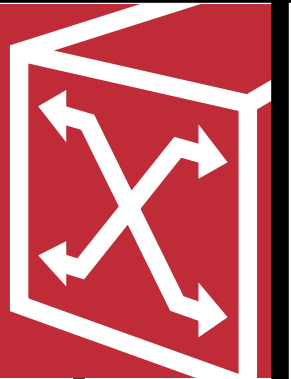


Catalyst 6500 Series Service Provider Features



The Catalyst® 6500 Series delivers feature-rich solutions for a variety of service provider applications such as gigabit aggregation, Web and application hosting, and WAN edge services by integrating high bandwidth, performance, availability, and services into a single platform. This white paper describes some of the key features that are most applicable in a service provider environment, specifically Cisco Express Forwarding (CEF), the FlexWAN Module, stateful high availability, policing, intrusion detection, and Private VLANs.



Catalyst 6500 Family Overview

The Catalyst 6500 Series has been designed as a solution for a wide variety of network environments. It is available in six- and nine-slot horizontal chassis, has a nine-slot vertical chassis designed for NEBS environments, and can be configured with redundant power supplies, switching fabrics, supervisor engines, and so forth. The Supervisor Engine provides the central processing and intelligence of the system. A Policy Feature Card (PFC) adds hardware-accelerated quality of service (QoS) and access control list (ACL) features. PFC is optional on a Supervisor 1A, and PFC2 is standard on a Supervisor 2. An optional Multilayer Switch Feature Card (MSFC) provides Cisco IOS® Software Layer 3 services. Line-card offerings include the standard 10/100 (RJ-45 and RJ-21) and gigabit connections (GBIC, MT-RJ), Network Analysis Modules, Intrusion Detection Modules, and WAN interface support. Software features provide intelligent network services including Border Gateway Patrol (BGP4), Intermediate System to Intermediate System (IS-IS), and Cisco IOS Server Load Balancing (IOS-SLB).

This paper focuses on the hardware and software specifically designed for service provider environments.



The new Supervisor Engine 2 uses the latest application-specific integrated circuit (ASIC) technology, building upon the Supervisor Engine 1A, to deliver next-generation features and services. The new Supervisor Engine 2 is a key component of the new Cisco Express Forwarding (CEF)-based architecture that enables distributed forwarding. The Supervisor Engine 2 also acts as the control module for the new Catalyst 6500 Series architecture that is built around the crossbar fabric that scales the bandwidth to 256 Gbps. Supervisor Engine 2 is required to enable a Catalyst 6500 Series with the 256-Gbps crossbar architecture of the Switch Fabric Module. The Supervisor Engine 2 allows connections to both the bus and switch fabric making it a suitable product for the entire Catalyst 6500 Series. This also allows full investment protection for customers who plan to transition to the new architecture by allowing them to populate the chassis with classic (non fabric-enabled cards) as well as new fabric-enabled cards.

Both Supervisor Engines 2 and 1A deliver the most advanced intelligent switching available in the industry, providing per-port application recognition, admission control, prioritization, and policing with multiple hardware queues to minimize network congestion and packet latency. This advanced QoS support is essential for enabling network-wide deployment of mission-critical applications and enterprise voice services and solutions.

The Supervisor Engine 2 further enhances performance by offering many features, such as multicast replication, that are required for video streaming applications in hardware. This allows both enterprises and services providers to use multiservice applications without sacrificing performance.

The new Supervisor Engine 2 and Switch Fabric Module (SFM) double centralized forwarding performance for both Layer 2 and Layer 3 to 30 Mpps. The new fabric-enabled gigabit ethernet cards have a connection to the crossbar fabric for high-speed switching between cards. They can also be used to enable distributed CEF via a distributed forwarding engine. This architecture enables local switching performance of up to 24 Mpps per card allowing the switch to scale to a forwarding rate of over

100 Mpps. The Supervisor Engine 2 connects to both the crossbar and bus allowing another level of redundancy by switching to the bus when there is a failure in the crossbar.

