BGP Fundamentals

Border Gateway Protocol

Agenda

- **Concepts**
- **Message Types and Operation**
- **Attribute Details**
- **Information Resources**

BGP-4

- **Border Gateway Protocol (BGP)**
  - is the Exterior Gateway Protocol used in the Internet nowadays
  - was developed to overcome limitations of EGP-2
  - RFC 1267 (BGP-3) older version
    - classful routing only
  - RFC 1771 (BGP-4) current version, DS
    - classless routing
- **primary function**
  - exchange of reachability information with other autonomous systems

BGP-4 Concepts   1

- reachability information exchanged between BGP routers carries a sequence of AS numbers
  - indicates the path of AS´s a route has traversed

- path vector protocol
  - extension of distance vector protocol
  - basic metrics is still the number of hops (AS´s traversed)
  - no simple cost metrics because of lack of global metrics coordination
  - however, other attributes might effect decisions
  - similar split horizon rules

BGP-4 Concepts 2

– <u>AS path information</u> allows BGP to construct a graph of autonomous systems
  • loop prevention (without SPF calculation)
    – checking AS number appearance in the AS path
    – assumes a full routing information for AS´s
    – depending on the actual topology loops might arise when an AS does not receive information about all other AS´s
  • no restriction on the underlying topology

– <u>incremental update</u> (triggered)
  • after first full exchange of reachability information between BGP routers only changes are reported
    – BGP Update message

BGP-4 Concepts 3

– <u>description</u> of <u>reachability information</u> by <u>BGP attributes</u>
  • used for establishing routing policy between ASes

– a <u>BGP route</u> is a unit of information that pairs a destination with the path attributes to that destination
  • destination is the network (IP prefix) reported in the NLRI (Network Layer Reachability Information) field
  • path is the information reported in the attributes field

BGP-4 Concepts 4

– IP prefix concept
  • supports VLSM
  • supports classless routing
  • supports aggregation (CIDR) and supernetting
    – full routing requirement on the backbone networks
    – aggregation might be enforced by backbone networks
    – so BGP-4 historically became as required for ISP peering

– allow smooth experimental and vendor extensions
    – optional attributes in the updates

BGP-4 Concepts 5

• **BGP is based on relationship between neighboring BGP-routers**
  – called BGP session or BGP connection
  – peer to peer (both sides can initiate actions)
• **BGP session runs on top of TCP**
  – reliable transport connection
  – well known port 179
  – TCP takes care of fragmentation, sequencing, acknowledgement and retransmission (error recovery)
  – hence these procedures must not be done by the BGP protocol itself

L46 - BGP Fundamentals

BGP-4 Limitations

- **BGP and associated tools cannot express arbitrary policies**
  - only hop-by-hop / destination based routing paradigm
  - once we sent the packets to the neighboring AS, we cannot fully influence the forwarding direction of this traffic behind the neighboring AS
    - because we just manipulate destination based routing tables
    - it will take the same route as the traffic originated from the neighboring AS to the same destination
    - so the destination will get all the aggregated traffic through a single path without possible preferential treatment of the senders
  - source IP address based policy routing might be available by some vendors to handle such needs of differential treatment of path selection

© 2006, D.I. Manfred Lindner   BGP Fundamentals, v4.7   9

Agenda

- **Concepts**
- **Message Types and Operation**
- **Attribute Details**
- **Information Resources**

© 2006, D.I. Manfred Lindner   BGP Fundamentals, v4.7   10

© 2006, D.I. Manfred Lindner

Page 46 - 5

L46 - BGP Fundamentals

BGP Message Types

- **Open (type 1)**
  - to establish relationship between BGP neighbors
- **Update (type 2)**
  - to advertise reachability information with its corresponding path attributes
  - path attributes are used for BGP route decision process and supports establishing of routing policy between AS´s
- **Notification (type 3)**
  - to report errors to the neighbor
    - after notification is sent relationship will be terminated
- **Keepalive (type 4)**
  - to constantly monitor reachability of BGP neighbor
- **Route Refresh (type 5, RFC 2918)**
  - to enforce a re-advertisement from the Adj-RIB-out from a BGP neighbor
  - Adj-RIB-out = storage place for all BGP-routes already sent to BGP neighbors

