BGP – The Backbone of the Internet

Michael Kehoe

Sr Staff Site Reliability Engineer

Today’s agenda

1	Introduction
2	Basics of BGP
3	History of BGP
4	Details of BGP
6	Conclusion
7	Q&A

Introduction

Michael Kehoe

$ WHOAMI

- Sr Staff Site Reliability Engineer @ LinkedIn
- Infrastructure-SRE Team
- Worked on:
  - Networks
  - Micro-services
  - Traffic Engineering
  - Databases

What is BGP

“Postal service of the internet”

Cloudflare

“The slowest routing protocol in the world”

Jeremy Cioara

“Border Gateway Protocol (BGP) is a standardized exterior gateway protocol designed to exchange routing and reachability information among autonomous systems (AS) on the Internet. The Border Gateway Protocol makes routing decisions based on paths, network policies, or rule-sets configured by a network administrator and is involved in making core routing decisions.”

Wikipedia

https://en.wikipedia.org/*wiki*/Border_Gateway_Protocol

Basics of BGP

Basics of BGP

What is BGP

- Exterior Gateway Protocol (EGP)
- Exchange routing & reachability information among AS on the public internet
- Can also be used within an AS – Interior Border Gateway Protocol (iBGP)
- Path Vector Protocol
- Layer 7 OSI Protocol (TCP/179)

https://www.routerfreak.com/ggp-egp-and-25-years-of-bgp-a-brief-history-of-internet-routing/
https://datapath.io/resources/blog/the-history-of-border-gateway-protocol/

History of BGP

History of BGP

Before BGP

- **First Internet message (1969)**
- **ARPRANET** – (1971)
- **GGP** – RFC 823 (1982)
- **EGP** – RFC 904 (1984)
- **RIP** –RFC 1058 (1988)

https://www.routerfreak.com/ggp-egp-and-25-years-of-bgp-a-brief-history-of-internet-routing/
https://datapath.io/resources/blog/the-history-of-border-gateway-protocol/

History of BGP

Main BGP RFC’s

- **BGPv1** – RFC 1105 (1989)
- **BGPv2** – RFC 1163 (1990)
- **BGPv3** – RFC 1267 (1991)
- **BGPv4** –
  - RFC 1771 (1995) – Original v4
  - RFC 1883/ 2283 IPv6 support (1995/ 1998)
  - RFC 4271 (2006) – Current v4

https://www.routerfreak.com/ggp-egp-and-25-years-of-bgp-a-brief-history-of-internet-routing/
https://datapath.io/resources/blog/the-history-of-border-gateway-protocol/

History of BGP

Key BGP Extensions

- **Communities** – RFC 1997 (1996)
- **Multiprotocol Ext**– RFC 2283 (1998)
- **MD5 Hashing** - RFC 2385 (1998)
- **Flap Damping** – RFC 2439 (1998)
- **32-bit AS Number**– RFC 4893 (2007)

https://www.routerfreak.com/ggp-egp-and-25-years-of-bgp-a-brief-history-of-internet-routing/
https://datapath.io/resources/blog/the-history-of-border-gateway-protocol/

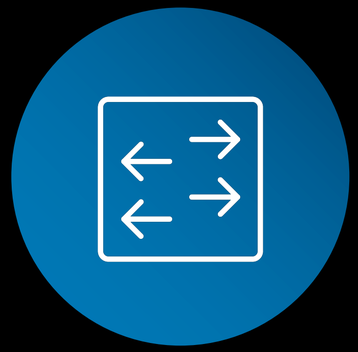
Details of BGP

Terminology

Terminology

- **BGP ID** - Indicates the BGP ID of the sender of BGP messages
- **BGP speaker** – A router that implements BGP
- **Exchange** – Physical network access point where major providers connect & exchange traffic
- **Neighbor/ Peer** – Two BGP speakers configured to connect with each other
- **Route** – A path
- **Transit** – A paid BGP session that provides a full route table
- **RIB** – Routing Information Base

Autonomous Systems



Autonomous Systems

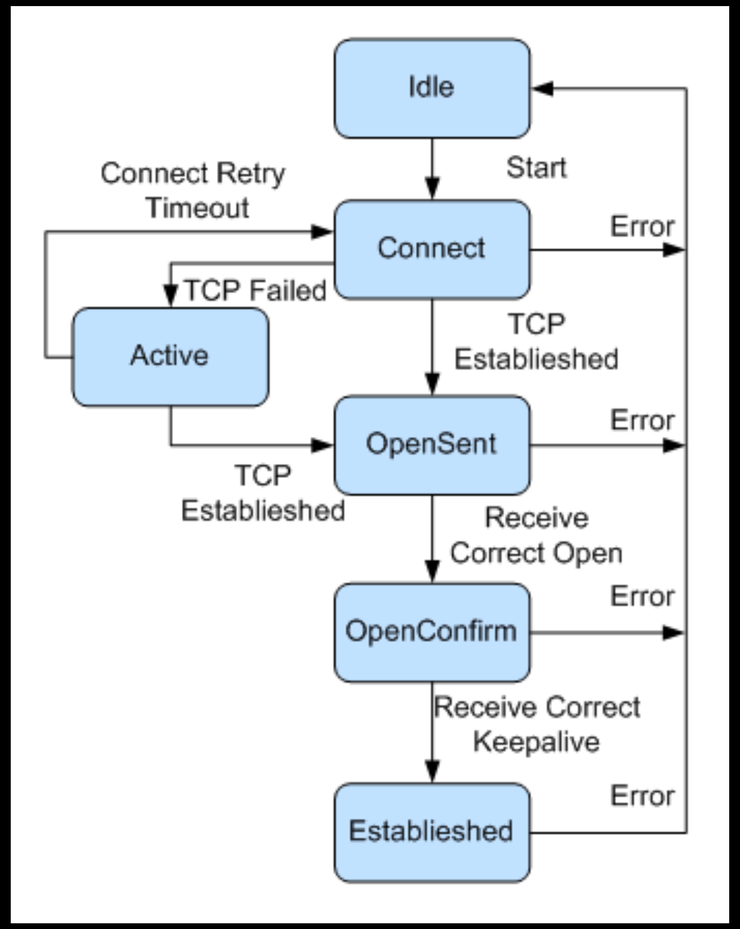
AS

- Set of routers under a single technical administration.
- Collection of IP prefixes
- Common routing policy (to other ASs)
- Registered by a RIR

https://www.routerfreak.com/ggp-egp-and-25-years-of-bgp-a-brief-history-of-internet-routing/