Cisco IOS Software Layer 3 services and features are available via the MSFC. This is a daughtercard for the Supervisor Engine that provides the control plane functions of the system. The MSFC works with the PFC for Layer 3 switching. The MSFC2, which has recently been announced for the Catalyst 6000 Family, triples the control plane performance from the original MSFC. All traffic that uses the MSFC2 for switching will experience a direct performance enhancement. As a result, features such as IOS-SLB and Web Cache Communication Protocol 2 (WCCP2) are also tripled in performance. The MSFC2 offers up to 512 MB of Error-Correcting Code (ECC) DRAM for full Internet routing table support as well as offering BGP4, IS-IS, and full multiprotocol routing.

Cisco Express Forwarding

Cisco Express Forwarding (CEF) is a technology that allows for increased scalability and performance to meet the future requirements for large enterprise networks and the core of the Internet. The more general industry name for CEF is Forwarding Information Base (FIB). CEF has evolved to meet the ever-changing traffic patterns of today's networks. These networks are frequently characterized by an increasing number of short-duration flows.¹ Shorter flows are very common in environments with a high degree of Web-based or other highly interactive types of traffic. As these types of applications continue to proliferate, a higher-performing and more scalable forwarding methodology is required to meet the needs in the largest of these environments.

With a CEF-based mechanism, the challenges presented by flow-caching models are greatly reduced and scalability is greatly increased. With flow-cached models, a complete forwarding table must be maintained for the proper handling of first packets by the CPU or packets that for some other reason cannot be processed in hardware. After the CPU makes a forwarding decision

1. A flow is defined as a unidirectional stream of traffic from a given Source + TCP/UDP Port number to a given Destination + TCP/UDP Port number.



(part of which involves a lookup against the routing table), an entry is made into the flow cache table, and subsequent packets are handled in hardware. With a CEF-based forwarding model, all packets, including the first packet in a given flow, are handled in hardware. A routing table is still maintained by the router CPU, but two additional tables are created in the CEF-based model. These tables are populated before any actual user traffic is present in the network, such as would be the case with a cache-based model. The first of these tables is actually a copy of the relevant forwarding data points from the routing table, and is known as the *FIB table*.

The second table is called the *adjacency table*. The adjacency table maintains a database of node adjacencies (two nodes are said to be adjacent if they can reach each other via a single hop at the link layer of the Open System Interconnection [OSI] model), and their associated Layer 2 Media Access Control (MAC) rewrite or next-hop information. By performing a high-speed lookup against these two tables, a forwarding decision, along with the appropriate rewrite information, can be accessed in a highly efficient and consistent manner while also providing a mechanism that offers a high degree of scalability.

Performance

In a pure fabric-enabled implementation of the Catalyst 6500, whereby a Catalyst 6500 Family chassis is equipped with a Supervisor Engine 2, switch fabric module, and fabric-enabled line cards, raw forwarding performance of up to 30 Mpps can be attained in the central switching engine. By comparison, Supervisor Engine 1A can achieve performance levels as high as 15 Mpps. Additional throughput can also be achieved by adding the forthcoming Cisco distributed forwarding cards (DFCs), which are field-upgradable daughter cards. The DFC provides additional CEF-based forwarding capabilities on a per-slot basis, allowing performance to scale to more than 100 Mpps. This model is conceptually very similar to the distributed CEF model of the Cisco 7500 Router Family. A critical consideration with respect to CEF functionality on the Catalyst 6500 Series is that ACL and QoS mechanisms are implemented in hardware and thus do not have a negative impact on system performance.

Scalability

CEF-based forwarding increases scalability, not only by the raw number of available FIB entries, but also by the replication of CEF-based forwarding technology on a per-slot basis. This distributed CEF capability allows the Catalyst 6500 Family to provide sufficient forwarding capabilities for the largest of networks. The Catalyst 6500 Family when equipped with a Supervisor Engine 2 can provide as many as 128,000 entries in the FIB table, and 128,000 entries in the adjacency table. Note that these entries are not stored via a hashing algorithm, as with the NetFlow table. The FIB and adjacency tables use Ternary CAM (TCAM) technology for high-speed lookups.

Availability

While the Catalyst 6500 family offers a breadth of features to support high availability, the addition of CEF further deepens the availability and stability aspects that are demanded in today's Internet infrastructures. As networks increase in size, a natural side effect is the increased chance of instability and change. Network instability, whether it is caused by failures, configuration changes, or bursty traffic patterns, can have a tremendous impact on routing implementations, which rely upon heavy CPU computations for routing-table maintenance. Because CEF employs a mechanism by which the forwarding information is constructed based on the topology of the network rather than a representation of traffic within the network, the CPU is no longer burdened by having to set up large numbers of entries. This also means that the consequent network availability is not directly linked to the actual size of the network.

