SDN Workshop

Contact: training@apnic.net

Issue Date: [Date]

Revision: [xx]

APNIC

BGP-LS

SDN Workshop

APNIC

WSDN01_v0.1

Issue Date: [Date]

Revision: [xx]

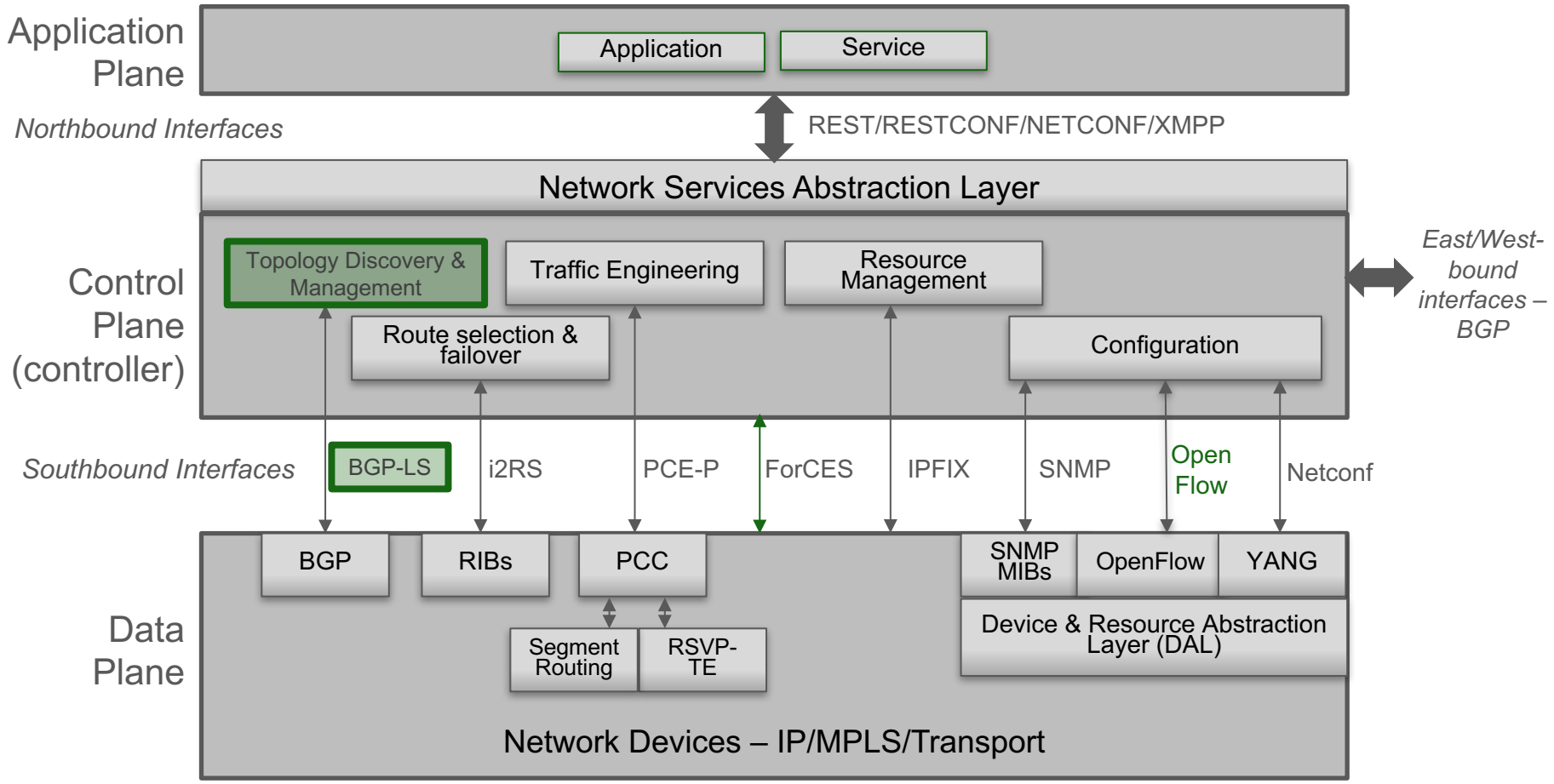
Overview

- In a nutshell
- Motivations
- Introduction
- BGP extensions
- Example
- SR extensions
- Extensions for BGP Egress Peer Engineering (EPE)
- Operational considerations

In a nutshell

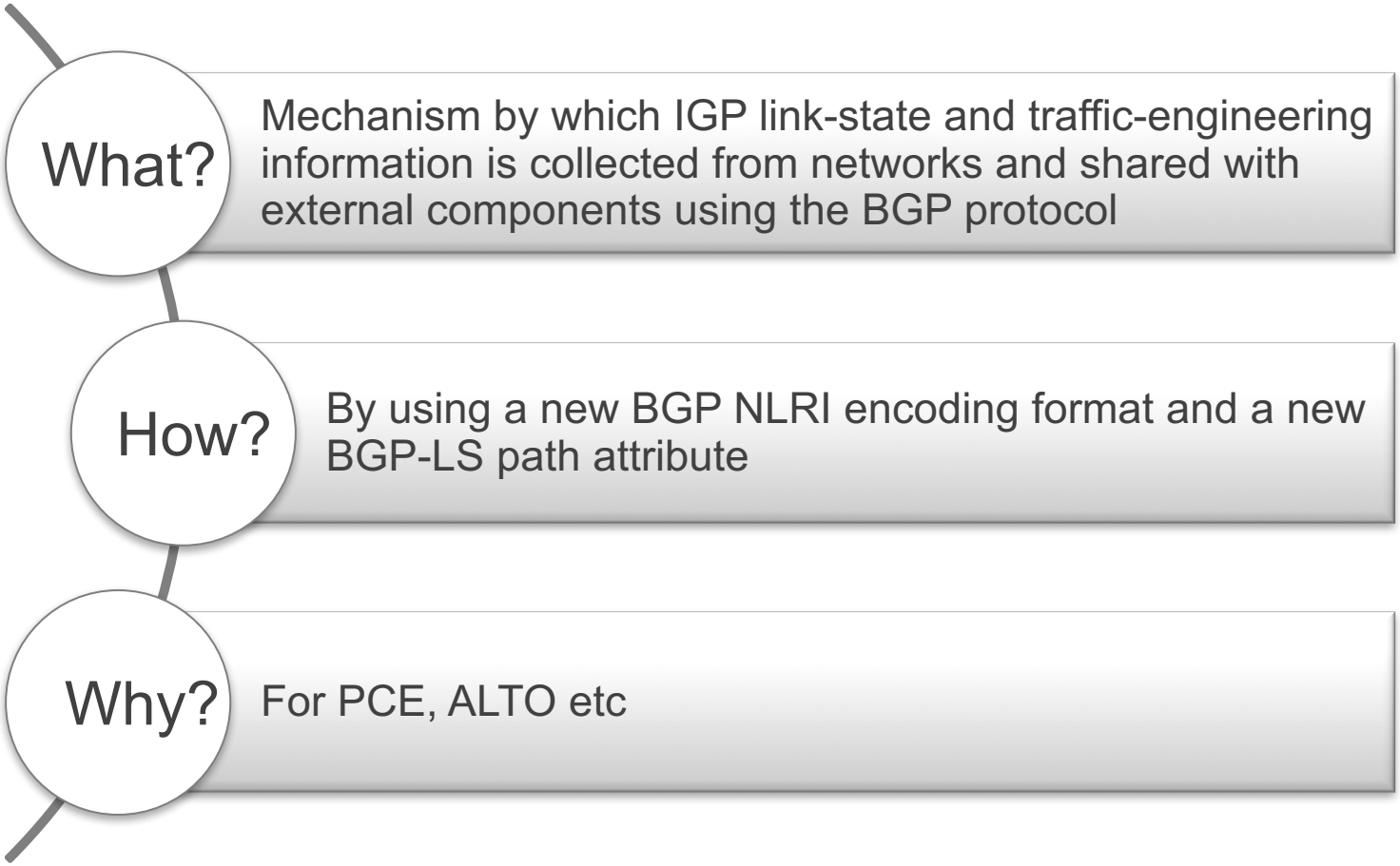
APNIC

SDN architectural framework



Note: designations of north-bound and south-bound are relative to the control plane ("controller")

BGP-LS: in a nutshell…



Motivations

APNIC

Complete topology visibility

- The contents of a router's link-state database (LSDB) or traffic-engineering database (TED) only describes links and nodes within a single area

- End-to-end network visibility is important for certain applications:
  - End-to-end traffic-engineered paths that cross IGP areas or autonomous systems (ASs)
  - Path Computation Element (PCE) function requires topology visibility

MPLS-TE with PCE

- A Path Computation Element (PCE) can be used to compute MPLS-TE paths:
  - within a “domain” (e.g. an IGP area)
    - May be used to provide greater computational capability, ability to impose additional policy and more complex algorithms
  - across multiple domains (such as a multi-area AS or multiple ASs)
    - Since the TED only includes information for a single area, routers cannot make optimal decisions for constructing end-to-end traffic-engineered paths.
    - One solution to this problem is per-domain path computation (RFC5152). However, it suffers from several limitations.

Topology visibility for PCE

- There is no standards-defined mechanism mandated for a PCE to acquire network topology
- One option that is used by many implementations for the PCE is to use the services of a passive IGP listener
- BGP-LS provides an alternative option

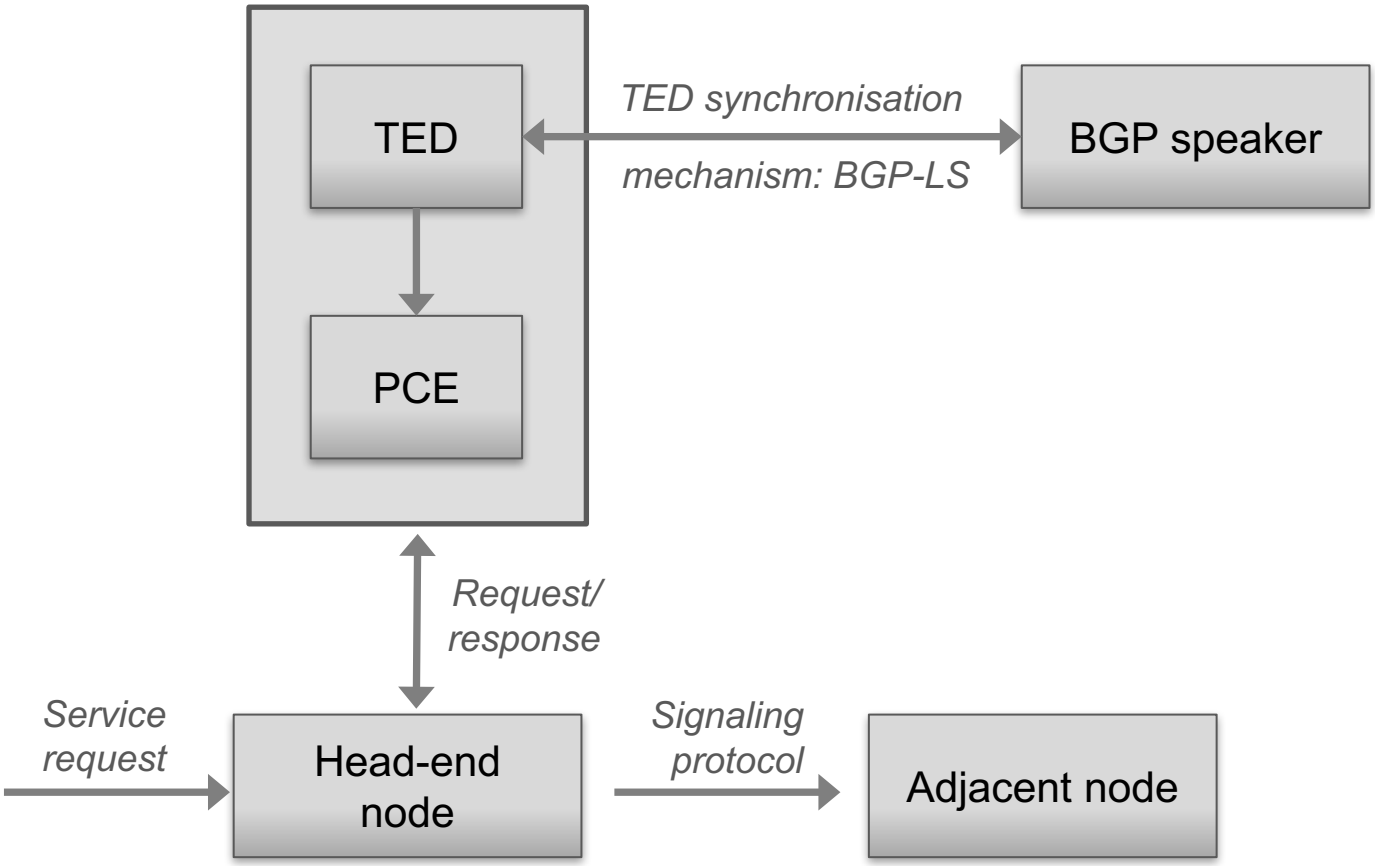
Challenges with using IGP listeners

- Controller needs to hear all IGP updates and therefore needs direct links to every area/domain in the network
- Operators tend to be uncomfortable with the approach of extending the IGP to external controllers
- Controllers need to support all relevant IGP protocols and extensions

Advantages of using BGP

- Limited to a single protocol
- BGP is already well-known and widely deployed; not a new protocol
- BGP has extensive policy mechanisms to control the exchange of information and protect the network
- BGP sessions between the controller and the BGP speakers can be multi-hop i.e. there is no need for direct connectivity
- Route reflectors can be used to reduce the points of contact between the controller and the network

TED synchronisation with PCE



Introduction

Router databases

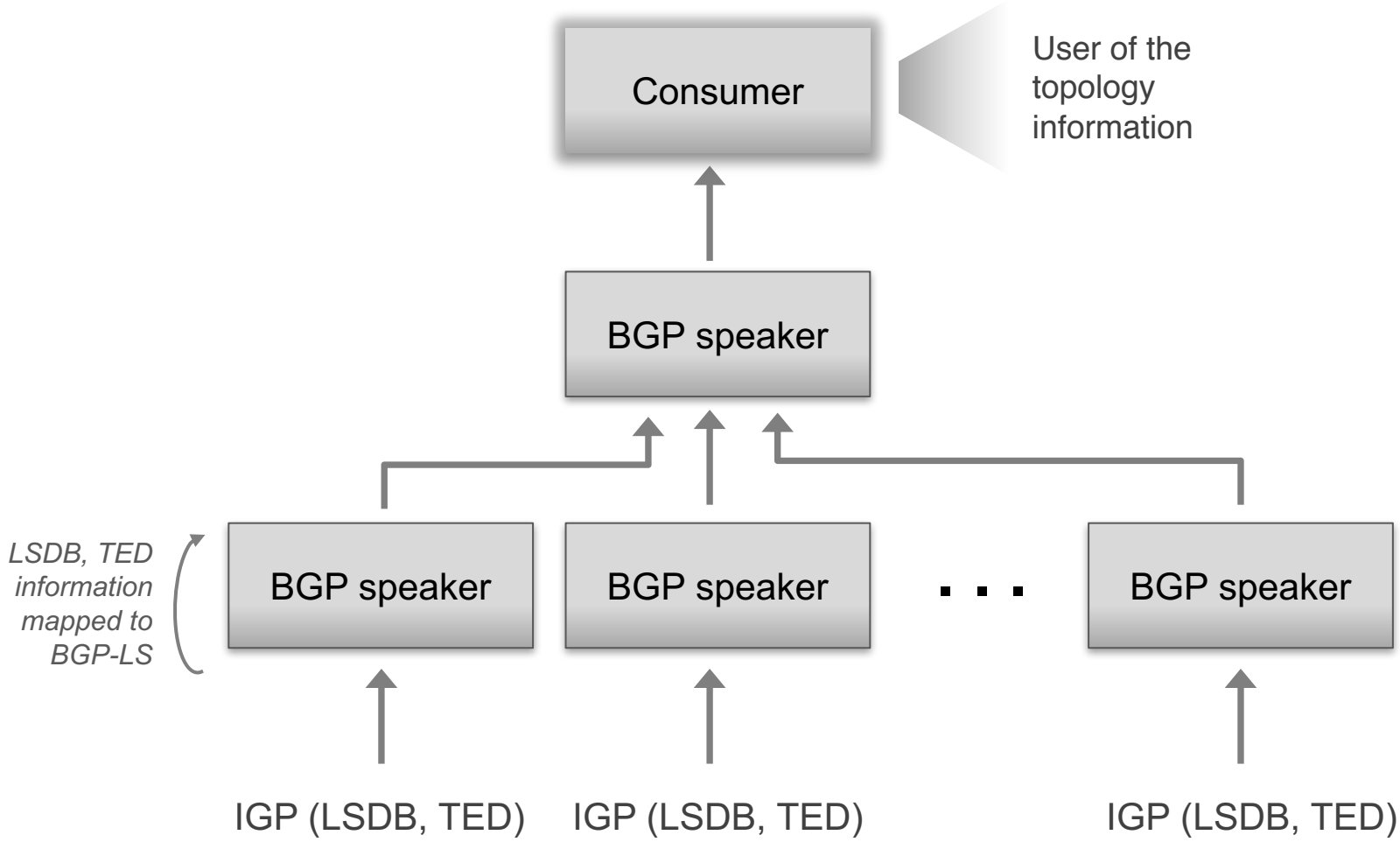
Link-state database (LSDB)

- Router IP addresses
- Router link attributes:
  - Neighbor router identifier
  - Local Interface IP address
  - Remote interface IP address
  - IP subnet address of link
  - IGP link metric

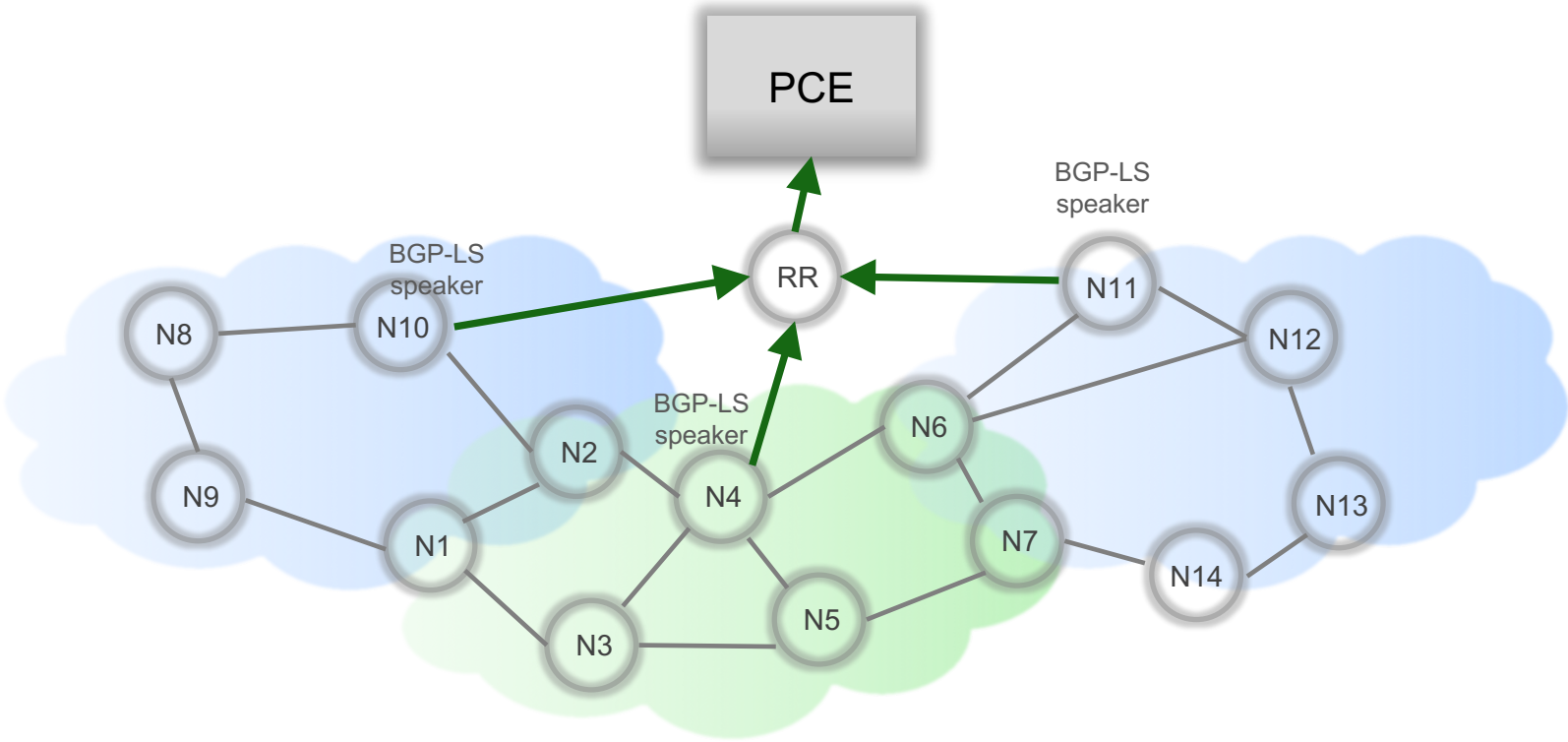
Traffic-engineering database (TED)

- Router IP addresses
- Router link attributes:
  - Local Interface IP address
  - Remote interface IP address
  - Traffic-engineering metric
  - Maximum bandwidth
  - Maximum reservable bandwidth
  - Unreserved bandwidth
  - Administrative group
  - Per-CoS reservation state
  - Pre-emption
  - SRLGs

Advertising IGP info to consumers



BGP-LS peering example



BGP extensions

New BGP definitions

- New BGP link-state NLRI:
  - To describe link, node and prefixes corresponding to IGP link-state data
- New BGP link-state path attribute:
  - Link, node and prefix properties and attributes (e.g. link metrics)
- Information carried by the new link-state NLRI and BGP-LS path attribute is represented in an IGP-neutral way as far as possible