© 2006, D.I. Manfred Lindner   BGP Fundamentals, v4.7   11
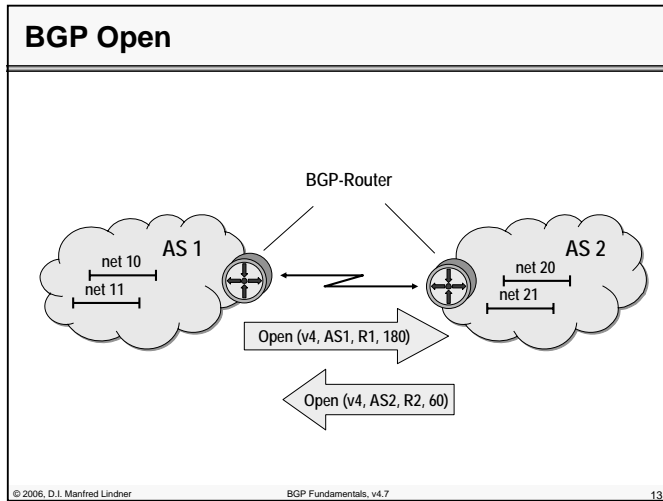
BGP Open

- **initial exchange of parameters**
  - BGP version number (3 or 4)
  - AS number of sending router
  - identifier of sending router (BGP Router - ID)
  - hold time
    - maximum time in seconds between successive receipt of keepalive or update messages
      - 2-byte unsigned integer
    - if time is exceeded neighbor would be considered dead
    - negation is done in direction whatever value is lower
    - hold time = 0 means that timer never expires
  - optional parameters
    - e.g. for authentication (MD5) and BGP Multiprotocol Extensions
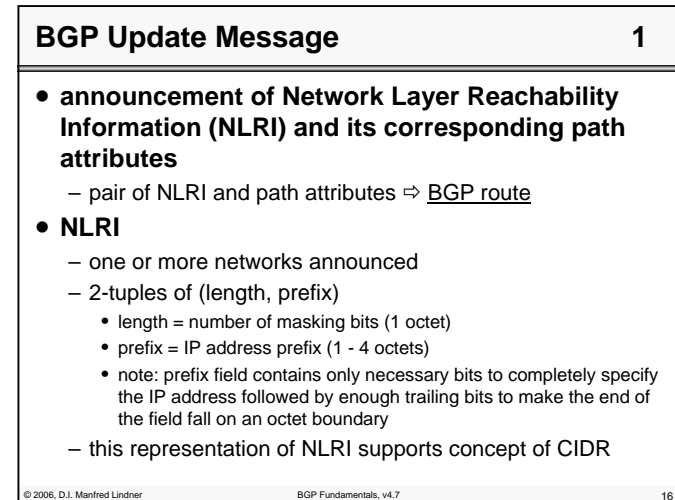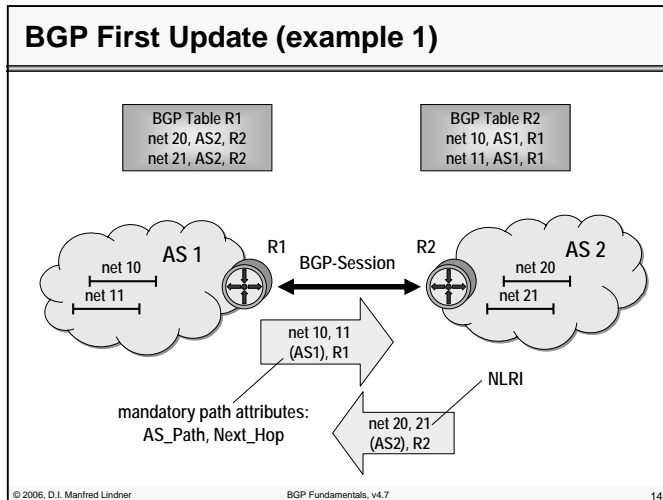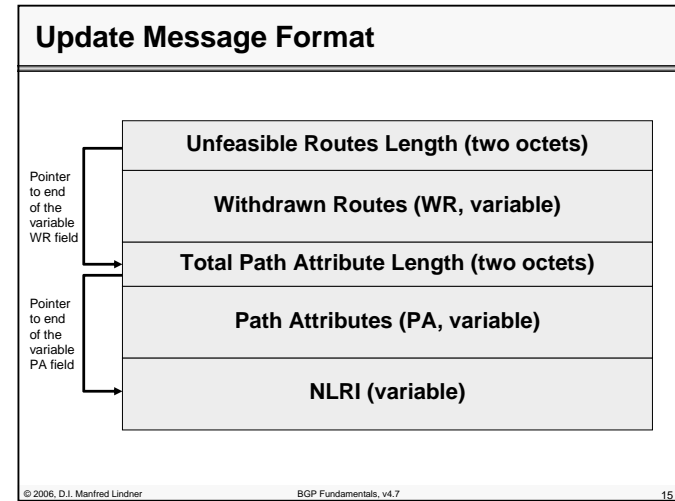
© 2006, D.I. Manfred Lindner   BGP Fundamentals, v4.7   12

