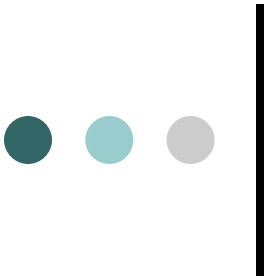


Exterior Gateway Protocols
BGP-4

Border Gateway Protocol

- RFC 1771 – Version 4
- Routing between AS
- Path vector based
- Runs on TCP port 179
- Peers/neighbors – two BGP routers that form a TCP connection

Basic Operation

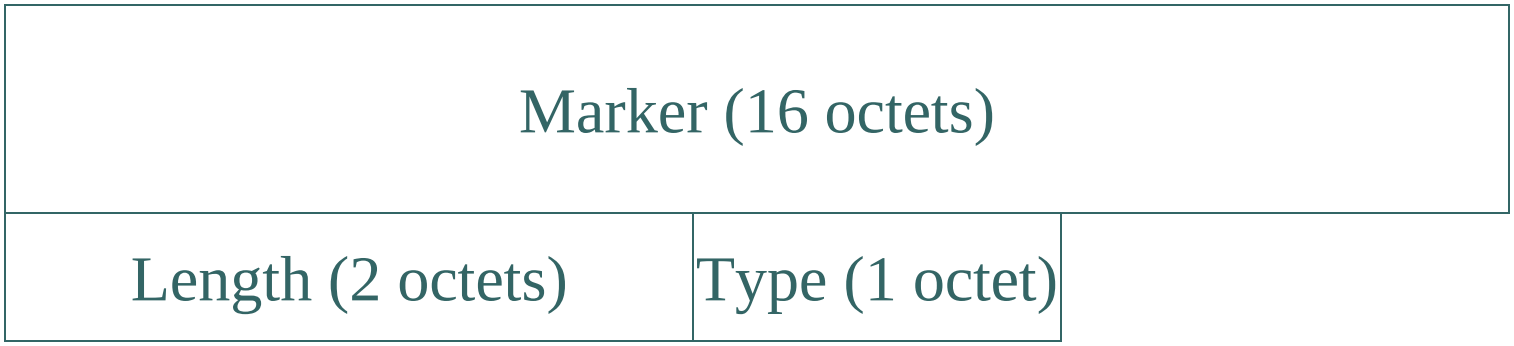
- Peers open a connection and negotiate connection parameters with OPEN messages (both ends send it)
- OPEN messages ack'ed by KEEPALIVE messages
- Initially, entire routing table is transmitted by UPDATE messages
- Afterwards, only incremental updates are sent (add, modify, withdraw), again by UPDATE message
- No periodic refresh of the whole routing table at all (so initial routing tables of all peers stored)
- KEEPALIVE messages sent periodically to maintain connection
- On error, NOTIFICATION message sent. Connection closed down immediately.
- Messages can be authenticated

BGP Route

- A set of networks reachable by that route
- A sequence of AS no.s through which to get to the destinations
- Next hop
- Other information like preference etc.

