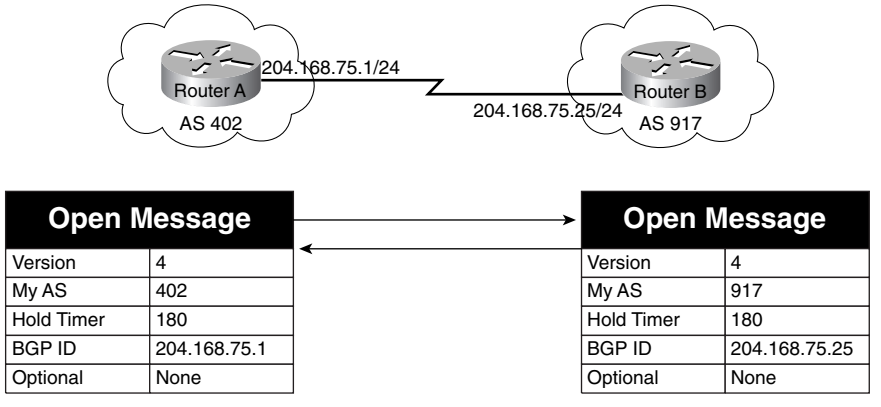
Table 7-3 *BGP OPEN Message Parameters (Continued)*

Message Parameter	Description
Optional	Contains optional BGP parameters, such as the Marker field, which contains authentication information; if authentication is not configured, the Marker field will contain all 1s. The optional *Capabilities* field contains information that allows for BGP feature negotiation; it is either supported or unsupported between BGP peers. If a Capability option is not supported, it will be ignored by the remote peer, and the session will be renegotiated without the capability.

Figure 7-14 *Opening a BGP Session*



Example 7-1 shows a packet capture that contains a BGP OPEN message. BGP uses the IP precedence value of Internetwork Control, shown as 110000, which is used for high-priority routing traffic. For more detailed information on the type of service (ToS) bits, refer to Chapter 5, "Integrated and Differentiated Services." Notice in this message that the TCP session is using the destination port 179, the BGP destination port. The BGP header for this OPEN message (BGP message type 1) includes a Marker field containing all 1s, which indicates that MD-5 authentication is not in use, with a 45-byte header; the Version field specifies that the sending host is using BGP-4. The host belongs to AS number 1 and the hold time is 180 seconds, and the sending host's BGP ID is 192.168.5.1.

Example 7-1 *BGP OPEN Message*

```
Frame Status Source Address Dest. Address Size Rel. Time Delta Time Abs. Time
Summary
8 [10.50.4.1] [10.50.4.2] 99 0:00:37.326 0.003.216 04/28/2002 03:14:50 PM
BGP: type = Open
DLC: -----
DLC Header -----
```

Example 7-1 *BGP OPEN Message (Continued)*

```
DLC:
DLC: Frame 8 arrived at 15:14:50.2341; frame size is 99 (0063 hex) bytes.
DLC: Destination = Station 000427228197
DLC: Source = Station 0004272281D8
DLC: Ethertype = 0800 (IP)
DLC:
IP: ----- IP Header -----
IP:
IP: Version = 4, header length = 20 bytes
IP: Type of service = C0
IP: 110. .... = internetwork control
IP: ...0 .... = normal delay
IP: .... 0... = normal throughput
IP: .... .0.. = normal reliability
IP: .... ..0. = ECT bit - transport protocol will ignore the CE bit
IP: .... ...0 = CE bit - no congestion
IP: Total length = 85 bytes
IP: Identification = 2
IP: Flags = 0X
IP: .0.. .... = might fragment
IP: ..0. .... = last fragment IP: Fragment offset = 0 bytes
IP: Time to live = 1 seconds/hops
IP: Protocol = 6 (TCP)
IP: Header checksum = 9C7B (correct)
IP: Source address = [10.50.4.1]
IP: Destination address = [10.50.4.2]
IP: No options
IP:
TCP: ----- TCP header -----
TCP:
TCP: Source port = 11002
TCP: Destination port = 179 (BGP)
TCP: Sequence number = 3817488861
TCP: Next expected Seq number= 3817488906
TCP: Acknowledgment number = 3816595146
TCP: Data offset = 20 bytes
TCP: Flags = 18
TCP: ..0. .... = (No urgent pointer)
TCP: ...1 .... = Acknowledgment
TCP: .... 1... = Push
TCP: .... .0.. = (No reset)
TCP: .... ..0. = (No SYN)
TCP: .... ...0 = (No FIN)
TCP: Window = 16384
TCP: Checksum = 97C3 (correct)
TCP: No TCP options
TCP: [45 Bytes of data]
TCP:
BGP: ----- BGP Message -----
BGP: BGP: 16 byte Marker (all 1's)
BGP: Length = 45
```