© 2006, D.I. Manfred Lindner

Page 46 - 6

L46 - BGP Fundamentals

BGP Open

BGP First Update (example 1)

Update Message Format

BGP Update Message   1

- **announcement of Network Layer Reachability Information (NLRI) and its corresponding path attributes**
  - pair of NLRI and path attributes ⇨ <u>BGP route</u>
- **NLRI**
  - one or more networks announced
  - 2-tuples of (length, prefix)
    - length = number of masking bits (1 octet)
    - prefix = IP address prefix (1 - 4 octets)
    - note: prefix field contains only necessary bits to completely specify the IP address followed by enough trailing bits to make the end of the field fall on an octet boundary
  - this representation of NLRI supports concept of CIDR
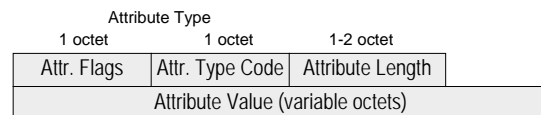
BGP Update Message 2

- **path attributes provide information about a NLRI**
  - to be used in the BGP filtering and BGP manipulation process (routing policy)
  - to be used in the route decision process

- **path attributes are composed of**
  - triples of (type, length, value) -> TLV notation
    - attribute type (two octets)
      - 8 bit attribute flags, 8 bit attribute type code
    - attribute length (one or two octets)
      - one or two octest signaled by attribute flag-bit nr.4
    - attribute value (variable length)
      - content depends on meaning signaled by attribute type code
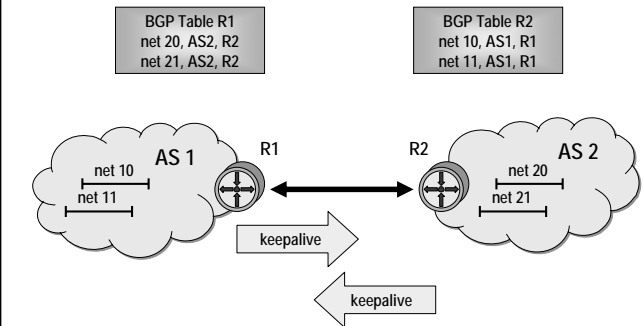
Path Attribute Format / NLRI Format

