

Load balancing with Cisco Express Forwarding

Introduction

The use of parallel T1 or E1 links provides higher aggregate bandwidth connections between routers where an upgrade to a high bandwidth DS3 or E3 link is not feasible due to cost, usage patterns or availability. It is also an approach that scales well as links can be added as the bandwidth requirements grow. In IOS software release 11.1(17)CC a new forwarding mechanism called Cisco Express Forwarding (CEF) has been introduced for Cisco 7200 and 7500 series routers. CEF can efficiently use multiple parallel links without additional hardware multiplexers. The purpose of this document is to describe how CEF uses parallel links and how to configure it. For a more general description of techniques to increase bandwidth with parallel links please see the White Paper "Alternatives for high bandwidth connections using parallel T1/E1 links."

Load balancing

The term Load balancing describes a functionality in a router that distributes packets across multiple links based on layer 3 routing information. If a router discovers multiple paths to a destination the routing table is updated with multiple entries for that destination:

```
router> show ip route
[... ]
I    192.168.25.0/24 [115/10] via 192.168.24.6
                               [115/10] via 192.168.24.10
                               [115/10] via 192.168.24.14
[... ]
```

Usually the paths have the same metric, as is the case with parallel links of the same type, however there are routing protocols that allow unequal cost load balancing. For the following sections only parallel links of the same metric are discussed, although the same principles apply to load balancing over unequal cost paths.

The load balancing function itself is inherent to the forwarding mechanism of a router. The basic definition of a layer 3 forwarding device does not require that load balancing needs to be supported, a router may forward all packets to a destination over one path, even if additional paths with equal cost exist. However a good implementation of a routing device will try to make the most efficient use of the available bandwidth and distribute traffic across multiple paths. This is a per-box implementation benefit, and does not require any special configurations in other devices in the network or the use of load balancing-specific communication mechanisms between the routers. A router learns about the existence of parallel paths through the standard routing protocols and should build its routing table accordingly, as pictured in the above example. It then may use this information to distribute the load across the paths. The number of paths used is limited by the number of entries the routing protocol puts in the routing table, the default in IOS is 4 entries for most IP routing protocols with the exception of BGP, where it is one entry. The maximum number that can be configured is 6 different paths.

As indicated above the algorithm to distribute the traffic is depending on the individual implementation, and as such the efficiency can vary. Cisco IOS software basically supports two modes of load balancing: On per-destination or per-packet basis.

In per-destination mode all packets for a given destination are forwarded along the same path. This preserves packet order, with potential unequal usage of the links. If one host receives the majority of the traffic all packets will use one link, leaving bandwidth on other links unused. A larger number of destination addresses lead to more equally used links. In IOS software this is achieved by building a route-cache entry for every destination address, instead of every destination network as done when only a single path exist. Therefor traffic for different hosts on the same destination network can use different paths. The downside of this approach is that for core backbone routers carrying traffic for 10000s of destination hosts memory and processing requirements for maintaining the cache become very demanding.

Per-packet load balancing guarantees equal load across all links, however potentially the packets may arrive out-of-order at the destination as differential delay may exist within the network. In Cisco IOS software, except the release 11.1CC, per packet load balancing does disable the forwarding acceleration by a route cache, as the route cache information includes the outgoing interface. For per-packet load balancing the forwarding process determines the outgoing interface for each packet by looking up the route table and picking the least used interface. This ensures equal utilization of the links, but is a processor intensive task and impacts the overall forwarding performance. This form of per-packet load balancing is not well suited for higher speed interfaces.

Load balancing with CEF for IP

In IOS software 11.1CC a new forwarding mechanism for IP packets was introduced: Cisco Express Forwarding (CEF). The design of Cisco Express Forwarding includes enhancements that allow to use load balancing without sacrificing forwarding performance even when using per packet load balancing. Previously per packet load balancing required disabling of route-caching mechanisms like fast switching or optimum switching (see above paragraph). CEF is available in IOS software version 11.1CC for Cisco 7200 and 7500 series routers only. CEF currently supports the following encapsulations: ATM/AAL5snap, ATM/AAL5mux, ATM/AAL5nlpid, Frame Relay, Ethernet, FDDI, PPP, HDLC, and tunnels.