Accounting

The flow cache model that the Catalyst 6500 Series has relied upon so successfully has provided much-needed network-usage accounting data. This data has been in the form of NetFlow accounting. NetFlow data export has been used to extract this flow accounting data from a Catalyst 6500 Series chassis, and to provide highly detailed accounting data, which provided usage information on an individual flow basis.



When equipped with a Supervisor Engine 2, the Catalyst 6500 Series chassis no longer makes flow-based forwarding decisions, but can still provide NetFlow accounting data. Although the CEF-based forwarding mechanism is used on Supervisor Engine 2 for actual Layer 3 forwarding, a NetFlow forwarding table continues to be maintained in parallel in order to provide the necessary accounting data.

Load Balancing

Cisco Express Forwarding also delivers additional functionality with respect to load balancing of traffic across multiple equal-cost parallel paths. Traditionally, routing protocols such as OSPF and Enhanced Interior Gateway Routing Protocol (EIGRP) have limited equal-cost load-balancing paths to four paths. With CEF, this has been increased to six equal-cost parallel paths. With routing protocols, more than four parallel paths can exist, but a maximum of four can be installed in the actual routing table. Routing protocols such as EIGRP and OSPF support the concept of a Routing Information Base (RIB). If a route cannot be installed into the routing table, it can be maintained in the RIB. Because CEF relies upon a FIB and an adjacency table for forwarding decisions, it is not limited to that which is actually present in the routing table, and, in fact, can reference the RIB, which enables CEF to install a greater number of parallel paths (with entries gleaned from the RIB). It is important to note that the Catalyst 6500 Family offers the ability to load balance across parallel paths on a per-flow basis only. This means that per-packet load balancing is not supported on the Catalyst 6500 Family of Switches when using CEF-based forwarding.

Access Control Lists

Another manner in which various CEF implementations differ is with respect to their support of ACLs. In the case of the Catalyst 6500 Family, ACLs can be handled in hardware for most common configurations. CEF-based forwarding does not have any impact upon this functionality. The most notable ACL options that disable hardware processing for a given ACL are the use of the log keyword or enabling *ip-unreachables* on a given interface. Note that these ACL options are independent of CEF on the Catalyst 6500 Family Switches.

CEF Support for IP Multicast

The CEF implementation on the Catalyst 6500 Family also includes support for IP multicast. The FIB table can hold as many as 16,000 entries for IP multicast, including both (S,G) and (*,G) entries (for a total of 16,000 entries).

Summary

CEF provides high performance and high scalability for the largest and most demanding networks. Cisco has enabled the Catalyst 6500 family of switches with CEF forwarding embedded in hardware with the introduction of the Supervisor Engine 2. The Supervisor Engine 2 offers the scalability and performance requirements necessary for both large enterprise and service provider environments. Cisco has also managed to maintain the rich accounting support of the multilayer switching forwarding model while introducing CEF-based forwarding.

FlexWAN Module

The FlexWAN module adds wide-area network (WAN) and metropolitan-area network (MAN) capabilities to the Catalyst 6000 Family of Switches ranging from Fractional T1/E1 to 155 Mbps. The module enables the use of Cisco 7200 and 7500 Series single-wide port adapters in the Catalyst 6000 Family.

FlexWAN allows for the integration of WAN and MAN access with local-area network (LAN) switching, simplifying network design and consolidating the network infrastructure. This results in fewer network elements to manage in the total solution. FlexWAN was designed to deliver WAN access and sophisticated QoS features such as Class-Based Weighted Fair Queuing. It delivers a broad range of WAN media options by using existing Cisco 7200 and 7500 Series port adapters. The module allows the Catalyst 6000 Family to serve the purpose that an external WAN router would normally perform.

With new network deployments, FlexWAN can be used to converge LAN, MAN, and WAN connections from other campuses. Catalyst 6000 Family Switches acting as the core or distribution elements with FlexWAN modules can terminate connections from other campuses in the MAN and WAN. The central Catalyst 6000 Family Switches use DS3 and ATM to terminate remote sites and OC-3 Packet over Synchronous Optical Network (SONET).