Attribute Type
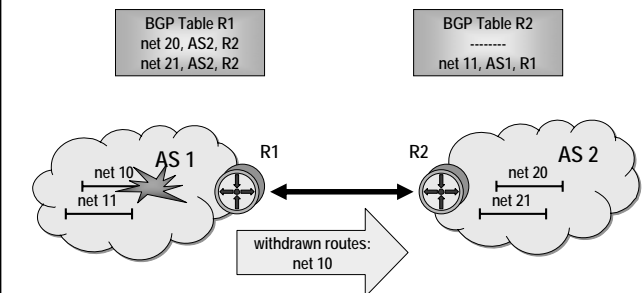
1 octet	1 octet	1-2 octet
Attr. Flags	Attr. Type Code	Attribute Length
Attribute Value (variable octets)		

Path Attribute Format

1 octet

Length	Prefix (1 - 4 octets)

NLRI

BGP Keepalive (example 1)

BGP Incremental Update (example 1)

BGP Update Message (Withdrawn)

- **advertising of unreachable routes**
  - located at the beginning of an Update message
  - called withdrawn or unfeasible routes
  - none, one or more withdrawn routes can be announced
  - fields
    - Unfeasible Routes Length (2 octets)
    - Withdrawn Routes (variable)
      - 2-tuples of (length, prefix)
      - <u>same way as NLRI but no attributes</u> !!!
  - in principle there are three ways to delete a BGP route
    - withdraw it
    - announce a new route for same destination network
    - close BGP session
    - route refresh (RFC 2918)

BGP Routing and BGP Policy

- **in example 1 (two AS´s connected point-to-point)**
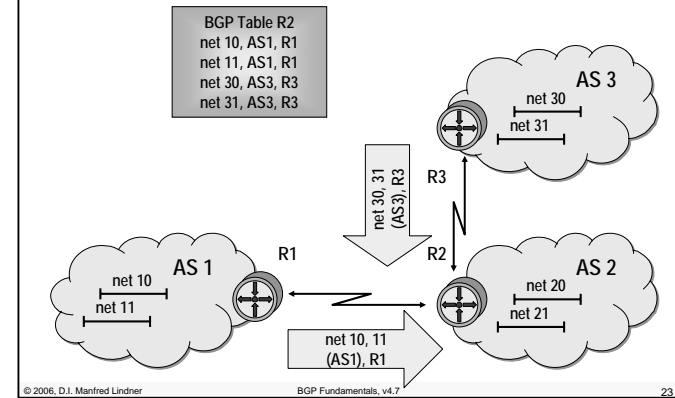 - routing policy is reduced to a minimal function
    - a BGP router can decide only which networks within the own AS should be announced to the neighbor and which learned networks should be advertised into the own AS
  - because of lack of redundancy
    - no route decision (selecting the best path) must be taken