BGP Message Format

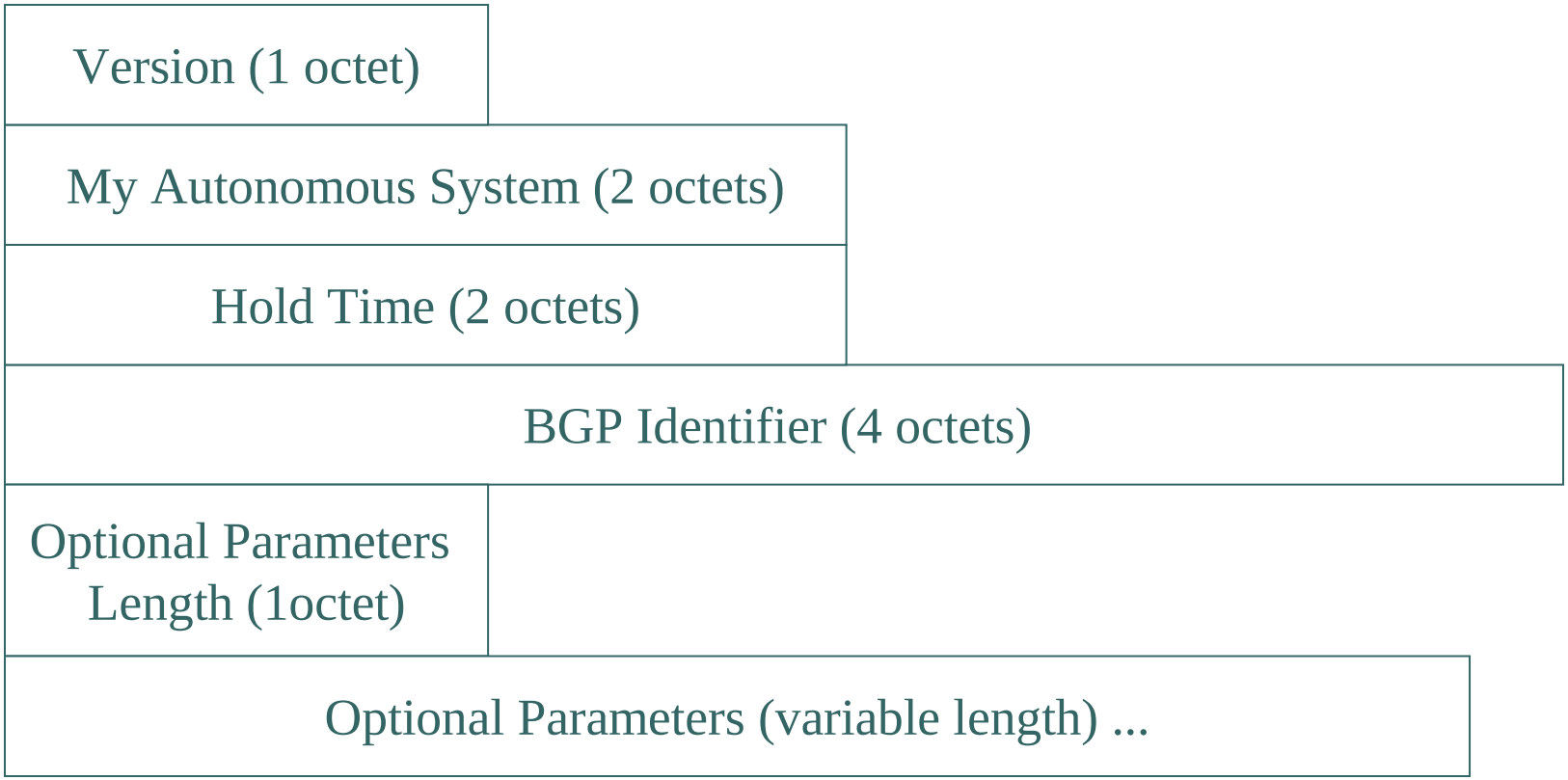
- Common header



  - Marker – authentication/synchronization
  - Length – length of message
  - Type – OPEN, UPDATE, NOTIFICATION, KEEPALIVE

OPEN Message

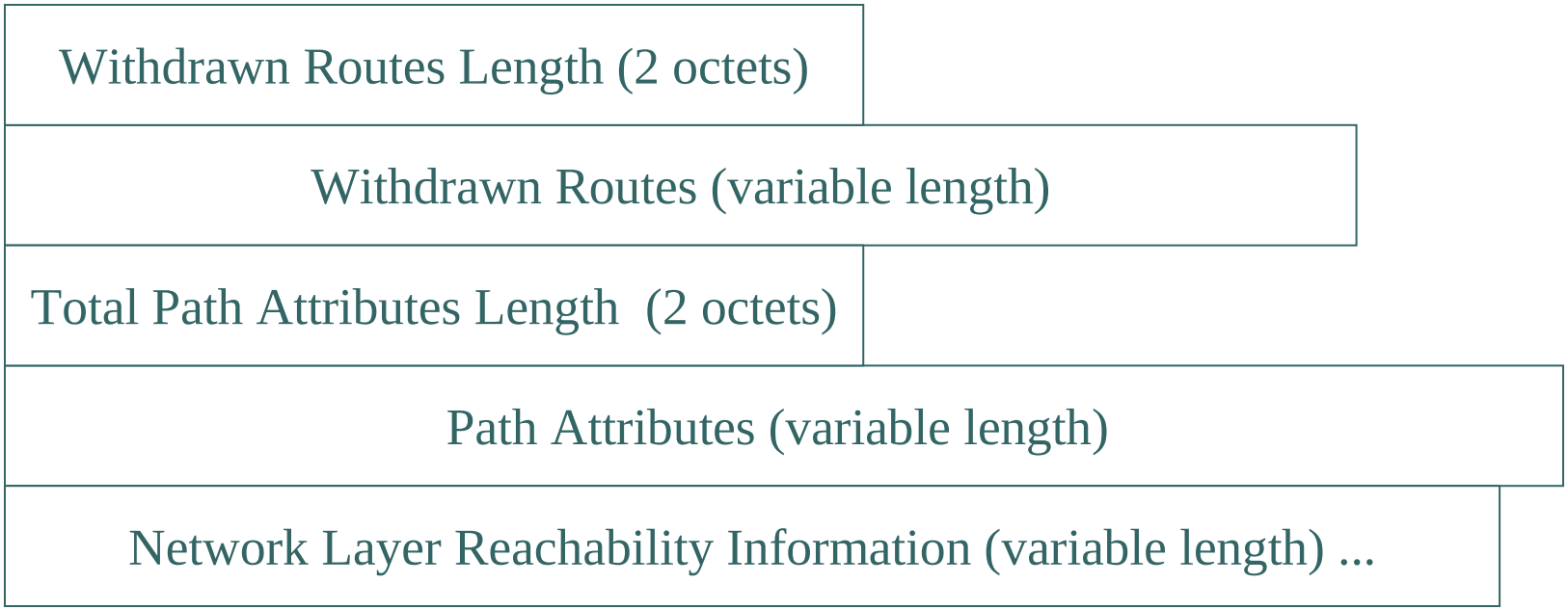
- Sets up a session, first message sent after the TCP connection is made