Link-state NLRI

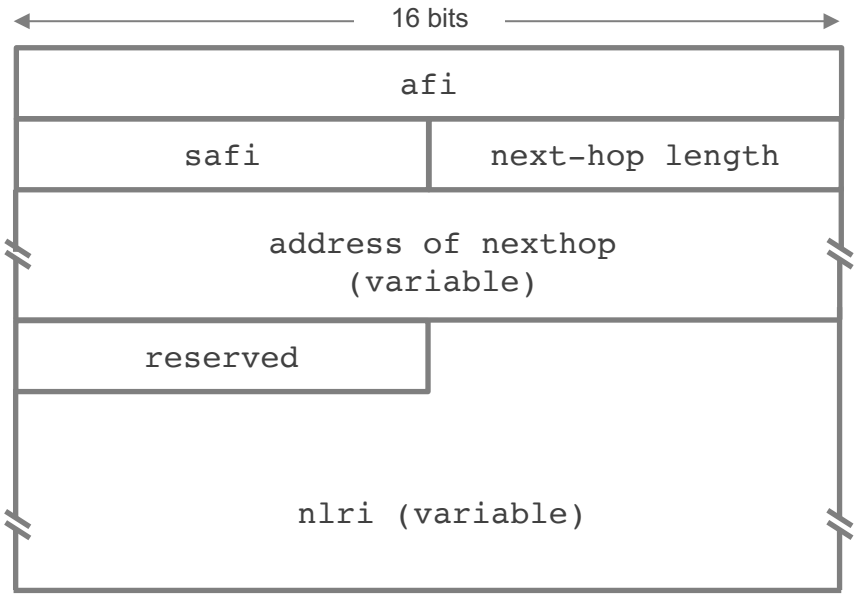
- Node NLRI
- Link NLRI
- Prefix NLRI

MP-BGP attributes

- The MP-BGP attributes MP_REACH_NLRI and MP_UNREACH_NLRI are used for advertising BGP-LS information

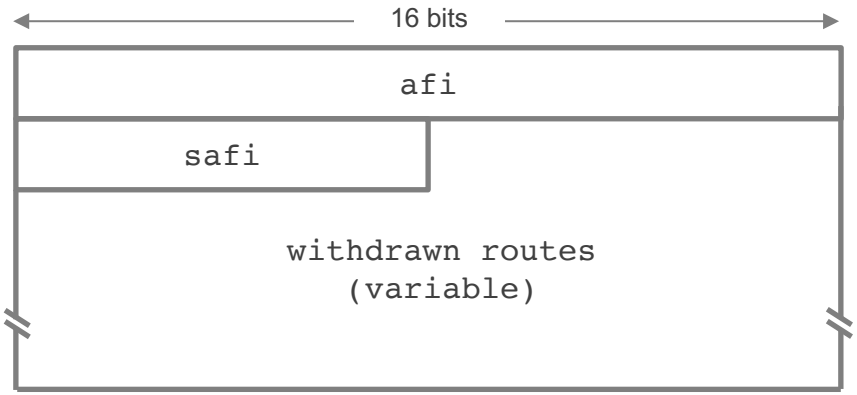
MP_REACH_NLRI (Type Code 14)

Used to advertise feasible routes and next-hop address



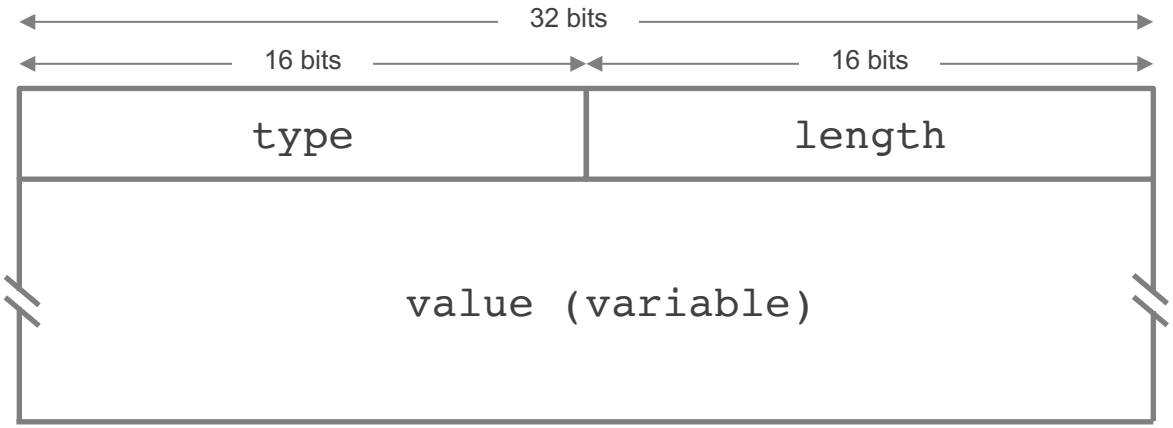
MP_UNREACH_NLRI (Type Code 15)

Used for withdrawing multiple unfeasible routes



Generic TLV format

- All information carried in the BGP-LS NLRI and path attribute uses this generic TLV format:



Link-state NLRI

- Each link-state NLRI describes either a node, link or a prefix

- Non-VPN link, node and prefix information:
  - AFI: 16388, SAFI: 71

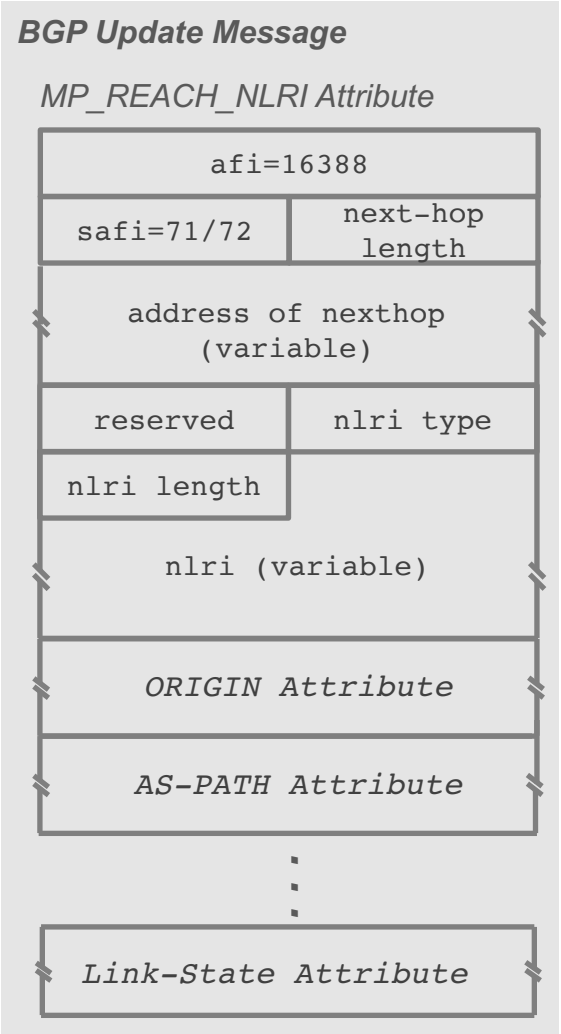
- VPN link, node and prefix information:
  - AFI: 16388, SAFI: 72

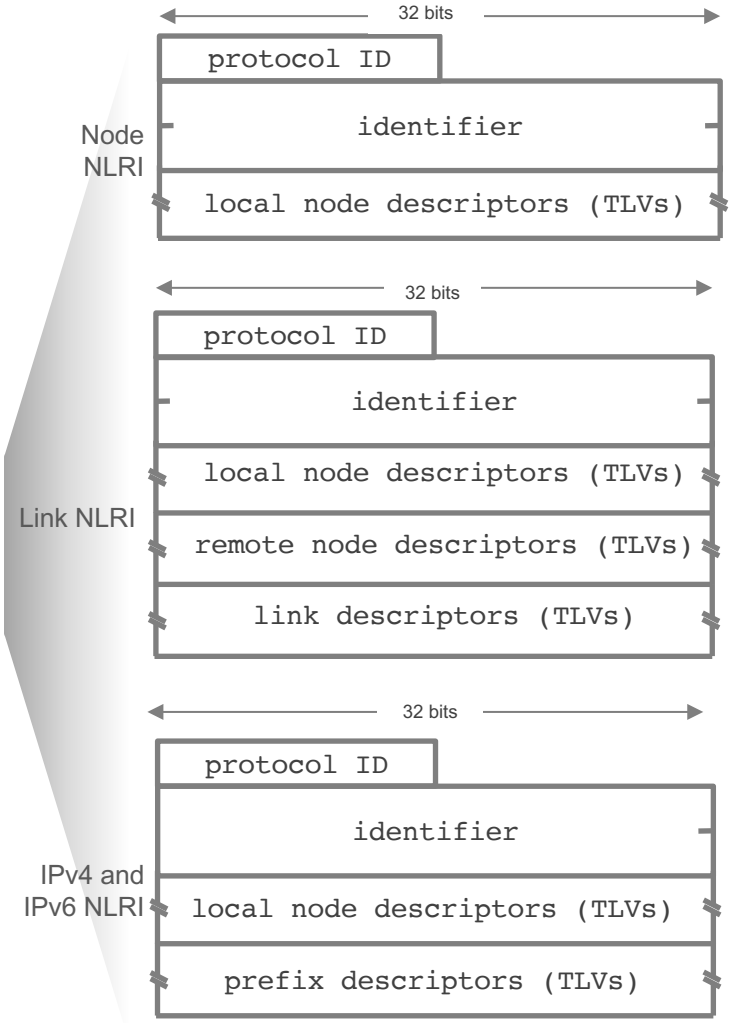
Compatibility

- For two BGP speakers to exchange link-state NRLIs, they must have advertised their ability to do so via BGP Capabilities Advertisement.

- Capabilities Advertisement:
  - Capability code: 1 (MP-BGP)
  - AFI 16388 / SAFI 71 for BGP-LS
  - AFI 16388 / SAFI 72 for BGP-LS-VPN

Link-state NLRI at a glance



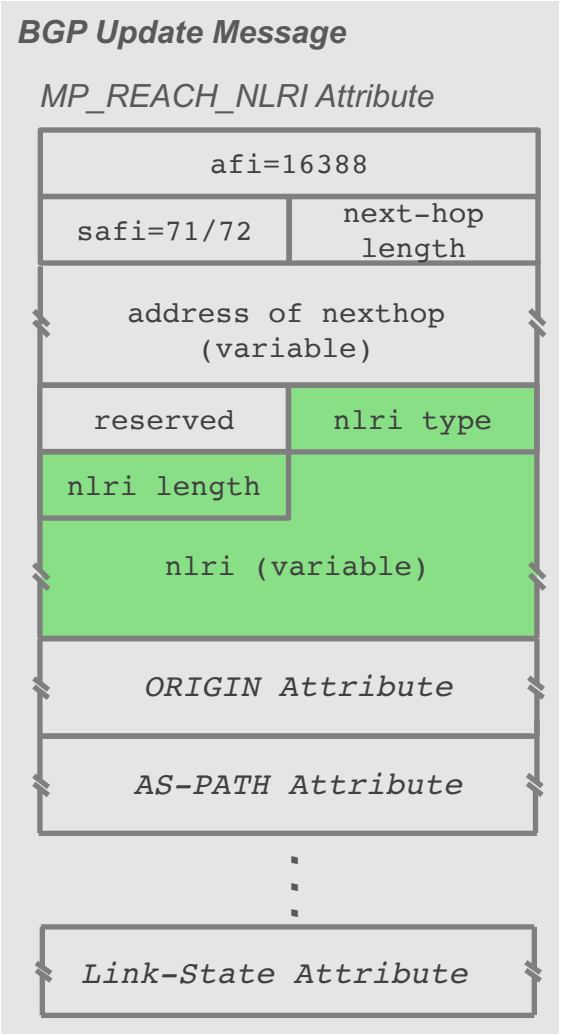


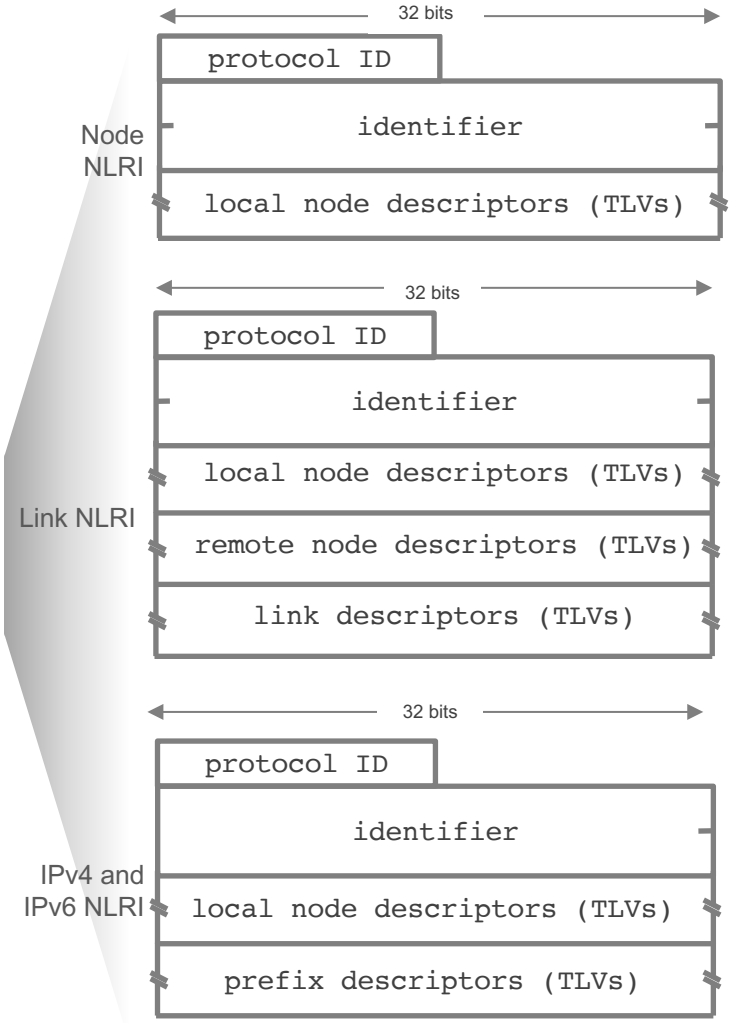
TLV	Description
512	Autonomous System
513	BGP-LS identifier
514	Area-ID
515	IGP Router-ID

TLV	Description
258	Link Local/Remote Identifiers
259	IPv4 interface address
260	IPv4 neighbour address
261	IPv6 interface address
262	IPv6 neighbour address
263	Multi-topology identifier

TLV	Description
263	Multi-topology identifier
264	OSPF route type
265	IP reachability information

Link-state NLRI



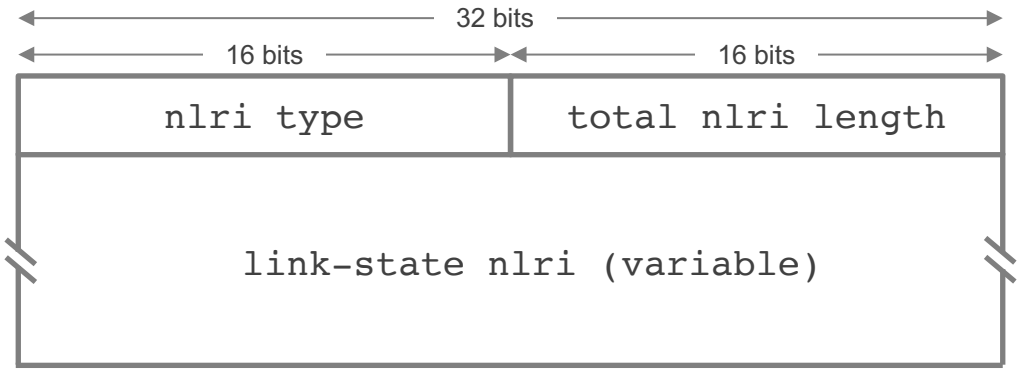


TLV	Description
512	Autonomous System
513	BGP-LS identifier
514	Area-ID
515	IGP Router-ID

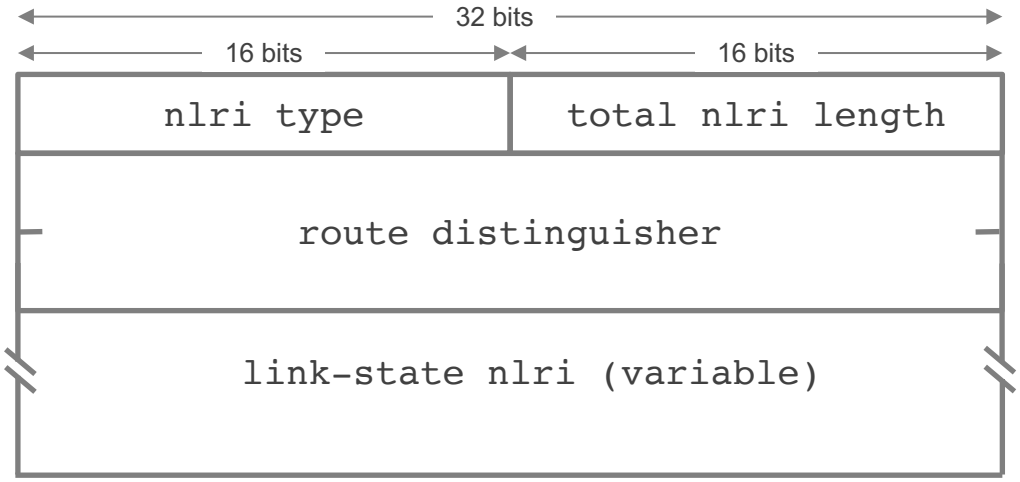
TLV	Description
258	Link Local/Remote Identifiers
259	IPv4 interface address
260	IPv4 neighbour address
261	IPv6 interface address
262	IPv6 neighbour address
263	Multi-topology identifier

TLV	Description
263	Multi-topology identifier
264	OSPF route type
265	IP reachability information

Link-state NLRI format (1)



Link-state NLRI (AFI 16388 / SAFI 71)
Non-VPN link-state



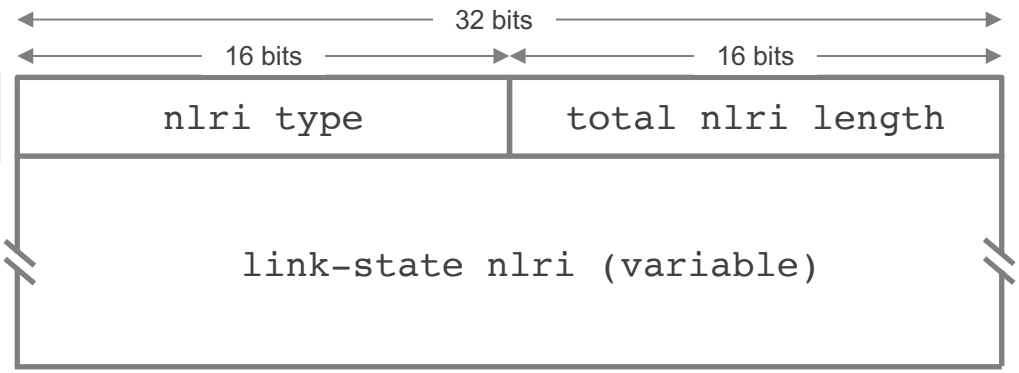
Link-state NLRI (AFI 16388 / SAFI 72)
VPN link-state

Link-state NLRI format (2)

NLRI types	
1	node nlri
2	link nlri
3	ipv4 topology prefix nlri
4	ipv6 topology prefix nlri

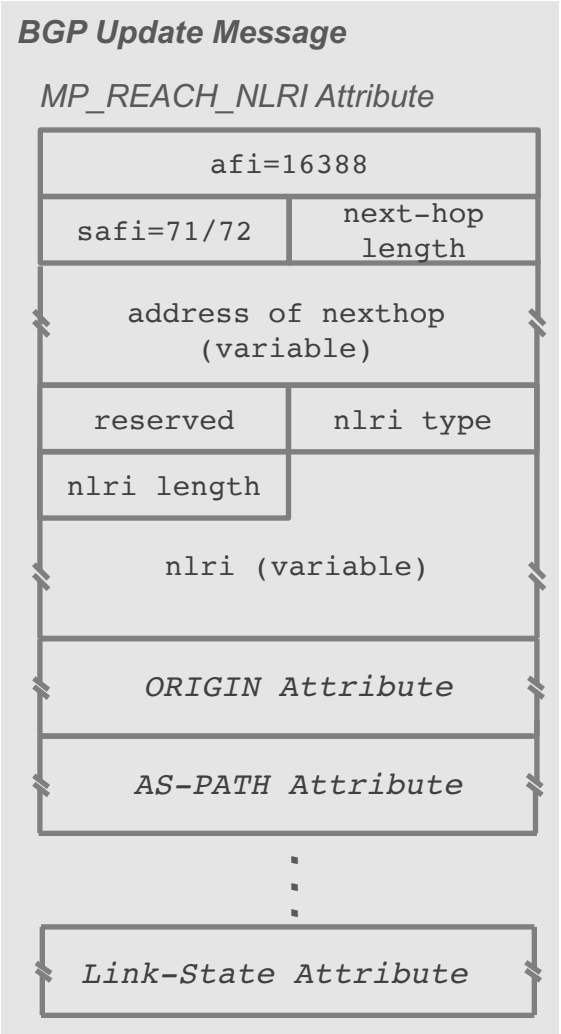
Link-state NLRI (AFI 16388 / SAFI 71)