continues

Example 7-1 *BGP OPEN Message (Continued)*

```
BGP: BGP type = 1 (Open)
BGP:
BGP: Version = 4
BGP: AS number = 1
BGP: Hold Time = 180 Second(s)
BGP:
BGP Identifier = C0A80501, [192.168.5.1]
BGP:
BGP: Optional Parameters Length = 16
BGP: Unknown Option Data
BGP:
ADDR HEX                                          ASCII 0000:
00 04 27 22 81 97 00 04 27 22 81 d8 08 00 45 c0 | ..'"....'"....E.
0010: 00 55 00 02 00 00 01 06 9c 7b 0a 32 04 01 0a 32 | .U.......{.2...2
0020: 04 02 2a fa 00 b3 e3 8a 41 dd e3 7c 9e ca 50 18 | ..*.....A..|..P.
0030: 40 00 97 c3 00 00 ff ff ff ff ff ff ff ff ff ff | @...............
0040: ff ff ff ff ff ff 00 2d 01 04 00 01 00 b4 c0 a8 | .......-........
0050: 05 01 10 02 06 01 04 00 01 00 01 02 02 80 00 02 | ................
0060: 02 02 00 | ...
```

BGP Capabilities Advertisement

Starting with BGP-4, BGP peer capabilities can be negotiated during session BGP initialization, using the Optional Capabilities parameter, which is contained in the OPEN message. BGP capabilities negotiation is described in RFC 2842. This element was added into BGP so that new features could be added into the BGP specification without requiring upgrades to newer versions of the protocol.

Using capabilities advertisement, peers can exchange capabilities and negotiate a session using the most agreed-upon features. If one of the peers does not support an optional parameter, it sends the advertiser a NOTIFICATION message with the error "Unsupported Optional Parameter." After receiving the NOTIFICATION message, the advertising peer resends the message without the unsupported parameter and so on, until both peers agree on a set of parameters. Table 7-4 describes the IANA-defined BGP capabilities codes.

Table 7-4 *BGP Capabilities Codes*

Capabilities Code	Description
0	Reserved
1	Multiprotocol extensions for BGP-4
2	ROUTE-REFRESH capability for BGP-4

Table 7-6 *BGP Attribute Flags*

Attribute Flag	Flag Name	Description
Highest bit	Optional bit	Defines whether an attribute is well known (0) or optional (1).
Second highest bit	Transitive bit	Defines whether an optional attribute is nontransitive (0) or transitive (1).
Third highest bit	Partial bit	Defines whether an optional transitive attribute is complete (0) or partial (1).
Fourth highest bit	Extended Length bit	Defines whether the attribute length is 1 octet (0) or 2 octets (1). This flag is only used (set to 1) when the attribute length is greater than 255 octets.

Example 7-2 shows a protocol analysis of an UPDATE message. Notice in the example that this message is a 68-byte BGP type 2 UPDATE message, with a Marker field of all 1s, indicating no authentication is taking place. This update does not contain any withdrawn routes, indicated by the 0 Unfeasible Routes Length. The first attribute in this message is the well-known transitive type 1 ORIGIN attribute value of 0-IGP, indicating that the message came from an I-BGP session. The next well-known transitive attribute is the type 2 AS_PATH attribute; this attribute lists the ASs through which the route has passed. The Path Segment Type field value of 2 (AS-SEQUENCE) means that this update contains an ordered list of autonomous systems. The Path Segment Length field value of 1 indicates that there is only one AS in the path, and the AS Identifier field value indicates that the packet originated from AS 2. The next well-known transitive attribute is the type 3 NEXT-HOP attribute that contains the next hop of 10.50.4.2. The final optional nontransitive attribute is the type 4 MED attribute. This attribute is used to determine which route to take if there are multiple exit points to an AS. The MED for this update is 0.

The next field in this update contains the NLRI information. The NLRI field contains new or changed routes that are being advertised in this message. This message contains routes to the networks 192.168.11.0/24, 192.168.12.0/24, 192.168.13.0/24, 192.168.14.0/24, and 192.168.15.0/24. Each of these routes is presented in [*prefix length*, *subnet mask*, *IP address*] format.

