systems

The following systems will need to be revised:

- Internet Route Registries and RPSL
- WHOIS databases, WHOIS query syntax and WHOIS report formats
- Protocol, log and dump analysers

- And anything else that wants to manipulate AS numbers, including your local support systems, scripts and databases



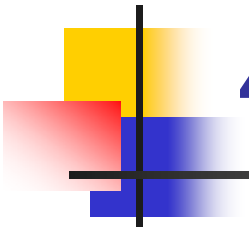
Changing BGP

- Known 32-bit BGP implementations:
  - Quagga (patches to 0.99.6)
  - OpenBGPD (patches to 3.9, 4.0)
  - JUNOSe 4-1-0 and later
  - Redback
  - Cisco IOS-XR 3.4.

32-bit AS Implementations

Patches to OpenBGPD and Quagga:

1. Convert BGP to be internal 32-bit AS in all data structures
2. Alter parser and output routines to support the various notational forms of AS numbers
3. Alter OPEN processing to negotiate 4 Byte AS Capability with BGP Peer
4. Alter UPDATE processing changes to support 16-bit peers
   - Generated updates include a generated NEW_AS_PATH attribute and a dynamically created 16-bit AS_PATH (and AGGREGATOR changes)
   - Received updates need to merge NEW_AS_PATH with AS_PATH to form a stored 32-bit AS PATH (and AGGREGATOR merges) and remove NEW_AS_PATH attribute



4 Byte AS Testing

- Tests have been undertaken using closed BGP networks, and over the public Internet
- Tests of 16-bit/32-bit transition boundaries in various permutations of transits and loops
- Current announcement of 203.10.62.0/24 originating from AS 2.2 to assist others in local testing of 32-bit BGP

The Route-Views View

```
route-views.oregon-ix.net>show ip bgp 203.10.62.0/24
BGP routing table entry for 203.10.62.0/24, version 177310093
Paths: (43 available, best #39, table Default-IP-Routing-Table)
  Not advertised to any peer
  3277 3216 3549 4637 1221 23456
    194.85.4.55 from 194.85.4.55 (194.85.4.16)
      Origin IGP, localpref 100, valid, external
      Community: 3216:3000 3216:3004 3277:3216 3549:2141 3549:30840
  7500 2497 4637 1221 23456
    202.249.2.86 from 202.249.2.86 (203.178.133.115)
      Origin IGP, localpref 100, valid, external
  2493 3602 812 812 4637 1221 23456
    206.186.255.223 from 206.186.255.223 (206.186.255.223)
      Origin IGP, localpref 100, valid, external
  2905 701 1239 4637 4637 4637 4637 4637 4637 1221 23456
    196.7.106.245 from 196.7.106.245 (196.7.106.245)
      Origin IGP, metric 0, localpref 100, valid, external

   ...
```

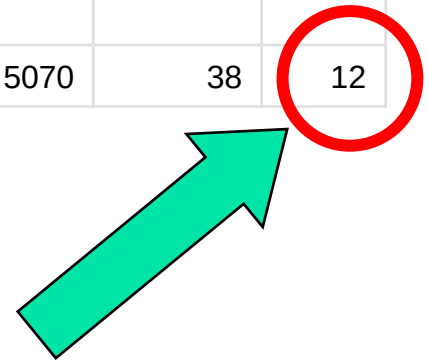
32-bit Path Reconstruction

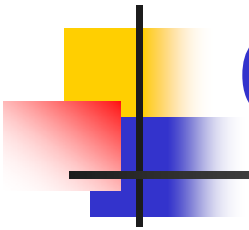
```
srv0# bgpctl show rib 203.10.62.0/24
flags: * = Valid, > = Selected, I = via IBGP, A = Announced
origin: i = IGP, e = EGP, ? = Incomplete

flags destination          gateway           lpref   med aspath origin
*>    203.10.62.0/24       147.28.0.1          100     0 0.3130 0.1239 0.4637
                                                         0.4637 0.4637 0.4637
                                                         0.4637 0.4637 0.1221
                                                         1.202 i
```

Experiment performed on January 11 2007, with the assistance of Randy Bush and George Michaelson, using OpenBGPD 3.9 with 4Byte AS support patches as the origin and the observer points.

32 bit AS Numbers in use

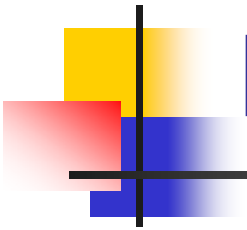
RIR	RIR Pool	Unadv	Adv	16-bit	Unadv	Adv	32-bit	Unadv	Adv
AFRINIC	1921	155	212	901	152	211	1020	3	1
APNIC	2207	1522	3074	1211	1499	3069	996	23	5
ARIN	2996	7754	11695	1976	7751	11694	1020	3	1
RIPE NCC	2211	4166	10477	1200	4157	10473	1011	8	5
LACNIC	1391	539	878	368	538	878	1023	1	0
TOTAL	10726	14136	26336	5656	14097	26325	5070	38	12





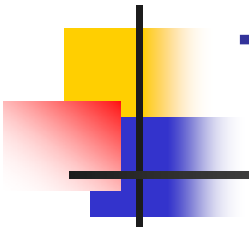
Conclusions

- Deployed BGP appears to be entirely capable of supporting incremental 32-bit AS deployments
  - No, the Internet is probably not going to crash and burn just because of this change in BGP!
- BUT: Will your OSS do the right thing when you need to use 4 Byte AS numbers?
  - What happens if you have 2 or more eBGP customers with 4 byte AS numbers?
  - What happens if you have 2 or more transits and/or peers using 4 byte AS numbers?



Resources

- IETF Specification
  - RFC4893
- OpenBGPD patches
  - http://www.potaroo.net/tools/bgpd
- Quagga patches
  - http://quagga.ncc.eurodata.de/



Thank You

Questions?