BGP Finite State Machine

BGP Finite State Machine

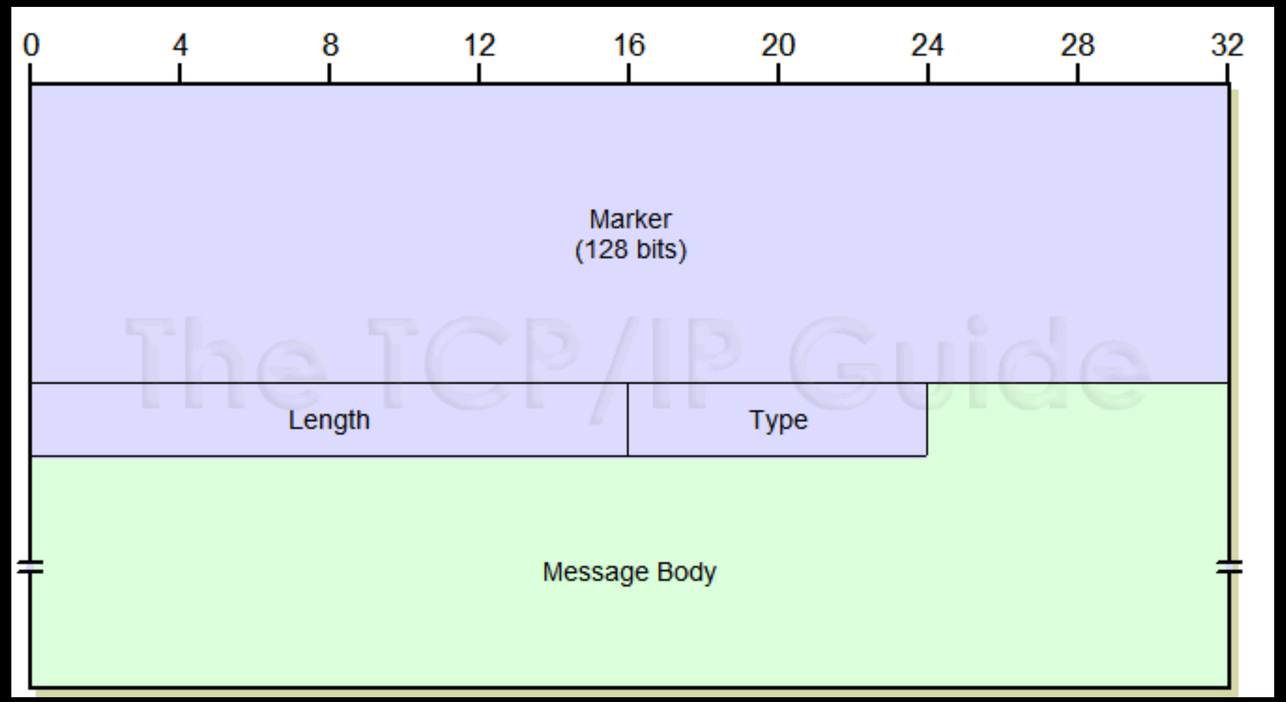


- FSM has 6 states
- 5 types of BGP Messages
  - Open
  - Keepalive
  - Notification
  - Update
  - Route Refresh (RFC 2918)

https://roshanznet.blogspot.com/2017/09/bgp-finite-state-machine.html

BGP Protocol Format

BGP Protocol Format

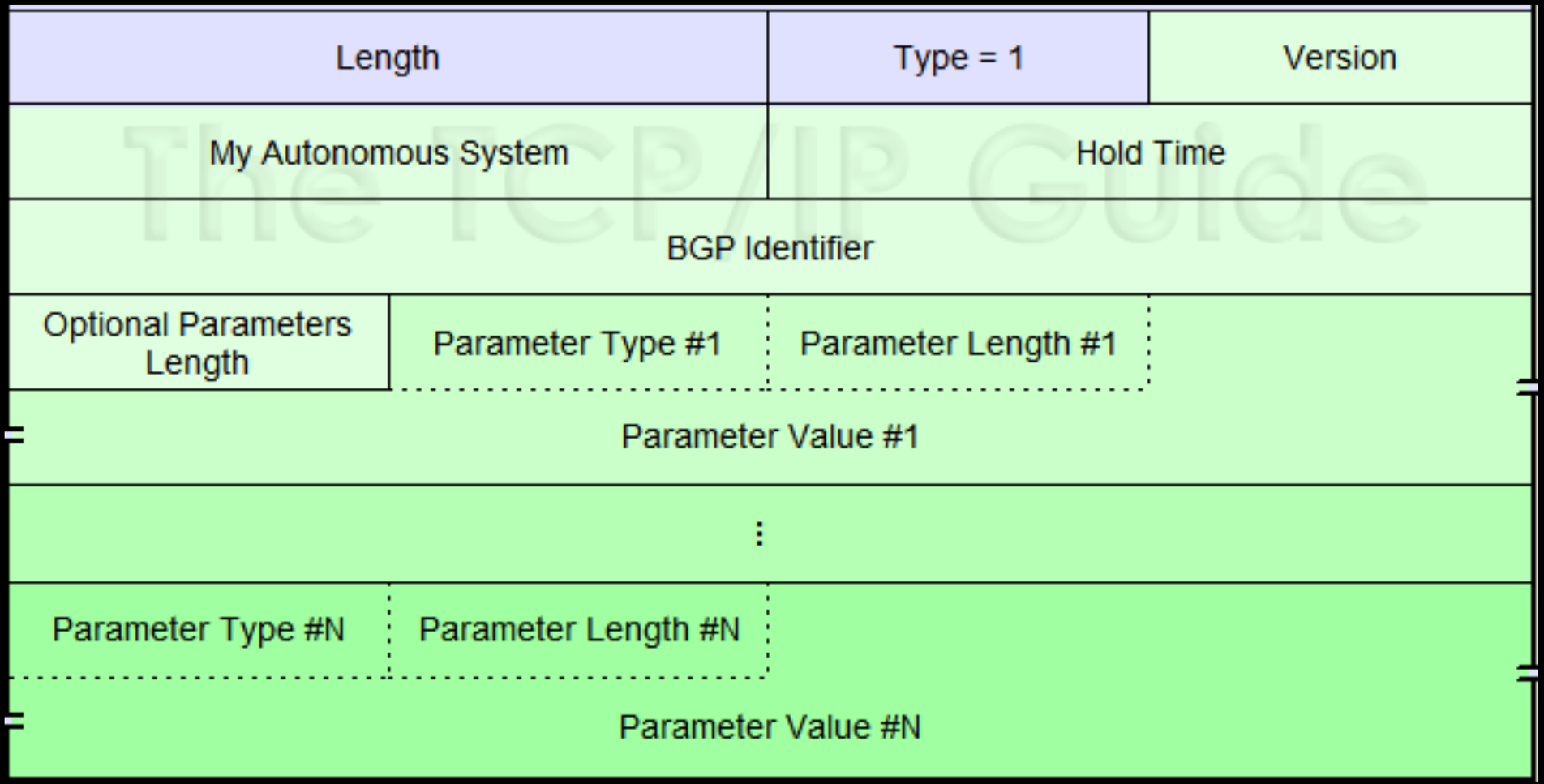


- Marker – Used for sync
- Length – Total length of message
- Type – BGP Message Type
  - Open/ Update/ Notification/ Keepalive/ Route-Refresh
- Message Body – Specific fields used to implement message types

http://www.tcpipguide.com/free/t_BGPMessageGenerationandTransportandGeneralMessageF-2.htm

