Cisco 12000 Series Internet Router Architecture: Switch Fabric

Document ID: 47240

Contents

Introduction

This document examines some of the hardware components of the Cisco 12000 Series Internet Router, namely the Backplane, the Switch Fabric, the Clock and Scheduler Card (CSC), the Switch Fabric Card (SFC), and Cisco Cells.

Prerequisites

Requirements

There are no specific requirements for this document.

Components Used

The information in this document is based on the Cisco 12000 Series Internet Router.

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, make sure that you understand the potential impact of any command.

Conventions

Refer to Cisco Technical Tips Conventions for more information on document conventions.

Backplane

Before looking at the Cisco 12000 switch fabric, let's look at the backplane.

Gigabit Route Processors (GRPs) and Line Cards (LCs) are installed from the front of the chassis and plug into a passive backplane. This backplane contains serial lines that interconnect all of the line cards to the switch fabric cards, as well as other connections for power and maintenance functions. On 120xx models, each 2.5 Gbps chassis slot has up to four 1.25 Gbps serial line connections, one to each of the switch fabric cards to provide a total capacity of 5 Gbps per slot or 2.5 Gbps full duplex. On 124xx models, each 10 Gbps chassis slot uses four sets of four serial line connections, providing each slot with a switching capacity of 20 Gbps full duplex.

All models of line cards also have a fifth serial line that can connect to a redundant Clock and Scheduler Card (CSC).

Switch Fabric

At the heart of the Cisco 12000 Series Internet Router is a multi−gigabit crossbar switch fabric that is optimized to provide high capacity switching at gigabit rates. The crossbar switch enables high performance for two reasons:

- Connections from the line cards to a centralized fabric are point−to−point links that can operate at very high speeds
- Multiple bus transactions can be supported simultaneously, increasing the aggregate bandwidth of the system. The Switch Fabric Card (SFC) receives the scheduling information and clocking reference from the Clock Scheduler Card (CSC), and performs the switching functions. You can imagine the SFC as an NxN matrix where N is the number of slots.

This architecture allows multiple line cards to transmit and receive data simultaneously. The CSC is responsible for selecting which line cards transmit and which line cards receive data during any given fabric cycle.

The switch fabric provides a physical path for the following traffic:

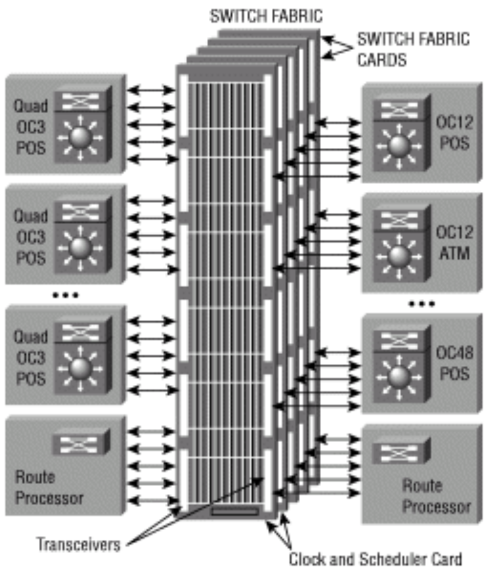
- Initial fabric downloader from the Route Processor (RP) to the line cards on power up
- Cisco Express Forwarding updates
- Statistics from the line cards
- Traffic switching

These functions are described in more detail below.

The switch fabric is an NxN non−blocking crossbar switch fabric where N stands for the maximum number of LCs that can be supported in the chassis (this includes the GRP). This allows each slot to simultaneously send and receive traffic over the fabric. In order to have a non−blocking architecture to allow multiple line cards to send to other line cards simultaneously, each LC has an N+1 virtual output queuing (VOQ) (one for each possible line card destination and one for multicast).