How CEF load balancing works

CEF is an advanced Layer 3 switching technology inside a router. Usually a router uses a route cache to speed up packet forwarding. The route cache is filled on demand when the first packet for a specific destination needs to be forwarded. If the destination is on a remote network reachable via a next hop router, the entry in the route cache is consisting of the destination network. If parallel paths exist this does not provide load balancing, as only one path would be used. Therefor the entry in the route cache now relates to a specific destination address, or host. If multiple hosts on the destination network are receiving traffic a route cache entry for each individual host is made, balancing the hosts over the available paths. This provides per destination load balancing. The problem that arises is that for a backbone router carrying traffic for several thousands of destination hosts a respective number of cache entries is needed. This consumes memory and makes cache maintenance a demanding task. In addition the decision about which path to use is done at the time the route-cache is filled, and it is based on the utilization of the individual links at that point in time. However the amount of traffic on individual connections can change over time, possibly leading to a situation where some links carry mostly idle connections and others are congested. CEF takes a different approach as it calculates all information necessary for the forwarding task in advance and decouples the forwarding information from the next hop adjacency, which allows for effective load balancing.

The two main components of CEF operation are the

- Forwarding Information Base
- Adjacency Tables

Forwarding Information Base

CEF uses a Forwarding Information Base (FIB) to make IP destination prefix-based switching decisions. The FIB is conceptually similar to a routing table or information base. It maintains a mirror image of the forwarding information contained in the IP routing table. When routing or topology changes occur in the network, the IP routing table is updated, and those changes are reflected in

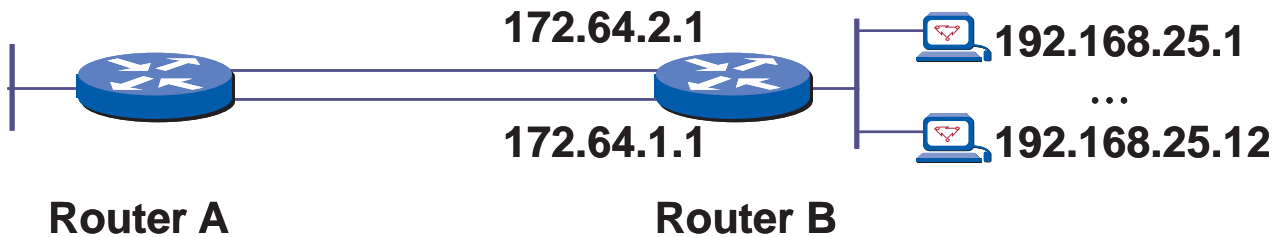
the FIB. The FIB maintains next-hop address information based on the information in the IP routing table. Because there is a one-to-one correlation between FIB entries and routing table entries, the FIB contains all known routes and eliminates the need for route cache maintenance that is associated with earlier switching paths such as fast switching and optimum switching.

Adjacency Tables

Network nodes in the network are said to be adjacent if they can reach each other with a single hop across a link layer. In addition to the FIB, CEF uses adjacency tables to prepend Layer 2 addressing information. The adjacency table maintains Layer 2 next-hop addresses for all FIB entries.

The adjacency table is populated as adjacencies are discovered. Each time an adjacency entry is created (such as through the ARP protocol), a link-layer header for that adjacent node is precomputed and stored in the adjacency table. Once a route is determined, it points to a next hop and corresponding adjacency entry. It is subsequently used for encapsulation during CEF switching of packets. A route might have several paths to a destination prefix, such as when a router is configured for simultaneous load balancing and redundancy. For each resolved path a pointer is added for the adjacency corresponding to the next-hop interface for that path. This mechanism is used for load balancing across several paths. For per destination load balancing a hash is computed out of the source and destination IP address. This hash points to exactly one of the adjacency entries in the adjacency table, providing that the same path is used for all packets with this source/destination address pair. If per packet load balancing is used the packets are distributed round robin over the available paths. In either case the information in the FIB and adjacency tables provide all the necessary forwarding information, just like for non-load balancing operation. The additional task for load balancing is to select one of the multiple adjacency entries for each forwarded packet.

