

Cisco's EIGRP Protocol Described from the Web, edited by J. Scott, Feb 2007

Introduction to EIGRP

Traditional routing protocols are inherently prone to loops as they flood routing information throughout the network, hence why techniques such as Split Horizon, Poison Reverse and Hold Down timers are used. Also, traditional routing protocols have to recalculate their algorithms before advertising routes out, and each router has to do this, thereby making convergence slow.

Enhanced Inter-Gateway Routing Protocol (EIGRP) is designed to give all the flexibility of routing protocols such as OSPF but with much faster convergence. In addition, EIGRP has Protocol-Dependent Modules that can deal with AppleTalk and IPX as well as IP. The advantage with this is that only one routing process needs run instead of a routing process for each of the protocols. EIGRP provides loop-free operation and almost instant simultaneous synchronizations of all routers. Redistribution between EIGRP and other routing protocols is generally automatic. For example, if IGRP and EIGRP routers use the same AS number then by default, routes are redistributed one to the other.

Whereas other routing protocols use a variant of the Bellman-Ford algorithm and calculate routes individually, EIGRP uses the Diffusing Update Algorithm (DUAL) (SRI International) where routers share the route calculations (hence 'diffuse'). A router only sends routing

updates as distance vectors of directly connected routes, rather than every route that is in the network. Also, the router only sends an update of a particular if a topology change has occurred to that specific route. In addition, this update is only sent to relevant neighbor routers, not to all routers. This makes EIGRP a bandwidth-efficient routing protocol. Other routing protocols have regular routing updates that contain all route information by default.

EIGRP packet delivery is handled using Reliable Transport Protocol (RTP) which ensures delivery in order using Reliable Multicast on the multicast address 224.0.0.10. EIGRP uses IP protocol number 88.

Unlike IGRP, in the IP environment, EIGRP is a Classless routing protocol since updates carry subnet mask information. Although EIGRP automatically summarizes on the network boundary, it can be configured to summarize on any bit boundary. EIGRP can also be used when aggregating routes i.e. when summarizing major networks.

EIGRP uses the Neighbor Table to list adjacent routers. The Topology Table lists all the learned routes to a destination whilst the Routing Table contains the best route to a destination, which is known as the Successor. The Feasible Successor is a backup route to a destination which is kept in the Topology Table.

MD5 authentication can be used to authorize EIGRP packets.

EIGRP Metric Values – the five K's

Cisco's EIGRP is similar to IGRP only in the sense that it uses the same metrics; Delay, Bandwidth, Reliability and Load. Be aware that the MTU is NOT used in the calculation of the metric, however the MTU is tracked through the path to find the smallest MTU.