BGP Protocol Format

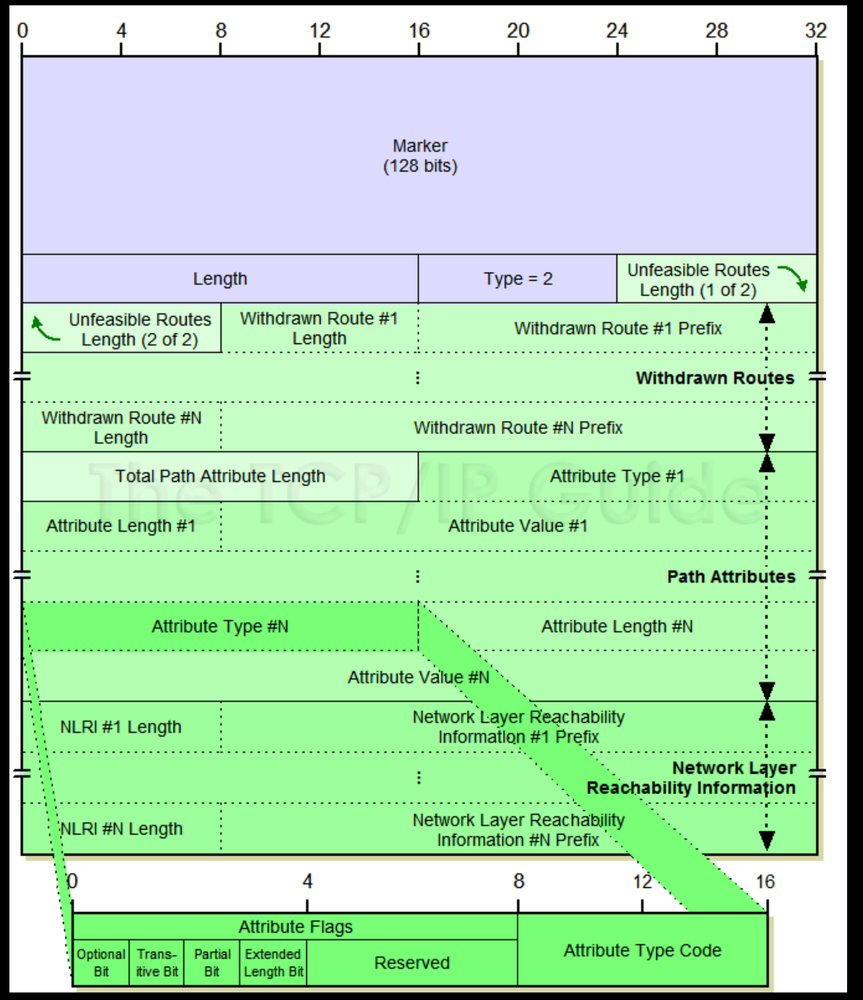
OPEN MESSAGE



http://www.tcpipguide.com/free/t_BGPConnectionEstablishmentOpenMessages-2.htm

BGP Protocol Format

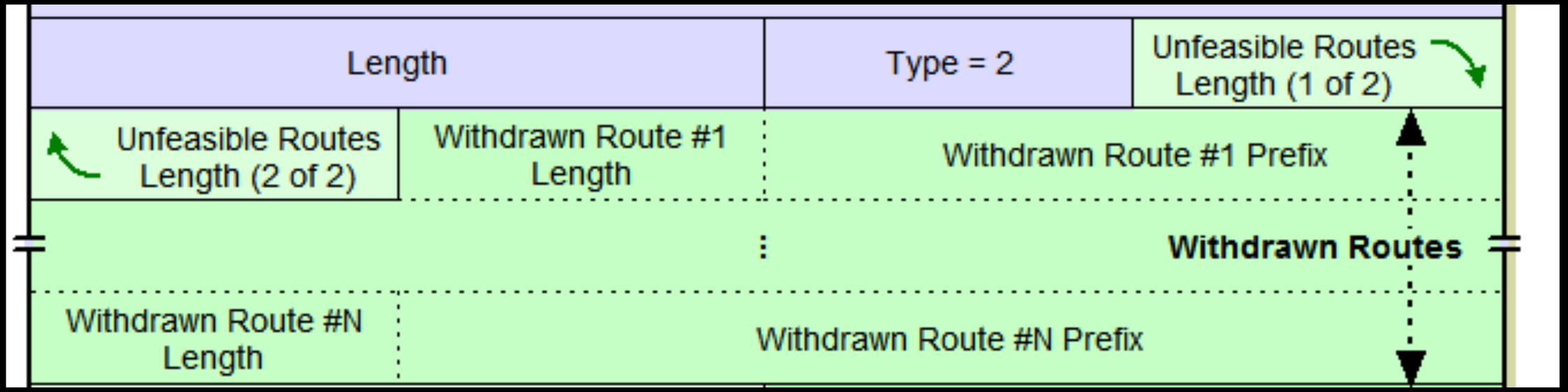
UPDATE MESSAGE



http://www.tcpipguide.com/free/t_BGPRouteInformationExchangeUpdateMessages-2.htm

BGP Protocol Format

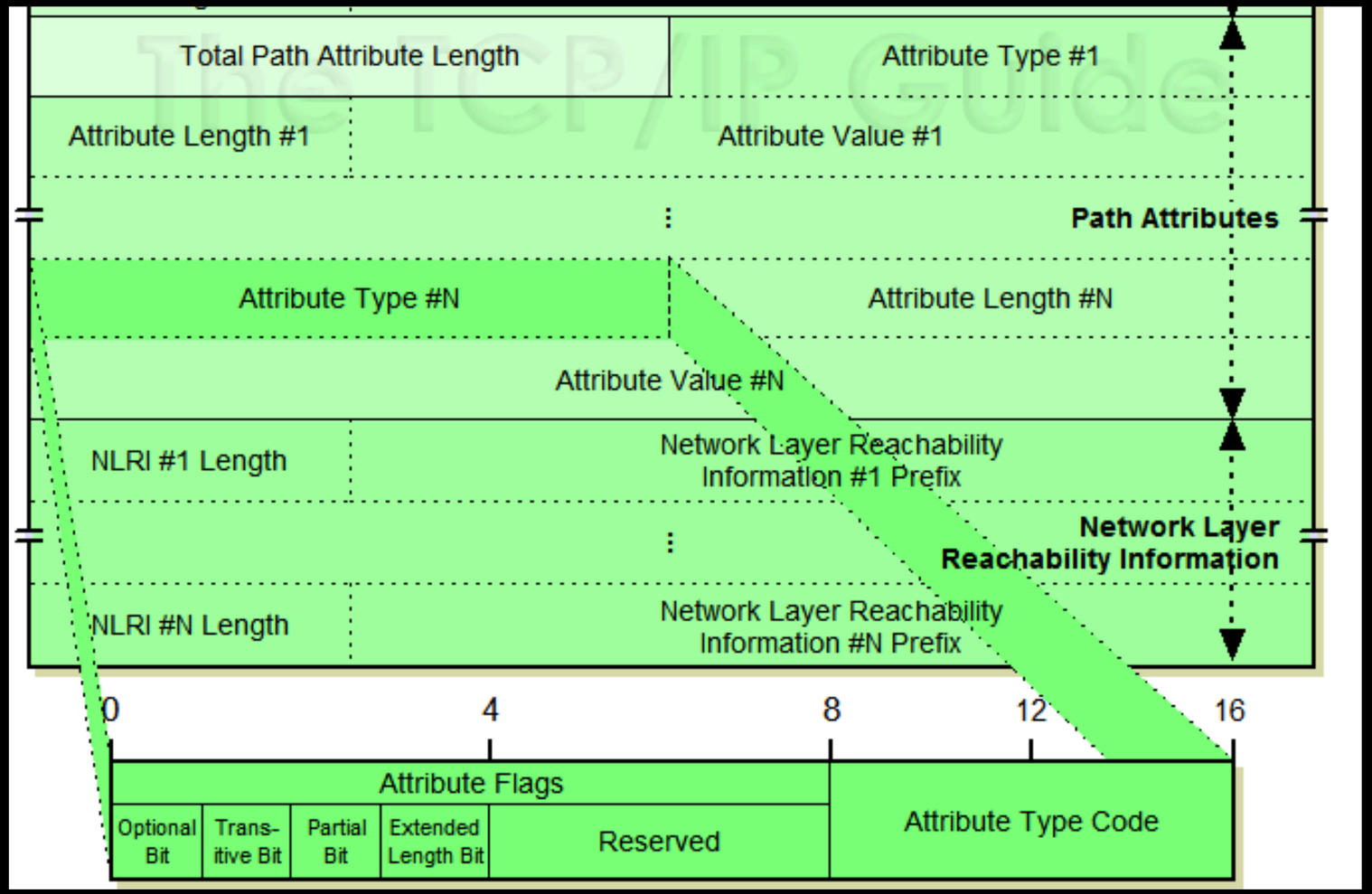
UPDATE MESSAGE