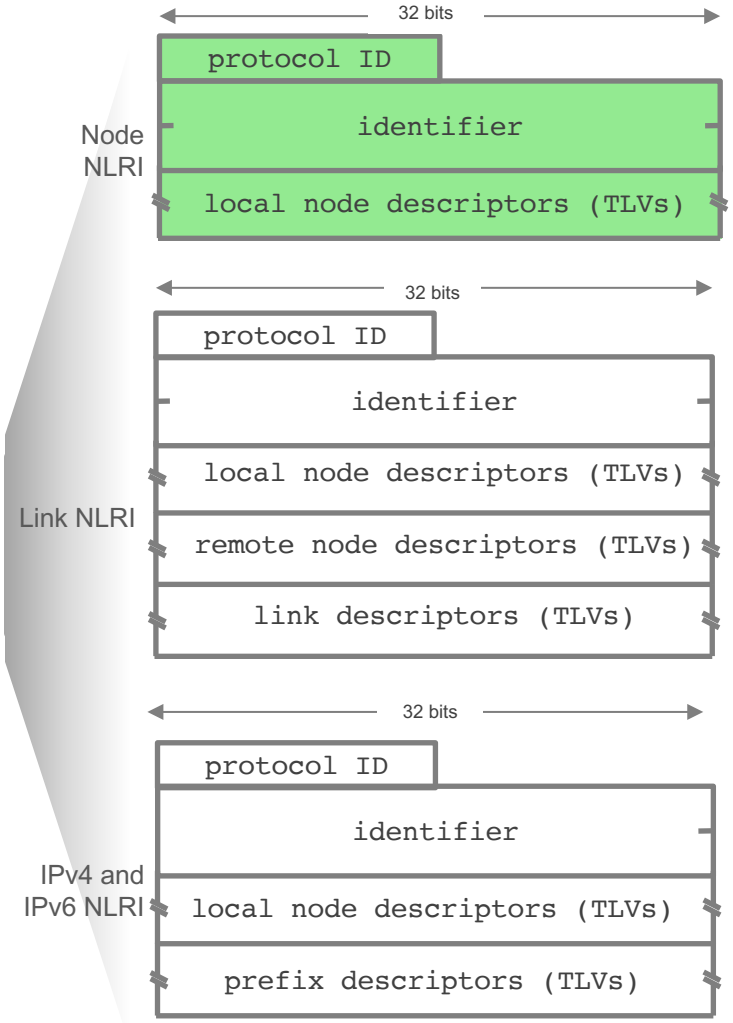
Non-VPN link-state



node nlri, link nlri or prefix nlri
(see following slides)

Node NLRI





TLV	Description
512	Autonomous System
513	BGP-LS identifier
514	Area-ID
515	IGP Router-ID

TLV	Description
258	Link Local/Remote Identifiers
259	IPv4 interface address
260	IPv4 neighbour address
261	IPv6 interface address
262	IPv6 neighbour address
263	Multi-topology identifier

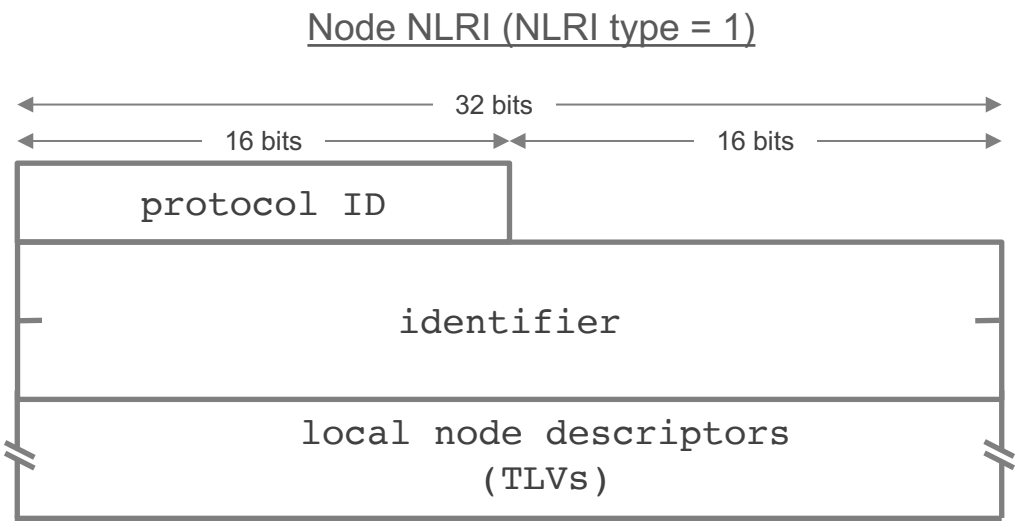
TLV	Description
263	Multi-topology identifier
264	OSPF route type
265	IP reachability information

Node NLRI

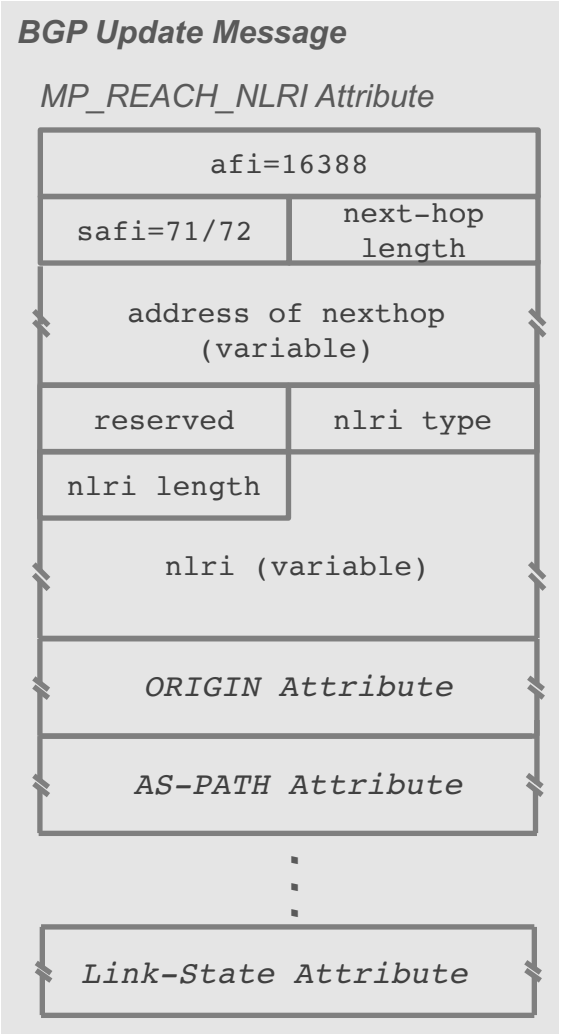
Protocol ID	
1	IS-IS L1
2	IS-IS L2
3	OSPFv2
4	Direct
5	Static
6	OSPFv3

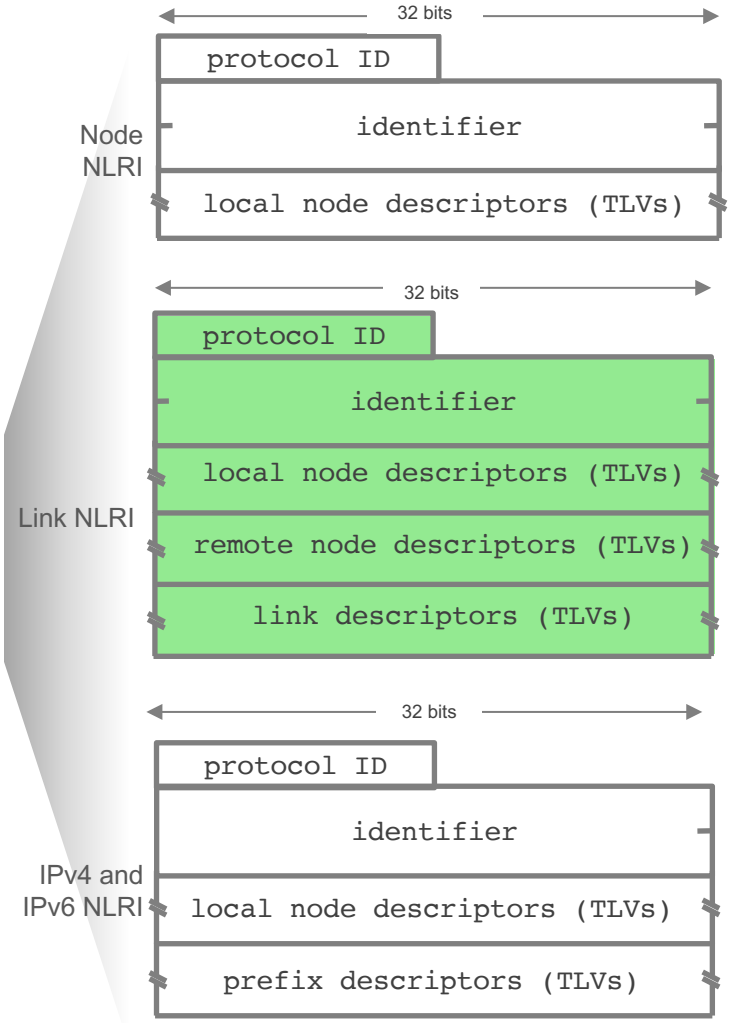
Identifies the IGP instance

identifier	
0	Default layer 3 routing topology



Link NLRI





TLV	Description
512	Autonomous System
513	BGP-LS identifier
514	Area-ID
515	IGP Router-ID

TLV	Description
258	Link Local/Remote Identifiers
259	IPv4 interface address
260	IPv4 neighbour address
261	IPv6 interface address
262	IPv6 neighbour address
263	Multi-topology identifier

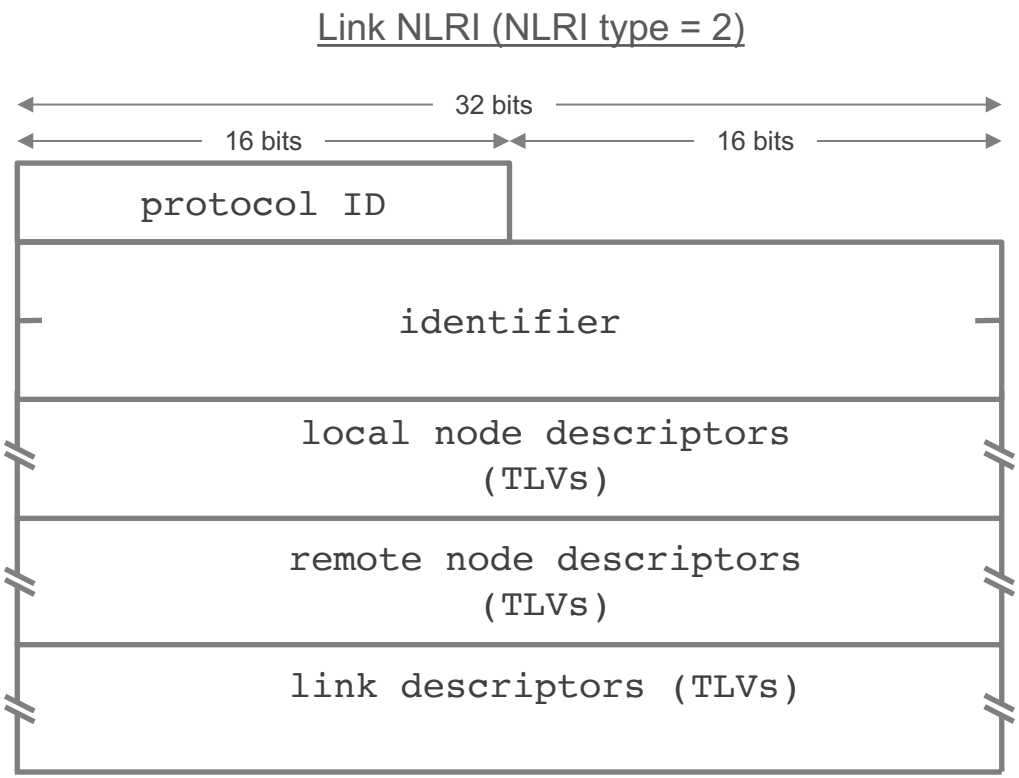
TLV	Description
263	Multi-topology identifier
264	OSPF route type
265	IP reachability information

Link NLRI format

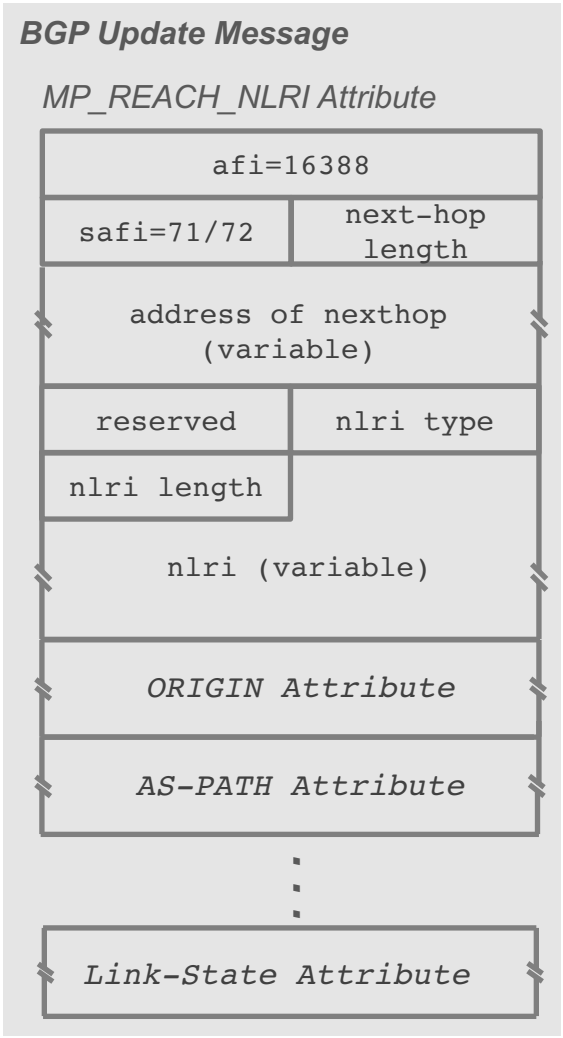
Protocol ID	
1	IS-IS L1
2	IS-IS L2
3	OSPFv2
4	Direct
5	Static
6	OSPFv3

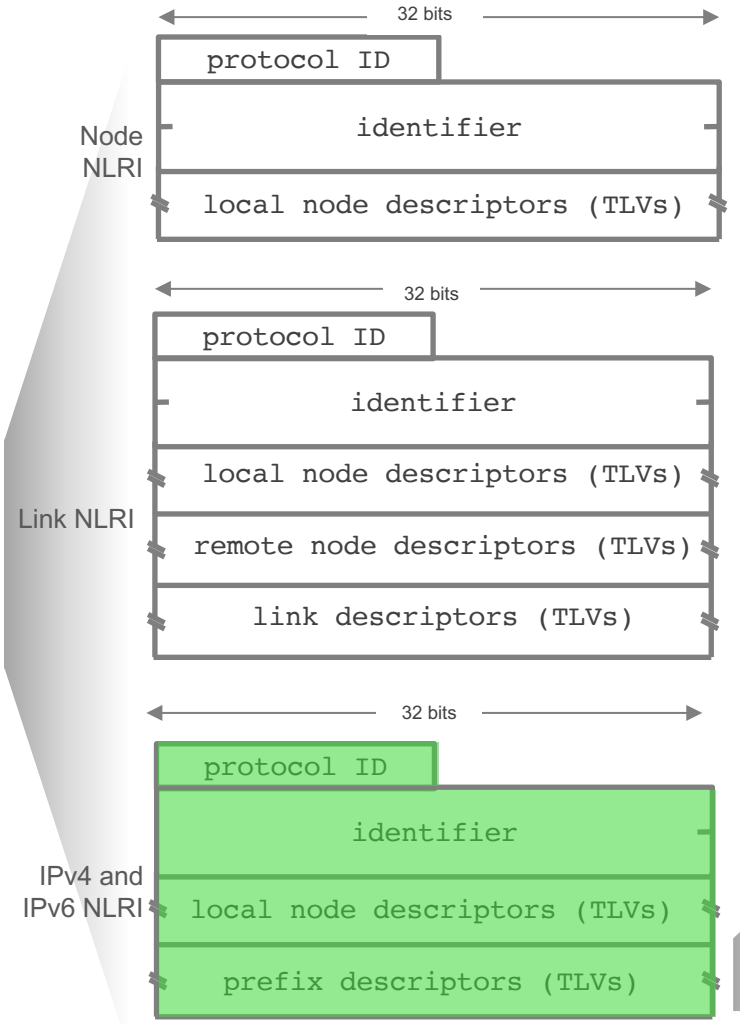
Identifies the IGP instance

identifier	
0	Default layer 3 routing topology



Prefix NLRI





TLV	Description
512	Autonomous System
513	BGP-LS identifier
514	Area-ID
515	IGP Router-ID

TLV	Description
258	Link Local/Remote Identifiers
259	IPv4 interface address
260	IPv4 neighbour address
261	IPv6 interface address
262	IPv6 neighbour address
263	Multi-topology identifier

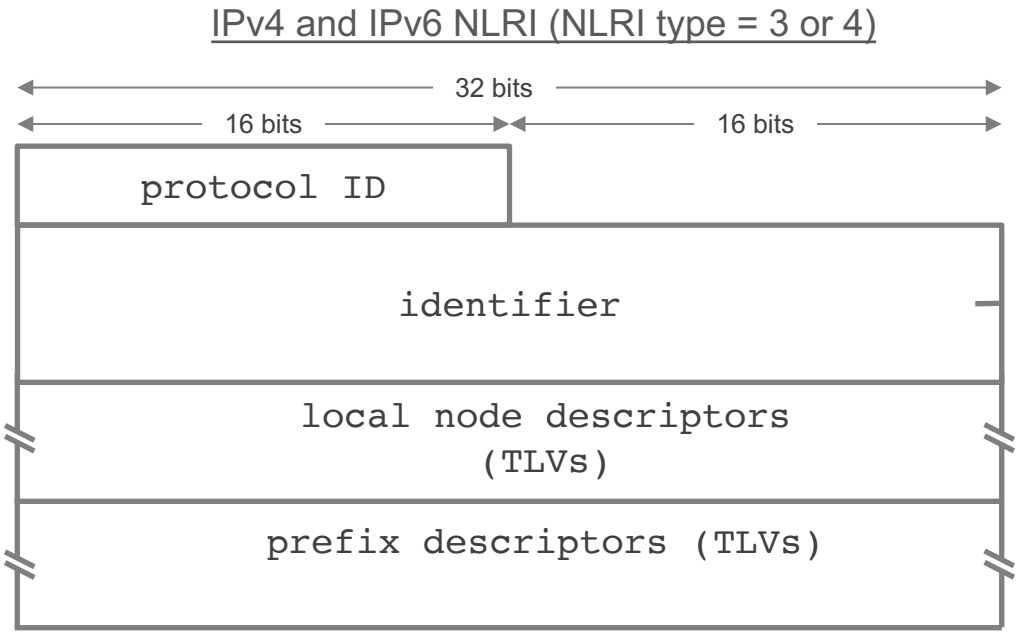
TLV	Description
263	Multi-topology identifier
264	OSPF route type
265	IP reachability information

IPv4 and IPv6 prefix NLRI format

Protocol ID	
1	IS-IS L1
2	IS-IS L2
3	OSPFv2
4	Direct
5	Static
6	OSPFv3

Identifies the IGP instance

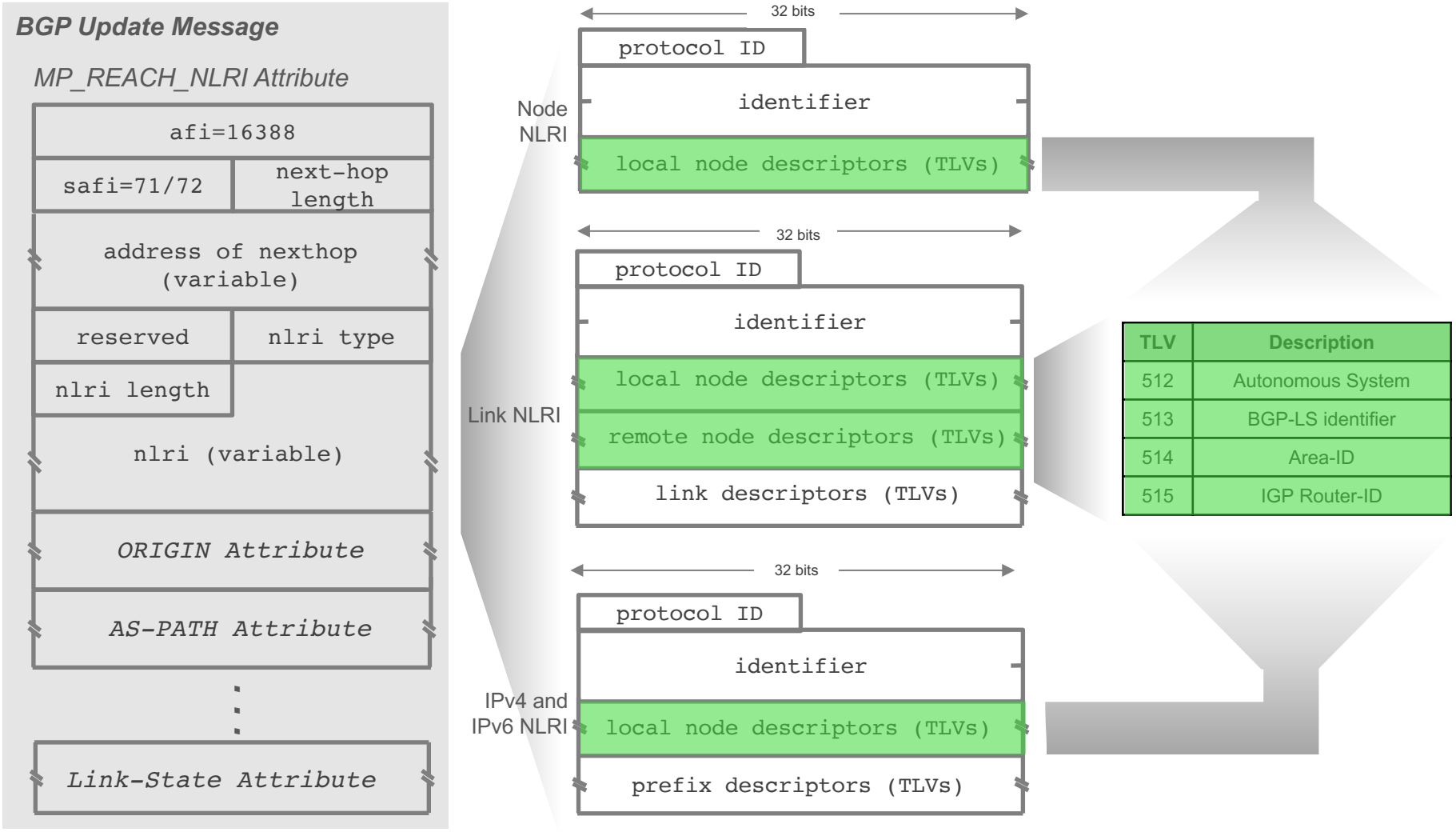
identifier	
0	Default layer 3 routing topology