Example 7-2 *BGP UPDATE Message*

```
Frame Status Source Address Dest. Address Size Rel. Time Delta Time Abs. Time
 Summary
13 [10.50.4.2] [10.50.4.1] 141 0:00:37.537 0.001.028 04/28/2002 03:14:50 PM
 BGP: type = Update
DLC: ----- DLC Header -----
DLC:
DLC: Frame 13 arrived at 15:14:50.4449; frame size is 141 (008D hex) bytes.
DLC: Destination = Station 0004272281D8
DLC: Source = Station 000427228197
DLC: Ethertype = 0800 (IP)
```

Example 7-2 *BGP UPDATE Message (Continued)*

```
DLC:
IP: ----- IP Header -----
IP:
IP: Version = 4, header length = 20 bytes
IP: Type of service = C0
IP: 110. .... = internetwork control
IP: ...0 .... = normal delay
IP: .... 0... = normal throughput
IP: .... .0.. = normal reliability
IP: .... ..0. = ECT bit - transport protocol will ignore the CE bit
IP: .... ...0 = CE bit - no congestion
IP: Total length = 127 bytes
IP: Identification = 3
IP: Flags = 0X
IP: .0.. .... = might fragment
IP: ..0. .... = last fragment
IP: Fragment offset = 0 bytes
IP: Time to live = 1 seconds/hops
IP: Protocol = 6 (TCP)
IP: Header checksum = 9C50 (correct)
IP: Source address = [10.50.4.2]
IP: Destination address = [10.50.4.1]
IP: No options
IP:
TCP: ----- TCP header -----
TCP:
TCP: Source port = 179 (BGP)
TCP: Destination port = 11002
TCP: Sequence number = 3816595210
TCP: Next expected Seq number= 3816595297
TCP: Acknowledgment number = 3817488925
TCP: Data offset = 20 bytes
TCP: Flags = 18
TCP: ..0. .... = (No urgent pointer)
TCP: ...1 .... = Acknowledgment
TCP: .... 1... = Push
TCP: .... .0.. = (No reset)
TCP: .... ..0. = (No SYN)
TCP: .... ...0 = (No FIN)
TCP: Window = 16320
TCP: Checksum = 19F9 (correct)
TCP: No TCP options
TCP: [87 Bytes of data]
TCP:
BGP: ----- BGP Message -----
BGP:

BGP: 16 byte Marker (all 1's)
BGP: Length = 68
BGP:
```

continues

Example 7-2 *BGP UPDATE Message (Continued)*

```
BGP type = 2 (Update)
BGP:
BGP: Unfeasible Routes Length = 0
BGP: No Withdrawn Routes in this Update
BGP: Path Attribute Length = 25 bytes
BGP: Attribute Flags = 4X
BGP: 0... .... = Well-known
BGP: .1.. .... = Transitive
BGP: ..0. .... = Complete
BGP: ...0 .... = 1 byte Length
BGP: Attribute type code = 1 (Origin)
BGP: Attribute Data Length = 1
BGP: Origin type = 0 (IGP)
BGP: Attribute Flags = 4X
BGP: 0... .... = Well-known
BGP: .1.. .... = Transitive
BGP: ..0. .... = Complete
BGP: ...0 .... = 1 byte Length
BGP: Attribute type code = 2 (AS Path)
BGP: Attribute Data Length = 4
BGP: Path segment type = 2 (AS_SEQUENCE)
BGP: Path segment length = 1
BGP: AS Identifier = 2
BGP: Attribute Flags = 4X
BGP: 0... .... = Well-known
BGP: .1.. .... = Transitive
BGP: ..0. .... = Complete
BGP: ...0 .... = 1 byte Length
BGP: Attribute type code = 3 (Next Hop)
BGP: Attribute Data Length = 4
BGP: Next Hop = [10.50.4.2]
BGP: Attribute Flags = 8X
BGP: 1... .... = Optional
BGP: .0.. .... = Non-transitive
BGP: ..0. .... = Complete
BGP: ...0 .... = 1 byte Length
BGP: Attribute type code = 4 (Multi Exit Disc)
BGP: Attribute Data Length = 4
BGP: Multi Exit Disc Attribute = 0
BGP:
BGP: Network Layer Reachability Information:
BGP: IP Prefix Length = 24 bits, IP subnet mask [255.255.255.0]
BGP: IP address [192.168.11.0]
BGP: IP Prefix Length = 24 bits, IP subnet mask [255.255.255.0]
BGP: IP address [192.168.12.0]
BGP: IP Prefix Length = 24 bits, IP subnet mask [255.255.255.0]
```