When a packet comes in an interface, a lookup is performed (this may be in the hardware or the software, depending on the LC and which features are configured). The lookup determines the output LC, interface, and appropriate Media Access Control (MAC) layer re−write information. Before the packet is sent to the output LC through the fabric, the packet is chopped into Cisco Cells. A request is then made to the clock scheduler for permission to transmit a Cisco Cell to the given output LC. One cell is transmitted every fabric clock cycle by E0 LCs and every four fabric clock cycles by E1 and higher LCs. The output LC then re−assembles these Cisco Cells into a packet, uses the MAC rewrite information sent with the packet to perform the MAC layer rewrite, and queues the packet for transmission on the appropriate interface.

Remember that even if a packet arrives on an interface on an LC and is supposed to go out another interface (or on the same interface in case of sub−interfaces) on the same LC, it is still segmented into Cisco Cells and sent over the fabric back to itself.



Clock and Scheduler Card (CSC)

The CSC accepts tranmission requests from line cards, issues grants to access the fabric, and provides a reference clock to all the cards in the system to synchronize data transfer across the crossbar. Only one CSC is active at any time.

The CSC can be removed and replaced, without disrupting normal system operations, only if a second (redundant) CSC is installed in the system. One CSC must be present and operational at all times to maintain normal system operations. A second CSC provides data path, scheduler, and reference clock redundancy. The interfaces between the line cards and the switch fabric are monitored constantly. If the system detects a Loss of Synchronization (LoS), it automatically activates the data paths of the redundant CSC, and data flows across the redundant path. The switch to the redundant CSC usually occurs in the order of seconds (the actual switch time depends on your configuration and its scale), during which time there can be a loss of data on some/all LCs.

Switch Fabric Card (SFC)

On the Cisco 12008, 12012, and 12016, an optional set of three SFCs can be installed in the router at any time to provide additional switch fabric capacity to the router. This configuration is called full bandwidth. The SFC cards increase the data handling capacity of the router. Any one or all of the SFCs can be removed and replaced at any time without system operations being disrupted or the router being powered down. For the length of time that any SFC is not functional, its data carrying capacity is lost to the router as a potential data path for the router's data handling and switching functions.

Redundancy and Bandwidth

The switch fabric card (SFC) and clock scheduler card (CSC) provide the physical switch fabric for the system as well as the clocking for the Cisco cells that carry data and control packets among the line cards and route processors.

On the 12008, 12012, and 12016, you must have at least one CSC card for the router to run. Having only one CSC card and no SFC cards is called quarter bandwidth, and only works with Engine 0 line cards. If other line cards are in the system, they will automatically be shut down. If you require line cards other than Engine 0, full bandwidth (three SFCs and one CSC) must be installed in the router. If redundancy is required, a second CSC is necessary. This redundant CSC only functions if either the CSC or an SFC goes bad. The redundant CSC can function as either the CSC or SFC.

The 12416, 12406, 12410, and 12404 require full bandwidth.

Other important details about switch fabric redundancy and bandwidth are:

- All 12000 Series routers have a maximum of three SFCs and two CSCs, except for the 12410 Series which has five dedicated SFCs and two dedicated CSCs, and the 12404 which has one board that contains all the CSC/SFC functionality. For the 12404, there is no redundancy.
- In the 12008, 12012, 12016, 12406, and 12416, the CSC cards also function as switch fabric cards. That is why, to get a full bandwidth redundant configuration, you only need three SFCs and two CSCs. In the 12410, there are dedicated clock and scheduler cards and switch fabric cards. To get a full bandwidth redundant configuration, you need two CSCs and five SFCs.
- Quarter bandwidth configurations can only be used on the 12008, 12012, and the 12016 if you have nothing but Engine 0 LCs in the chassis. The CSC192 and SFC192, which reside in the 12400 series chassis, do not support quarter bandwidth configurations.

Below are some interesting switch fabric−related links for all the platforms:

Cisco 12008 Internet Router

The CSCs are installed in the upper card cage and the SFCs are installed in the lower card cage which is located directly behind the air filter assembly (see Figure 1−22: Components in the Lower Card Cage under Product Overview Documentation).