High Availability

The Catalyst 6500 Series of multilayer switches has become a key component of a sound network design in today's service provider environments. Playing such a critical role, the Catalyst 6500 Series must provide a reliable switching platform while, at the same time, offering high performance and intelligent network services. The Catalyst 6500 Series has several features and options for increasing system high availability. These include optional redundant switching fabrics, the stateful protocol redundancy of the high availability feature, and image versioning support for hitless software upgrades. The following sections of this paper describe these high availability features in greater detail.

Redundant Switching Fabrics

Since its introduction, the Catalyst 6500 Series has been built on a single bus switching architecture, which provides the data path for all packets through the system. The next generation for the Catalyst 6500 Series includes a crossbar switching fabric (the Switch Fabric Module, or SFM) as an alternative switching architecture for higher bandwidth requirements. The SFM also provides another level of hardware redundancy to the system. The first generation of the fabric-enabled line cards (for example, WS-X6516-GBIC) will provide a connection to both the switching fabric as well as the existing system bus backplane. This allows the Catalyst 6500 system to use the switching fabric as the primary means of data transfer for fabric-enabled line cards. If the switch fabric fails, the system bus backplane will take over to ensure packet switching continues, albeit at 15 Mpps, and the network remains online. Note that this change in switching performance is applicable only if the system is initially forwarding at greater than 15 Mpps. If a system is running at 15 Mpps or below, then the fabric to system bus failover does not affect performance. Additionally, a Catalyst 6500 series can be configured with dual switch fabric modules (in slots 5 and 6), which provide a third level of fabric redundancy. In this configuration, a failure on the primary fabric module would result in a switchover to the secondary fabric module for continued operation at 30 Mpps. Any further fabric module failures, however unlikely, would still have the ability to switchover yet again to the system bus.

Redundant Supervisor Engines

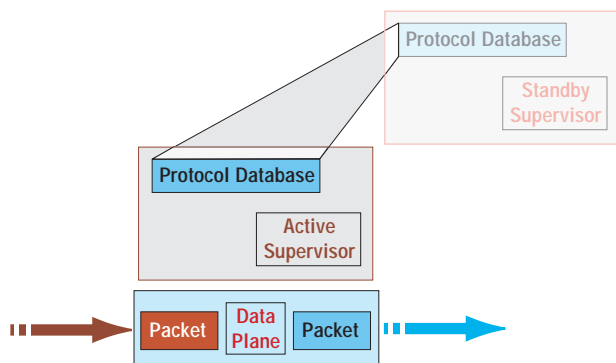
Dual supervisor engines provide hardware redundancy for the forwarding intelligence of the Catalyst 6500 Series. The Catalyst 6500 Series can support up to two supervisors (slots 1 and 2 only). One supervisor serves as the running, or active, supervisor and the other serves as the standby supervisor. The active supervisor engine is the first one to come online. This can be confirmed by the "Active" LED externally on the supervisor or by typing the show module command from the console. Both supervisor engines *must* be the same hardware model (that is, if a PFC and MSFC are on a Supervisor 1A in slot 1, then a PFC and MSFC must also be on a Supervisor 1A in slot 2, or if a Supervisor 2 is in slot 1, then a Supervisor 2 must also be in slot 2). Both Supervisor Engines 1A and 2 can be used in the Catalyst 6000 and 6500 Series. If an active supervisor is taken offline or fails, the standby supervisor takes over control of the system.

The two supervisors in a redundant supervisor configuration have different responsibilities. The active supervisor controls the system bus and all line cards. All protocols run on the active supervisor, which performs all packet forwarding. The standby supervisor does not communicate with the line cards. It receives packets from the network and populates its forwarding tables with this information, but does not participate in any packet forwarding. The relevant protocols on the system are initialized, but not active, on the standby supervisor. The Catalyst 6500 Series Supervisor Engines are hot-swappable, and the standby supervisor engine can be installed into an active system without affecting network operation. Note that redundant supervisors do not perform load sharing. The active supervisor provides the entire packet forwarding intelligence for the system (N+1 redundancy). If the active supervisor fails, the standby supervisor can still maintain the same system load.