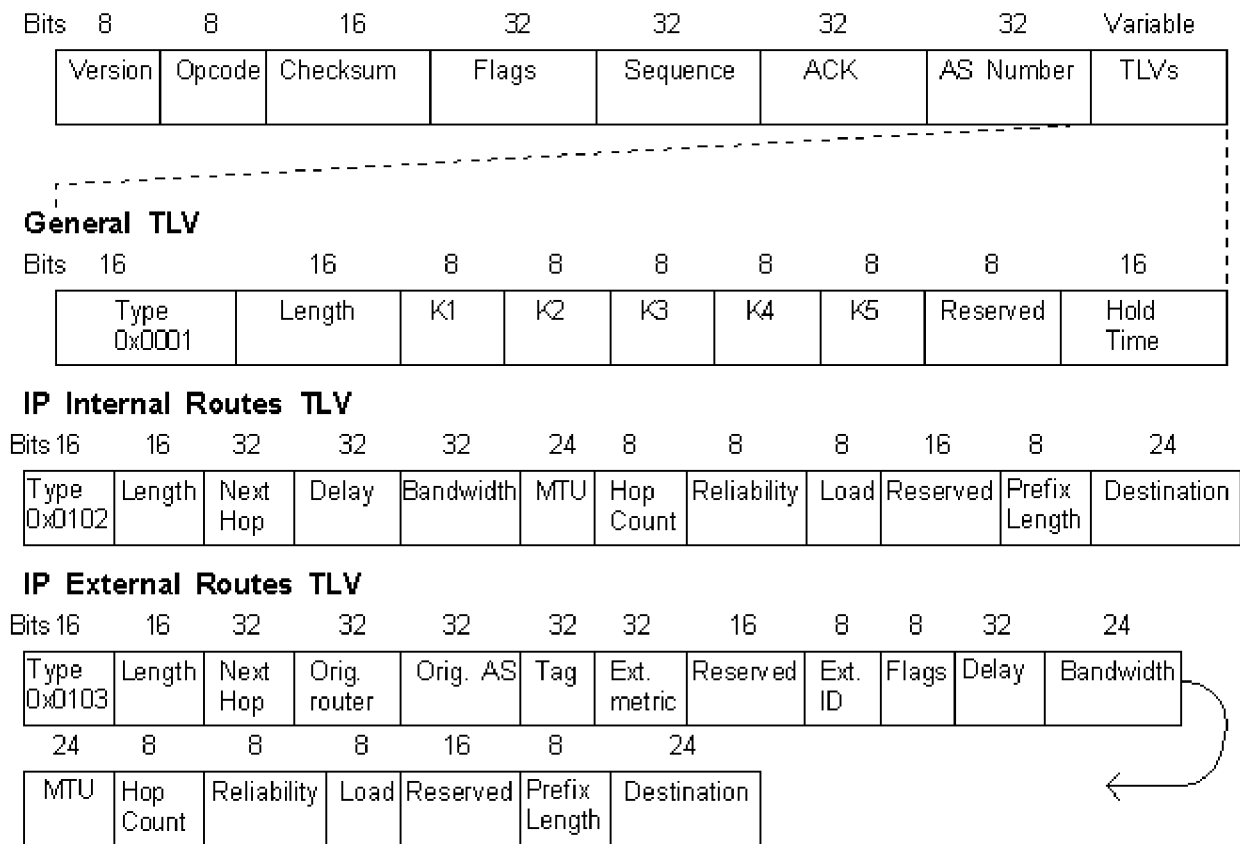
As with IGRP, the 'K' values for the last three are defaulted to '0'. Only the military use 'Reliability'. Most configurations use the first two metrics Delay and Bandwidth, with Bandwidth taking precedence. The metric for EIGRP is calculated by multiplying the IGRP metric by 256. So the formula used to calculate the metric is EIGRP Metric =

$256 * ([K_1 * Bw + K_2 * Bw / (256 - Load) + K_3 * Delay] * [K_5 / (Reliability + K_4)])$.

The default values for weights are: $K_1 - 1$, $K_2 - 0$, $K_3 - 1$, $K_4 - 0$, $K_5 - 0$. This makes the default formula of $256 * (Bw + Delay)$ for the EIGRP metric. The term $[K_5 / (Reliability + K_4)]$ is completely ignored if $K_5 = 0$! You can change the weights if you want to. However, just as with IGRP, these weights must be the same on all the routers in your autonomous system!

Taking the example we used when looking at IGRP, a link where the bandwidth to a particular destination is 128k and the delay is 84000 microseconds. Using the cut down formula EIGRP metric = $256 * (Bw + Delay)$, we obtain the value $256 * (10^7 / 128 + 84000 / 10)$ which gives $256 * 86525 = 22150400$.

EIGRP Packet Format



- Version - there has only been one version
- Opcode - this is the EIGRP packet type:
 - 1 - Update
 - 3 - Query
 - 4 - Reply
 - 5 - Hello
 - 6 - IPX SAP
- Checksum - this is calculated for the whole EIGRP portion of the IP datagram
- Flags - The LSB (0x00000001) is the Init bit meaning that the route in this packet is the first in a new neighbor relationship.

The next bit (0x00000002) is the Conditional Receive bit used in Cisco's Reliable Multicasting algorithm.

- Sequence - the 32-bit sequence number used by RTP.
- ACK - the 32-bit sequence last heard from the neighbor. A Hello packet with a non-zero value is an ACK.
- AS Number - the Autonomous System number of the EIGRP domain.
- Type/Length/Value (TLV) - There are a number of TLVs, all of which begin with a 16 bit Type field and a 16 bit Length field. There then follows a number of fields that vary depending on the type as given below.
 - General TLVs
 - § 0x0001 - General EIGRP parameters (applies to any EIGRP packet regardless of protocol)
 - § 0x0003 - Sequence (used by Cisco's Reliable Multicast)
 - § 0x0004 - EIGRP software version, the original version being 0 and the current version being 1 (used by Cisco's Reliable Multicast)
 - § 0x0005 - Next Multicast Sequence (used by Cisco's Reliable Multicast)
 - IP TLVs
 - § 0x0102 - IP internal routes
 - § 0x0103 - IP external routes
 - §
 - AppleTalk TLVs
 - § 0x0202 - AppleTalk internal routes
 - § 0x0203 - AppleTalk external routes
 - § 0x0204 - AppleTalk cable setup