http://www.tcpipguide.com/free/t_BGPRouteInformationExchangeUpdateMessages-2.htm

BGP Protocol Format

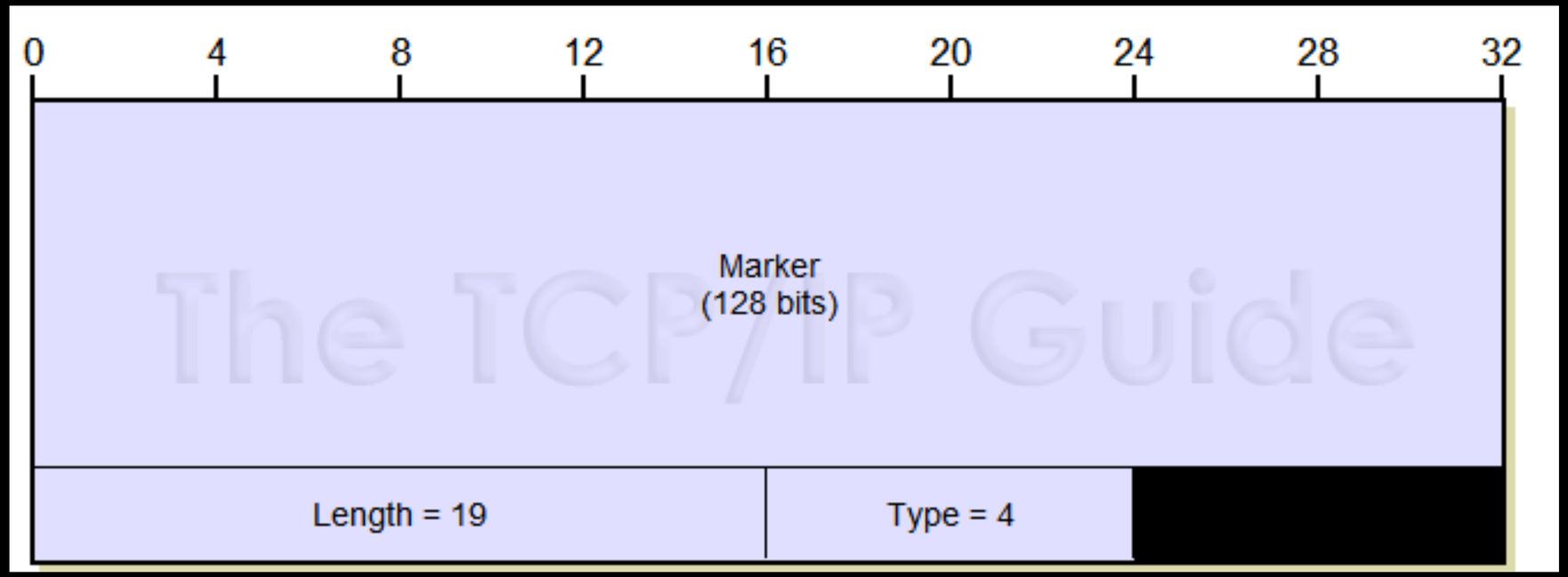
UPDATE MESSAGE



http://www.tcpipguide.com/free/t_BGPRouteInformationExchangeUpdateMessages-2.htm

BGP Protocol Format

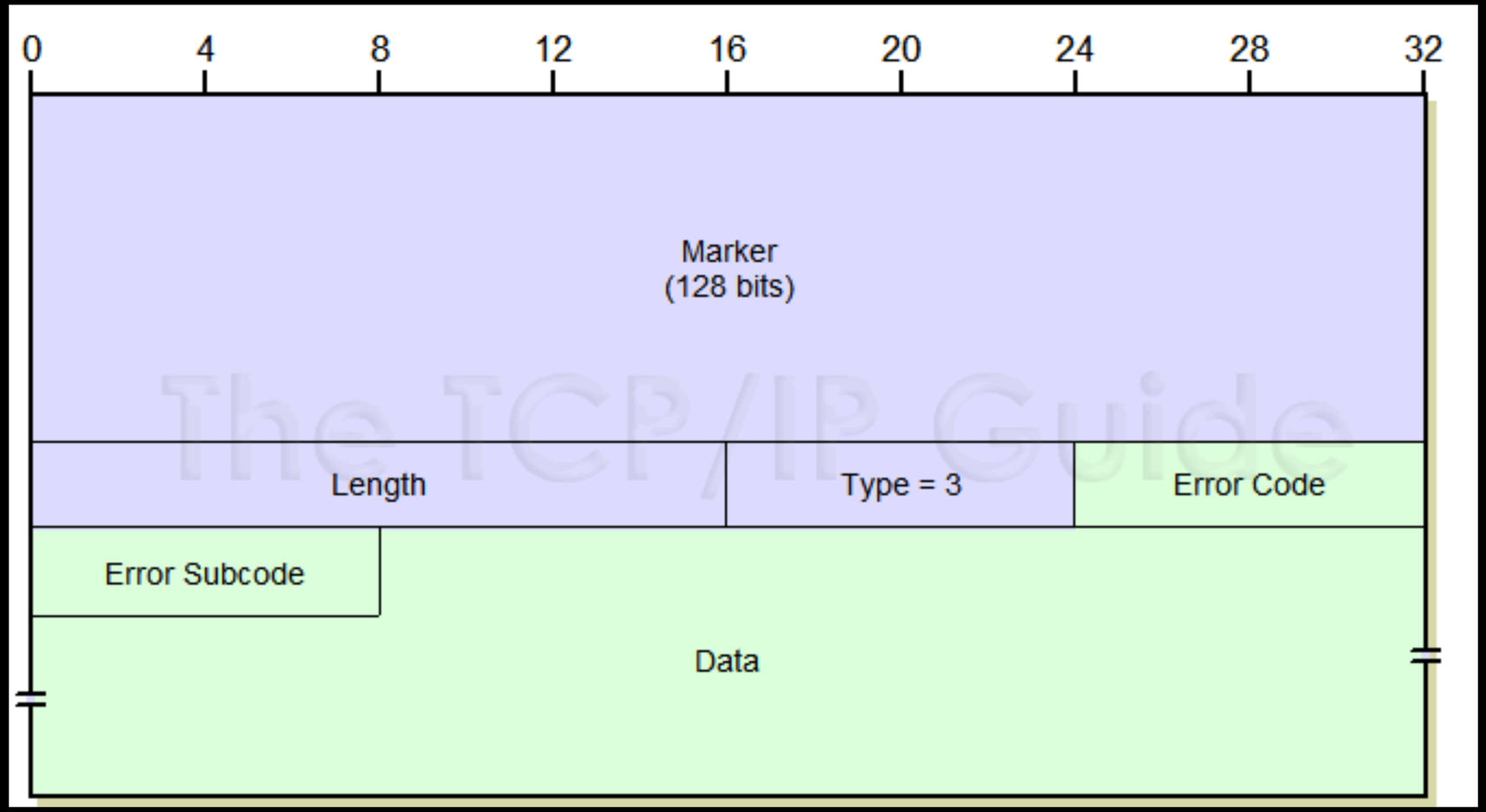
KEEPALIVE MESSAGE



http://www.tcpipguide.com/free/t_BGPConnectivityMaintenanceKeepaliveMessages-2.htm