The high availability feature decreases the switchover time from the active to the standby supervisor to less than 3 seconds for return to normal operation. This minimized downtime is achieved by synchronizing many of the Layer 2, 3, and 4 protocols between the active and standby supervisor engines. This is referred to as maintaining protocol state and is depicted in Figure 1.



Figure 1 Redundant Supervisor Stateful Protocol Redundancy



For stateful protocol redundancy between dual supervisor engines, a protocol state database is maintained on each supervisor engine for all protocols and features requiring high availability support. Most of these protocols only run on the active supervisor. In the event of a high availability switchover, the new active supervisor can start the protocols from the updated database state, rather than the initialization state. In this way, a redundant supervisor system can maintain stateful protocol redundancy and minimal network downtime when the active supervisor goes offline.

Image Versioning

Image versioning is the second portion of the high availability feature. This feature is dependent on having the high availability feature enabled in a dual supervisor configuration. As such, it allows different but compatible images to run on the active and standby supervisors, thus disabling the default supervisor image synchronization process. This feature enables upgrade of the supervisor software version “on the fly” by using the stateful supervisor switchover of the high availability feature. This process has been referred to as “hitless software upgrades” but in reality a small “hit” (under 3 seconds) is taken. This allows not only the upgrading of software without rebooting the entire box, but also the ability to maintain a previously used and tested version of on the standby supervisor as a fallback plan should anything go wrong with the software upgrade.

MSFC Redundancy

The Multilayer Switch Feature Card 2 (MSFC 2) routing engine is an optional daughtercard on the supervisor engine. A redundant supervisor hardware configuration

can also include redundant MSFC 2 routing engines. As such, the proper operation of the MSFC 2 is predicated by proper operation of the supervisor engine. A supervisor reset or failover also resets the MSFC routing engine.

While the CatOS high availability feature maintains protocol state between redundant supervisors, Hot Standby Routing Protocol (HSRP) needs to be configured for failover between redundant MSFCs. HSRP is used to provide the first-hop, unicast redundancy. HSRP needs to be configured on both MSFCs for each virtual LAN (VLAN) where first hop redundancy is required. In a dual Supervisor Engine and MSFC configuration, both MSFCs are active routers. So, HSRP between MSFCs works in the same manner as HSRP between physically separate boxes by using a polling mechanism between the routing engines. Further HSRP information can be found in the Catalyst 6000 Family Software User Guide available on Cisco.com.

Before the MSFC IOS Software Release 12.1(3a)E4, each MSFC had to be configured separately. More specifically, configuration parameters such as access lists, QoS features, and so forth, had to be individually configured on both MSFCs in an identical manner. Parameters that cannot be duplicated, such as IP addresses and HSRP settings, still had to be configured differently on each MSFC.

Additionally, before this new software release, the interfaces of a FlexWAN module only belonged to the designated MSFC. This means that those interfaces did not show up in the nondesignated MSFC configuration and, thus, were not configurable on the nondesignated MSFC. During a Supervisor/MSFC failover, the MSFC that becomes the new designated MSFC will not have the proper configuration of the FlexWAN interfaces. For this reason, a redundant Supervisor/MSFC configuration was not supported with the FlexWAN module installed. The MSFC config-sync feature removes this limitation. Thus, a FlexWAN module is now supported in a redundant Supervisor configuration with the config-sync feature enabled.

Beginning in the MSFC IOS release 12.1(3a)E4, a MSFC redundancy feature called config-sync is available to streamline the redundant MSFC configuration process for both MSFC and MSFC2. This feature synchronizes both



the startup and running configurations between the designated (primary) and nondesignated (secondary) MSFCs. Specifically, whenever a write memory or copy *source startup-config* command is issued on the designated MSFC, the startup configurations in NVRAM of both MSFCs are updated. This allows the configurations on both the designated and nondesignated MSFC to maintain the same configuration without having to manually type each command twice. All configurations for the designated and nondesignated MSFCs are done through the command-line interface (CLI) of the designated MSFC.

The use of a dual supervisor and dual MSFC configuration combined with the high availability feature provides an added level of redundancy to a network design.