- IPX TLVs
 - § 0x0302 - IPX internal routes
 - § 0x0303 - IPX external routes

The above diagram illustrates the General TLV (containing the 'K' values) and the IP TLVs (containing details such as the five metrics). Of most interest are the IP TLVs, and these are detailed below:

Type 0x0102 IP internal routes TLV

- Type 0x0102
- Length - Length of the TLV
- Next Hop - The next hop route for this route
- Delay - The number of 10 microsecond chunks which is the sum of delays
- Bandwidth - $256 * \text{IGRP bandwidth}$
- MTU - The smallest MTU encountered along the route to this particular destination network.
- Hop Count - A number between 0x00 (directly connected network) and 0xFF.
- Reliability - A number between 0x01 and 0xFF to indicate the error rates totaled along the route. 0xFF is reliable.
- Load - A number between 0x01 and 0xFF expressing the total load along a route where 0xFF is totally loaded.
- Reserved - 0x0000 and not used.
- Prefix Length - The number of bits used for the mask
- Destination - Destination network

Type 0x0103 IP external routes TLV

- Type 0x0103
- Length - Length of the TLV
- Next Hop - The next hop route for this route
- Originating Autonomous System - The AS from where the route came
- Tag - Used with Route Maps to track routes
- External Protocol Metric - The metric for this route used by the external routing protocol e.g. IGRP, OSPF, RIP
- Reserved - 0x0000 and not used.
- External Protocol ID - identifies the external protocol advertising this particular route
 - 0x01 - IGRP
 - 0x02 - EIGRP (a different AS)
 - 0x03 - Static Route
 - 0x04 - RIP
 - 0x05 - Hello
 - 0x06 - OSPF
 - 0x07 - IS-IS
 - 0x08 - EGP
 - 0x09 - BGP
 - 0x0A - IDRP
 - 0x0B - directly connected
- Flags - 0x01 means the route is an external route whereas 0x02 means that the route could be a default route.
- Delay - The number of 10 microsecond chunks which is the sum of delays
- Bandwidth - $256 * \text{IGRP bandwidth}$

- MTU - The smallest MTU encountered along the route to this particular destination network.
- Hop Count - A number between 0x00 (directly connected network) and 0xFF.
- Reliability - A number between 0x01 and 0xFF to indicate the error rates totaled along the route. 0xFF is reliable.
- Load - A number between 0x01 and 0xFF expressing the total load along a route where 0xFF is totally loaded.
- Reserved - 0x0000 and not used.
- Prefix Length - The number of bits used for the mask
- Destination - Destination network

EIGRP Neighbor Discovery and Adjacencies

Neighbor discovery is achieved via the periodic use of Hello packets. The Hello Interval is every 5 seconds on LANs and fast WANs using multicast Hellos, and every 60 seconds on slow WAN links (except point-to-point sub-interfaces), using Unicast Hellos. The multicast Hello packets are sent to the multicast address 224.0.0.10 since it is not necessary to send unicast packets specifically to each neighbor. These neighbor/peer relationships only occur over primary interface addresses NOT via any secondary addresses that may be configured!

EIGRP uses Reliable Transport Protocol to provide guaranteed, ordered packet delivery to all the neighbors with mixed unicast and multicast packets. On multi-access networks, Hellos are multicast without the requirement for Acknowledgements i.e. Unreliable Multicast. Updates on the other hand, DO require Acknowledgements. The Hellos are the only regular communication.

Once a neighbor has been discovered, the router attempts to form an adjacency with that neighbor whereby routing Updates are sent.

Routing Updates are NOT regularly sent, thereby minimizing bandwidth usage, instead Updates are sent when there are changes to routes, and even then, the Updates are only sent to those routers that need them. If one router requests an Update, the Update is unicast, but if a number of routers require an Update (e.g. because of a topology change), then the Update is multicast.

The Hello packet contains a Hold Time which is 3 times the Hello Interval. This Hold Time is the time that the receiving router should expect to wait before declaring the neighbor as unreachable. For most networks, this Hold Time is $3 \times 5 = 15$ seconds which is much faster than traditional routing protocols Hold time (e.g. 180 seconds for RIP).

