Border Gateway Protocol (BGP)

Background

Routing involves two basic activities: determination of optimal routing paths and the transport of information groups (typically called packets) through an internetwork. The transport of packets through an internetwork is relatively straightforward. Path determination, on the other hand, can be very complex. One protocol that addresses the task of path determination in today's networks is the *Border Gateway Protocol* (BGP). This chapter summarizes the basic operations of BGP and provides a description of its protocol components.

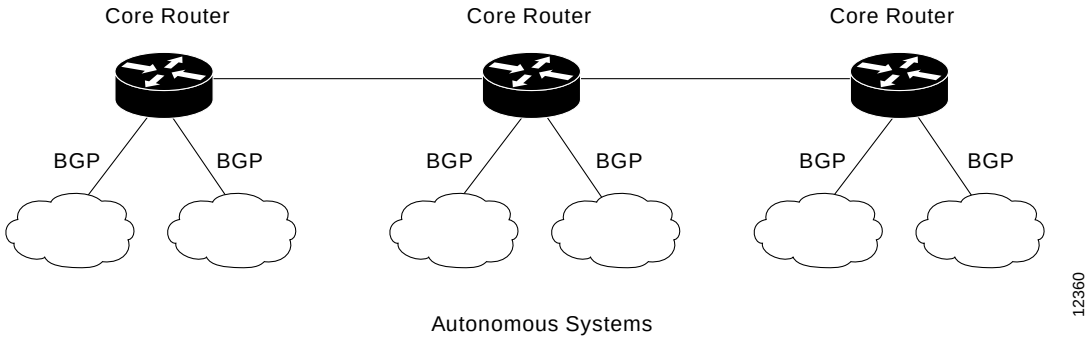
BGP performs interdomain routing in Transmission-Control Protocol/Internet Protocol (TCP/IP) networks. BGP is an exterior gateway protocol (EGP), which means that it performs routing between multiple autonomous systems or domains and exchanges routing and reachability information with other BGP systems.

BGP was developed to replace its predecessor, the now obsolete *Exterior Gateway Protocol* (EGP), as the standard exterior gateway-routing protocol used in the global Internet. BGP solves serious problems with EGP and scales to Internet growth more efficiently.

Note EGP is a particular instance of an exterior gateway protocol (also EGP)—the two should not be confused.

Figure 35-1 illustrates core routers using BGP to route traffic between autonomous systems.

Figure 35-1 Core routers can use BGP to route traffic between autonomous systems.



BGP is specified in several *Request For Comments* (RFCs):

- RFC 1771—Describes BGP4, the current version of BGP

- RFC 1654—Describes the first BGP4 specification
- RFC 1105, RFC 1163, and RFC 1267—Describes versions of BGP prior to BGP4

BGP Operation

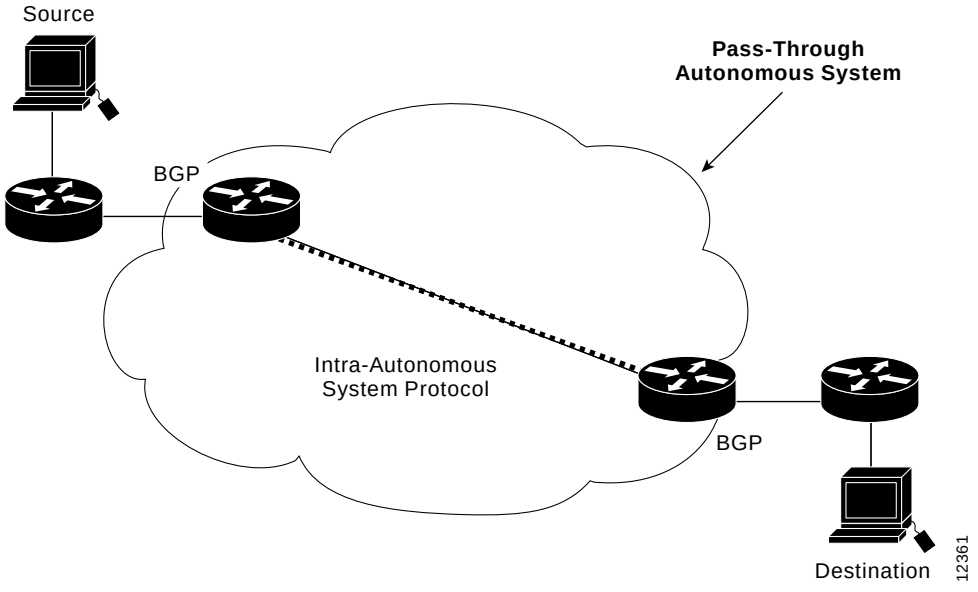
BGP performs three types of routing: *interautonomous system routing*, *intra-autonomous system routing*, and *pass-through autonomous system routing*.