High Availability via EtherChannel

EtherChannel® technology provides increased bandwidth and redundancy from standard single-link connections. An EtherChannel bundle is a group of up to eight Fast Ethernet or Gigabit Ethernet links. The bundle functions as a single, logical connection between switches and routers. EtherChannel is convenient because it scales the bandwidth without adding to the complexity of the design. Spanning-Tree Protocol treats the EtherChannel bundle as a single link, so no spanning-tree loops are introduced. Routing protocols also treat the EtherChannel bundle as a single, routed interface with a common IP address, so no additional IP subnets are required, and no additional router peering relationships are created. The load balancing is handled by the interface hardware. An EtherChannel link can be created across supervisor or line-card ports, thereby reducing the impact of a single line-card failure.

Policing

There is great interest today by service providers to sell incremental bandwidth by provisioning 10/100/1000 Mbps Ethernet connections to their customers, in both Web-hosting data centers and MANs. Previously the only way to provision incremental bandwidth in WANs and MANs was through Frame Relay, Asynchronous Transfer Mode (ATM), or serial connections. Data centers or Web-hosting facilities “gave away” bandwidth by using Ethernet connection to data center hosts without any means of selling incremental bandwidth in this manner.

Today there is a solution to this problem. The Catalyst 6000 Family offers a robust policing feature for 10/100/1000Mbps Ethernet connections. Policing is the process by which the Catalyst 6000 Family Multilayer Switch limits the bandwidth consumed by a flow of traffic. Policing can either mark traffic up or down, or drop traffic outright.

Configuring policing on the Catalyst 6000 Family Ethernet interfaces allows service providers to sell bandwidth at varying quantities on these connections. This feature now gives service providers a way to sell their customers bandwidth with the ability to provision or upgrade that service incrementally by software configuration on the switch.

Policing is performed on the PFC hardware of the Catalyst 6000 Family with no performance impact. The Catalyst 6000 Family Switches can police Ethernet interfaces at a minimum rate parameter of 32 Kbps, up to a maximum of 8 Gbps.

Policing policy is configured to either drop all out-of-profile packets or mark all out-of-profile packets with a lower IP precedence or differentiated services control point (DSCP) value.

With a PFC2, you can specify a dual rate aggregate policing rule with a normal rate and an excess rate.

- Normal rate—Packets exceeding this rate are marked down.
- Excess rate—Packets exceeding this rate are either marked down or dropped as specified by the drop indication flag.
 - Packets are dropped if the excess rate of the policer returns an out-of-profile decision and a drop action is configured.
 - Packets are marked-down if the excess rate of the policer returns an out-of-profile decision and a markdown action is configured.

Additionally, the PFC2 offers per-policer statistics that display the following:

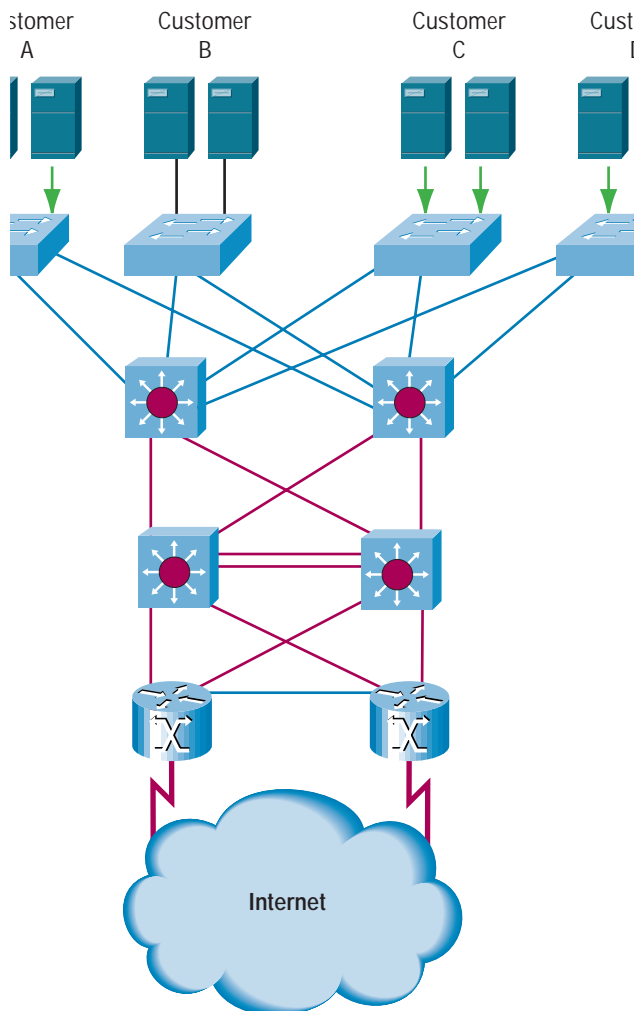
- Total traffic policed
- Traffic exceeding normal rate
- Traffic exceeding excess rate



Figure 2 illustrates an example of using policing in a data center or Web-hosting facility. This allows a service provider to incrementally “sell” and control bandwidth to customers or servers within a data center.