BGP Protocol Format

NOTIFICATION MESSAGE

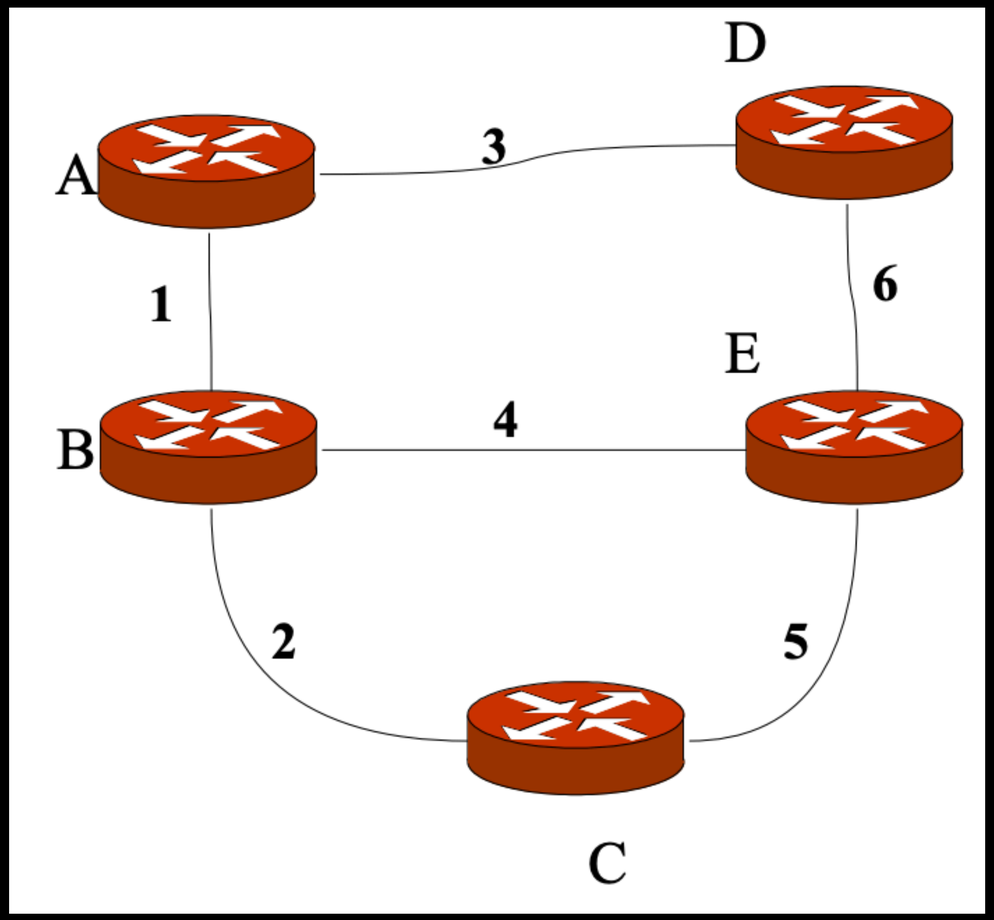


http://www.tcpipguide.com/free/t_BGPErrorReportingNotificationMessages-2.htm

Route Building

Route Selection

BELLMAN FORD DIAGRAM



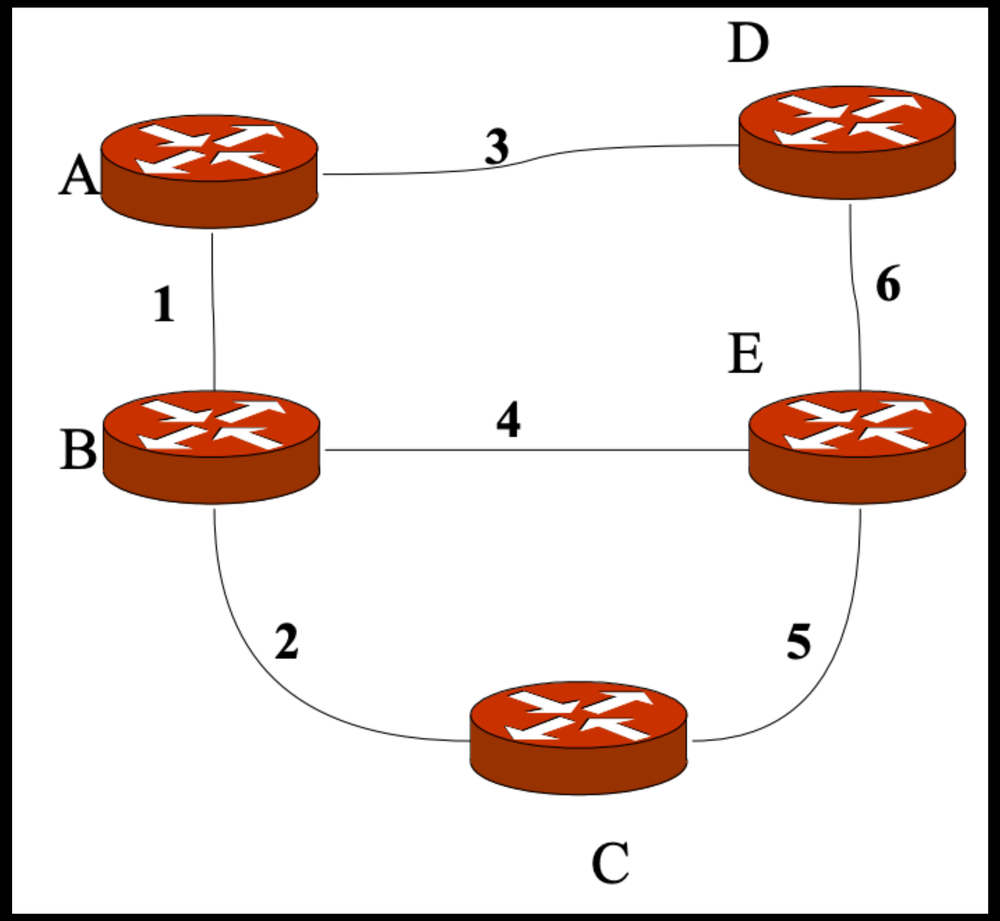
A's directly-connected networks

Dest	Cost	NextHop
B	1	-
D	3	-

http://www.csc.kth.se/utbildning/kth/kurser/DD2393/protok11/material/lectures/L11_Routing.pdf

Route Selection

BELLMAN FORD DIAGRAM



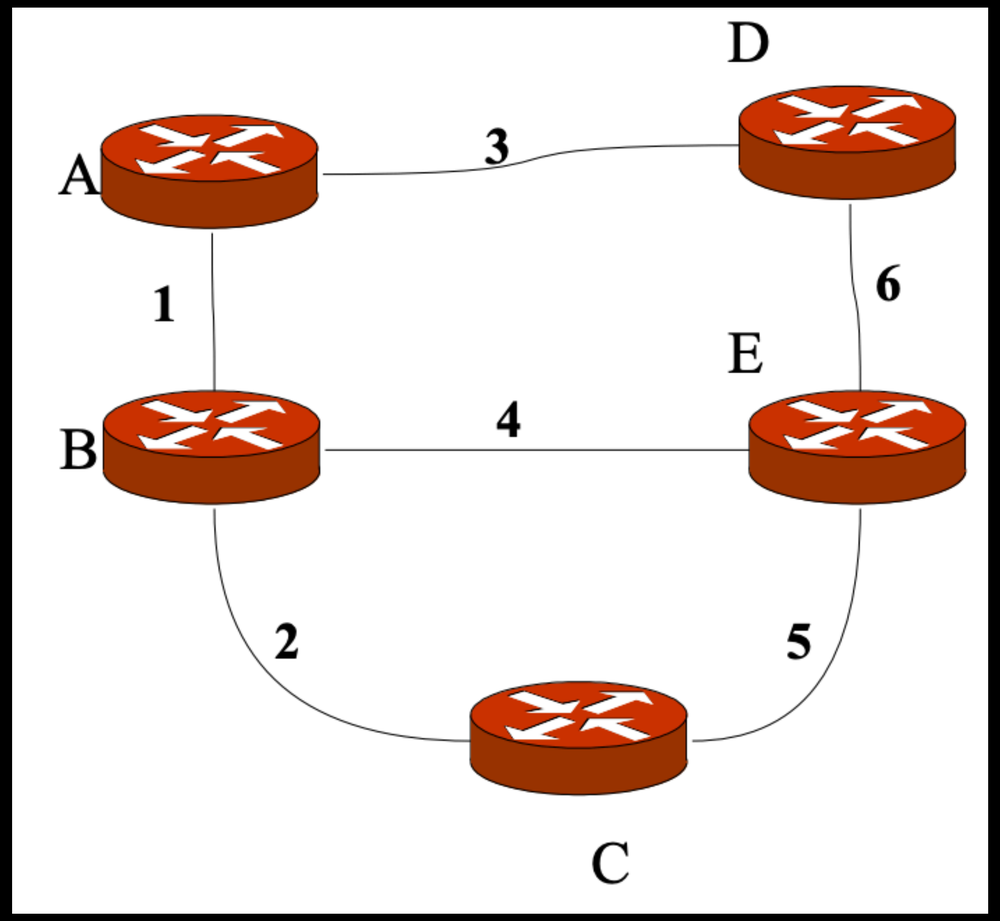
B's directly-connected networks

Dest	Cost
A	1
C	2
E	4

http://www.csc.kth.se/utbildning/kth/kurser/DD2393/protok11/material/lectures/L11_Routing.pdf

Route Selection

BELLMAN FORD DIAGRAM



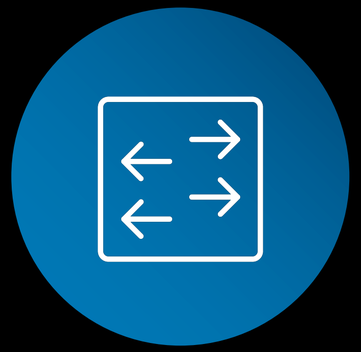
A merge's it's initial state with B's

Dest	Cost	NextHop
B	1	-
C	3	B
D	3	-
E	5	B

http://www.csc.kth.se/utbildning/kth/kurser/DD2393/protok11/material/lectures/L11_Routing.pdf