Interautonomous system routing occurs between two or more BGP routers in different autonomous systems. Peer routers in these systems use BGP to maintain a consistent view of the internetwork topology. BGP neighbors communicating between autonomous systems must reside on the same physical network. The Internet serves as an example of an entity that uses this type of routing because it is comprised of autonomous systems or administrative domains. Many of these domains represent the various institutions, corporations, and entities that make up the Internet. BGP is frequently used to provide path determination to provide optimal routing within the Internet.

Intra-autonomous system routing occurs between two or more BGP routers located within the same autonomous system. Peer routers within the same autonomous system use BGP to maintain a consistent view of the system topology. BGP also is used to determine which router will serve as the connection point for specific external autonomous systems. Once again, the Internet provides an example of interautonomous system routing. An organization, such as a university, could make use of BGP to provide optimal routing within its own administrative domain or autonomous system. The BGP protocol can provide both inter- and intra-autonomous system routing services.

Pass-through autonomous system routing occurs between two or more BGP peer routers that exchange traffic across an autonomous system that does not run BGP. In a pass-through autonomous system environment, the BGP traffic did not originate within the autonomous system in question and is not destined for a node in the autonomous system. BGP must interact with whatever intra-autonomous system routing protocol is being used to successfully transport BGP traffic through that autonomous system. Figure 35-2 illustrates a pass-through autonomous system environment:

Figure 35-2 In pass-through autonomous system routing, BGP pairs with another intra-autonomous system-routing protocol.



BGP Routing

As with any routing protocol, BGP maintains routing tables, transmits routing updates, and bases routing decisions on routing metrics. The primary function of a BGP system is to exchange network-reachability information, including information about the list of autonomous system paths, with other BGP systems. This information can be used to construct a graph of autonomous system connectivity from which routing loops can be pruned and with which autonomous system-level policy decisions can be enforced.

Each BGP router maintains a routing table that lists all feasible paths to a particular network. The router does not refresh the routing table, however. Instead, routing information received from peer routers is retained until an incremental update is received.

BGP devices exchange routing information upon initial data exchange and after incremental updates. When a router first connects to the network, BGP routers exchange their entire BGP routing tables. Similarly, when the routing table changes, routers send the portion of their routing table that has changed. BGP routers do not send regularly scheduled routing updates, and BGP routing updates advertise only the optimal path to a network.

BGP uses a single routing metric to determine the best path to a given network. This metric consists of an arbitrary unit number that specifies the degree of preference of a particular link. The BGP metric typically is assigned to each link by the network administrator. The value assigned to a link can be based on any number of criteria, including the number of autonomous systems through which the path passes, stability, speed, delay, or cost.

BGP Message Types

Four BGP message types are specified in RFC 1771, *A Border Gateway Protocol 4 (BGP-4)*: open message, update message, notification message, and keep-alive message.

The *open message* opens a BGP communications session between peers and is the first message sent by each side after a transport-protocol connection is established. Open messages are confirmed using a keep-alive message sent by the peer device and must be confirmed before updates, notifications, and keep-alives can be exchanged.

An *update message* is used to provide routing updates to other BGP systems, allowing routers to construct a consistent view of the network topology. Updates are sent using the Transmission-Control Protocol (TCP) to ensure reliable delivery. Update messages can withdraw one or more unfeasible routes from the routing table and simultaneously can advertise a route while withdrawing others.

The *notification message* is sent when an error condition is detected. Notifications are used to close an active session and to inform any connected routers of why the session is being closed.

The keep-alive message notifies BGP peers that a device is active. Keep-alives are sent often enough to keep the sessions from expiring.