Figure 2 Policing Example in a Data Center Intrusion Detection



Policing on Server Links

links allows individual configuration of bandwidth per port, per class of traffic that exceeds the configured bandwidth rate will be dropped (or marked). For example:

server connections are policed at 5 Mbps even though each server has 10-Mbps connections can be defined with one policer policy statement.

server connections are not policed because they purchased 10-Mbps connections.

server connections are policed at 2 Mbps even though each server has 10-Mbps connections can be defined with one policer policy statement.

server connections are policed at 2 Mbps even though each server has 10-Mbps connections can be defined with one policer policy statement.

Content and intellectual property are the greatest assets that most companies have, and securing that intelligence has become a major challenge. Security is no longer perceived primarily as an insurance policy. In the e-business environment, security is rightly viewed as a business-enabling infrastructure. Customers are increasingly using security products and services to help them dramatically scale revenue, transactions, and customers at high double-digit rates while confining cost increases to single-digit or low double-digit rates. If implemented successfully, this strategy ensures profitable growth.

The Catalyst 6000 Intrusion Detection System (IDS) Module provides around-the-clock network surveillance while analyzing the packet data streams within the network, searching for unauthorized activity, such as attacks by hackers, and enabling users to respond to security breaches before systems are compromised. When unauthorized activity is detected, the system can send alarms to a management console with details of the activity and can often dynamically tune policies to cut off the unauthorized sessions.

Figure 3 Catalyst 6000 Family IDS Module



Using the comprehensive signature database and custom signature, the IDS Module can detect both context-based (packet header only) and content-based (payload) attacks, for example, smurf, land, syn attack, port sweep, ping of death, to name a few. The IDS Module analyzes the captured packets and compares them against the signature database for typical intrusion activity. If the captured packets match a defined intrusion pattern in the rule set, the module sends an alarm to the management console and automatically responds (if configured to do so). The alarms are sent out on a separate management interface to avoid impeding continual packet capture by the monitoring interface.

Private VLANs

Background

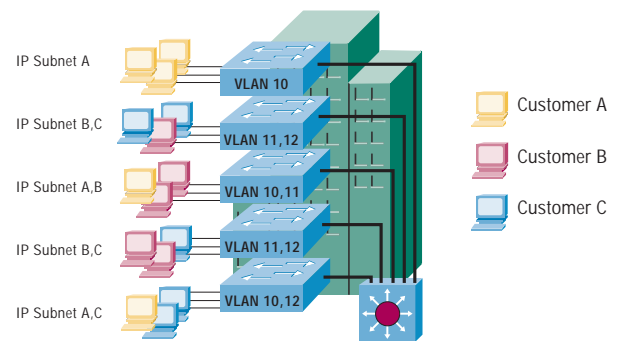
Within a switched Ethernet environment, a VLAN is used to provide any-to-any connectivity for all hosts within the VLAN. As segregation is required for hosts within a switched Ethernet environment, more VLANs are created. Layer 3 devices provide interconnectivity between VLANs, namely routers or multilayer switches. As each VLAN terminates on a Layer 3 device, each VLAN is also assigned its own network subnet from an IP perspective. Regardless of whether a single or multiple VLANs exist within a single switch, Layer 3 devices are used to interconnect them.

A common application of VLANs is for security purposes. For example, if different customers were within a shared network environment, it would be required to group those stations belonging to a single customer to their own

Many complications arise with this model of mobility. The complexity continues to increase as more customers and more users are added to the network. The more users that are added over time, the less likely users belonging to a common customer will have the ability to be in close proximity to one another.

respective VLAN and IP subnets. Figure 4 shows the VLAN and IP subnet layout for such a configuration. Because the network users of each customer could potentially be dispersed within the building, the individual VLANs might tend to span multiple floors.