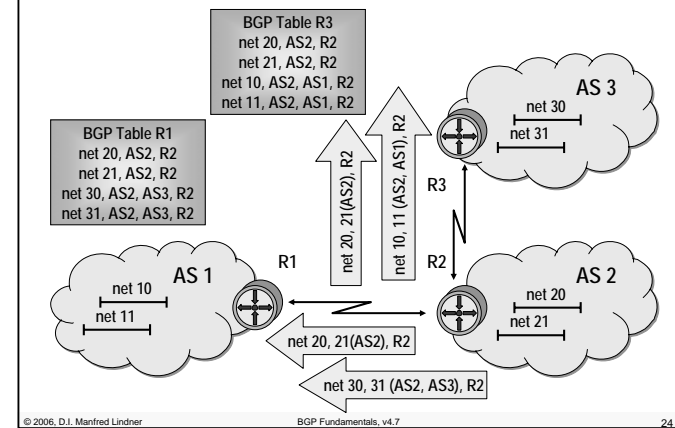- **in a simple transit topology of AS´s (example 2)**
  - reachability information is propagated hop-by-hop
    - split horizon technique
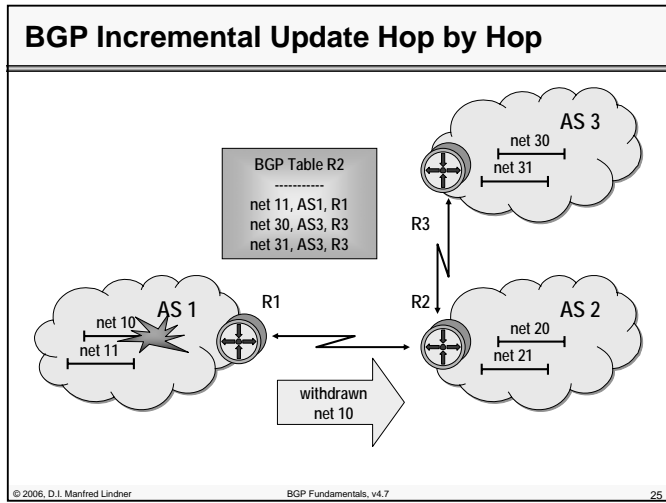 - routing policy can be used to decide which routes should be propagated to other peers
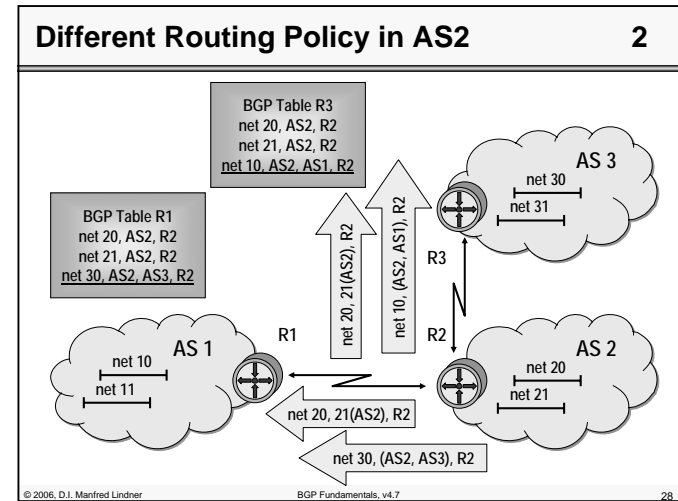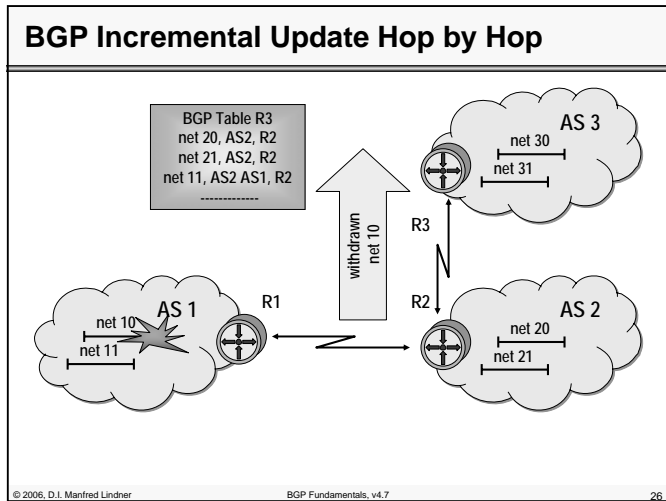
BGP Update R1, R3 (example 2)

BGP Update R2 - Split Horizon (example 2)