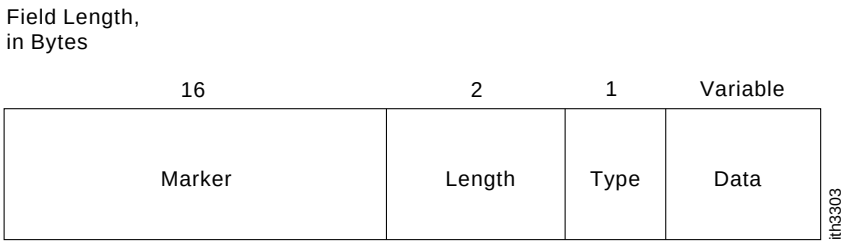
BGP Packet Formats

The sections that follow summarize BGP open, updated, notification, and keep-alive message types, as well as the basic BGP header format. Each is illustrated with a format drawing, and the fields shown are defined.

Header Format

All BGP message types use the basic packet header. Open, update, and notification messages have additional fields, but keep-alive messages use only the basic packet header. Figure 35-3 illustrates the fields used in the BGP header. The section that follows summarizes the function of each field.

Figure 35-3 A BGP packet header consists of four fields.



BGP Packet-Header Fields

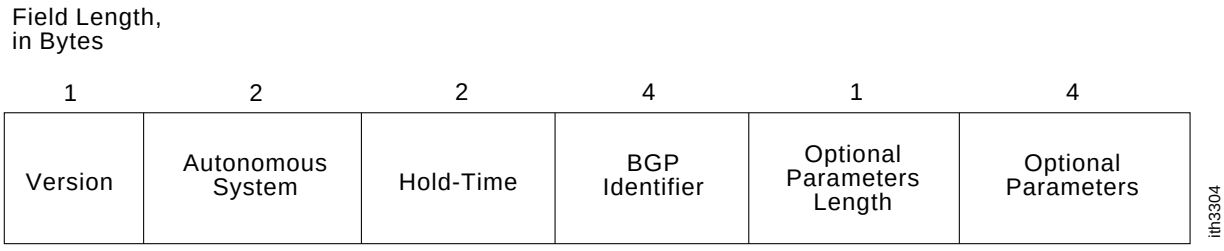
Each BGP packet contains a header whose primary purpose is to identify the function of the packet in question. The following descriptions summarize the function of each field in the BGP header illustrated in Figure 35-3.

- *Marker*—Contains an authentication value that the message receiver can predict.
- *Length*—Indicates the total length of the message in bytes.
- *Type—Type* — Specifies the message type as one of the following:
  - Open
  - Update
  - Notification
  - Keep-alive
- *Data*—Contains upper-layer information in this optional field.

Open Message Format

BGP open messages are comprised of a BGP header and additional fields. Figure 35-4 illustrates the additional fields used in BGP open messages.

Figure 35-4 A BGP open message consists of six fields.



BGP Open Message Fields

BGP packets in which the type field in the header identifies the packet to be a BGP open message packet include the following fields. These fields provide the exchange criteria for two BGP routers to establish a peer relationship.

- *Version*—Provides the BGP version number so that the recipient can determine whether it is running the same version as the sender.
- *Autonomous System*—Provides the autonomous system number of the sender.
- *Hold-Time*—Indicates the maximum number of seconds that can elapse without receipt of a message before the transmitter is assumed to be nonfunctional.
- *BGP Identifier*—Provides the BGP identifier of the sender (an IP address), which is determined at startup and is identical for all local interfaces and all BGP peers.
- *Optional Parameters Length*—Indicates the length of the optional parameters field (if present).
- *Optional Parameters*—Contains a list of optional parameters (if any). Only one optional parameter type is currently defined: authentication information.