Route Selection

Route Selection



Routing Selection

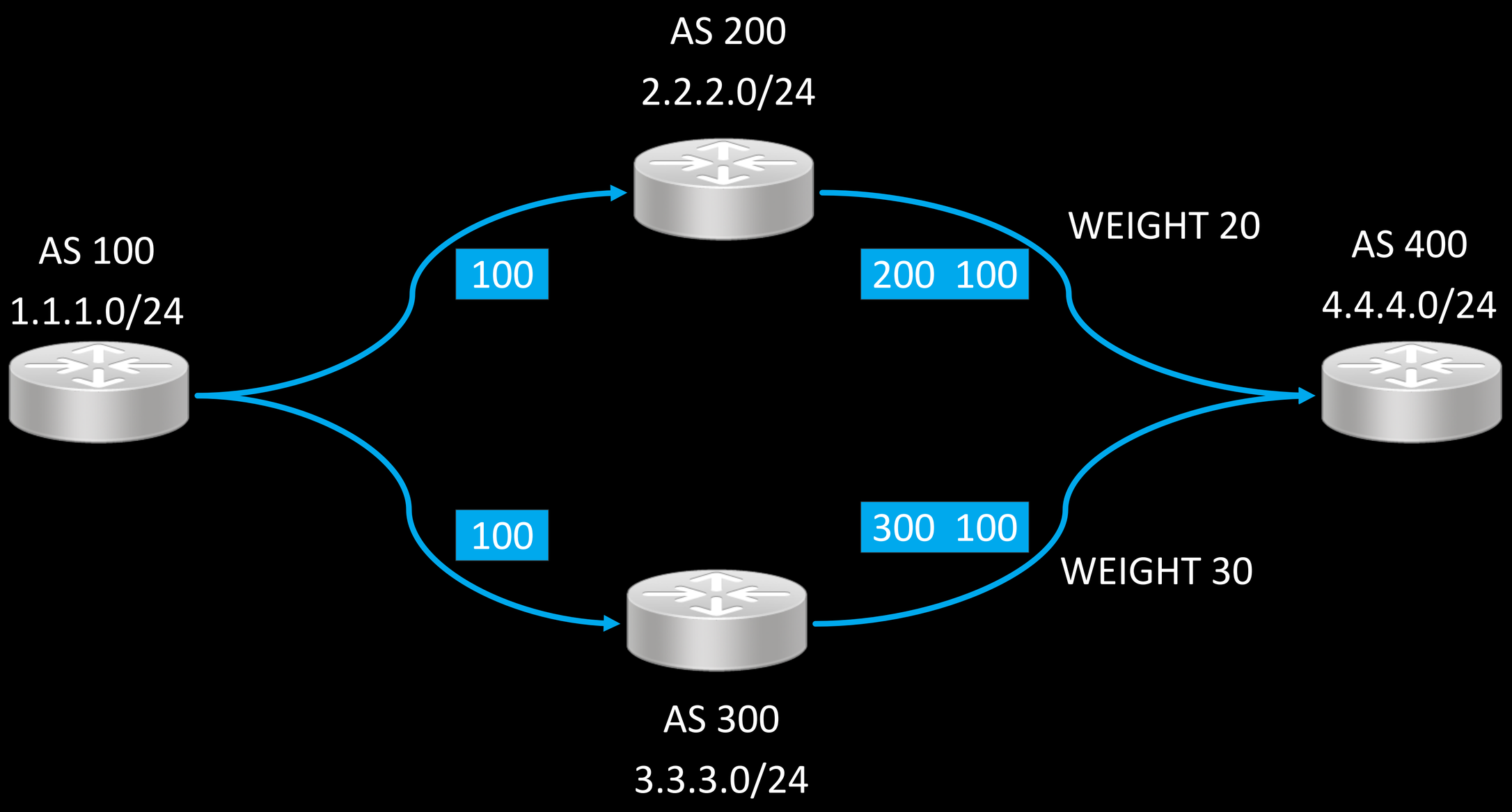
AS

- Highest weight
- Highest local preference
- Shortest AS path
- Origin type
- Multi-Exit Discriminator (MED)
- Route Age
- Other tiebreaking & multipath criteria

https://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13753-25.html
https://www.routerfreak.com/ggp-egp-and-25-years-of-bgp-a-brief-history-of-internet-routing/

How it works in practice?

BGP in Practice



BGP in Practice

AS 400’S ROUTING TABLE

	Network	Next Hop	Path
>*	4.4.4.0/24	i	i
>*	2.2.2.0/24	2.2.2.2	200 i
>*	3.3.3.0/24	3.3.3.2	300 i
*	1.1.1.0/24	2.2.2.2	200 100 i
>*	1.1.1.0/24	3.3.3.2	300 100 i

Conclusion

Conclusion

BGP

- Message Bus of the internet
- Scales reasonably well
  - Full-Convergence is never possible
- Getting implementations that support all features can be difficult

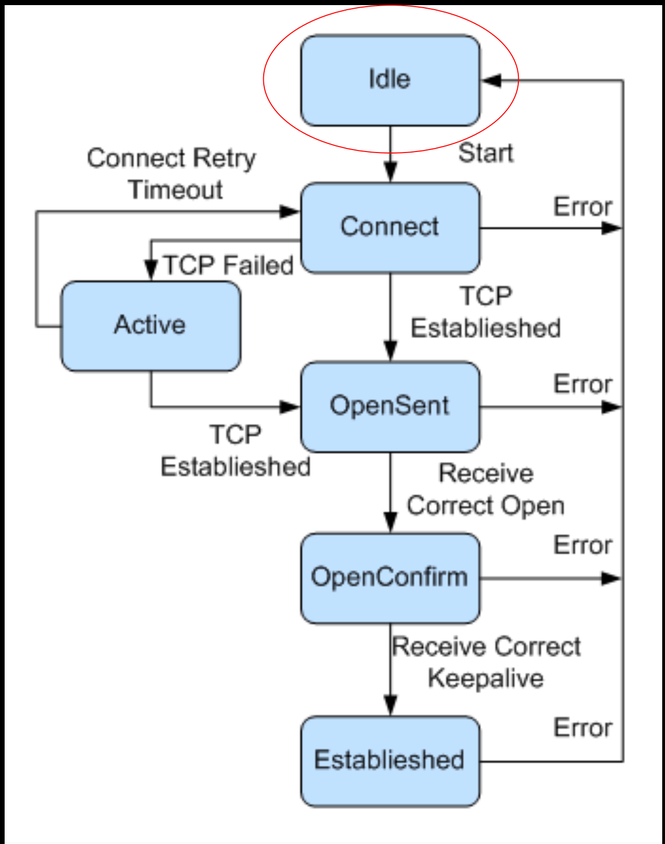
Q & A


LinkedIn

Appendix A: BGP FSM

BGP Finite State Machine

IDLE

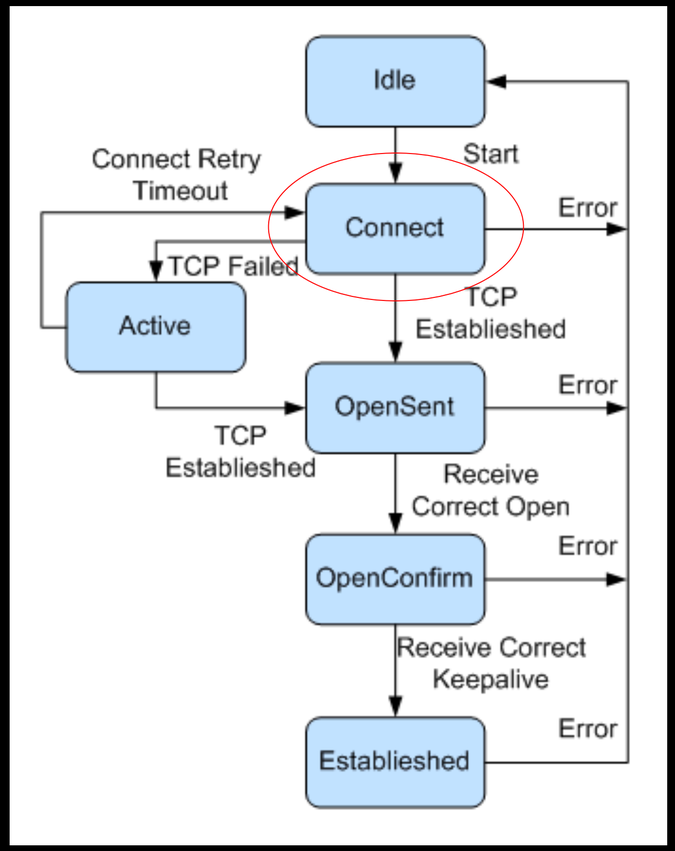


- Initial BGP state
- Router refuses BGP connections
- Goes to CONNECT state after receiving a “Start” event