Example 7-2 *BGP UPDATE Message (Continued)*

```
BGP: IP address [192.168.13.0]
BGP: IP Prefix Length = 24 bits, IP subnet mask [255.255.255.0]
BGP: IP address [192.168.14.0]
BGP: IP Prefix Length = 24 bits, IP subnet mask [255.255.255.0]
BGP: IP address [192.168.15.0]
BGP:
BGP: 16 byte Marker (all 1's)
BGP: Length = 19
BGP:
BGP type = 4 (KEEPALIVE)
BGP:
DLC: --- Frame too short
ADDR HEX                                                ASCII
0000: 00 04 27 22 81 d8 00 04 27 22 81 97 08 00 45 c0 | ..'"....'"....E.
0010: 00 7f 00 03 00 00 01 06 9c 50 0a 32 04 02 0a 32 | .........P.2...2
0020: 04 01 00 b3 2a fa e3 7c 9f 0a e3 8a 42 1d 50 18 | ....*..|....B.P.
0030: 3f c0 19 f9 00 00 ff ff ff ff ff ff ff ff ff ff | ?..ù............
0040: ff ff ff ff ff ff 00 44 02 00 00 00 19 40 01 01 | .......D.....@..
0050: 00 40 02 04 02 01 00 02 40 03 04 0a 32 04 02 80 | .@......@...2...
0060: 04 04 00 00 00 00 18 c0 a8 0b 18 c0 a8 0c 18 c0 | ................
0070: a8 0d 18 c0 a8 0e 18 c0 a8 0f ff ff ff ff ff ff | ................
0080: ff ff ff ff ff ff ff ff ff ff 00 13 04 | .............
```

In Figure 7-15, for example, Routers A and B have an established BGP session and are now exchanging routing information using UPDATE messages. Router A sends an update removing two routes: one to 50.1.1.0/24, and one to 50.2.2.0/24. This routing update also contains four new routes: 51.3.3.0/24, 51.4.4.0/24, 51.5.5.0/24, and 60.1.1.0/24. These routes are sent out as routes learned through E-BGP, but originating from an I-BGP session (indicated by the Type 1 IGP path attribute), with an AS path of AS 402, AS 10, and AS 30, with a next hop of 51.5.2.4. Router B receives the UPDATE message, removes the routes to 50.1.1.0/24 and 50.2.2.0/24 from its Adj-RIB-In table, and then adds the routes to the 51.3.3.0/24, 51.4.4.0/24, 51.5.5.0/24, and 60.1.1.0 networks to its Adj-RIB-In table to be processed by its BGP decision process.

Router B then takes its routes from the local Adj-RIB-Out table, and sends an update to Router A containing new routes to networks 197.62.59.0/24, 197.63.59.0/24, and 197.64.59.0/24. The new routes all came from an E-BGP session, but originated from an I-BGP session, using an AS path of AS 917, AS 40, and AS 29, and have the next hop of 197.61.1.1. Router A takes these new routes and adds them to its Adj-RIB-In table to be processed by the BGP decision process, and then adds the best routes to its local BGP routing table Loc-RIB. Until there are any route changes, Routers A and B will not send any further routing updates; they will only send KEEPALIVE messages back and forth, notifying each other that the BGP session is still active.

KEEPALIVE Message

After the BGP session has been successfully established, and BGP updates have been sent and received, the BGP peers send each other periodic KEEPALIVE messages. KEEPALIVE messages are sent by the peering routers every 60 seconds, by default, to notify neighboring peers that the BGP connection is active. The KEEPALIVE message interval can be changed from the default value to any other value between 3 and 4,294,967,295 or set to 0 to signify that KEEPALIVE messages will not be exchanged. KEEPALIVE values of 1 or 2 seconds are not valid. If invalid KEEPALIVE values are used, the BGP session will fail with the NOTIFICATION message "Open failed: Connection refused by remote host." KEEPALIVE timers might also be set to 1/3 the negotiated hold-timer value, which is, by default, 180 seconds. Figure 7-16 shows the process, followed by each of the three BGP messages, including the KEEPALIVE messages sent during a successful BGP session.