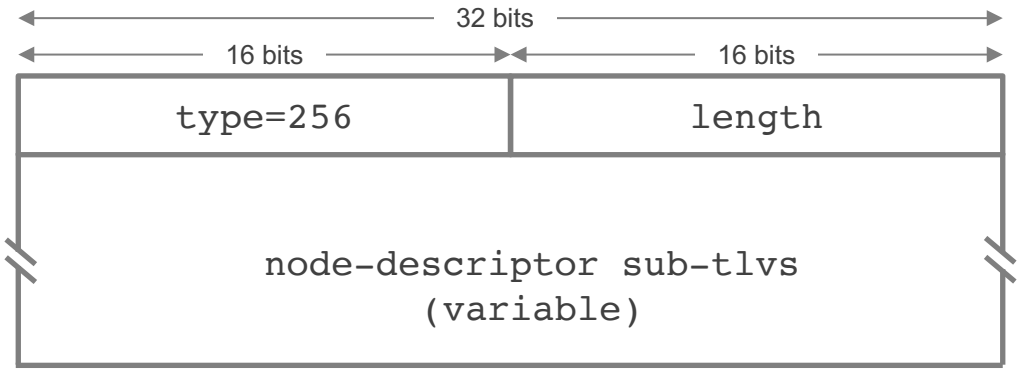
Node identification

- Router identification:
  - OSPF uses 32-bit Router-ID (does not even have to be routable)
  - IS-IS uses a 48-bit ISO System ID

Node descriptors



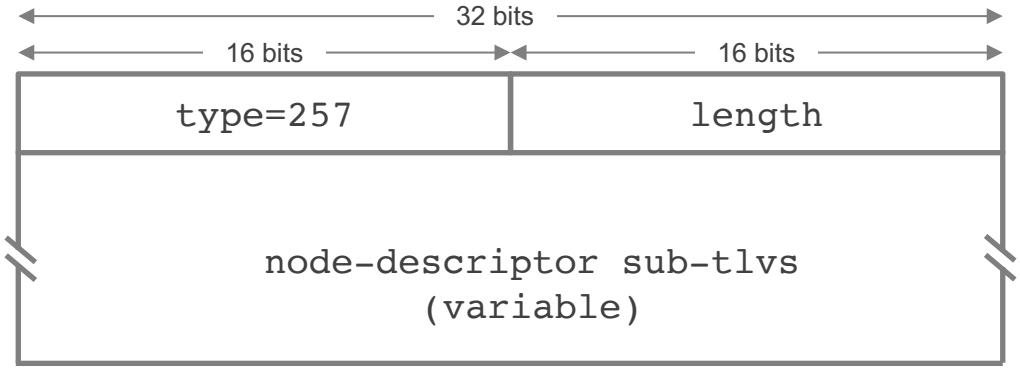
Node descriptor TLVs



Local Node Descriptors

Node descriptor for the local end of a link or the node itself

- Mandatory for all three types of NLRIs

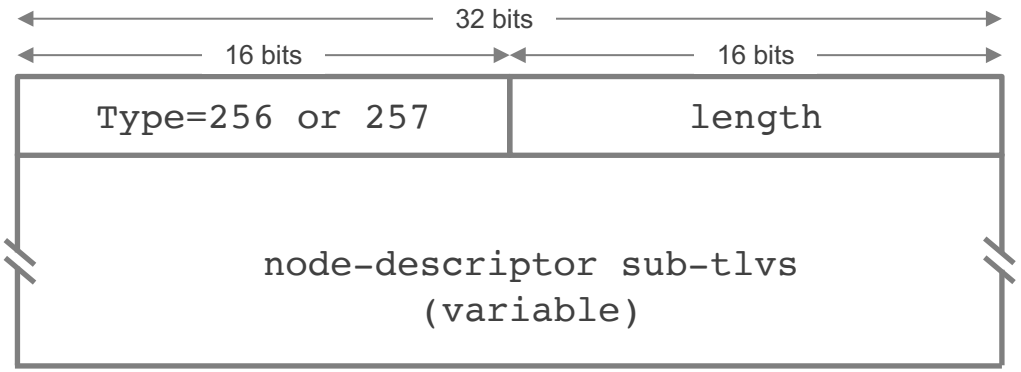


Remote Node Descriptors

Node descriptor for the node at the remote end of a link

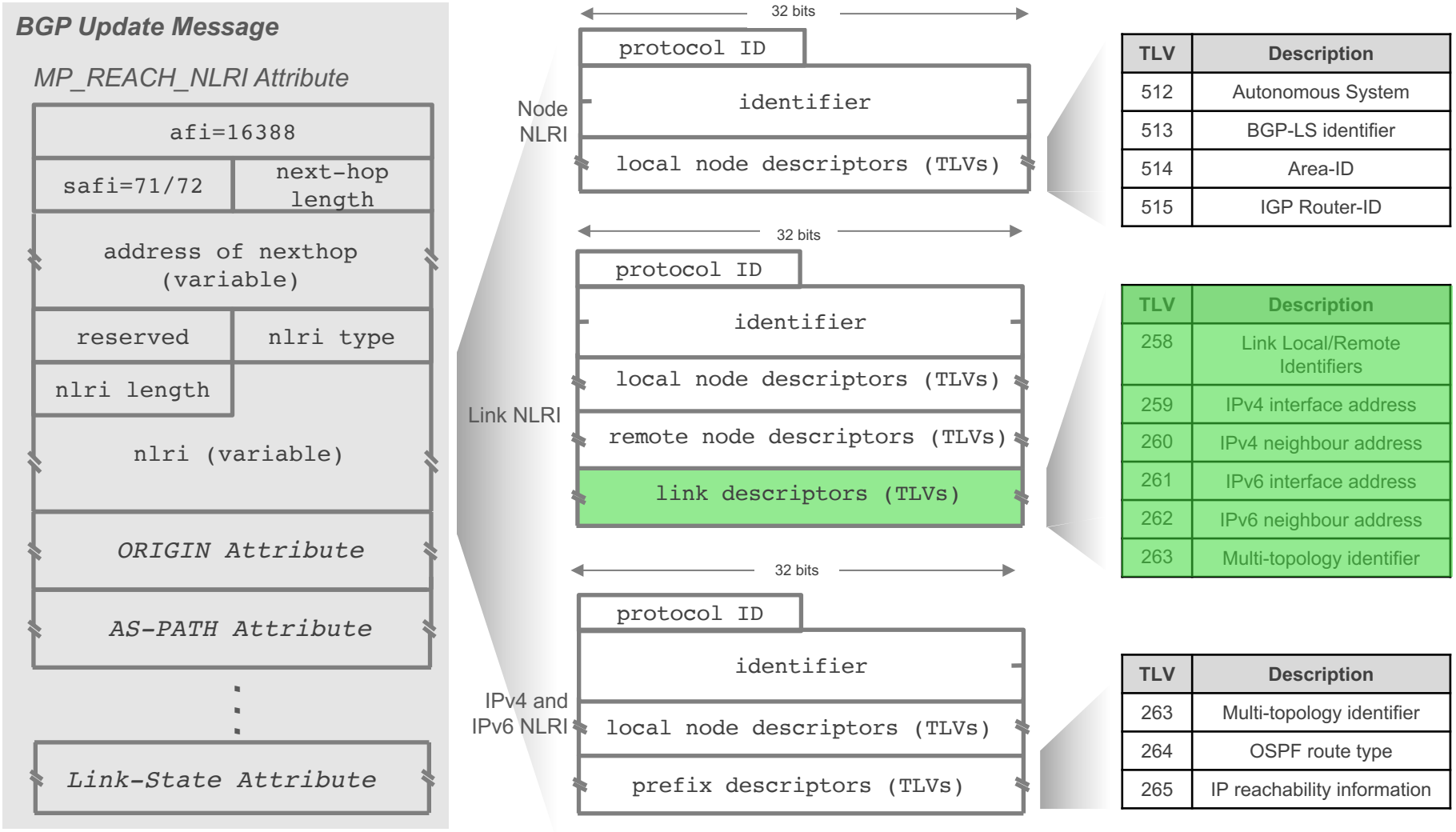
- Mandatory for Link NLRIs

Node descriptor sub-TLVs



Sub-TLV code point	Length	Description	Details
512	4	Autonomous System	32-bit ASN
513	4	BGP-LS identifier	Uniquely identifies BGP-LS domain; combination of ASN and BGP-LS ID has to be globally unique
514	4	Area-ID	IGP area ID
515	variable	IGP Router-ID	Either a 6-octet ISO node-ID or 4-octet OSPF router-ID

Link descriptors

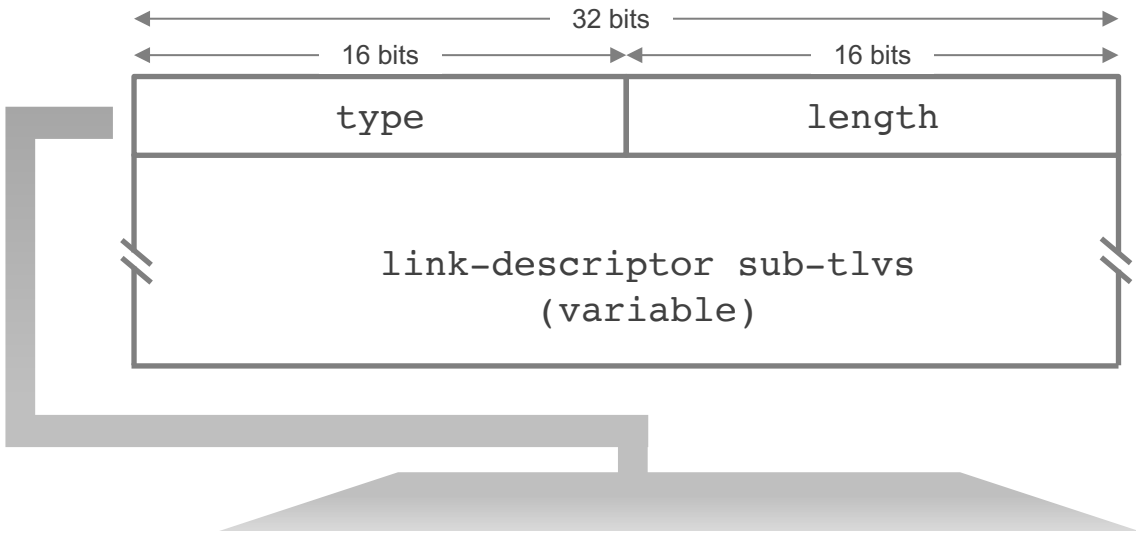


TLV	Description
512	Autonomous System
513	BGP-LS identifier
514	Area-ID
515	IGP Router-ID

TLV	Description
258	Link Local/Remote Identifiers
259	IPv4 interface address
260	IPv4 neighbour address
261	IPv6 interface address
262	IPv6 neighbour address
263	Multi-topology identifier

TLV	Description
263	Multi-topology identifier
264	OSPF route type
265	IP reachability information

Link descriptor TLVs (1)



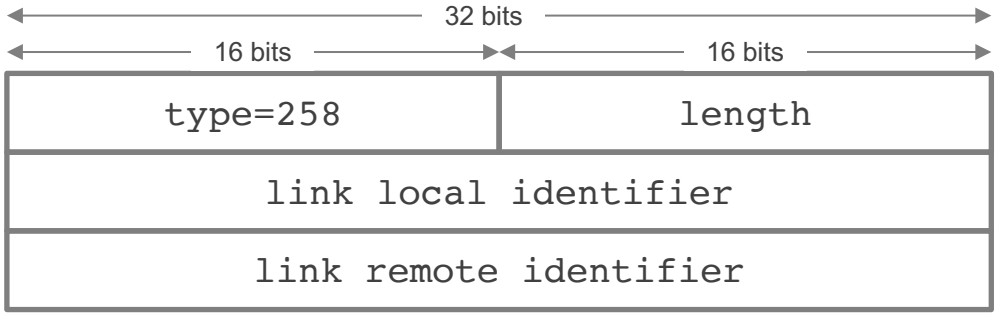
Link Descriptor

Uniquely identifies a link between multiple parallel links between a pair of routers. Actually a half-link: unidirectional logical link. The remote end advertises the other half.

TLV code point	Description	Corresponding ISIS TLV/sub-TLV
258	Link Local/Remote Identifiers	22/4
259	IPv4 interface address	22/6
260	IPv4 neighbour address	22/8
261	IPv6 interface address	22/12
262	IPv6 neighbour address	22/13
263	Multi-topology identifier	-

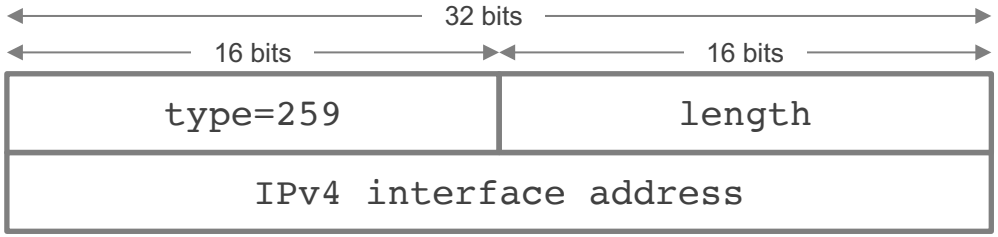
Format and semantics of the value fields of these TLVs is identical to ISIS Extended IS Reachability sub-TLVs (RFC5305, RFC5307, RFC6119). A subset of these TLVs is present.

Link descriptor TLVs (2)

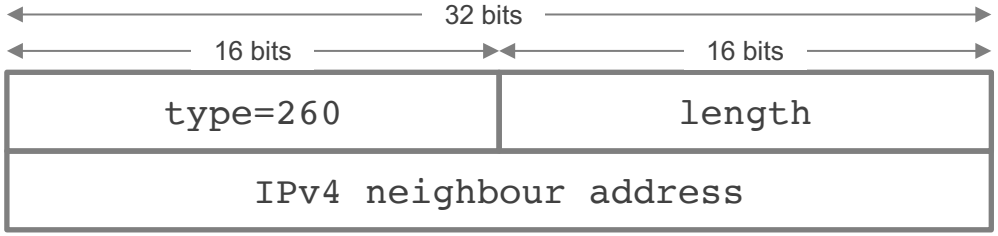


Link Local/Remote Identifiers TLV

Used for unnumbered interfaces

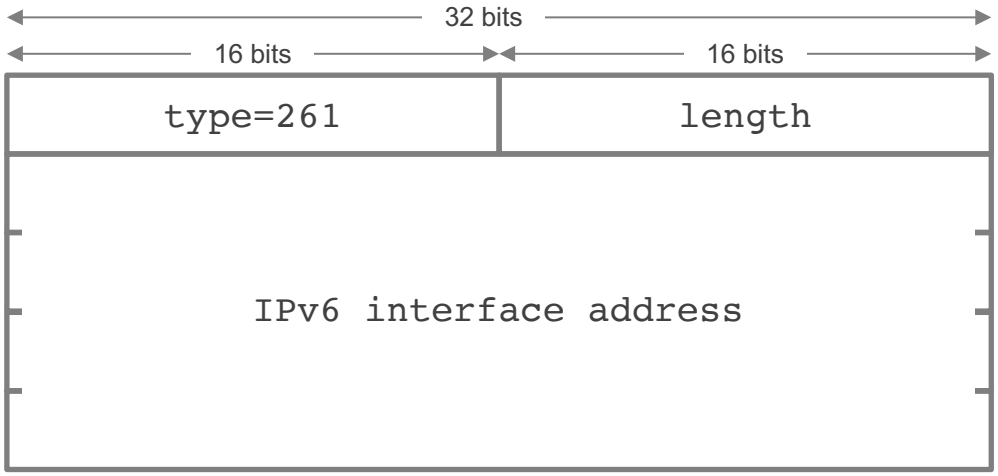


IPv4 Interface Address TLV

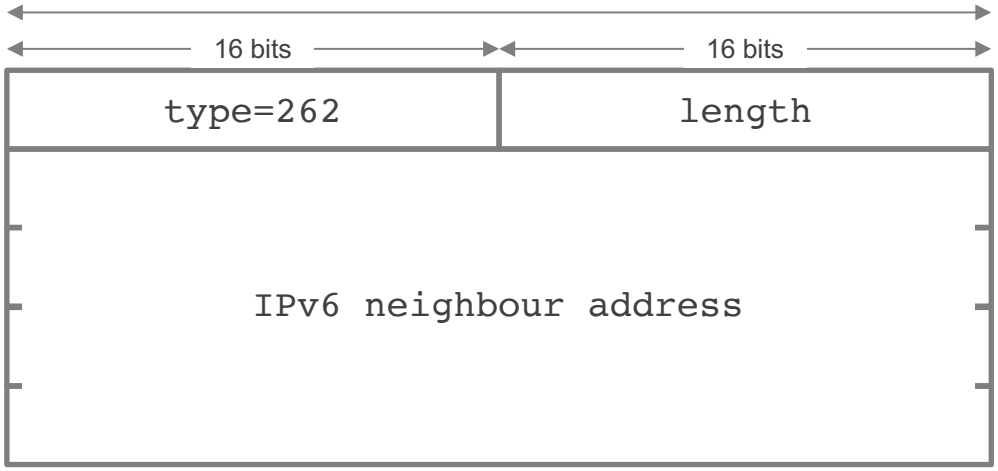


IPv4 Neighbour Address TLV

Link descriptor TLVs (3)

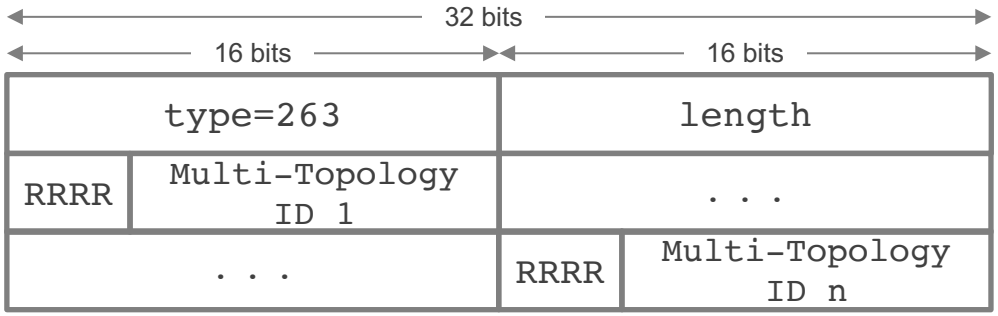


IPv6 Interface Address TLV



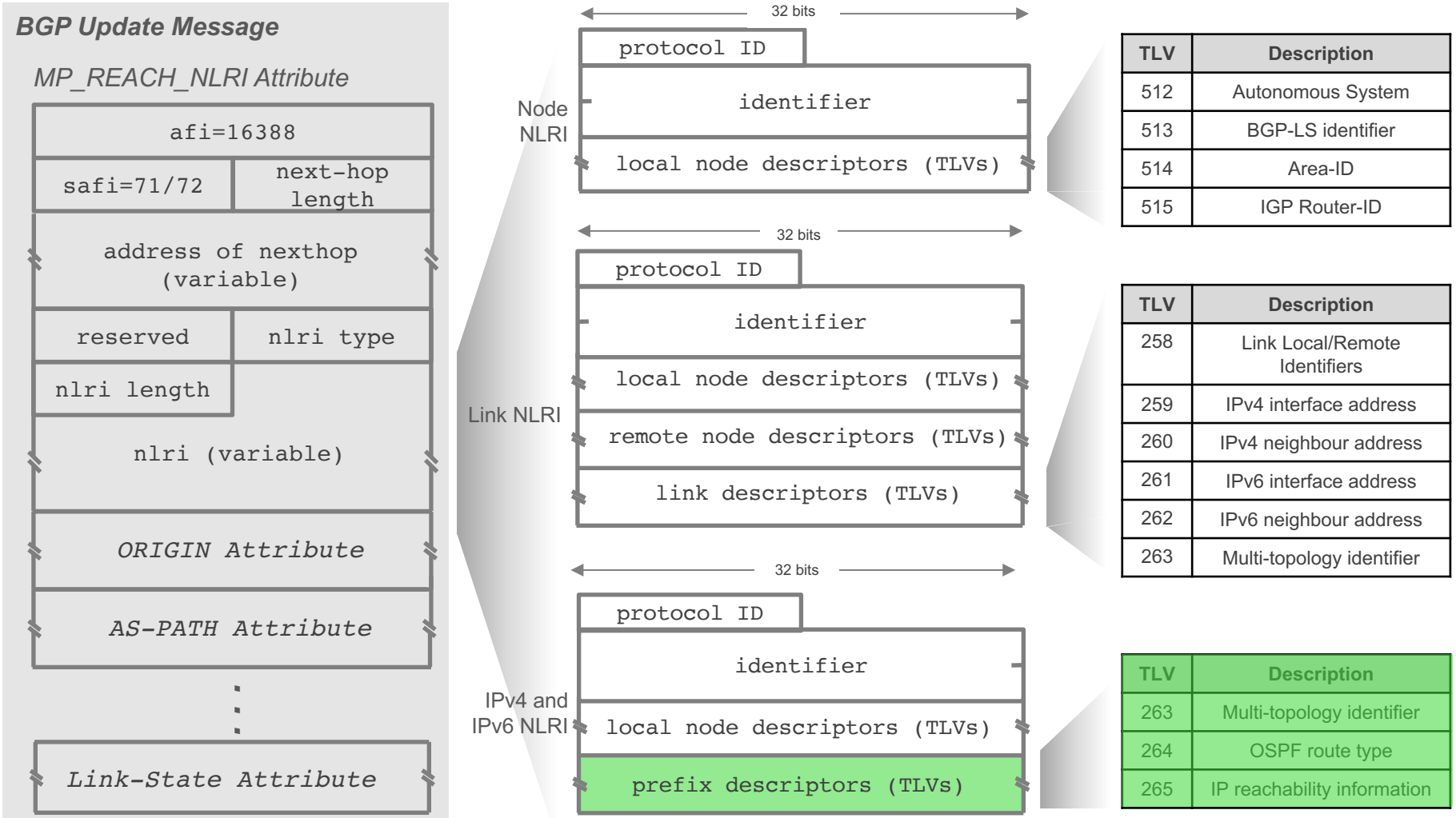
IPv6 Neighbour Address TLV

Link descriptor TLVs (4)