- Hold Time
  - max. time in sec. allowed between two successive KEEPALIVE or UPDATE message on a link (per-peer basis)
  - Value selected finally is the lower of the hold timer of router and its peer
- BGP Identifier – sender's BGP id, usually one IP address of router
- Optional Parameter Length – length of optional parameter field
- Optional Parameters – list of <type, length, value> for different parameter types. Ex. – Type 1 is authentication information

UPDATE Message

- Sent to advertise a single route or withdraw one or more routes
- Note that multiple networks may be reached by the single route



- Withdrawn routes and NLRI are lists of networks of the form

- All path attributes apply to the single route in NLRI (all networks specified in the NLRI must be reached through a single route)

- Path attributes
  - Well-known mandatory (must exist in packet)
    - ORIGIN (origin of route info – IGP or EGP)
    - AS_PATH (AS no.s through which this update has passed. Specifies the route in terms of AS numbers. Allows checking for routing loops)
    - NEXT_HOP (next router on path)
  - Well-known discretionary (may not exist in packet, but all BGP implementations must understand it)
    - LOCAL_PREF (degree of preference for a route)
  - Optional Transitive & Optional Non-transitive
    - May not be understood by all BGP routers.

NOTIFICATION Message

- Sent on an error before connection is closed

Error Code (1 octet)	Error Subcode (1 octet)	
Data		

KEEPALIVE Message

- Only the header
- Rate based on hold timer value and rate of send of update

Choosing a route

- Reject routes that have unreachable next hop or has this router's AS no. already in its AS_PATH (routing loop)
- Router assigns a degree of preference to each remaining route (can be based on LOCAL_PREF attribute or any other policy)
- For destination, keep route with highest preference. Add if not present
- Tie-breaking rules apply for equal preference
- Commercial routers allow lots of variation and policies

External & Internal BGP

- EBGP – peers in two different AS
- IBGP – peers in same AS
- Type determined by AS number in message header
- Main difference – routes learnt from an IBGP router are not re-advertised to other IBGP routers in the same AS. Advertised to EBGP routers
- BGP routers in 2 ASs usually connected by direct link
- IBGP routers may or may not be directly connected. Require underlying IGP routing table to be set up properly for the TCP connection to work

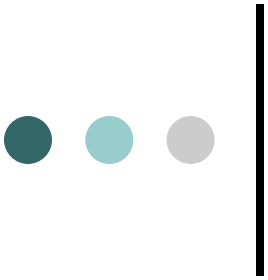
Interactions Between BGP and IGP

- Routes injected from BGP Tables to IGP tables or vice-versa
- Can be dynamic (full redistribution of all routes), semi-dynamic (inject specific networks from the whole table), or static (statically add networks to be advertised)
  - Ex. Cisco routers “redistribute” and “network” command for dynamic and semi-dynamic respectively
- Full injection of BGP into IGP bad.
  - Inside AS, use defaults in IGP table to BGP router
  - Default route can be injected by AS boundary router into AS through IGP
- Partial injection of BGP routes into IGP table helpful sometimes to better direct traffic

Types of AS and Routing

- Stub AS – reaches outside networks through one exit point (ex., a company taking a single connection from a provider)
  - AS has default route to provider
  - Provider can advertise networks in AS through static routes
  - Or, IGP can be used between AS and provider (but customer link instability cause problems in provider)
  - Or run BGP between AS and provider
    - Private AS number given by provider to customer as public AS number hard to obtain in this case

- Multihomed Nontransit AS – more than one exit point to outside and no transit traffic (transit = source and destination both outside AS)
  - Ex. a company taking more than one connection from one or more providers
  - Nontransit AS advertise only its own routes and not routes it learns from other ASs (since no traffic should come from outside and go to these other ASs through this AS)
  - Can work with only static routes set up at routers at customer and providers (ex. IIT Kharagpur, though we are not really an AS)
  - Can run BGP with providers



- Transit AS – provides only transit service, source and destination both outside the AS
  - Ex., ISPs
  - Usually all routers within a transit AS run BGP with each other

Kernel IP Routing Table

- Final table that the IP software uses for routing
- Can have static entries, and dynamic entries from IGPs and EGPs
- Routing daemons (ex. routed, gated, zebra, or proprietory s/w inside routers) inject entries into the routing table from the respective protocols (ex. from BGP table etc.)
- What routes to inject depends on policy