A Neighbor Table is built up and contains the following information:

- H is the order in which the neighbors were discovered
- IP Address of neighbor
- Interface on which the Hello was received
- Hold Time in seconds
- Up Time i.e. how long the neighbor has been up
- Smooth Round Trip Time (SRTT) - the average time in milliseconds between the transmission of a packet to a neighbor and the receipt of an acknowledgement.
- Retransmission Timeout (RTO) - if a multicast has failed, then a unicast is sent to that particular router, the RTO is the

time in milliseconds that the router waits for an acknowledgement of that unicast.

- Queue - shows the number of queued packets.
- Sequence Number of the last EIGRP packet received.

The SRTT generally indicates the speed of the link(s) along the path to that particular neighbor. The RTO defaults to 200ms and increases if a neighbor fails to respond to a query. You can test this by clearing one neighbor and seeing the increase in the RTO on the other neighbor(s). Over time as and when updates are sent, the RTO starts to come down, this only happens if changes occur in the network since EIGRP only sends updates when changes occur.

EIGRP uses Split Horizon and Poison Reverse to ensure that routes learned on a particular interface are not re-advertised out of that same interface, or if they are, that they are advertised as unreachable. If a router has an interface with a secondary address configured say on a LAN, then other routers on that LAN will not learn of that subnet from that router because of Split Horizon being enabled (by default).

EIGRP Topology Table and the DUAL Algorithm

Once a neighbor relationship has been formed, called an Adjacency, the routers exchange routing update information and each router builds its own topology table. The Updates contain all the routes known by the sender. For each route, the receiving router calculates a distance for that route based on the distance that is conveyed and the cost to that neighbor that advertised the particular route. If the receiving router sees several routes to a particular network with

different metrics, then the route with the lowest metric becomes the Feasible Distance (FD) to that network. The Feasible Distance is the metric of a network advertised by the connected neighbor plus the cost of reaching that neighbor. This path with the best metric is entered into the routing table because this is the quickest way to get to that network.

With the other possible routes to a particular network with larger metrics, the receiving router also receives the Reported Distance (RD) to this network via other routers. The Reported Distance being the total metric along a path to a destination network as advertised by an upstream neighbor. The Reported Distance for a particular route is compared with the Feasible Distance that it already has for that route. If the Reported Distance is larger than the Feasible Distance then this route is not entered into the Topology Table as a Feasible Successor. This prevents loops from occurring. If the Reported Distance is smaller than the Feasible Distance, then this path is considered to be a Feasible Successor and is entered into the Topology table. The Successor for a particular route is the neighbor/peer with the lowest metric/distance to that network.

If the receiving router has a Feasible Distance to a particular network and it receives an update from a neighbor with a lower advertised distance (Reported Distance) to that network, then there is a Feasibility Condition. In this instance, the neighbor becomes a Feasible Successor for that route because it is one hop closer to the destination network. There may be a number of Feasible Successors in a meshed network environment, up to 6 of them are entered into the Topology table thereby giving a number of next hop

choices for the local router should the neighbor with the lowest metric fail. What you should note here, is that the metric for a neighbor to reach a particular network (i.e. the Reported Distance) must always be less than the metric (Feasible Distance) for the local router to reach that same network. This way routing loops are avoided. This is why routes that have Reported Distances larger than the Feasible Distance are not entered into the Topology table, so that they can never be considered as successors, since the route is likely to loop back through that local router.

DUAL therefore uses distance information to select the optimum routes that do not create loops. There could be a number of routers that can lead to a particular destination network with the potential for loops. DUAL uses this concept of Feasible Successor, which is a router that has a least cost path to a network and therefore does not form part of a loop since the router will not choose a path that runs back through itself again.

The Topology Table consists of the advertised metric to reach a network by a neighbor and the Feasible Distance to that destination network, via that particular EIGRP neighbor. A network could have a number of entries. Each entry will have the following information:

- The Feasible Distance
- Feasible Successors
- Each Feasible Successor's distance to the network
- The locally calculated metric to the network via each Feasible Successor.
- The interface on which each Feasible Successor is discovered.

For example the composite metric 327168/326912 would mean that the locally calculated metric is 327168 and the advertised Feasible Successor's distance (RD) to the network is 326912.

For each network listed in the Topology table the one with the lowest metric is added to the Route table and the neighbor that advertises that route becomes the Successor.