Multi-Topology ID TLV

Prefix descriptors

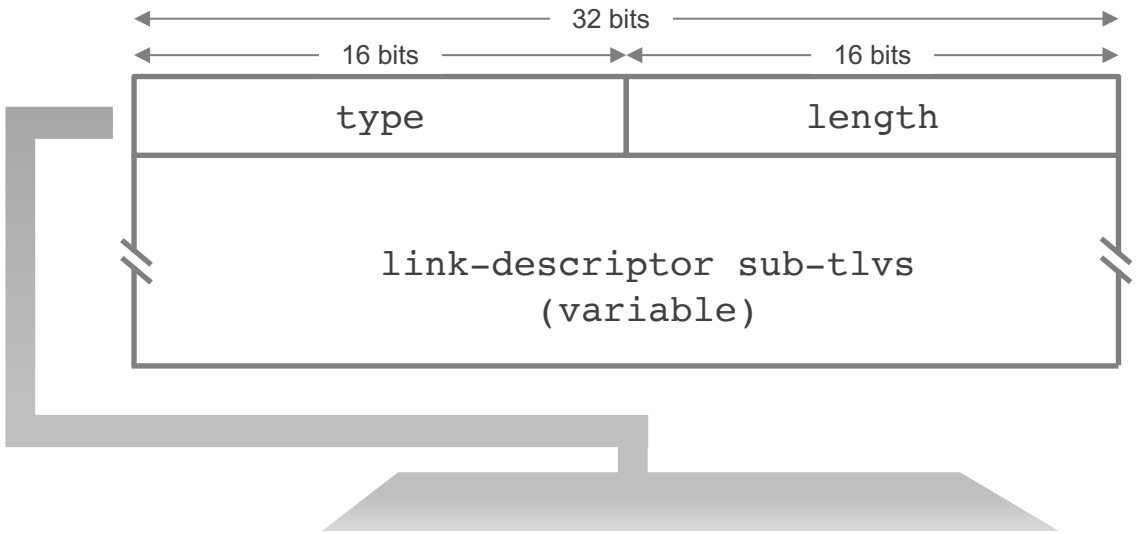


TLV	Description
512	Autonomous System
513	BGP-LS identifier
514	Area-ID
515	IGP Router-ID

TLV	Description
258	Link Local/Remote Identifiers
259	IPv4 interface address
260	IPv4 neighbour address
261	IPv6 interface address
262	IPv6 neighbour address
263	Multi-topology identifier

TLV	Description
263	Multi-topology identifier
264	OSPF route type
265	IP reachability information

Prefix descriptor TLVs (1)

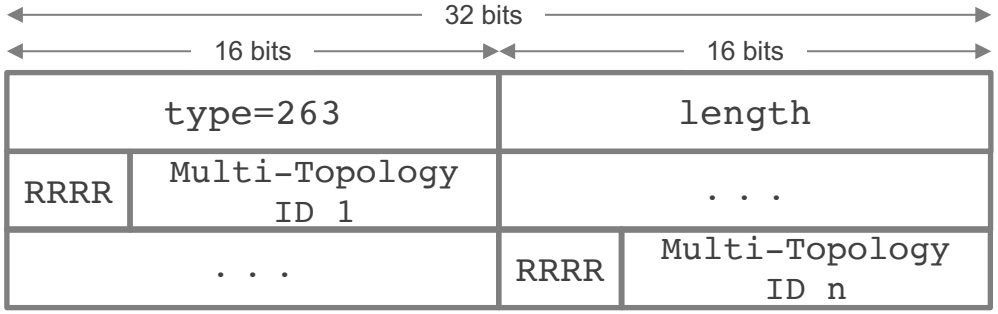


Prefix Descriptor

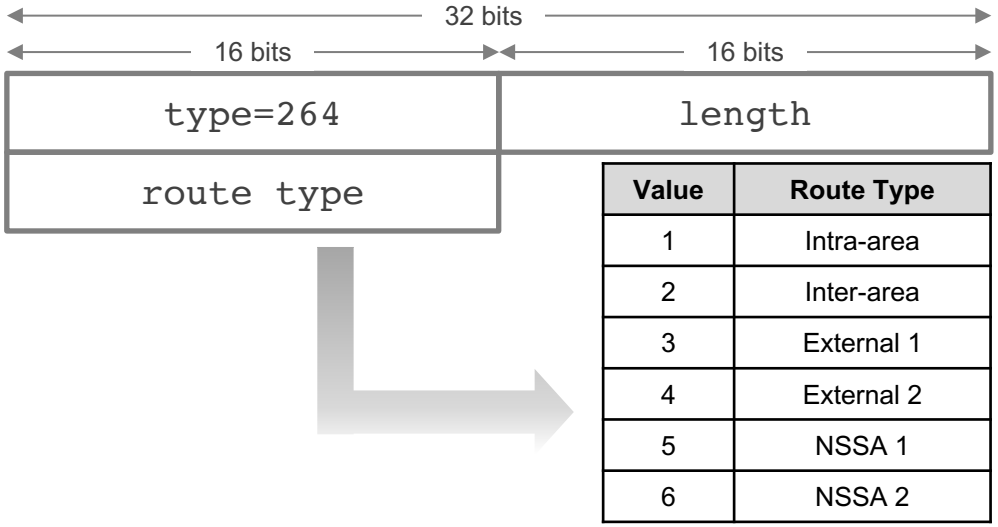
Uniquely identifies an IPv4 or IPv6 prefix originated by a node

TLV code point	Description
263	Multi-topology identifier
264	OSPF route type
265	IP reachability information

Prefix descriptor TLVs (2)



Multi-Topology ID TLV

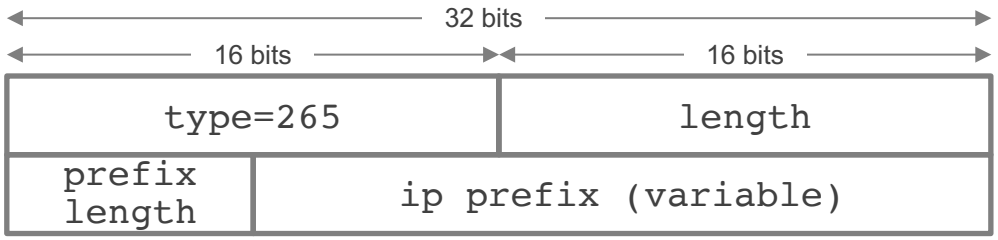


Value	Route Type
1	Intra-area
2	Inter-area
3	External 1
4	External 2
5	NSSA 1
6	NSSA 2

OSPF route type TLV

Identifies the OSPF route type of the prefix

Prefix descriptor TLVs (3)



IP reachability information TLV

Contains one IP prefix originally advertised in the IGP

Next-hop

- BGP-LS is supported over both IPv4 and IPv6 BGP sessions.

```
IF address_family(BGP session) = IPv6 THEN
    address_family(next-hop in MP_REACH_NLRI) = IPv6_address
ELSE /* must be IPv4 */
    address_family(next-hop in MP_REACH_NLRI) = IPv4_address
```

- Generally, the next-hop will be set to the local endpoint address of the BGP session

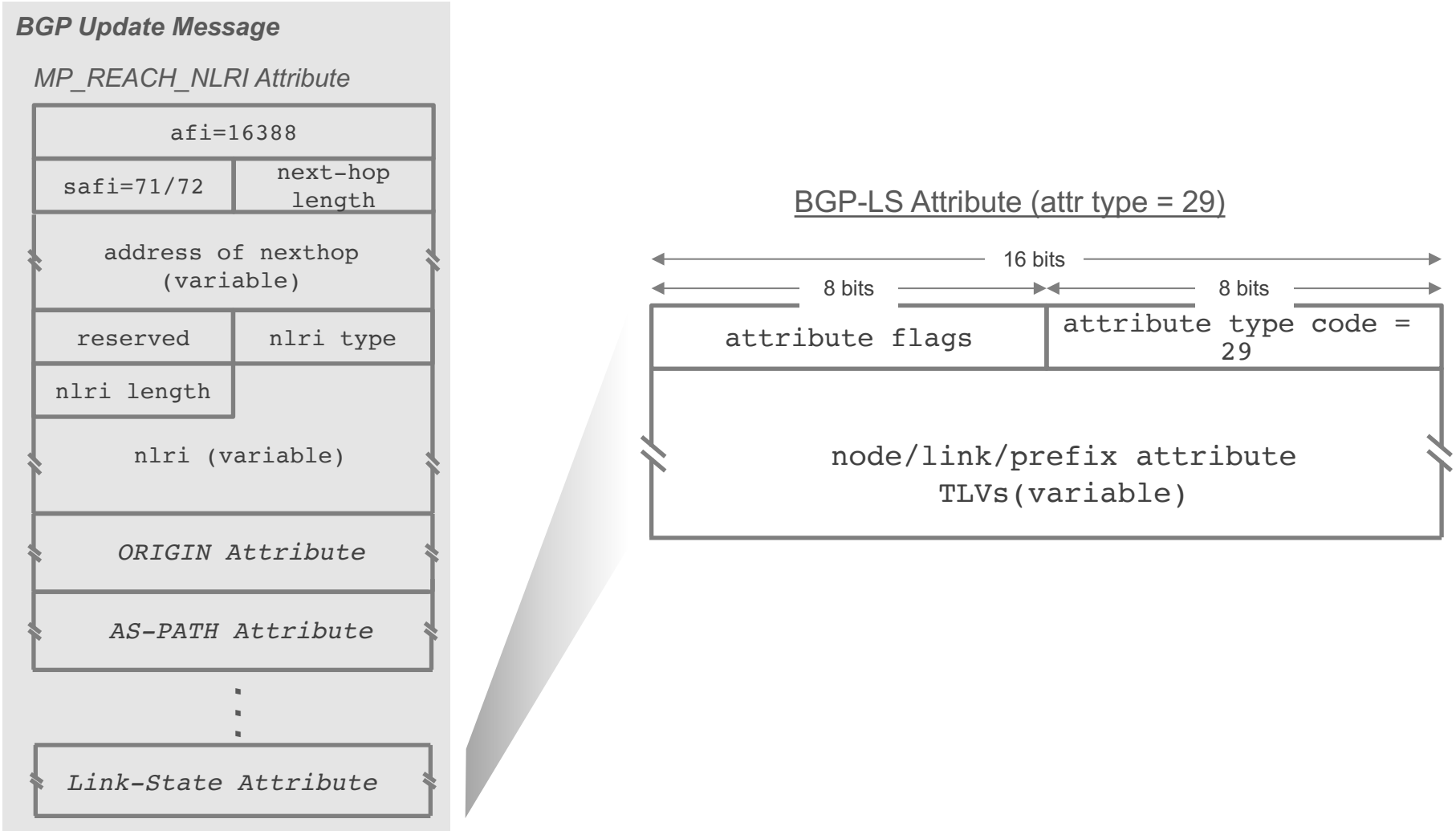
Link-state Path Attribute

- Node attribute TLVs
- Link attribute TLVs
- Prefix attribute TLVs

BGP-LS attribute

- Optional, non-transitive BGP attribute
- Attribute type value assigned by IANA is 29
- Carries link, node and prefix parameters and attributes corresponding to the nodes, links or prefixes carried in the link-state NLRI
- Attribute is only applicable when included with link-state NLRIs

Link-state attribute



Node attribute TLVs

- Node attribute TLVs that may be encoded in the BGP-LS path attribute when advertised together with a Node NLRI.

TLV code point	Name	Description
263	Multi-topology identifier	Carries MT-IDs of all topologies where the node is reachable
1024	Node flag bits	Overload, Attached, External, ABR, Router, V6 bites
1025	Opaque node attribute	Used for possible future IGP extensions not natively supported by BGP-LS
1026	Node name	Symbol name (e.g. FQDN) of the router
1027	IS-IS area identifier	All of the IS-IS areas a node is part of
1028	IPv4 router-ID of local node	Auxiliary router-ID that a node may be using e.g. for TE
1029	IPv6 router-ID of local node	Auxiliary router-ID that a node may be using e.g. for TE

Link attribute TLVs (1)

- Link attribute TLVs that may be encoded in the BGP-LS path attribute when advertised together with a Link NLRI.

TLV code point	Name	Description
1028	IPv4 router-ID of local node	Auxiliary router-ID that the local node may be using e.g. for TE
1029	IPv6 router-ID of local node	Auxiliary router-ID that the local node may be using e.g. for TE
1030	IPv4 router-ID of remote node	Auxiliary router-ID that the remote node may be using e.g. for TE
1031	IPv6 router-ID of remote node	Auxiliary router-ID that the remote node may be using e.g. for TE
1088	Admin group (colour)	Bitmask of supported admin groups
1089	Max link bandwidth	Maximum bandwidth of the link in this direction
1090	Max reservable link bandwidth	Maximum bandwidth of the link that can be reserved in this direction

Link attribute TLVs (2)

- Link attribute TLVs that may be encoded in the BGP-LS path attribute when advertised together with a Link NLRI.

TLV code point	Name	Description
1091	Unreserved bandwidth	Amount of bandwidth of the link that is available for reservation in this direction
1092	TE default metric	TE metric for the link
1093	Link protection type	Bitmask of supported protection types
1094	MPLS protocol mask	Bitmask describing which MPLS protocols are enabled: LDP, RSVP-TE
1095	IGP metric	IGP metric for the link
1096	Shared risk link group	List of SRLG values
1097	Opaque link attribute	Used for possible future IGP extensions not natively supported by BGP-LS
1098	Link name	Symbol name (e.g. FQDN) for the link

Prefix attribute TLVs

- Prefix attribute TLVs that may be encoded in the BGP-LS path attribute when advertised together with a Prefix NLRI.

TLV code point	Name	Description
1152	IGP flags	IS-IS up/down bit, OSPF “no unicast bit”, OSPF “local address” bit, OSPF “propagate NSSA bit”
1153	IGP route tag	Original IGP tags of the prefix
1154	IGP extended route tag	IS-IS extended route tags
1155	Prefix metric	Metric of the prefix as known in the IGP topology
1156	OSPF forwarding address	OSPF forwarding address as known in the original OSPF advertisement
1157	Opaque prefix attribute	Used for possible future IGP extensions not natively supported by BGP-LS

TLV Summary

Summary: TLV codepoints (1)

TLV code point	Description	Usage
256	Local node descriptors	Node NLRI, Link NLRI, Prefix NLRI
257	Remote node descriptors	Link NLRI
258	Link Local/remote identifiers	Link NLRI
259	IPv4 interface address	Link NLRI
260	IPv4 neighbour address	Link NLRI
261	IPv6 interface address	Link NLRI
262	IPv6 neighbour address	Link NLRI
263	Multi-topology identifier	Link NLRI, Prefix NLRI, Link-state attribute with node NLRI
264	OSPF route type	Prefix NLRI
265	IP reachability information	Prefix NLRI
512	Autonomous System	Remote/Local node descriptor TLVs
513	BGP-LS identifier	Remote/Local node descriptor TLVs
514	Area-ID	Remote/Local node descriptor TLVs
515	IGP Router-ID	Remote/Local node descriptor TLVs

Summary: TLV codepoints (2)

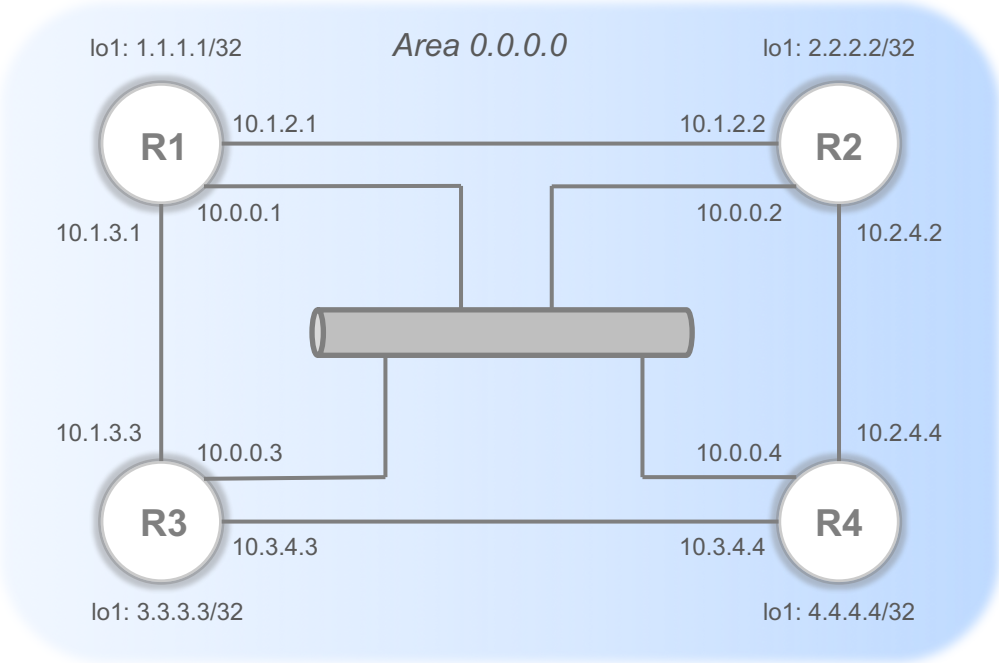
TLV code point	Description	Usage
1024	Node flag bits	Link-state attribute with node NLRI
1025	Opaque node attribute	Link-state attribute with node NLRI
1026	Node name	Link-state attribute with node NLRI
1027	IS-IS area identifier	Link-state attribute with node NLRI
1028	IPv4 router-ID of local node	Link-state attribute with node/link NLRI
1029	IPv6 router-ID of local node	Link-state attribute with node/link NLRI
1030	IPv4 router-ID of remote node	Link-state attribute with link NLRI
1031	IPv6 router-ID of remote node	Link-state attribute with link NLRI
1088	Admin group (colour)	Link-state attribute with link NLRI
1089	Max link bandwidth	Link-state attribute with link NLRI
1090	Max reservable link bandwidth	Link-state attribute with link NLRI
1091	Unreserved bandwidth	Link-state attribute with link NLRI

Summary: TLV codepoints (3)

TLV code point	Description	Usage
1093	Link protection type	Link-state attribute with link NLRI
1094	MPLS protocol mask	Link-state attribute with link NLRI
1095	IGP metric	Link-state attribute with link NLRI
1096	Shared risk link group	Link-state attribute with link NLRI
1097	Opaque link attribute	Link-state attribute with link NLRI
1098	Link name	Link-state attribute with link NLRI
1152	IGP flags	Link-state attribute with prefix NLRI
1153	IGP route tag	Link-state attribute with prefix NLRI
1154	IGP extended route tag	Link-state attribute with prefix NLRI
1155	Prefix metric	Link-state attribute with prefix NLRI
1156	OSPF forwarding address	Link-state attribute with prefix NLRI
1157	Opaque prefix attribute	Link-state attribute with prefix NLRI

Example

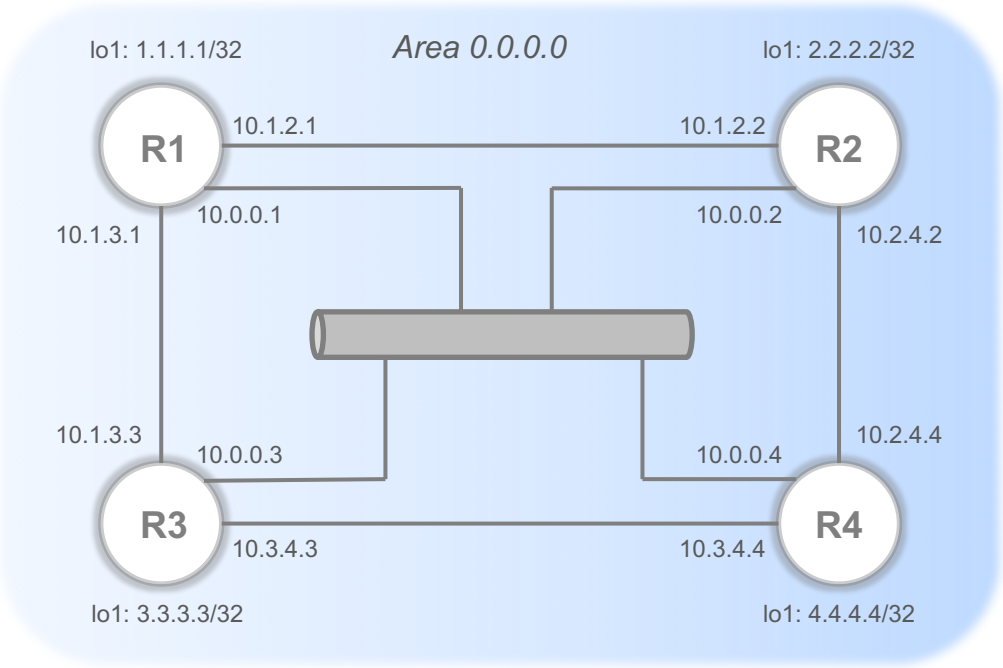
OSPF network



OSPF network design

- *Single area:0.0.0.0*
- *All links identical*
- *Default metrics*
- *4 point-to-point links*
- *1 broadcast interface on each router*
- *Router-ID is lowest configured IP address*
- *R4 is the DR for the broadcast network*