L46 - BGP Fundamentals

BGP Incremental Update Hop by Hop

BGP Table R2
----
net 11, AS1, R1
net 30, AS3, R3
net 31, AS3, R3

AS 3
net 30
net 31

AS 1
net 10
net 11

R1   R2   R3

AS 2
net 20
net 21

withdrawn
net 10

BGP Incremental Update Hop by Hop

BGP Table R3
net 20, AS2, R2
net 21, AS2, R2
net 11, AS2 AS1, R2

AS 3
net 30
net 31

AS 1
net 10
net 11

withdrawn
net 10

R1   R2   R3

AS 2
net 20
net 21

L46 - BGP Fundamentals

Different Routing Policy in AS2   1

R2 filters net 11 and net 31
before passing BGP Update
to next BGP peer

BGP Table R2
net 10, AS1, R1
net 11, AS1, R1
net 30, AS3, R3
net 31, AS3, R3

AS 3
net 30
net 31

net 30, 31
(AS3), R3

AS 1
net 10
net 11

R1   R2   R3

AS 2
net 20
net 21

net 10, 11
(AS1), R1

Different Routing Policy in AS2   2

BGP Table R3
net 20, AS2, R2
net 21, AS2, R2
net 10, AS2, AS1, R2

BGP Table R1
net 20, AS2, R2
net 21, AS2, R2
net 30, AS2, AS3, R2

AS 3
net 30
net 31

net 20, 21(AS2), R2

net 10, (AS2, AS1), R2

R1   R2   R3

AS 1
net 10
net 11

AS 2
net 20
net 21

net 20, 21(AS2), R2

net 30, (AS2, AS3), R2

L46 - BGP Fundamentals

Agenda

- **Concepts**
- **Message Types and Operation**
- **Attribute Details**
- **Information Resources**

L46 - BGP Fundamentals

Currently Defined Attributes   2

- **Basic attributes (cont.)**
  - Atomic_Aggregate
    - well-known discretionary; type 6
  - Aggregator
    - optional transitive; type 7

  - these are the attributes that you can rely on in a multi-vendor environment

---

Currently Defined Attributes   1

- **Basic attributes**
  - defined in RFC 1771 (Draft Standard)
  - Origin
    - well-known mandatory; type 1
  - AS_Path
    - well-known mandatory; type 2
  - Next_Hop
    - well-known mandatory; type 3
  - Multi_Exit_Discriminator MED
    - optional non-transitive; type 4
  - Local_Preference
    - well-known discretionary; type 5

Currently Defined Attributes   3

- **Advanced attributes**
  - Community
    - optional transitive; type 8
    - defined in RFC 1997 (Proposed Standard)
  - Originator_ID
    - optional non-transitive; type 9
    - defined in RFC 1966 (Experimental) and RFC 2796 (Proposed Standard) -> Route Reflector
  - Cluster_List
    - optional non-transitive; type 10
    - defined in RFC 1966 (Experimental) and RFC 2796 (Proposed Standard) -> Route Reflector

Currently Defined Attributes 4

- **Advanced attributes (cont.)**
  - Multiprotocol Reachable NLRI
    - MP_REACH_NLRI
    - optional non-transitive; type 14
    - defined in RFC 2858 (Proposed Standard) -> Multiprotocol Extensions
  - Multiprotocol Unreachable NLRI
    - MP_UNREACH_NLRI
    - optional non-transitive; type 15
    - defined in RFC 2858 (Proposed Standard) -> Multiprotocol Extensions

  - in a multi-vendor environment carefully check implementation details

© 2006, D.I. Manfred Lindner BGP Fundamentals, v4.7 33

Format of Attribute-Type

- **8 bit attribute flags**
  - 1. bit (MSB)
    - optional (1) or well-known (0)
  - 2. bit
    - transitive (1) or non-transitive (0)
    - only for optional; set to 1 for well-known
  - 3. bit
    - partial (1) or complete (0)
    - set to 0 for well-known and optional non-transitive
  - 4. bit
    - two octet (1) or one octet (0) attribute length field
- **8 bit attribute type code**
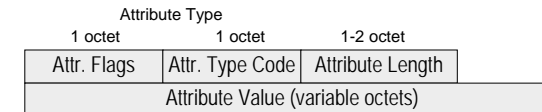 - values 1 - 16 currently defined

© 2006, D.I. Manfred Lindner BGP Fundamentals, v4.7 34

Path Attribute Format

Attribute Type

1 octet	1 octet	1-2 octet
Attr. Flags	Attr. Type Code	Attribute Length
Attribute Value (variable octets)		