Here is an example network with parallel serial links configured for PPP, there are 12 hosts on the subnet 192.168.25.0 connected to router B. Clients on router A are accessing the hosts via the serial links:



Router A's route table looks like this

Destination	Next hop	Interface
192.168.25.0/24	via 172.64.2.1	serial 0
	via 172.64.1.1	serial 1

If route-cache mechanisms are disabled, the route table is the only information used to forward a packet. For every packet the destination address is matched against the route table, and the packet is forwarded through the interface with the lower utilization.

With traditional route-cache based forwarding mechanisms like fast switching or optimum switching load balancing is supported on per-destination basis, and the route cache contains entries for each of the 12 destination hosts:

Destination	Interface	Encapsulation
192.168.25.1/32	serial 0	PPP
192.168.25.2/32	serial 1	PPP
...		
192.168.25.12/32	serial 0	PPP

For each packet a lookup of the destination address is done in the route cache to find a matching entry. Packets destined for new destination hosts on a known network create additional route cache entries. The assignment of a path to a destination is made on the basis of the current link utilization, the cache entry is generated using the interface with the lowest utilization.

If CEF is used as forwarding mechanism there are two tables, the Forwarding Information Base and the Adjacency Table, these replace the traditional route cache. The FIB entry has one or more pointers to the adjacency information (34,45 are arbitrary chosen numbers to represent the pointer relation):

FIB Adjacency Table

Destination	Adjacencies	#	Interface	Encapsulation
192.168.25.0/24	34, 45	34	Serial0	PPP
		45	Serial1	PPP

Note that it is common that multiple entries in the FIB are pointing towards the same adjacencies, e.g. if a network 192.168.30.0 is reachable via Router B the FIB for 192.168.30.0 is pointing to the same adjacencies 34 and 45.

For each packet a match of the destination address against the FIB is made. In this example the FIB entry has multiple adjacencies for the destination 192.168.25.0, one of the adjacencies needs to be selected. For per-packet load balancing packets are encapsulated and forwarded by alternating between entry #34 and entry #45. For load balancing on IP address basis a hashcode is generated out of the addresses. In this example the hash may have the values 1 or 2, selecting the first or second entry of the adjacencies listed for the destination network 192.168.25.0. For any given pair of source and destination addresses the hashcode is always the same, thus guaranteeing that always the same path is used.

CEF Configuration

CEF configuration is pretty straightforward, as the only configurable option is the load balancing mode:

The global command to enable CEF is

```
ip cef [distributed] switch
```

with the keyword `distributed` enabling `distributed` forwarding on VIP2-20, VIP2-40 and VIP2-50 interface processors in a Cisco 7500 series router. This command automatically enables CEF on all interfaces that use supported encapsulations. Interfaces that do not support CEF should be explicitly configured with the `ip route cache` command. The default load balancing mode of CEF is `per-destination`.

CEF operation can be verified with the command `show ip cef`:

```
router> show ip cef 192.168.25.0
192.168.25.0/24, version 22, per-destination sharing
2635 packets, 408375 bytes
  via 192.168.24.6, 0 dependencies, recursive
    traffic share 1, current path
    next hop 192.168.24.6, Serial1/0 via 192.168.24.4/30
    valid glean adjacency
  via 192.168.24.10, 0 dependencies, recursive
    traffic share 1
    next hop 192.168.24.10, Serial1/1 via 192.168.24.8/30
    valid glean adjacency
  via 192.168.24.14, 0 dependencies, recursive
    traffic share 1
    next hop 192.168.24.14, Serial1/2 via 192.168.24.12/30
    valid glean adjacency
2635 packets, 408375 bytes switched through the prefix
```

This example shows three entries for the destination 192.168.25.0, each being a serial link. In the first line the load balancing mode is shown: “per-destination sharing”