More details are available in the documentation below:

- Cisco 12008 Gigabit Switch Router Switch Card Replacement Instructions
- Switch Fabric of the Cisco 12008

Cisco 12012 Internet Router

Both the CSCs and SFCs are installed in the five−slot lower card cage. See Front View and Lower Card Cage.

More details can be found in the documentation below:

- Cisco 12012 Gigabit Switch Router Switch Fabric Cards Replacement Instructions
- Switch Fabric of the Cisco 12012

Cisco 12016/12416 Internet Routers

There are currently two switch fabric options available for the Cisco 12016:

- 2.5 Gbps switch fabric (80 Gbps switching system bandwidth) − This consists of the GSR16/80−CSC and the GSR16/80−SFC fabric set. Each SFC or CSC card provides a 2.5 Gbps full−duplex connection to each line card in the system. For a Cisco 12016 with 16 line cards, each with 2 x 2.5 Gbps capacity (full duplex), the system switching bandwidth is 16 x 5 Gbps = 80 Gbps. (The older switch fabric is sometimes referred to as the 80−Gbps switch fabric).
- 10 Gbps switch fabric (320 Gbps switching system bandwidth) − This consists of the GSR16/320−CSC and the GSR16/320−SFC fabric set. Each SFC or CSC card provides a 10 Gbps full−duplex connection to each line card in the system. For a Cisco 12016 with 16 line cards, each

with 2 x 10 Gbps capacity (full duplex), the system switching bandwidth is 16 x 20 Gbps = 320 Gbps. (The newer switch fabric is sometimes referred to as the 320 Gbps switch fabric).

When the Cisco 12016 router contains the 320 Gbps switching fabric, it is referred to as a Cisco 12416 Internet Router.

CSCs and SFCs are installed in the five−slot switch fabric card cage.

See the documents below for more details:

- Cisco 12016 Gigabit Switch Router Clock and Scheduler and Switch Fabric Card Replacement Instructions
- Multi−gigabit Crossbar Switch Fabric

Cisco 12404 Internet Router

The Cisco 12404 has one board called the Consolidated Switch Fabric (CSF) that provides synchronized speed interconnections for the line cards and the RP. The CSF circuitry is contained on one card and consists of a clock scheduler and switch fabric functionality. The CSF card is housed in the bottom slot labeled FABRIC ALARM in the Cisco 12404 Internet Router chassis.

For more details, see:

- Cisco 12404 Internet Router Consolidated Switch Fabric Replacement Instructions
- Clock and Scheduler, and Switch Fabric Cards

Cisco 12410 Internet Router

The switch fabric for the Cisco 12410 consists of two clock and scheduler cards (CSCs) and five switch fabric cards (SFCs) installed in the switch fabric and alarm card cage. One CSC and four SFCs are required for an active switch fabric; the second CSC and the fifth SFC provide redundancy. The two alarm cards that are also located in the switch fabric and alarm card cage are not part of the switch fabric.

Unlike other systems in the Cisco 12000 series, the Cisco 12410 supports only the latest 10 Gbps switch fabric. Each SFC or CSC card provides a 10 Gbps full−duplex connection to each line card in the system. Thus, for a Cisco 12410 with 10 line cards, each with 2 x 10 Gbps capacity (full duplex), the system switching bandwidth is 10 x 20 Gbps = 200 Gbps.

See the documents below for more details:

- Cisco 12410 Gigabit Switch Router Clock Scheduler and Switch Fabric Cards Replacement Instructions
- Switch Fabric and Alarm Card Cage

Cisco 12416 Internet Router

See the Cisco 12016 Internet Router.