Path Attribute Format

© 2006, D.I. Manfred Lindner BGP Fundamentals, v4.7 35

Classification of Attributes 1

- **well-known**
  - must be recognized by all BGP implementations
- **well-known mandatory**
  - must be included in every Update message
    - Origin, AS_Path, Next_Hop
- **well-known discretionary**
  - may or may not be included in every Update message
    - Local_Preference, Atomic_Aggregate
- **all well-known attributes must be passed along to other BGP peers**
  - some will be updated properly first, if necessary

© 2006, D.I. Manfred Lindner BGP Fundamentals, v4.7 36

Classification of Attributes 2

- **optional**
  - it is not required or expected that all BGP implementation support all optional attributes
  - may be added by the originator or any AS along the path
  - paths are accepted regardless whether the BGP peer understands an optional attribute or not
- **handling of recognized optional attributes**
  - propagation of attribute depends on meaning of the attribute
  - propagation of attribute is not constrained by transitive bit of attribute flags
    - but depends on the meaning of the attribute

Classification of Attributes 3

- **handling of unrecognized optional attribute**
  - propagation of attribute depends on transitive bit of attribute flags
  - transitive
    - paths are accepted (attribute is ignored) and attribute remains unchanged when path is passed along to other peers
    - attribute is marked as partial (bit 3 of attribute flags)
    - example: Community
  - non-transitive
    - paths are accepted, attribute is quietly ignored and discarded when path is passed along to other peers
    - example: Multi_Exit_Discriminator

Agenda

- **Concepts**
- **Message Types and Operation**
- **Attribute Details**
- **Information Resources**

BGP related documents 1

- **Draft Standard**
  - RFC 1771 - A Border Gateway Protocol 4 (BGP-4)
    - previous versions: RFC 1105, RFC 1163, RFC 1267, RFC 1654
  - RFC 1772 - Application of the BGP in the Internet
    - previous versions: RFC 1655
  - RFC 1657 - Definitions of Managed Objects for the Fourth Version of the Border Gateway Protocol (BGP-4) using SMIv2

BGP related documents 2

- **Proposed Standard**
  - RFC 1997 - BGP Communities Attribute
  - RFC 2385 - Protection of BGP Sessions via the TCP MD5 Signature Option
  - RFC 2439 - BGP Route Flap Damping
  - RFC 2545 - Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing
  - RFC 2796 - BGP Route Reflection
  - RFC 2858 - Multiprotocol Extensions for BGP-4
  - RFC 2918 - Route Refresh Capability for BGP-4
  - RFC 3107 - Carrying Label Information in BGP-4
  - RFC 3392 - Capabilities Advertisement with BGP-4

BGP related documents 3

- **Experimental**
  - RFC 1863 - A BGP/IDRP Route Server alternative to a full mesh routing
    - previous versions: RFC 1645
  - RFC 1965 - Autonomous System Confederations for BGP
  - RFC 1966 - BGP Route Reflection - An alternative to full mesh BGP
- **Historical**
  - RFC 1397 - Default Route Advertisement In BGP2 and BGP3 Version of The Border Gateway Protocol
  - RFC 1403 - BGP OSPF Interaction
    - previous versions: RFC 1364
  - RFC 1745 - BGP4/IDRP for IP - OSPF interaction

BGP related documents 4

- **Informational**
  - RFC 1773 - Experience with the BGP-4 protocol
    - previous versions: RFC 1266, RFC 1656
  - RFC 1774 - BGP-4 Protocol Analysis
    - previous versions: RFC 1265
  - RFC 1998 - An Application of the BGP Community Attribute in Multi-Home Routing
  - RFC 2042 - Registering New BGP Attribute Types
  - RFC 2547 - BGP / MPLS VPNs
- **Best Current Practice**
  - RFC 1930 - Guidelines for creation, selection, and registration of an Autonomous System