https://roshanznet.blogspot.com/2017/09/bgp-finite-state-machine.html

BGP Finite State Machine

CONNECT

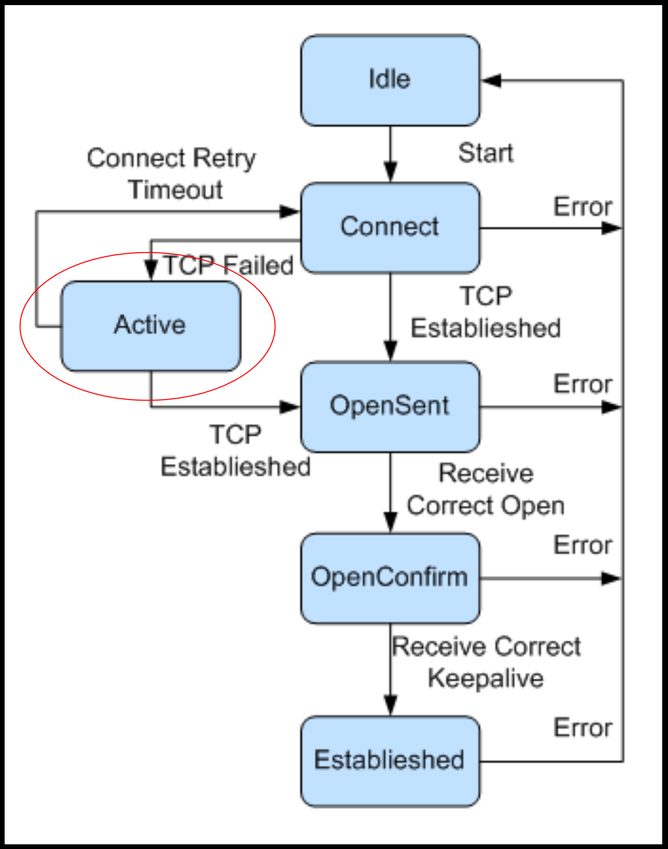


- Tries to create a TCP connection on TCP/179
- If TCP connection is established:
  - Send OPEN message, goes to OPENSENT state
- Else:
  - Goes to ACTIVE state

https://roshanznet.blogspot.com/2017/09/bgp-finite-state-machine.html

BGP Finite State Machine

ACTIVE

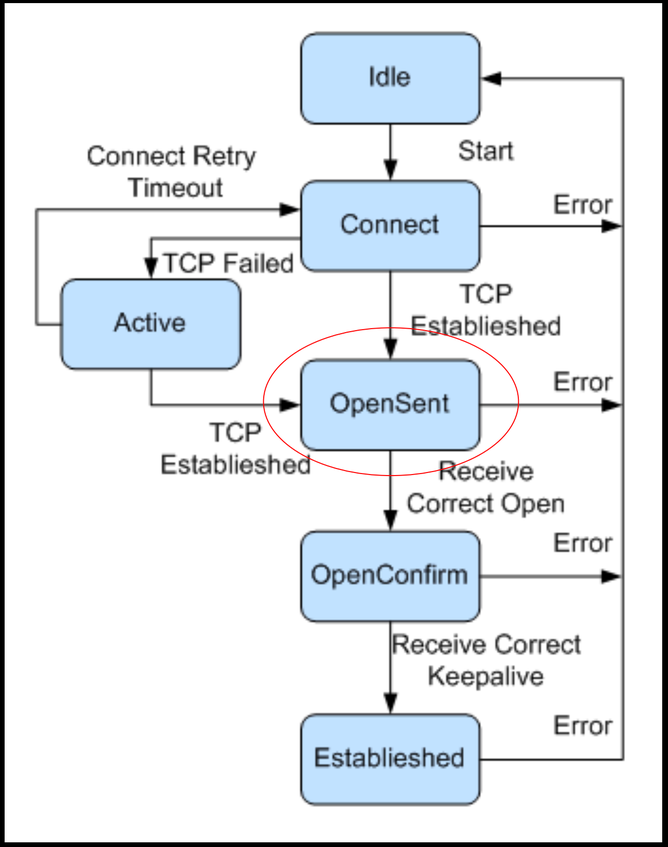


- Tries to establish TCP connection with peer
- If TCP connection is established:
  - Send OPEN message, goes to OPENSENT state
- Else:
  - Stays in ACTIVE state

https://roshanznet.blogspot.com/2017/09/bgp-finite-state-machine.html

BGP Finite State Machine

OPENSENT

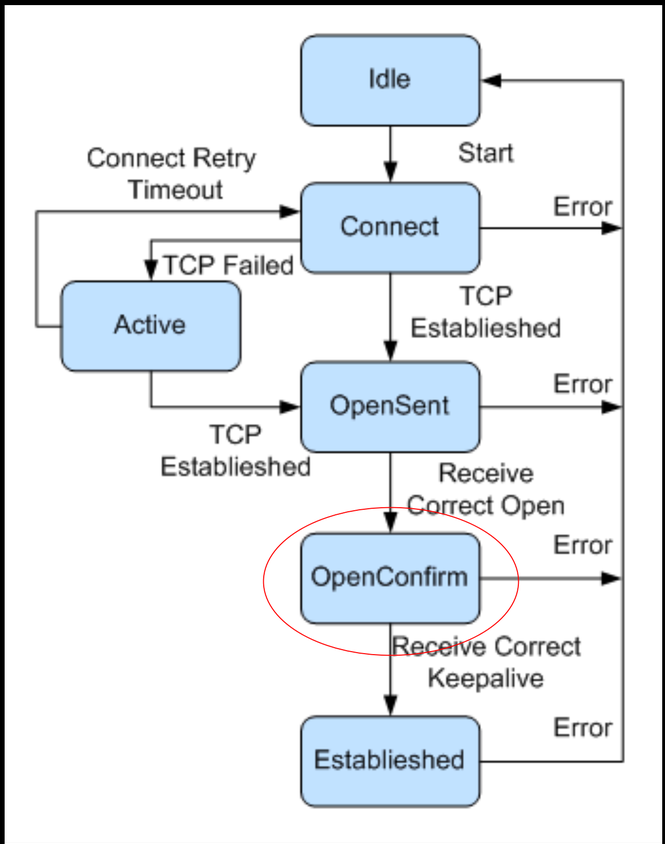


- Waits for OPEN message and checks message validity
- If valid:
  - Router sends KEEPALIVE message
- Else:
  - Return to IDLE state