Troubleshooting Tips for the Switch Fabric Cards

The switch fabric cards in the 12016 and 12416 are not easy to insert, and may require a little bit of force. If either of the CSCs are not seated properly, you may see this error message:

```
%MBUS−0−NOCSC: Must have at least 1 CSC card in slot 16 or 17
```

```
%MBUS−0−FABINIT: Failed to initialize switch fabric infrastructure
```

You may also get this error message if there are only enough CSCs and SFCs seated for quarter bandwidth configurations. In this case, none of the E1 or higher LCs will boot.

One sure way to tell if the cards are seated properly is that, on the CSC/SFC, you should see four lights "on" . If this is not the case, then the card is not seated correctly.

When dealing with problems related to the fabric and LCs not booting, it is important to verify that all necessary CSCs and SFCs are correctly seated and powered on. For instance, three SFCs and two CSCs are required on a 12016 to get a full bandwidth redundant system. Three SFCs and only one CSC are needed to get a full bandwidth non−redundant system.

The output from the **show version** and **show controller fia** commands tells you which hardware configuration is currently running in the box.

```
Thunder#show version
Cisco Internetwork Operating System Software
IOS (tm) GS Software (GSR−P−M), Experimental Version 12.0(20010505:112551)
[tmcclure−15S2plus−FT 118]
Copyright (c) 1986−2001 by cisco Systems, Inc.
Compiled Mon 14−May−01 19:25 by tmcclure
Image text−base: 0x60010950, data−base: 0x61BE6000

ROM: System Bootstrap, Version 11.2(17)GS2, [htseng 180] EARLY DEPLOYMENT
RELEASE SOFTWARE (fc1)
BOOTFLASH: GS Software (GSR−BOOT−M), Version 12.0(15.6)S, EARLY DEPLOYMENT
MAINTENANCE INTERIM SOFTWARE

Thunder uptime is 17 hours, 53 minutes
System returned to ROM by reload at 23:59:40 MET Mon Jul 2 2001
System restarted at 00:01:30 MET Tue Jul 3 2001
System image file is "tftp://172.17.247.195/gsr−p−mz.15S2plus−FT−14−May−2001"

cisco 12012/GRP (R5000) processor (revision 0x01) with 262144K bytes of memory.
R5000 CPU at 200Mhz, Implementation 35, Rev 2.1, 512KB L2 Cache
Last reset from power−on

2 Route Processor Cards
1 Clock Scheduler Card
3 Switch Fabric Cards
1 8−port OC3 POS controller (8 POs).
1 OC12 POs controller (1 POs).
1 OC48 POs E.D. controller (1 POs).
7 OC48 POs controllers (7 POs).
1 Ethernet/IEEE 802.3 interface(s)
17 Packet over SONET network interface(s)
507K bytes of non−volatile configuration memory.

20480K bytes of Flash PCMCIA card at slot 0 (Sector size 128K).
8192K bytes of Flash internal SIMM (Sector size 256K).

Thunder#show controller fia
Fabric configuration: Full bandwidth nonredundant
Master Scheduler: Slot 17
```

We recommend that you read How To Read the Output of the show controller fia Command for more detailed information.

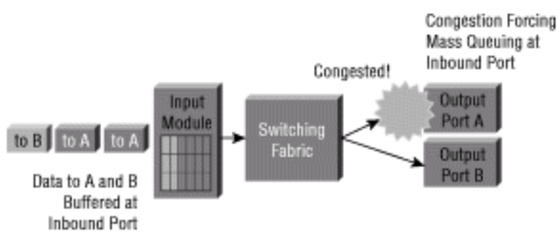
Switch Fabric Design

The 12000 switch fabric design includes innovative approaches resulting in a highly efficient system. The switch fabric uses the following key components to provide a highly efficient carrier class and scalable design:

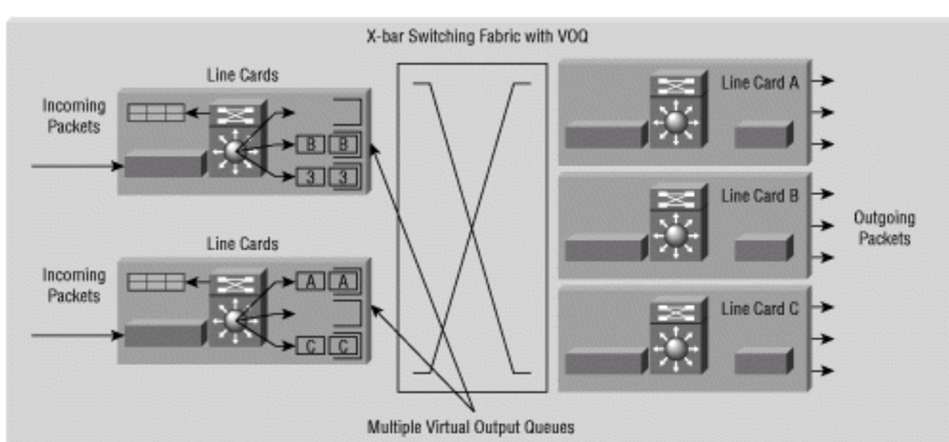
- Virtual output queues per line card to eliminate head of line blocking.
- An efficient scheduling algorithm in place of the traditional round robin approach to improve fabric efficiency.
- Hardware−based replication for multicast traffic; supports partial fulfillment to provide a highly efficient platform for multicast traffic.
- Pipelining to improve switch fabric performance.

Virtual Output Queues

Head of Line Blocking (HoLB) is a problem that occurs in any system where congestion exists at the output port (see the figure below). HoLB occurs when multiple packets, destined for multiple destinations, all share one queue. Packets destined for a specific location must wait until all packets ahead of it are processed before being passed through the switch fabric. An example of this is when several multiple lane highways are merged into a one lane highway. The best way to solve this is to have several multi−lane highways merge into one multi−lane highway.

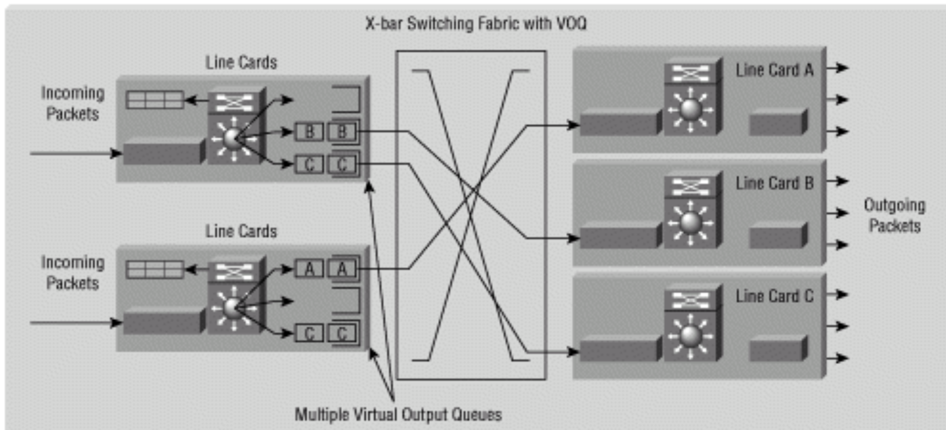


The Cisco 12000 Series Internet Router uses a unique multi−queue implementation to eliminate Head of Line Blocking. As packets arrive into the line card, they are arranged into one of multiple output queues categorized by slot, port, and Class of Service (CoS). These queues are referred to as virtual output queues (VOQs).



In the figure above, Virtual Output Queue (A) represents line card A, VOQ B represents line card B, and so on. Each packet is sorted and placed in the proper VOQ. The sorting and placement in the VOQ are based on the forwarding information contained in the Cisco Express Forwarding (CEF) table.