The KEEPALIVE message contains no data; it is just a 19-byte BGP header, as shown in the protocol analysis contained in Example 7-3.

Example 7-3 *BGP KEEPALIVE Message*

```
Frame Status Source Address  Dest. Address Size  Rel. Time Delta Time Abs. Time
 Summary
10 [10.50.4.1] [10.50.4.2] 73 0:00:37.336 0.008.155 04/28/2002 03:14:50 PM
 BGP: type =
KEEPALIVE
DLC: ----- DLC Header -----
DLC:
DLC: Frame 10 arrived at 15:14:50.2443; frame size is 73 (0049 hex) bytes.
DLC: Destination = Station 000427228197
DLC: Source = Station 0004272281D8
DLC: Ethertype = 0800 (IP)
DLC:
IP: ----- IP Header -----
IP: IP: Version = 4, header length = 20 bytes
IP: Type of service = C0
IP: 110. .... = internetwork control
IP: ...0 .... = normal delay
IP: .... 0... = normal throughput
IP: .... .0.. = normal reliability
IP: .... ..0. = ECT bit - transport protocol will ignore the CE bit
IP: .... ...0 = CE bit - no congestion
IP: Total length = 59 bytes
IP: Identification = 3 IP: Flags = 0X
IP: .0.. .... = might fragment
IP: ..0. .... = last fragment
IP: Fragment offset = 0 bytes
IP: Time to live = 1 seconds/hops IP: Protocol = 6 (TCP)
IP: Header checksum = 9C94 (correct)
IP: Source address = [10.50.4.1]
IP: Destination address = [10.50.4.2]
IP: No options
IP:
```

Example 7-3 *BGP KEEPALIVE Message (Continued)*

```
TCP: ----- TCP header -----
TCP:
TCP: Source port = 11002
TCP: Destination port = 179 (BGP)
TCP: Sequence number = 3817488906
TCP: Next expected Seq number= 3817488925
TCP: Acknowledgment number = 3816595191
TCP: Data offset = 20 bytes
TCP: Flags = 18 TCP: ..0. .... = (No urgent pointer)
TCP: ...1 .... = Acknowledgment
TCP: .... 1... = Push
TCP: .... .0.. = (No reset)
TCP: .... ..0. = (No SYN)
TCP: .... ...0 = (No FIN)
TCP: Window = 16339
TCP: Checksum = 7BB6 (correct)
TCP: No TCP options
TCP: [19 Bytes of data]
TCP: BGP: ----- BGP Message -----
BGP:
BGP: 16 byte Marker (all 1's)
BGP: Length = 19 BGP: BGP type = 4 (KEEPALIVE)
BGP:
BGP:
ADDR HEX                             ASCII
0000: 00 04 27 22 81 97 00 04 27 22 81 d8 08 00 45 c0 | ..'"....'"....E.
0010: 00 3b 00 03 00 00 01 06 9c 94 0a 32 04 01 0a 32 | .;.........2...2
0020: 04 02 2a fa 00 b3 e3 8a 42 0a e3 7c 9e f7 50 18 | ..*.....B..|..P.
0030: 3f d3 7b b6 00 00 ff ff ff ff ff ff ff ff ff ff | ?.{.............
0040: ff ff ff ff ff ff 00 13 04 | .........
```

ROUTE-REFRESH Message

Prior to Cisco IOS Software Release 12.0(6)T, all BGP-speaking routers used to require a manual BGP session reset each time the local routing policy changed. This session reset allowed peers to apply new policies as the routers processed and received the incoming routing updates from their remote peers. In legacy versions of Cisco IOS software, this problem was solved, on a peer-by-peer basis, using BGP *soft reconfiguration*. After BGP soft reconfiguration has been configured on a legacy peer, that router stores the full, unmodified copy of the incoming Adj-RIB-In table that it received from each remote peer in memory. Although this feature promotes network stability by preventing BGP session interruptions, it also consumes large amounts of system resources. Soft configuration is triggered each time a soft-reconfiguration request is issued using the **clear ip bgp** {***** | *ip-address* | *peer-group*} **soft** [**in** | **out**] command; the use of this command is covered later in Chapter 9, "Advanced BGP Configuration." When this command is issued, the local BGP peer acts as though it has just received a full routing update from the remote peer by refreshing routes stored in the Loc-RIB table using the Adj-RIB-In information stored in memory.