OSPF link-state database



Router (type 1) LSAs

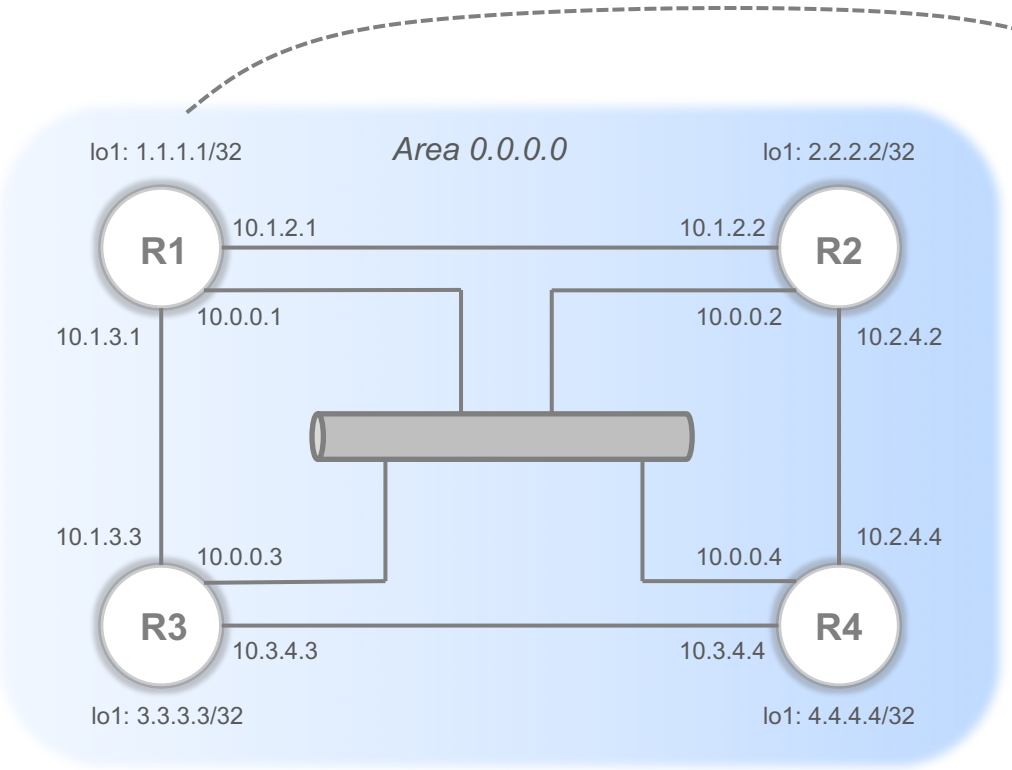
Link State ID	Advertising Router
1.1.1.1	1.1.1.1
2.2.2.2	2.2.2.2
3.3.3.3	3.3.3.3
4.4.4.4	4.4.4.4

Network (type 2) LSAs

Link State ID	Advertising Router
10.0.0.4	4.4.4.4

Identical database on all nodes within the area

R1's router-LSA



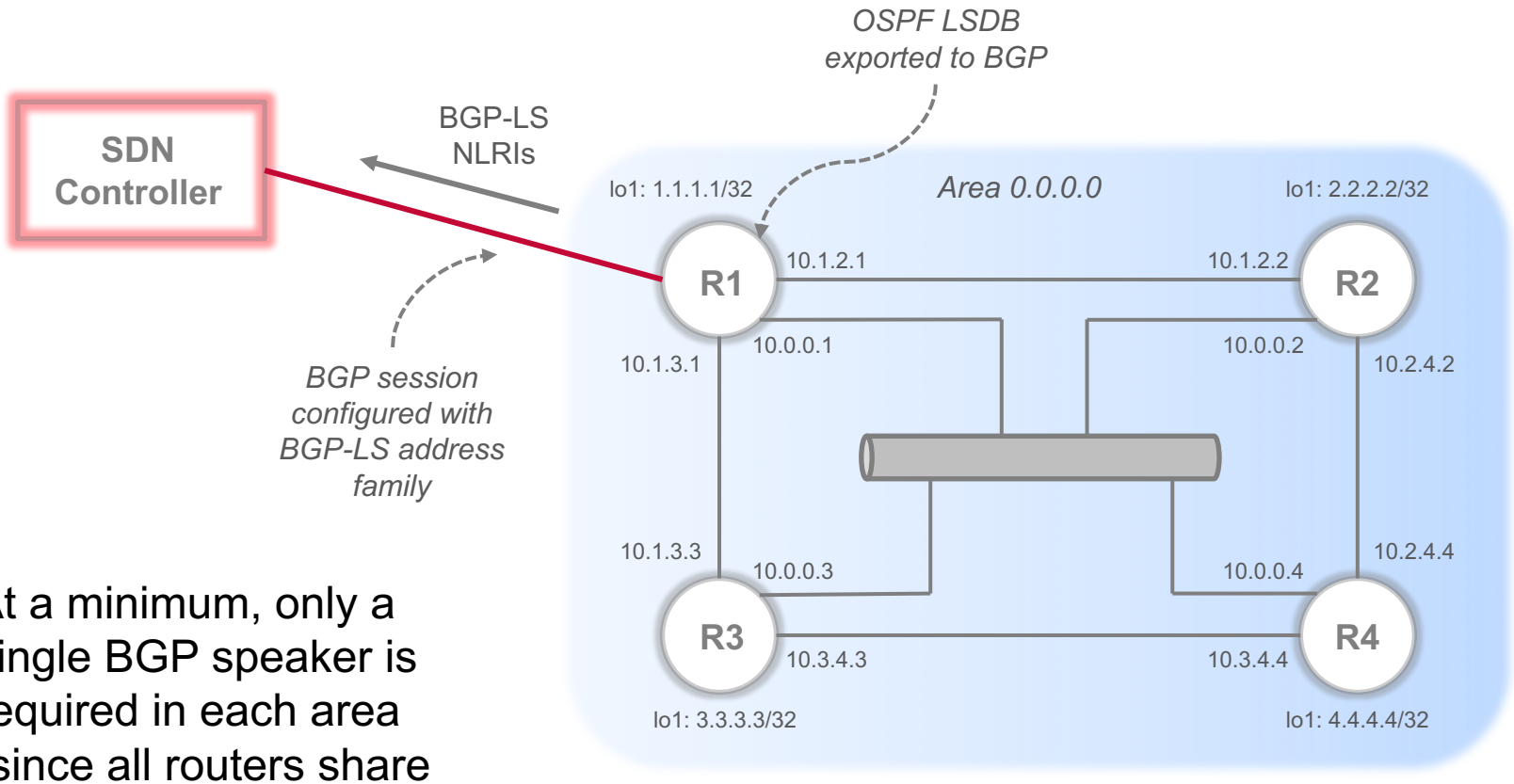
Router (type 1) LSAs

Link State ID	Advertising Router
1.1.1.1	1.1.1.1

links: 6

Type	Link ID	Link Data
Stub	1.1.1.1	255.255.255.255
Stub	10.1.2.0	255.255.255.0
Stub	10.1.3.0	255.255.255.0
Point-to-point	2.2.2.2 (neighbor router ID)	10.1.2.1 (intf addr)
Point-to-point	3.3.3.3 (neighbor router ID)	10.1.3.1 (intf addr)
Transit	10.0.0.4 (DR)	10.0.0.1 (intf addr)

Introducing BGP-LS into the network



- At a minimum, only a single BGP speaker is required in each area (since all routers share the same LSDB)

R1's Link-state NLRIs (1)

Router (type 1) LSAs

Link State ID	Advertising Router
1.1.1.1	1.1.1.1

links: 6

Type	Link ID	Link Data
Stub	1.1.1.1	255.255.255.255
Stub	10.1.2.0	255.255.255.0
Stub	10.1.3.0	255.255.255.0
Point-to-point	2.2.2.2 (neighbor router ID)	10.1.2.1 (intf addr)
Point-to-point	3.3.3.3 (neighbor router ID)	10.1.3.1 (intf addr)
Transit	10.0.0.4 (DR)	10.0.0.1 (intf addr)

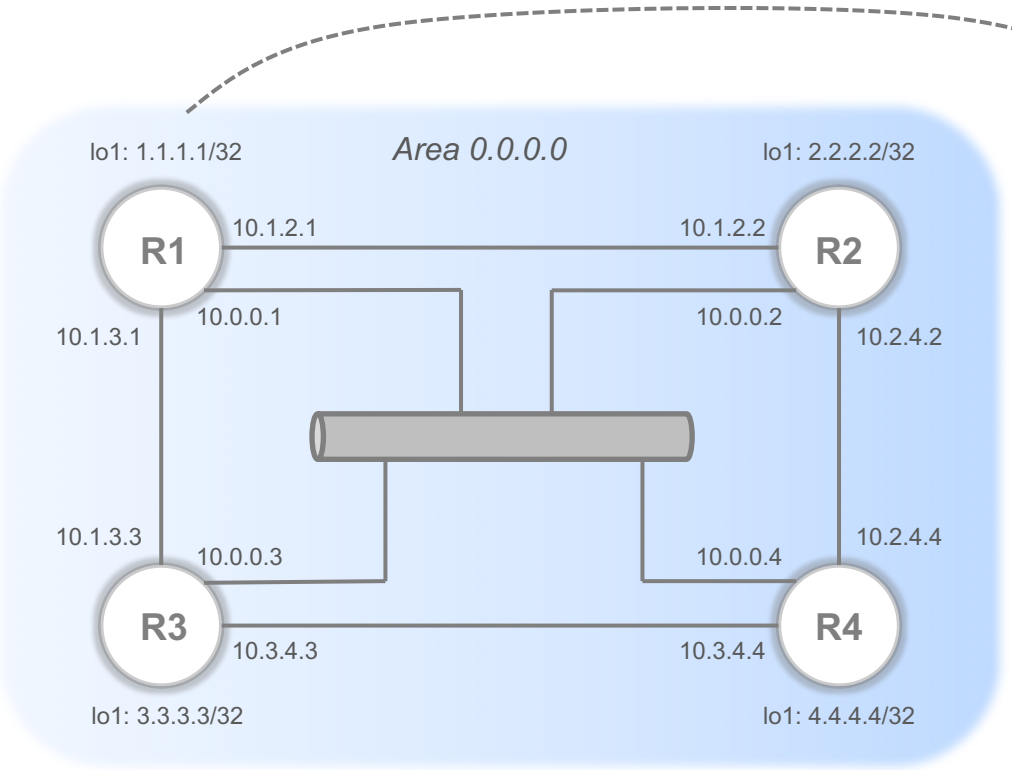
R1's BGP-LS NLRIs

Type	Description
Node	Local node descriptor

Type	Description
Link	Link to R2
Link	Link to R3
Link	Link to broadcast network

Type	Description
Prefix	10.1.2.0/24
Prefix	10.1.3.0/24
Prefix	1.1.1.1/32

R1's link-state NLRIs (2)



Type	Description
Node	Local node descriptor

Type	Description
Link	Link to R2
Link	Link to R3
Link	Link to broadcast network

Type	Description
Prefix	10.1.2.0/24
Prefix	10.1.3.0/24
Prefix	1.1.1.1/32

All link-state NLRIs (1)

- For each router:
  - 1 x Node NLRI
  - 3 x Link NLRIs:
    - 2 for the point-to-point links
    - 1 for the broadcast network
  - 3 x Prefix NLRIs
    - 2 for the point-to-point links
    - 1 for the loopback interface
  - TOTAL: 28 link-state NLRIs
- For the broadcast network:
  - 1 x Node NLRI
    - For the DR
  - 4 x Link NLRIs:
    - For each DR to node adjacency
  - TOTAL: 5 link-state NLRIs

Total of 33 BGP-LS NLRIs generated for this OSPF network

All link-state NLRIs (2)

Type	Local Node	Description
Node	R1	Local node descriptor
Node	R2	Local node descriptor
Node	R3	Local node descriptor
Node	R4	Local node descriptor
Node	R4 (DR)	Local node descriptor

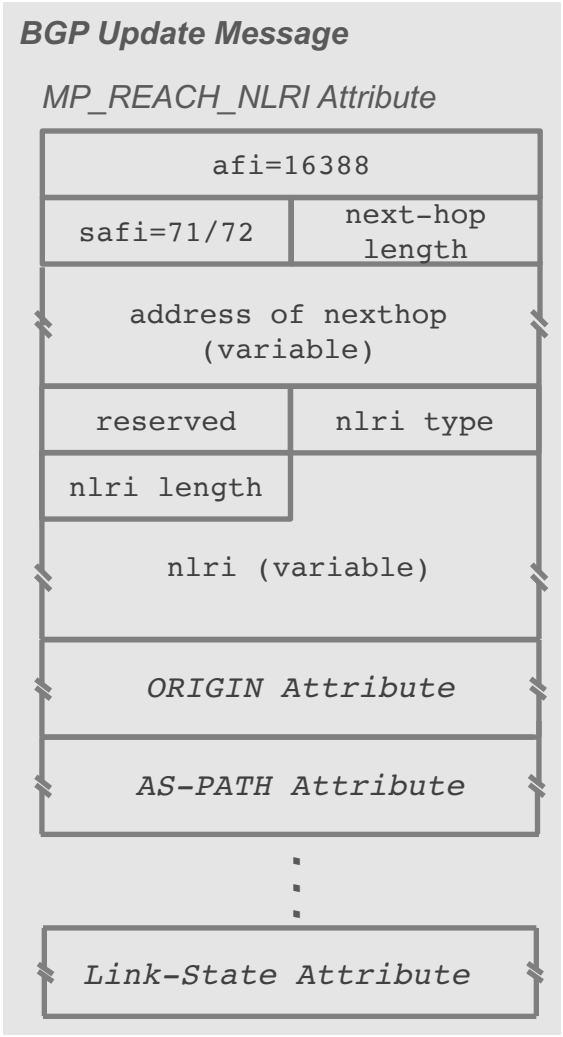
Type	Local Node	Description
Prefix	R1	10.1.2.0/24
Prefix	R1	10.1.3.0/24
Prefix	R1	1.1.1.1/32
Prefix	R2	10.1.2.0/24
Prefix	R2	10.2.4.0/24
Prefix	R2	2.2.2.2/32
Prefix	R3	10.1.3.0/24
Prefix	R3	10.3.4.0/24
Prefix	R3	3.3.3.3/32
Prefix	R4	10.2.4.0/24
Prefix	R4	10.3.4.0/24
Prefix	R4	4.4.4.4/32

Type	Local Node	Description
Link	R1	Link to R2
Link	R1	Link to R3
Link	R1	Link to broadcast network (DR)
Link	R2	Link to R1
Link	R2	Link to R4
Link	R2	Link to broadcast network (DR)
Link	R3	Link to R1
Link	R3	Link to R4
Link	R3	Link to broadcast network (DR)
Link	R4	Link to R2
Link	R4	Link to R3
Link	R4	Link to broadcast network (DR)
Link	R4(DR)	To R1
Link	R4(DR)	To R2
Link	R4(DR)	To R3
Link	R4(DR)	To R4

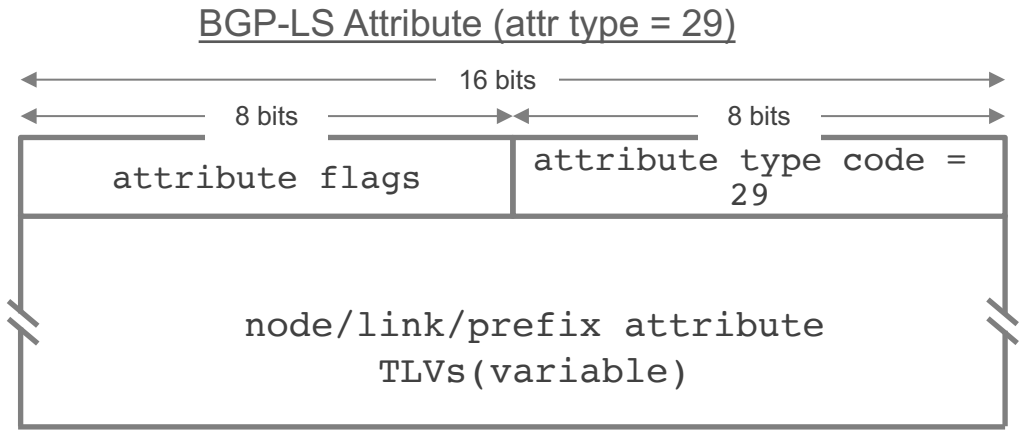
Extensions for Segment Routing

APNIC

Link-state attribute



There is no change to the link-state NLRIs.
Additional TLVs are defined for the BGP-LS attribute



SR: node attribute TLVs

- Node attribute TLVs that may be encoded in the BGP-LS path attribute when advertised together with a Node NLRI.

TLV code point	Name	Description
1161	SID/Label	Encodes a segment identifier (SID)
1034	SR capabilities	Advertises the node's SRGB
1035	SR algorithm	List of SR algorithms supported
1036	SR local block	Range of labels reserved for local SIDs
1037	SRMS preference	SRMS preference for advertising router

Link attribute TLVs

- Link attribute TLVs that may be encoded in the BGP-LS path attribute when advertised together with a Link NLRI.

TLV code point	Name	Description
1099	Adjacency Segment Identifier (Adj-SID)	Includes an adjacency SID
1100	LAN Adjacency Segment Identifier (Adj-SID)	Allows use of a single NLRI to announce all adjacencies on a LAN segmemt
1172	L2 Bundle Member	Identifies L2 bundle member links associated with a parent L3 link

Prefix attribute TLVs

- Prefix attribute TLVs that may be encoded in the BGP-LS path attribute when advertised together with a Prefix NLRI.

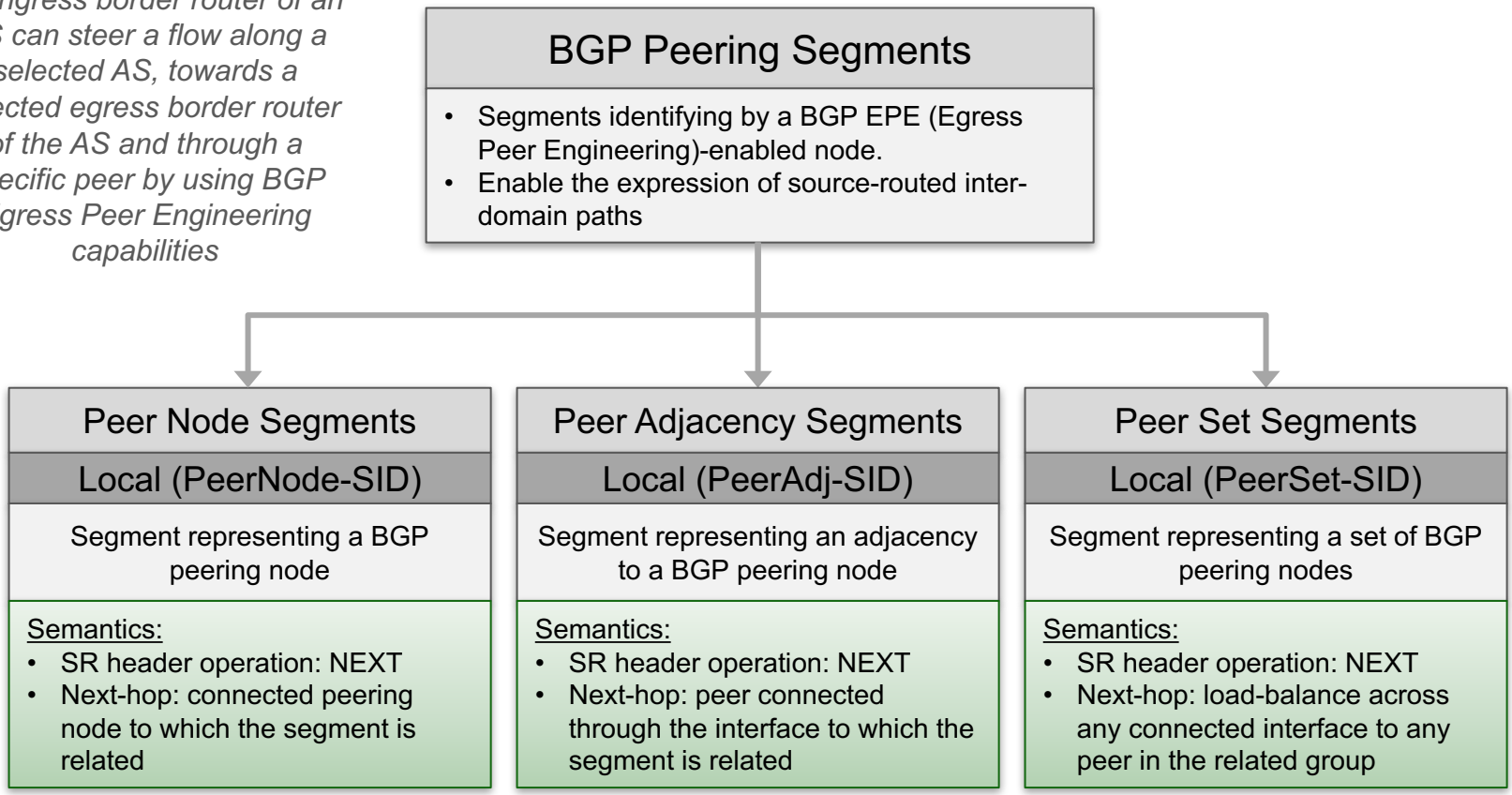
TLV code point	Name	Description
1158	Prefix SID	Advertises SID for a prefox
1159	Range	Range of prefix to SID mappings
1170	IGP prefix attributes	Routing protocol-specific flags
1171	Source router-ID	IPv4 or IPv6 router-ID of the originator of the prefix

Extensions for BGP Egress Peer Engineering (EPE)

APNIC

BGP peering segments

An ingress border router of an AS can steer a flow along a selected AS, towards a selected egress border router of the AS and through a specific peer by using BGP Egress Peer Engineering capabilities



Link NLRI format

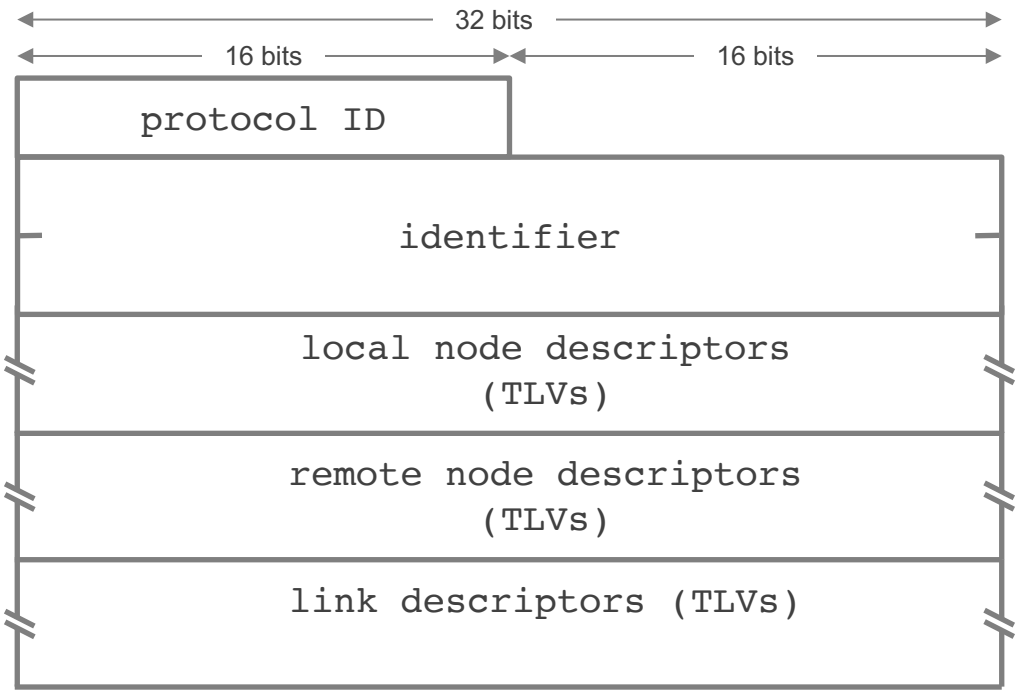
BGP-EPE segments describe links. Hence, the BGP-LS Link NLRI is used.