The following figure shows how the VOQ approach avoids the HoLB problem. As the figure indicates, packet placement minimizes the HoLB problem. Even if a series of packets is being sent to one line card, the other packets in the different VOQs can be sent across the switching fabric, avoiding the classic HoLB problem.



Scheduling

The SFC/CSC has an embedded scheduling algorithm. The scheduling algorithm, jointly developed by Cisco Systems and Stanford University, receives up to 13 input requests for the Cisco 12008 and Cisco 12012 (12 slots and 1 multicast) and 17 input requests for the Cisco 12016 (16 slots and 1 multicast). All requests are completed during a given clock interval. The algorithm calculates the best input−to−output match available in that interval. This high−speed algorithm, along with the VOQ innovation, enables the switching fabric to achieve very high levels of switching efficiency. This means the throughput of the switching fabric can reach up to 99 percent of the theoretical maximum versus the 53 percent achieved by earlier switch fabric designs (data based on research conducted at Stanford University).

Multicast Support

The switching fabric is also designed for next−generation applications, which use IP multicast. The switching fabric overcomes the traditional problems associated with IP multicast by:

- Using special hardware that performs intensive replication of IP packets on a distributed basis (in the fabric and line card)
- Dedicating separate queues (VOQs) for multicast traffic, so that other unicast traffic is not impacted
- Allowing for the creation of partial multicast segments

An interface can send both multicast and unicast requests to the switch fabric. When a multicast request is sent, it specifies all destinations for the data and the priority of the request. The CSC handles multicast and unicast requests together, giving precedence to the highest priority request, whether unicast or multicast.

When a multicast request is received, a request is sent to the Clock Scheduler Card. Once a grant is received from the CSC, the packet is then forwarded to the switch fabric. The switch fabric makes copies of the packet and sends the copies to all destination line cards simultaneously (during the same cell clock cycle). Each receiving line card makes additional copies of the packet if it must be sent to several ports.

In order to reduce blocking, the switching fabric supports partial allocation for multicast transmissions. This means that the switching fabric performs the multicast operation for all available cards. If a destination card is receiving a packet from another source, the multicast process is continued in subsequent allocation cycles.

These new enhancements avoid the bandwidth−wasting obstacles inherent in first generation crossbar switching fabrics, and enable Cisco Systems to deliver a switching fabric that achieves a very high level of switching efficiency without sacrificing reliability.

Pipelining

The switching fabric supports full−duplex operation, supplemented by advanced pipelining techniques. Pipelining allows the switch fabric to start allocating switch resources for future cycles before it has completed transmission of data for previous cycles. By eliminating dead time (wasted clock cycles), pipelining dramatically improves the overall efficiency of the switch fabric. Pipelining enables high performance in the switching fabric, allowing it to reach its theoretical maximum throughput.

Cisco Cells

The unit of transfer across the crossbar switch fabric is always fixed−size packets, also referred to as Cisco cells, which are easier to schedule than variable−size packets. Packets are broken into cells before being placed on the fabric, and are reassembled by the outbound LC before they are transmitted. Cisco cells are 64 bytes long, with an 8−byte header, a 48−byte payload, and an 8−byte cyclic redundancy check (CRC).

Related Information

- **Cisco 12000 Series Internet Router Architecture − Chassis**
- **Cisco 12000 Series Internet Router Architecture − Route Processor**
- **Cisco 12000 Series Internet Router Architecture − Line Card Design**
- **Cisco 12000 Series Internet Router Architecture − Memory Details**
- **Cisco 12000 Series Internet Router Architecture − Maintenance Bus, Power Supplies and Blowers, and Alarm Cards**
- **Cisco 12000 Series Internet Router Architecture − Software Overview**
- **Cisco 12000 Series Internet Router Architecture − Packet Switching**
- **Understanding Cisco Express Forwarding**
- **How To Read the Output of the show controller fia Command**
- **Technical Support − Cisco Systems**

Updated: Jul 07, 2005 Document ID: 47240