Maintaining a Topology Table allows a router to make sure that all its own metrics to destination networks are larger than its neighbors, thereby avoiding routing loops. EIGRP therefore does not need Hold Down or Flush timers since loops are avoided anyway.

If a route becomes unreachable e.g. the link to the Successor fails, then the router looks in its Topology Table for another route with a lower metric than its Feasible Distance i.e. a Feasible Successor, and that one becomes the Successor. This requires no neighbor querying and is therefore very fast.

If a neighbor fails, after three failed hello messages, the router sends an update. If the backup route fails, only then does it query its neighbors for an alternate route. When route information changes, the router sends an update about that link only, and only to the routers in the same AS that need the update. This is in contrast to OSPF where the whole link state database needs to be synchronized across the whole area.

In the routing table, because EIGRP relies on the Topology table for updating its routes, the routing entries can become very old. The

Topology table contains the known routes and the successors for each route with each interface indicated on which the successors are connected.

By default, if there are multiple equal-cost paths to a destination the router will load share across up to four paths. Generally with most routing protocols, you can change this in the routing process with the command `maximum-paths number` and have up to 6 paths. By default, on interfaces where fast switching is enabled, the router will perform per-destination load balancing. If fast-switching is turned off then all packets will be examined by the CPU and be load-balanced on a per-packet basis. The load on the CPU can be extensive. Using Cisco Express Forwarding (CEF), you can choose to load balance on a per-packet or per-destination basis with less impact on the CPU.

You can also load share over unequal cost paths. To do this we use the variance feature in the EIGRP routing process. The variance is defined with a multiplier that represents the difference between the metrics of the paths. The default variance is '1' which means that the multiple paths must have the same metrics.

EIGRP's Use of the DUAL Finite State Machine and Diffusing Computations

The principles of DUAL are:

- Neighbor loss or detection occurs within a finite time.
- Messages are correctly received and in order, within a finite time.
- Messages are processed in the order in which they are received, within a finite time.

In a steady state situation where the Successors for each network are known and the Feasible Distances are the lowest, then each network listed in the Topology Table will be in the Passive state meaning that no diffusing calculations are being performed.

The list of Feasible Successors for a particular route will be reassessed locally if there is a change to the cost of the link, a change of state or if update, query or reply packets are received. It could be that the Feasible Distance changes, or that the Feasible Successor takes over from the existing Successor. Provided that a Feasible Successor is found, this is advertised via Updates whilst all the while remaining in Passive state. The idea with this is that if a topology change occurs, the router should be able to find an alternate route without having to recompute the route.

If no neighbor exists with a metric for a particular network that is less than the Feasible Distance, i.e. no Feasible Successor exists, then the local router goes into Active state and queries its neighbors for routing information. If no Feasible Successor is available for a route, then a Diffusing Computation must be implemented, thereby slowing down re-convergence. The local router sets a Reply Status flag to track all the queries to its neighbors.

When performing the Diffusing Computation, queries are sent to all the neighbors and these contain the new locally calculated distance for the network. If a neighbor has feasible successors, it will recalculate its own local distance to the network and send this back. If a neighbor does not have a Feasible Successor, then it will itself move into Active state.

The originating router does not consider the Diffusing Computation to be complete until replies have been received from all the neighbors. There is an Active Timer that has a default value of 3 minutes. This timer is used to time how long it takes to perform the Diffusing Computation. On a large network where a chain of routers may end up performing the Diffusing Computation, it may be a while before the originating router completes. If all the replies are not received within the 3 minutes, then the route is said to be Stuck-in-Active (SIA). The neighbor involved is removed from the neighbor table and the metric for that route set to infinity so that another neighbor can meet the Feasibility Condition and become a Feasible Successor.

If an EIGRP network is particularly large, or there are a number of low bandwidth links such that it takes a while for replies to get back, then those neighbors that have yet to reply have their Reply Status flag set. If no reply is received from a particular neighbor before the Active timer times out, then the neighbor will be removed from the neighbor table. If a reply DOES come back after the Active timer has timed out then the neighbor gets reinstated. The disappearance and reappearance of neighbors causes extra Diffusing computations, and hence, changes are made to the routing table. Examining the Topology tables of the routers as you chase the SIA neighbor entries helps to track the issues causing the SIA.

Using DUAL, routers maintain up to six backup routes in case the main one fails, and this is carried out by storing neighbor's routing tables. Using the DUAL Finite State Machine results in very fast convergence as it keeps track of all routes advertised by all neighbors.