https://roshanznet.blogspot.com/2017/09/bgp-finite-state-machine.html

BGP Finite State Machine

OPENCONFIRM

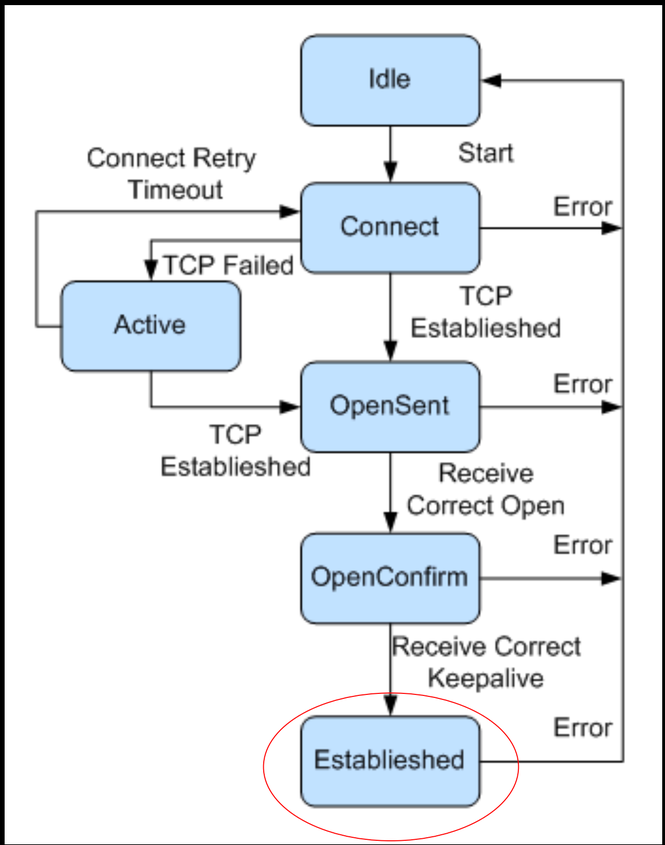


- Waits for KEEPALIVE or NOTIFICATION messages
- If KEEPALIVE is received:
  - Goes to ESTABLISHED
- If NOTIFICATION is received:
  - Goes to IDLE

https://roshanznet.blogspot.com/2017/09/bgp-finite-state-machine.html

BGP Finite State Machine

ESTABLISHED

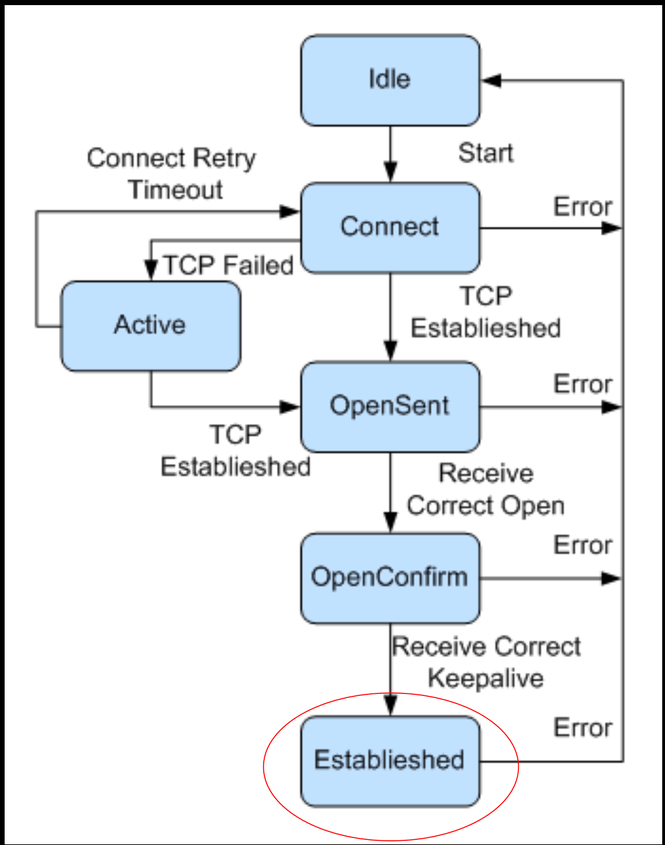


- Exchange of messages:
  - UPDATE
  - KEEPALIVE
  - ROUTEREFRESH
  - NOTIFICATION

https://roshanznet.blogspot.com/2017/09/bgp-finite-state-machine.html

BGP Finite State Machine

ESTABLISHED

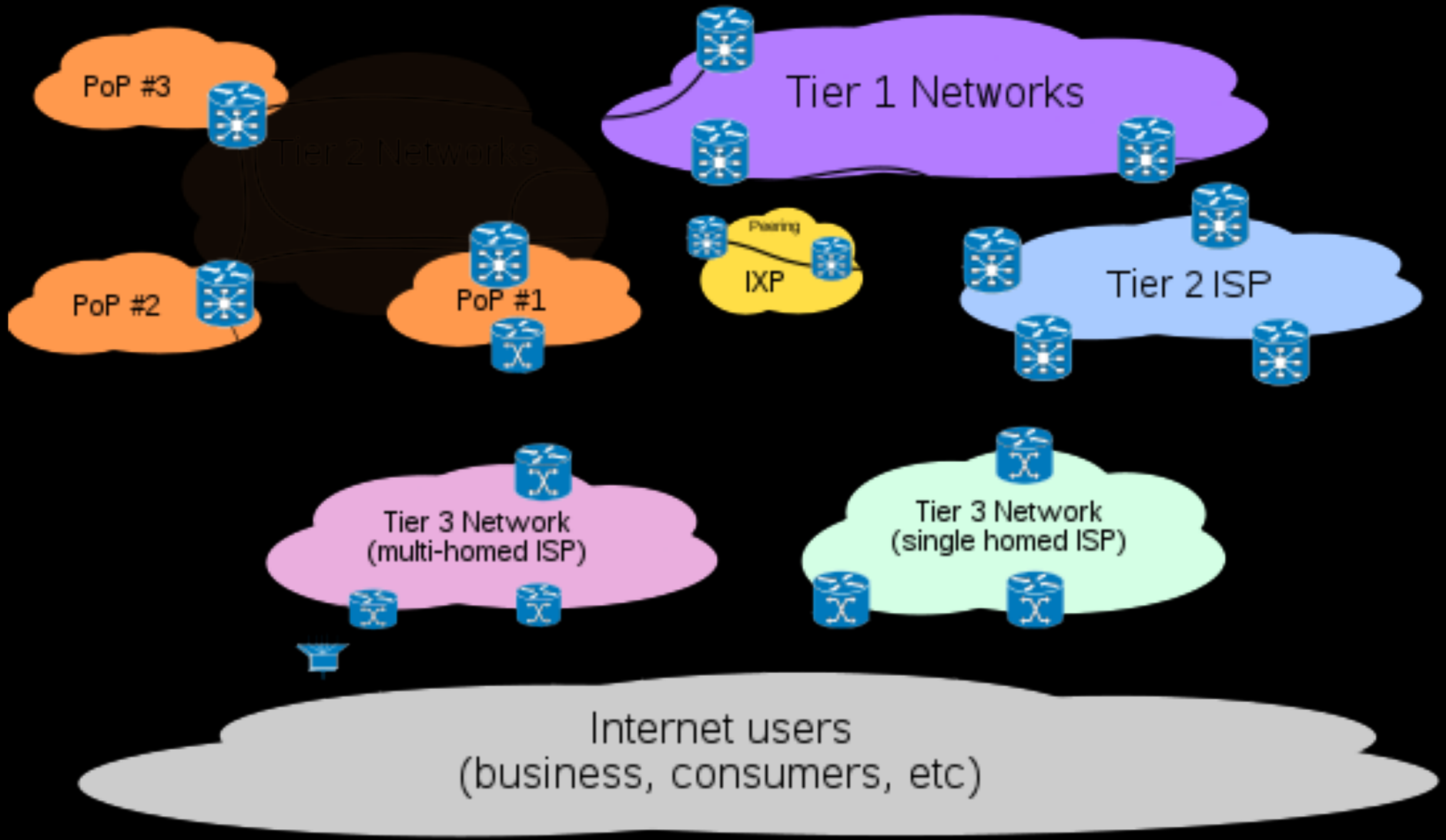


- If valid UPDATE or KEEPALIVE
  - Stays in ESTABLISHED state
- If invalid UPDATE or KEEPALIVE
  - Goes to IDLE state

https://roshanznet.blogspot.com/2017/09/bgp-finite-state-machine.html

Appendix B: ISP Tier Classification

ISP Tier Classification



https://en.wikipedia.org/wiki/Tier_1_network