To change the load balancing from per-destination to per-packet load balancing use the interface command:

```
ip load-sharing per-packet
```

on *all* interfaces that forward traffic to the destination. Again this can be verified with the show ip cef command:

```
router>sho ip cef 192.168.25.0
192.168.25.0/24, version 22, per-packet sharing
2635 packets, 408375 bytes
[...]
```

The simplest configuration only requires to enable CEF forwarding. This is sufficient for most cases, as the per-packet load balancing is only rarely needed with the new per-destination algorithm. There are a few scenarios where per-packet load balancing is more advisable, e.g. the majority of traffic is between two hosts.

Summary

Cisco Express Forwarding supports load balancing for TCP/IP over parallel links without impacting performance even if the traffic patterns require per-packet load balancing. This is a major enhancement compared to the route-cache based forwarding mechanisms used in other IOS versions. CEF is currently available for Cisco 7200 and 7500 series routers with IOS release 11.1CC

Additional Information

Release Note for Cisco IOS Release 11.1 CC and Feature Modules

White Paper "Alternatives for high bandwidth connections using parallel T1/E1 links"



Corporate Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
<http://www.cisco.com>
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 526-4100

European Headquarters
Cisco Systems Europe s.a.r.l.
Parc Evolic, Batiment L1/L2
16 Avenue du Quebec
Villebon, BP 706
91961 Courtaboeuf Cedex
France
<http://www-europe.cisco.com>
Tel: 33 1 6918 61 00
Fax: 33 1 6928 83 26

Americas
Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
<http://www.cisco.com>
Tel: 408 526-7660
Fax: 408 527-0883

Asia Headquarters
Nihon Cisco Systems K.K.
Fuji Building, 9th Floor
3-2-3 Marunouchi
Chiyoda-ku, Tokyo 100
Japan
<http://www.cisco.com>
Tel: 81 3 5219 6250
Fax: 81 3 5219 6001

Cisco Systems has more than 200 offices in the following countries. Addresses, phone numbers, and fax numbers are listed on the

Cisco Connection Online Web site at <http://www.cisco.com>.

Argentina • Australia • Austria • Belgium • Brazil • Canada • Chile • China (PRC) • Colombia • Costa Rica • Czech Republic • Denmark
England • France • Germany • Greece • Hungary • India • Indonesia • Ireland • Israel • Italy • Japan • Korea • Luxembourg • Malaysia
Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland • Portugal • Russia • Saudi Arabia • Scotland •
Singapore

Copyright © 1998 Cisco Systems, Inc. All rights reserved. Printed in USA. AccessPath, AtmDirector, the CCIE logo, CD-PAC, Centri, Centri Bronze, Centri Gold, Centri Security Manager, Centri Silver, the Cisco Capital logo, Cisco IOS, the Cisco IOS logo, CiscoLink, the Cisco NetWorks logo, the Cisco Powered Network logo, the Cisco Press logo, ClickStart, ControlStream, Fast Step, FragmentFree, IGX, JumpStart, Kernel Proxy, LAN²LAN Enterprise, LAN²LAN Remote Office, MICA, Natural Network Viewer, NetBeyond, Netsys Technologies, Packet, PIX, Point and Click Internetworking, Policy Builder, RouteStream, Secure Script, SMARTnet, StrataSphere, StrataSphere BILLder, StrataSphere Connection Manager, StrataSphere Modeler, StrataSphere Optimizer, Stratum, StreamView, SwitchProbe, The Cell, TrafficDirector, VirtualStream, VlanDirector, Workgroup Director, Workgroup Stack, and XCI are trademarks; Empowering the Internet Generation and The Network Works. No Excuses. are service marks; and BPX, Catalyst, Cisco, Cisco Systems, the Cisco Systems logo, EtherChannel, FastHub, FastPacket, ForeSight, IPX, LightStream, OptiClass, Phase/IP, StrataCom, and StrataView Plus are registered trademarks of Cisco Systems, Inc. in the U.S. and certain other countries. All other trademarks mentioned in this document are the property of their respective owners. 9802R