	Protocol ID
1	IS-IS L1
2	IS-IS L2
3	OSPFv2
4	Direct
5	Static
6	OSPFv3
7	**BGP**

Identifies the IGP instance

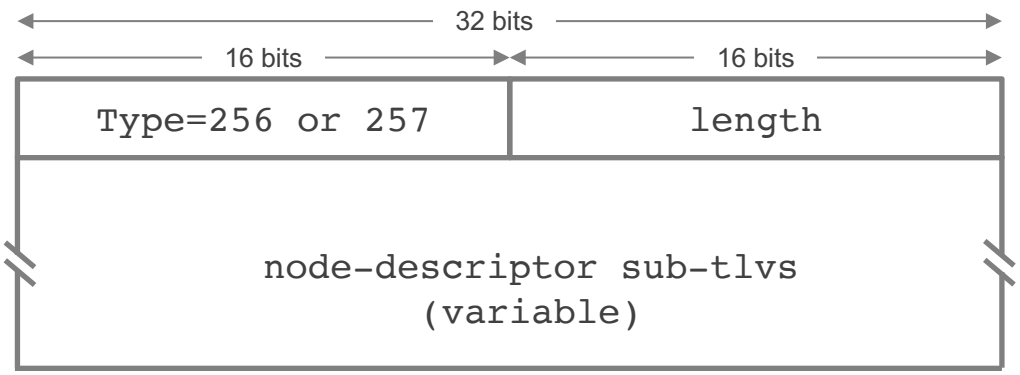
	identifier
0	Default layer 3 routing topology

Link NLRI (NLRI type = 2)



BGP-EPE-specific extension

Node descriptor sub-TLVs

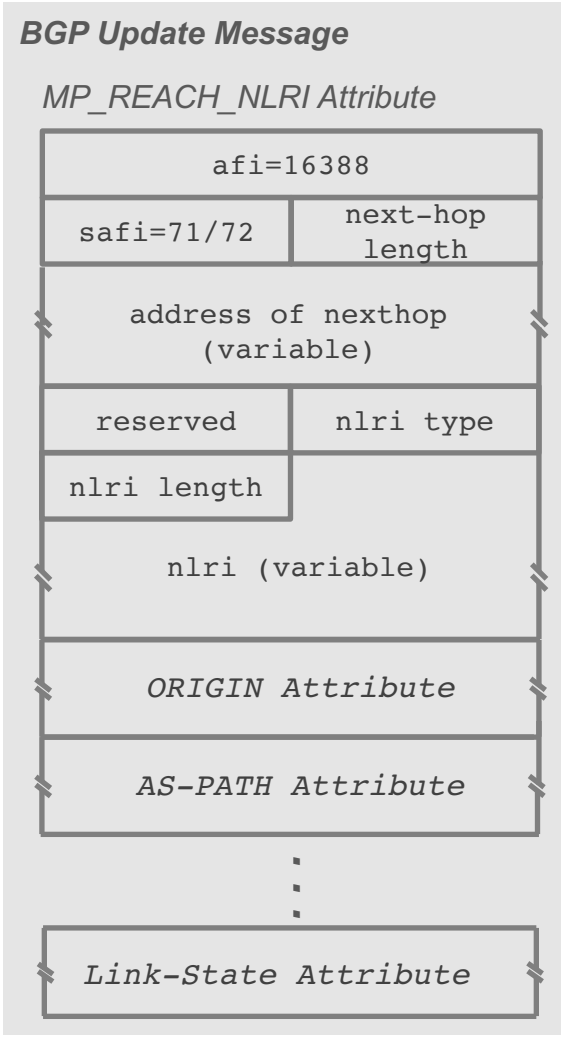


BGP-EPE-specific extension

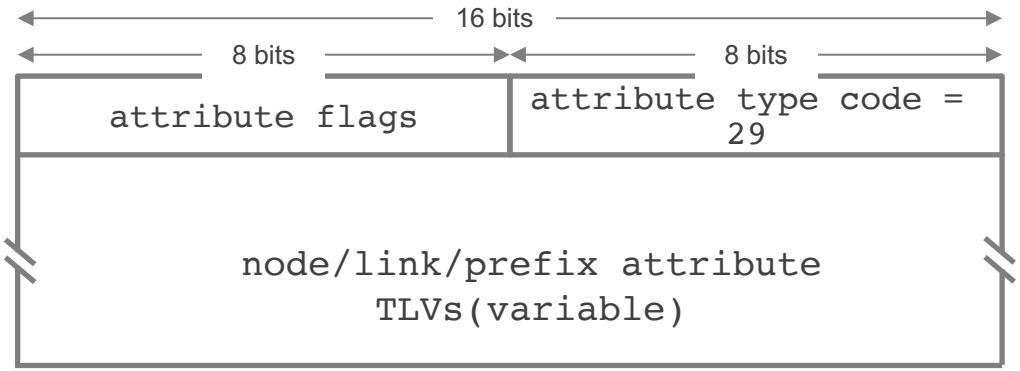
Sub-TLV code point	Length	Description	Details
512*	4	Autonomous System	32-bit ASN
513*	4	BGP-LS identifier	Uniquely identifies BGP-LS domain; combination of ASN and BGP-LS ID has to be globally unique
514	4	Area-ID	IGP area ID
515	variable	IGP Router-ID	Either a 6-octet ISO node-ID or 4-octet OSPF router-ID
516*	4	BGP Router-ID	BGP router identifier
517	4	Confed Member ASN	AS number of confederation member

* mandatory for Link NLRI Local Node Descriptors

Link-state attribute



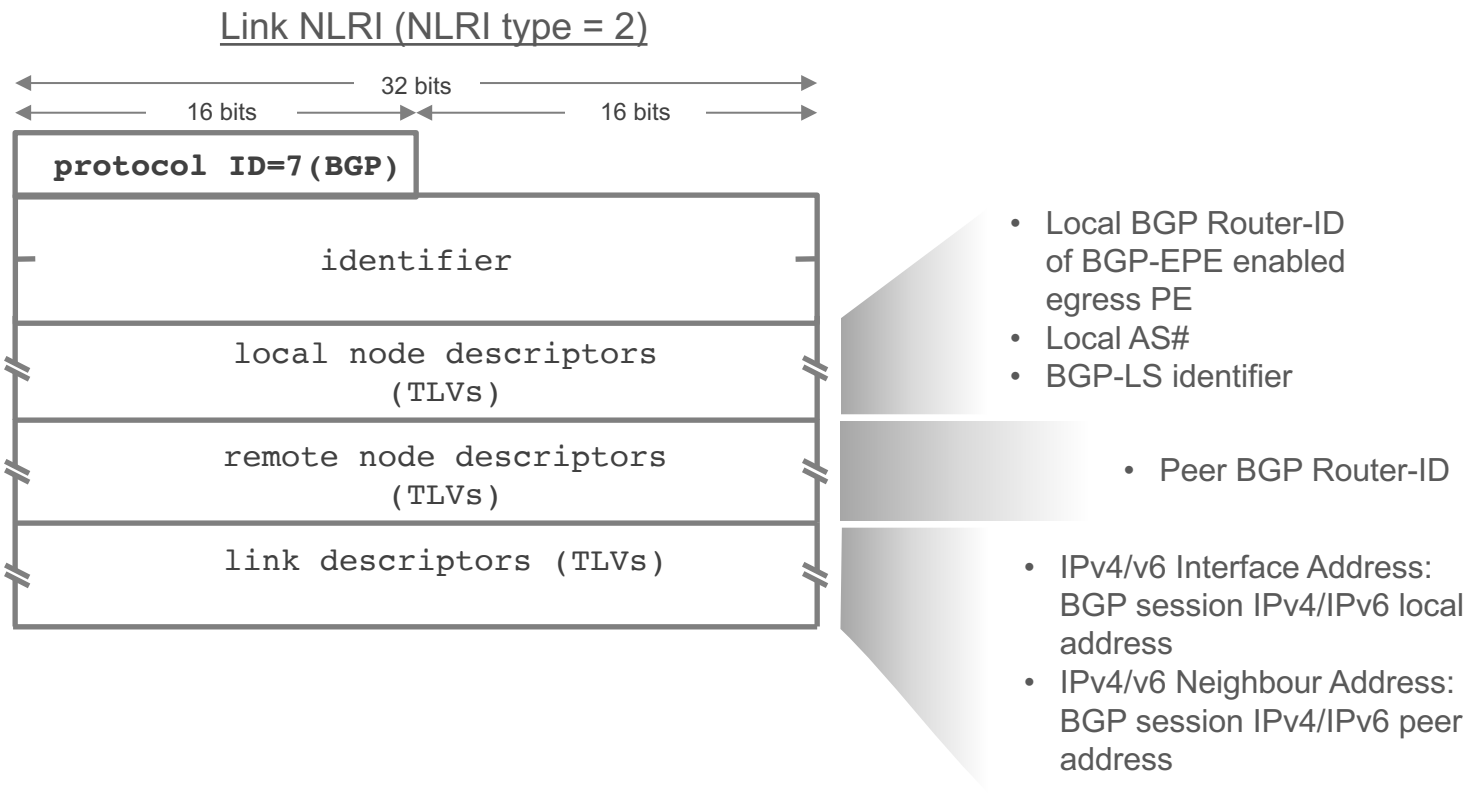
BGP-LS Attribute (attr type = 29)



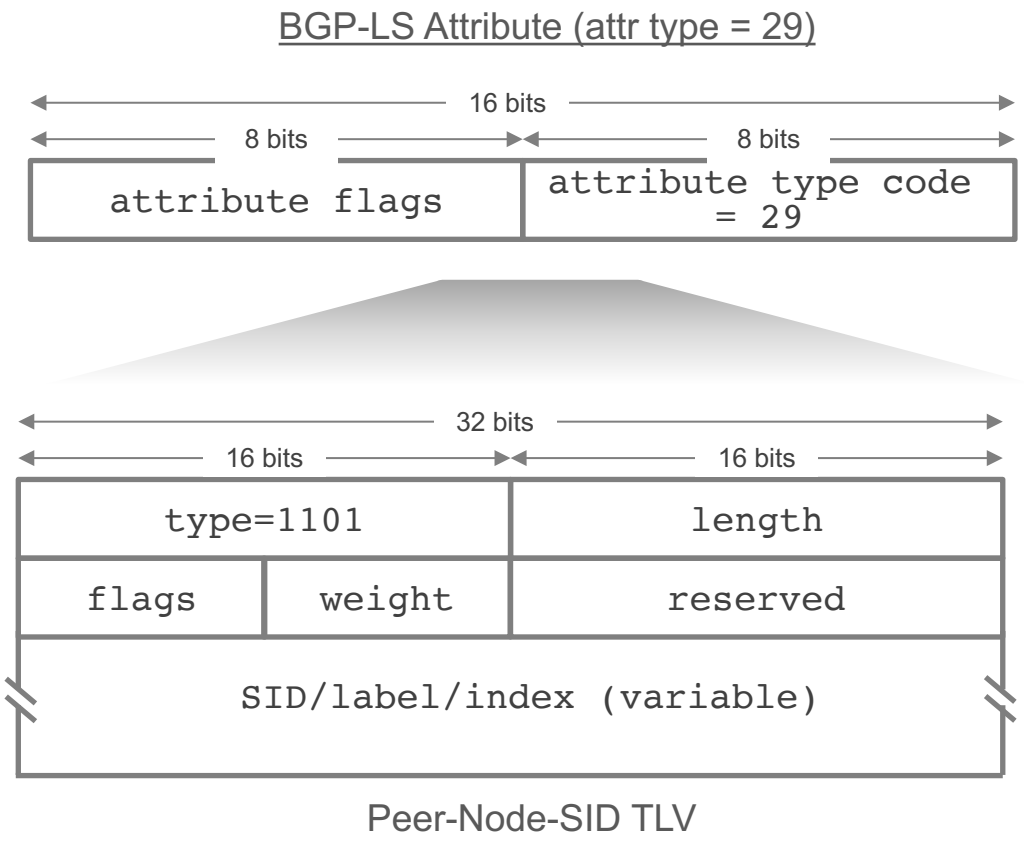
TLV	Name	Description
1101	Peer-Node-SID	Peer Node Segment Identifier (SID/index/label) + weight (for load-balancing)
1102	Peer-Adj-SID	Peer Adjacency Segment Identifier (SID/index/label) + weight (for load-balancing)
1103	Peer-Set-SID	Peer Set Segment Identifier (SID/index/label) + weight (for load-balancing)

BGP-EPE Link attribute TLVs

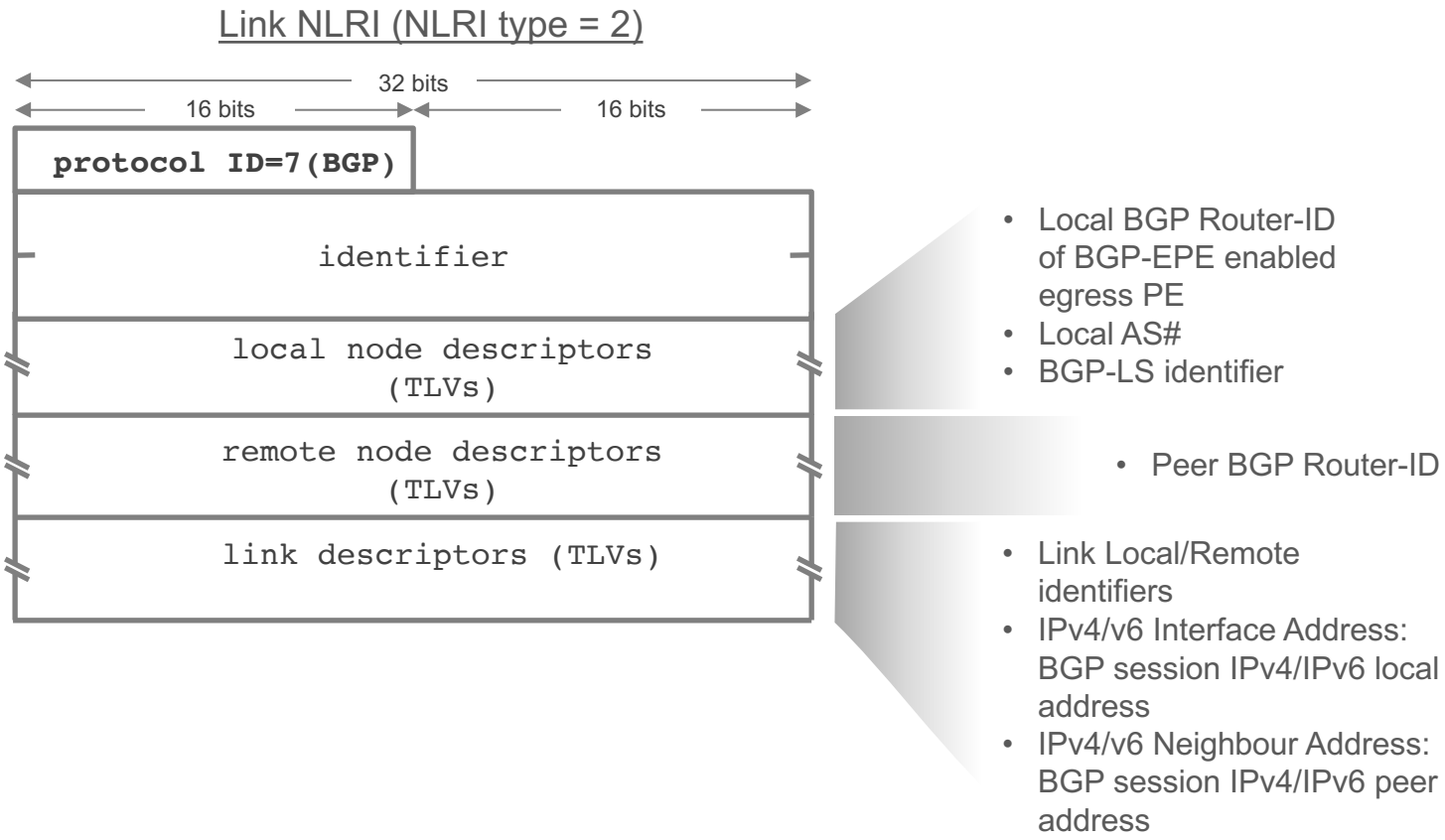
Peer-Node-SID advertisement (1)



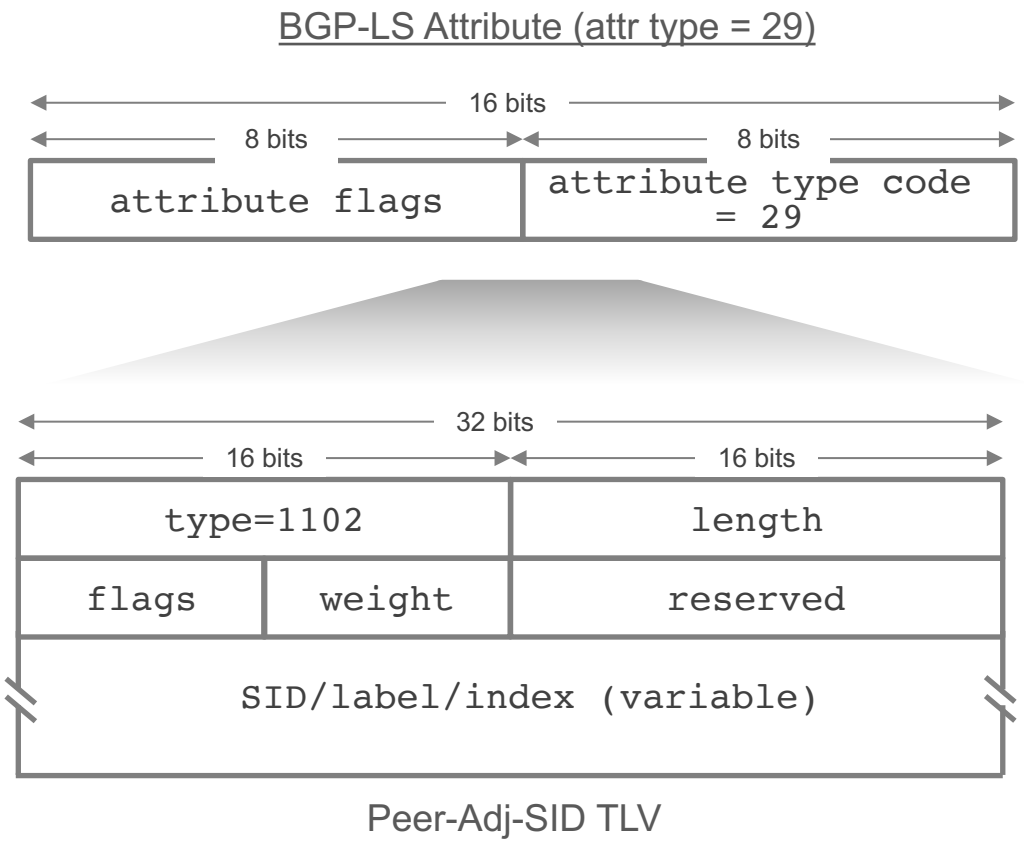
Peer-Node-SID advertisement (2)



Peer-Adj-SID advertisement (1)