  Authentication information consists of the following two fields:

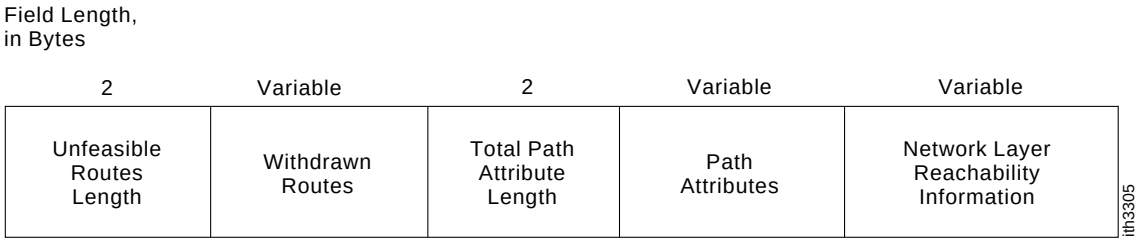
  — Authentication code: Indicates the type of authentication being used.

  — Authentication data: Contains data used by the authentication mechanism (if used).

Update Message Format

BGP update messages are comprised of a BGP header and additional fields. Figure 35-5 illustrates the additional fields used in BGP update messages.

Figure 35-5 A BGP update message contains five fields.



BGP Update Message Fields

BGP packets in which the type field in the header identifies the packet to be a BGP update message packet include the following fields. Upon receiving an update message packet, routers will be able to add or delete specific entries from their routing tables to ensure accuracy. Update messages consist of the following packets:

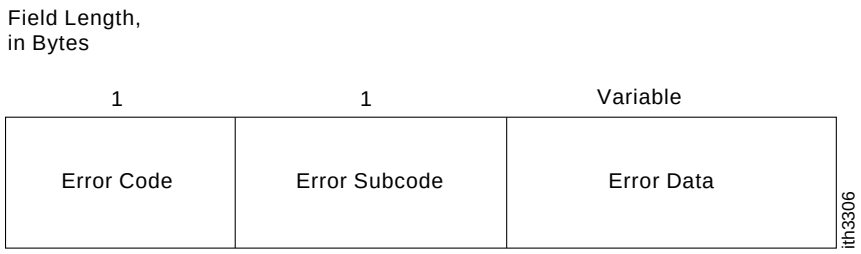
- *Unfeasible Routes Length*—Indicates the total length of the withdrawn routes field or that the field is not present.
- *Withdrawn Routes*—Contains a list of IP address prefixes for routes being withdrawn from service.
- *Total Path Attribute Length*—Indicates the total length of the path attributes field or that the field is not present.

- *Path Attributes*—Describes the characteristics of the advertised path. The following are possible attributes for a path:
  - Origin: Mandatory attribute that defines the origin of the path information
  - AS Path: Mandatory attribute composed of a sequence of autonomous system path segments
  - Next Hop: Mandatory attribute that defines the IP address of the border router that should be used as the next hop to destinations listed in the network layer reachability information field
  - Mult Exit Disc: Optional attribute used to discriminate between multiple exit points to a neighboring autonomous system
  - Local Pref: Discretionary attribute used to specify the degree of preference for an advertised route
  - Atomic Aggregate: Discretionary attribute used to disclose information about route selections
  - Aggregator: Optional attribute that contains information about aggregate routes
- *Network Layer Reachability Information*—Contains a list of IP address prefixes for the advertised routes

Notification Message Format

Figure 35-6 illustrates the additional fields used in BGP notification messages.

Figure 35-6 A BGP notification message consists of three fields.



BGP Notification Message Fields

BGP packets in which the type field in the header identifies the packet to be a BGP notification message packet include the following fields. This packet is used to indicate some sort of error condition to the peers of the originating router.

- *Error Code*—Indicates the type of error that occurred. The following are the error types defined by the field:
  - Message Header Error: Indicates a problem with a message header, such as unacceptable message length, unacceptable marker field value, or unacceptable message type.
  - Open Message Error: Indicates a problem with an open message, such as unsupported version number, unacceptable autonomous system number or IP address, or unsupported authentication code.
  - Update Message Error: Indicates a problem with an update message, such as a malformed attribute list, attribute list error, or invalid next-hop attribute.

  - Hold Time Expired: Indicates that the hold-time has expired, after which time a BGP node will be considered nonfunctional.
  - Finite State Machine Error: Indicates an unexpected event.
  - Cease: Closes a BGP connection at the request of a BGP device in the absence of any fatal errors.
- *Error Subcode*—Provides more specific information about the nature of the reported error.
- *Error Data*—Contains data based on the error code and error subcode fields. This field is used to diagnose the reason for the notification message.