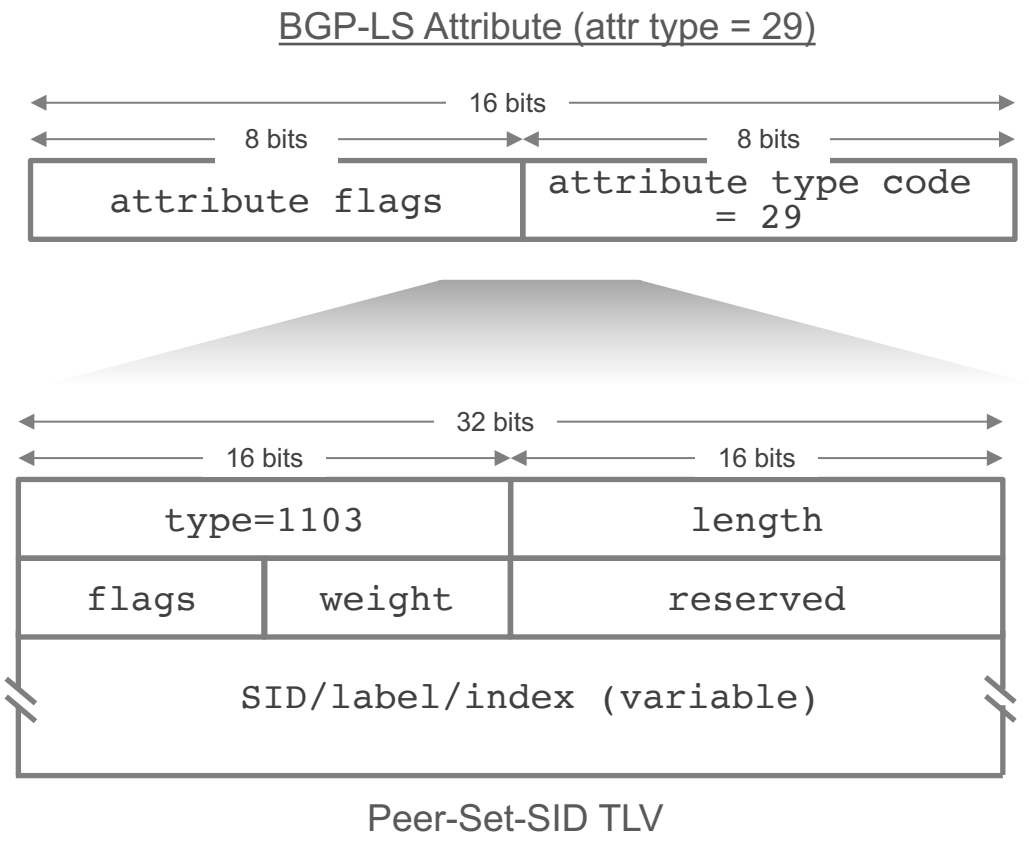


Peer-Adj-SID advertisement (2)

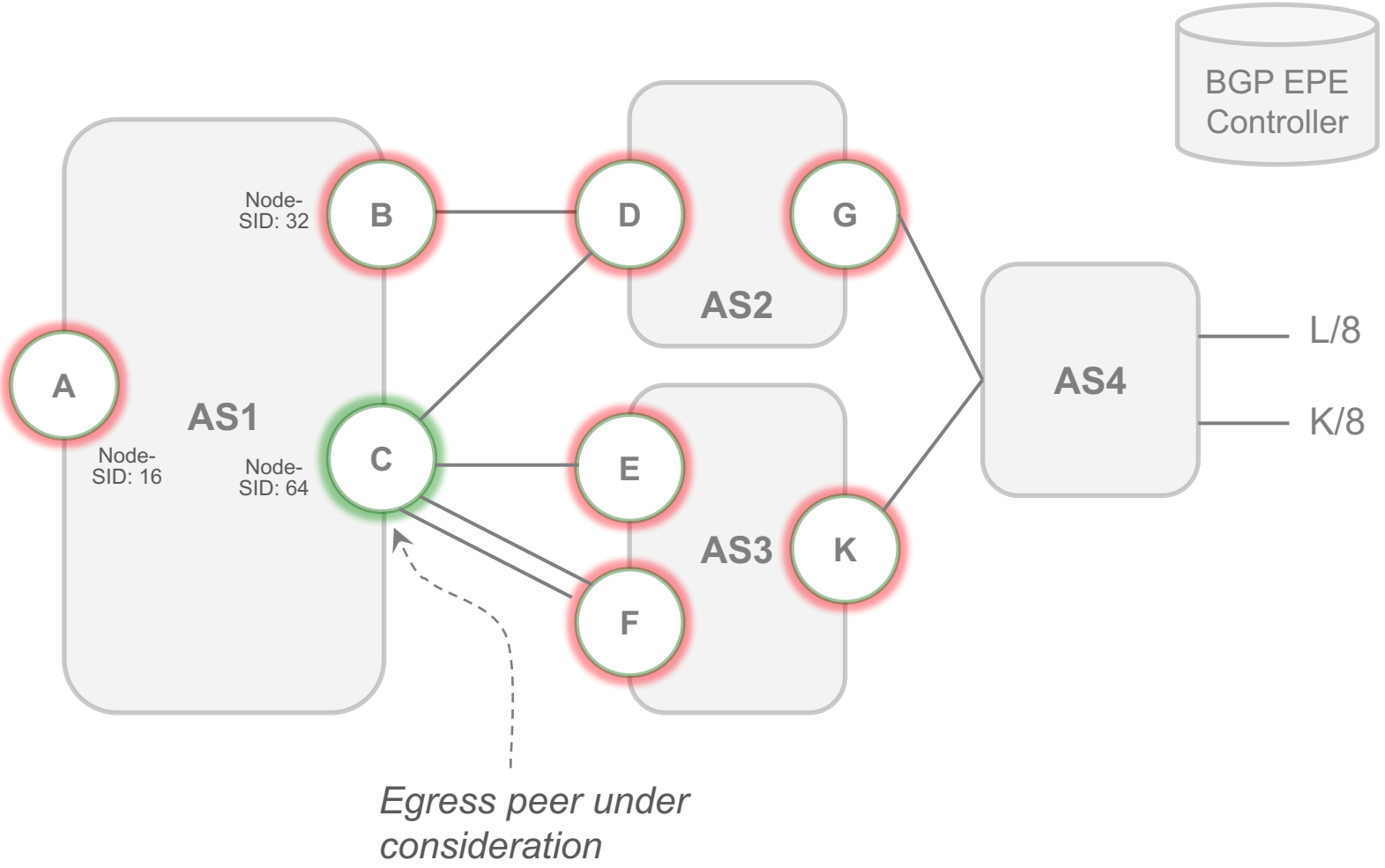


Peer-Set-SID advertisement

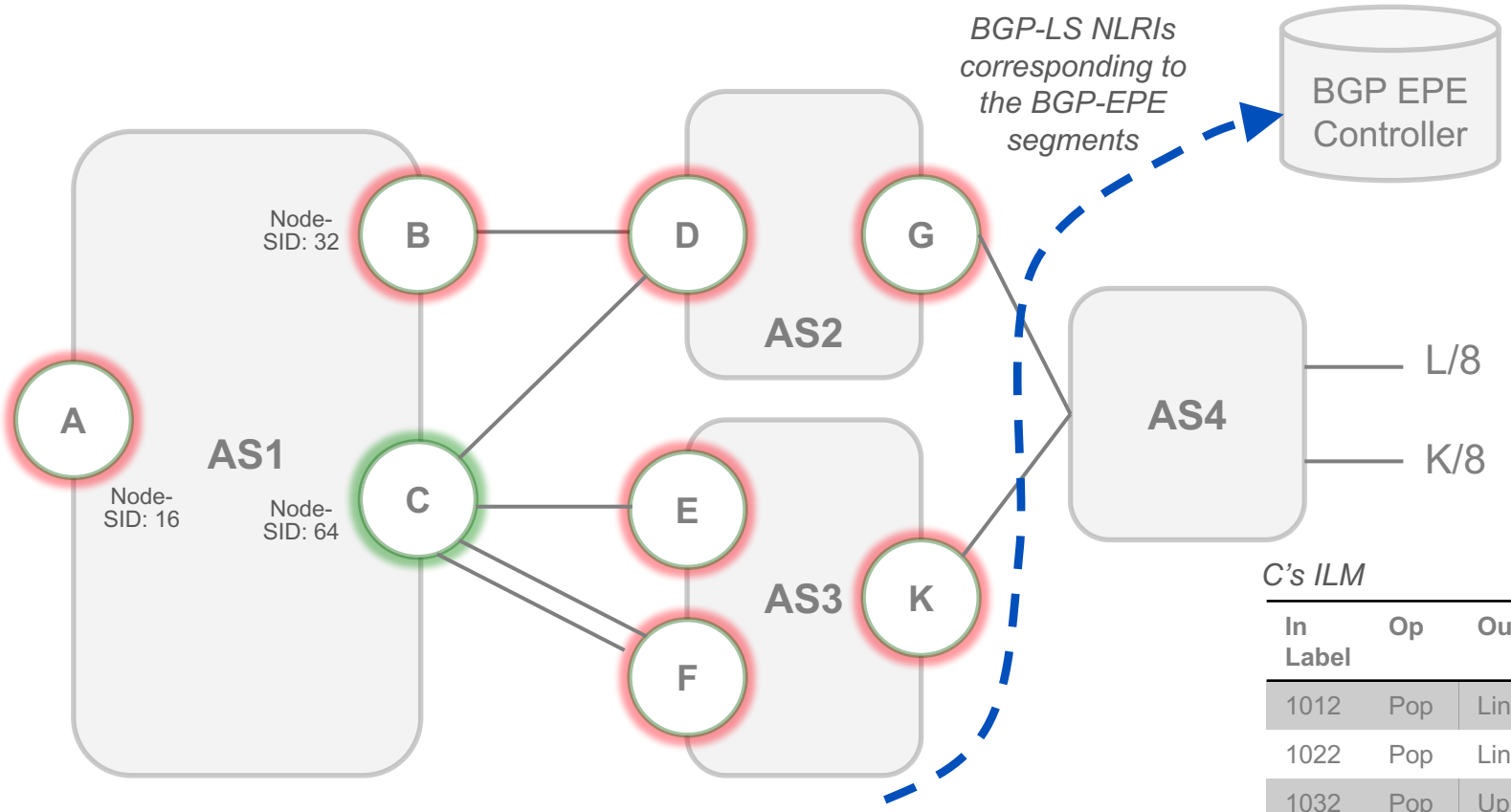
The Peer-Set-SID is advertised only as a BGP-LS attribute that is associated with Link NLRIs for Peer Node Segments or Peer Adjacency Segments that belong to the peer set.



Example: BGP-EPE network



Example: BGP-EPE Segments



BGP Peering segments allocated by PE C

Peer Node Segments	
D	1012
E	1022
F	1052

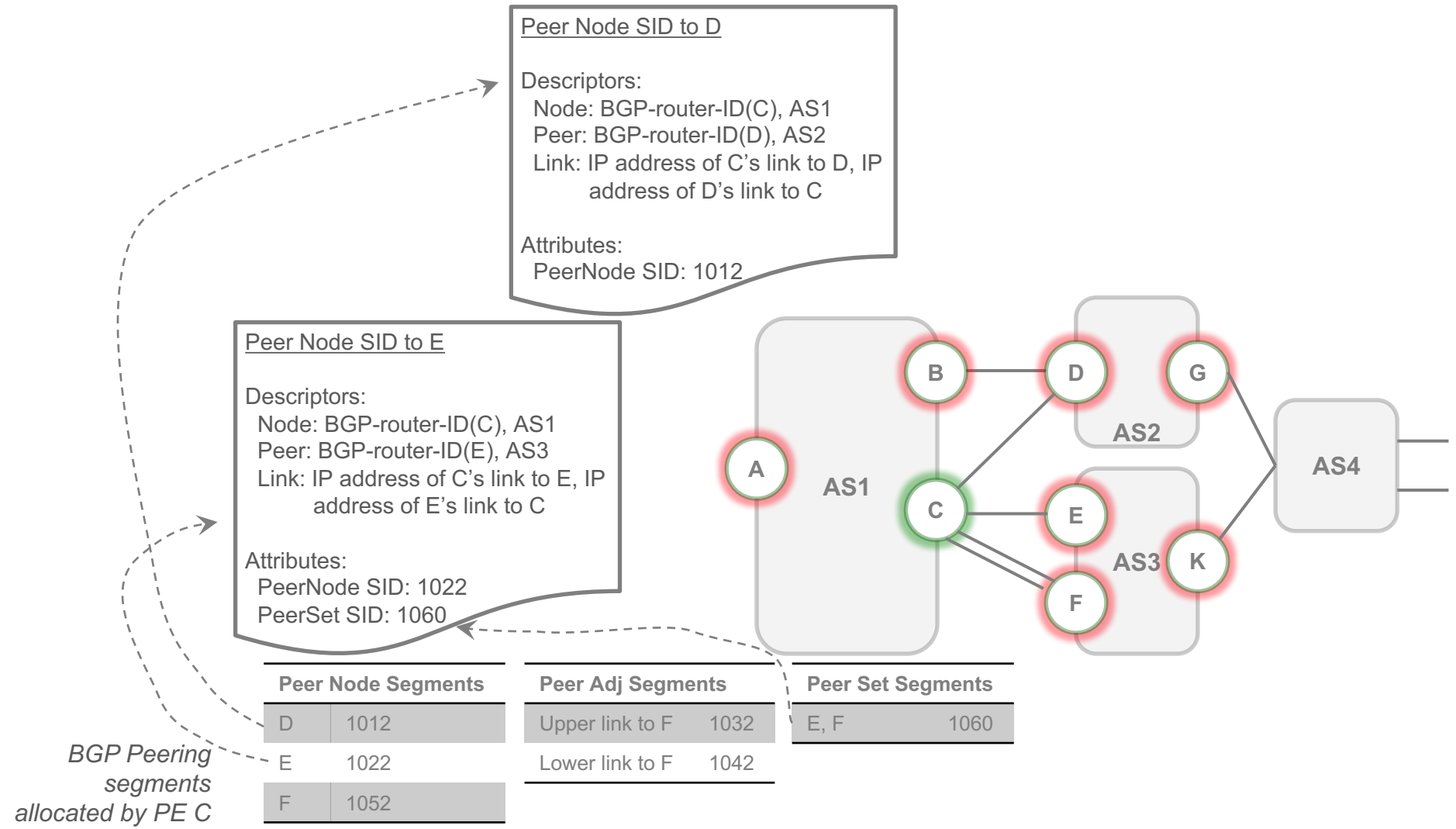
Peer Adj Segments	
Upper link to F	1032
Lower link to F	1042

Peer Set Segments	
E, F	1060

C's ILM

In Label	Op	Out intf
1012	Pop	Link to D
1022	Pop	Link to E
1032	Pop	Upper link to F
1042	Pop	Lower link to F
1052	Pop	Load balance on any link to F
1060	Pop	Load balance on any link to E or F

Example: BGP-LS advertisements (1)



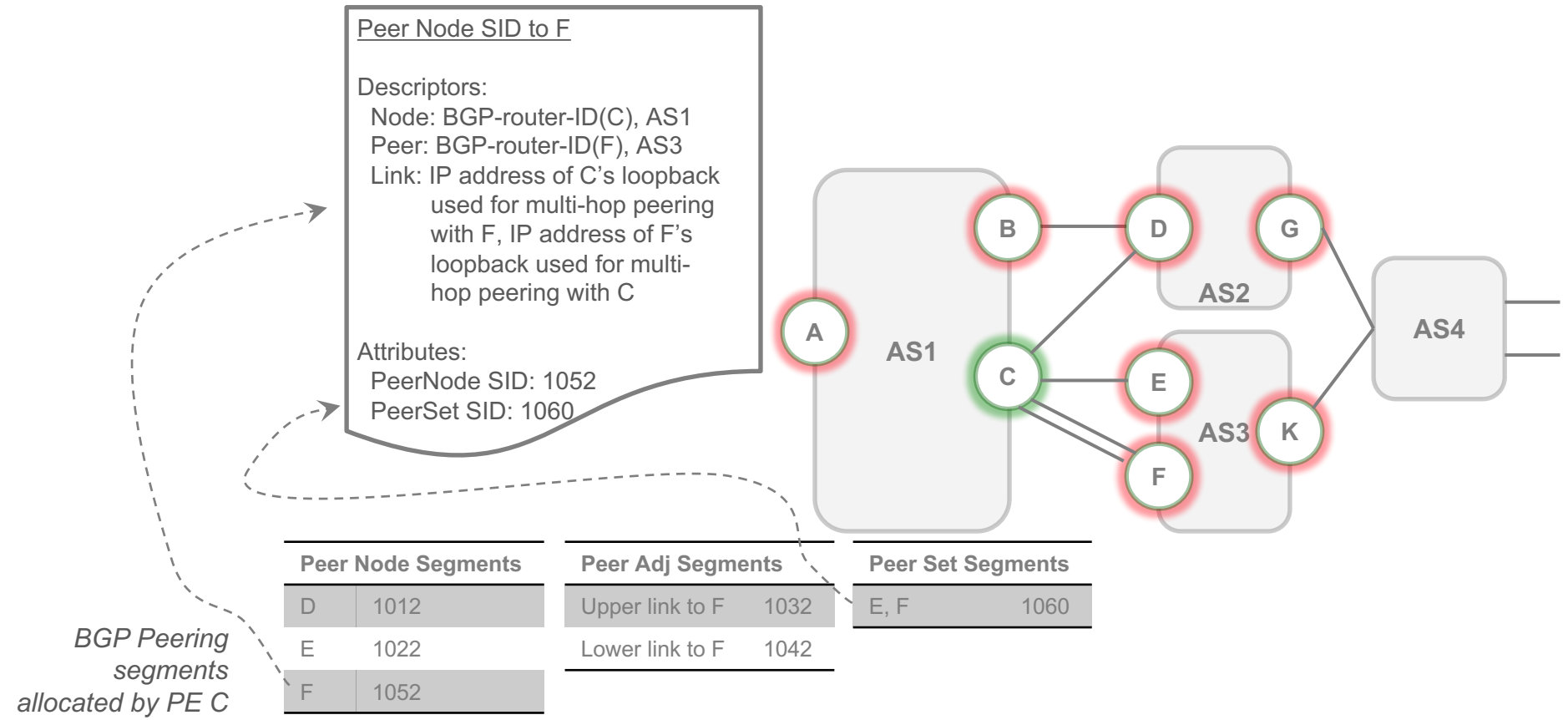
Peer Node Segments	
D	1012
E	1022
F	1052

Peer Adj Segments	
Upper link to F	1032
Lower link to F	1042

Peer Set Segments	
E, F	1060

BGP Peering segments allocated by PE C

Example: BGP-LS advertisements (2)



Example: BGP-LS advertisements (3)

Peer Adj SID – upper link to F

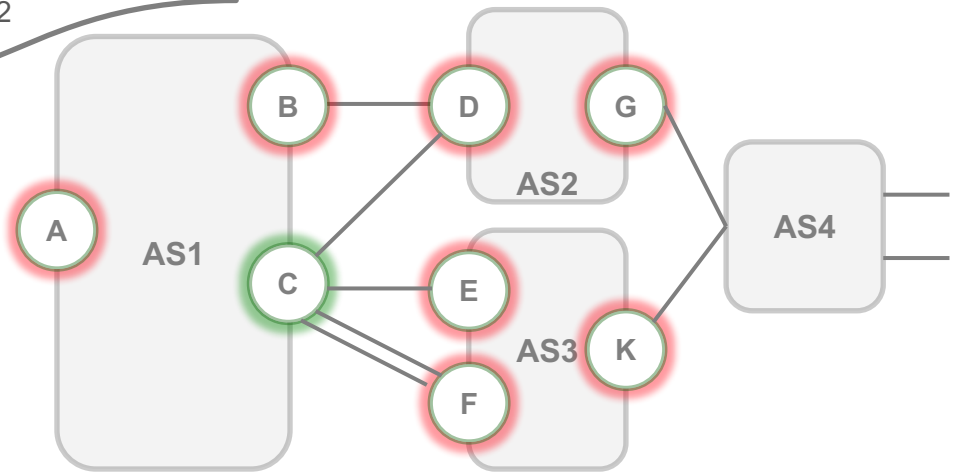
Descriptors:
Node: BGP-router-ID(C), AS1
Peer: BGP-router-ID(F), AS3
Link: IP address of C's upper link to F, IP address of F's upper link to C

Attributes:
PeerAdj SID: 1032

Peer Adj SID – lower link to F

Descriptors:
Node: BGP-router-ID(C), AS1
Peer: BGP-router-ID(F), AS3
Link: IP address of C's lower link to F, IP address of F's lower link to C

Attributes:
PeerAdj SID: 1042



BGP Peering segments allocated by PE C

Peer Node Segments	
D	1012
E	1022
F	1052

Peer Adj Segments	
Upper link to F	1032
Lower link to F	1042

Peer Set Segments	
E, F	1060

Operational Considerations

Operations (1)

- Unlike other BGP NLRIs, link-state NLRIs only carry application-level (i.e. topology) data which has no impact on forwarding state on BGP link-state speakers

- Intent is for dedicated route-reflectors to handle the distribution of this NLRI; not necessary for all routers to support this capability

Operations (2)

- Recommended maximum rate for advertising/withdrawing link-state NLRIs is 200 updates per second

- Distribution of link-state information is recommended to be restricted to a single administrative domain (multiple areas or multiple ASs)

- Flow of updates is one-way-only: from BGP speakers to consumers; any updates from consumers need to be dropped

References

References

- RFC4760
- RFC7752
- draft-ietf-idr-bgp-ls-segment-routing-ext-03
- draft-ietf-idr-bgpls-segment-routing-epe-13
- Using BGP-LS/PCE-P with XR and ODL (https://communities.cisco.com/community/developer/opendaylight/blog/2015/07/25/using-bgp-lspce-p-with-xrv-and-odl-or-cisco-osc)

Thank You !

End of session