Figure 4 Common VLAN Layout

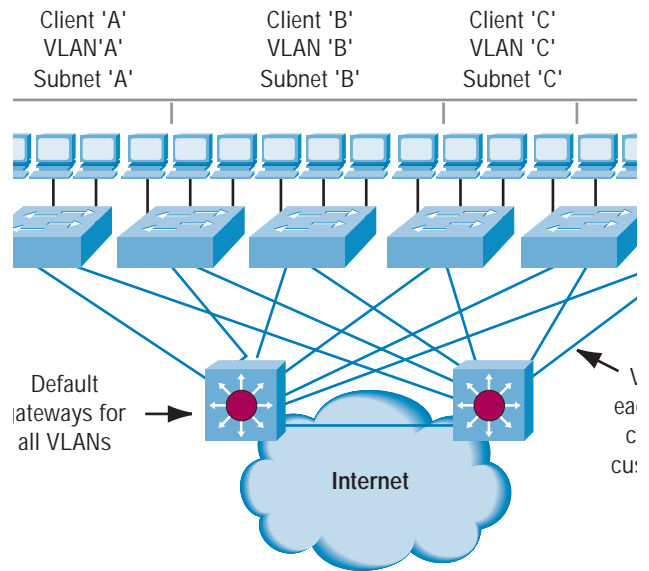


From a Layer 2 perspective, each customer that is added to such a network requires a new VLAN to be configured and managed. Spanning-tree becomes more complicated as more VLANs are added to the topology. In addition, to efficiently manage bandwidth within this environment, you must constantly track each VLAN trunk to ensure it is only transporting necessary VLANs. As new customers are added to the topology, or existing customers span to new switches, spanning-tree becomes a little more complicated.

From a Layer 3 perspective, there are more complications. Because each customer has its own VLAN, it must also have its own IP subnet. Within such a design, there are likely to be two separate exit points, namely Layer 3 switches. Therefore, for each customer added to the network, a separate IP subnet must be allocated. Within each allocated subnet, two addresses are consumed by broadcast addresses, and three addresses are consumed by those addresses associated with HSRP. The complexity is increased by the need to accurately estimate the future address space requirement so that an initial IP subnet can be allocated to accommodate expected growth. Because customers can vary in size, the routing tables within such a design can become cluttered by numerous variable-length subnets.

The most common example of this sort of network requirement exists within a Web-hosting service. A Web-hosting service consists of a large LAN infrastructure with the ability to house many customers and their associated Web servers. Figure 5 shows a common Web-hosting network with four customers. To ensure security between customers of the Web-hosting service, each customer is assigned its own VLAN and IP subnet. Each customer houses a different number of servers and potentially requires a different size of subnet.

Figure 5 Typical Web-Hosting Environment



One obvious limitation to the illustrated deployment is the scalability of spanning-tree. Within a modern data center that provides Web-hosting services, more than 10,000 servers are not uncommon. However, in the typical model, each customer is assigned its own VLAN regardless of its size. This practice results in a limitation imposed by spanning-tree scalability and performance within a large switched environment.

The typical hosting environment leads to the excessive waste of IP address space as well.

A common characteristic of environments like the Web-hosting design is the sole requirement for connectivity from the hosts themselves to the Internet through the default gateways. In many cases, connectivity among the servers is not a requirement. A common approach to facilitating communication between servers that belong to the same customer is to build a separate “back-end” network. Therefore, the sole purpose of allocating separate VLANs to unique customers is to prevent server communication from occurring between servers that belong to different customers while allowing communication between all servers and the default gateways enroute to the Internet.

To alleviate many of the difficulties associated with deployment and management of environments such as a typical hosting environment, Cisco has developed a feature called private VLANs.

Private VLAN Description

A Private VLAN is a Layer 2 network structure not uncommon to, but rather an extension, of the common VLAN. Within a Private VLAN there are three separate port designations. Each port designation has its own unique set of rules that regulate a connected endpoint's ability to communicate with other endpoints connected to ports within the Private VLAN. The three port designations are promiscuous, isolated, and community.

An endpoint connected to a promiscuous port has the ability to communicate with any endpoint within the Private VLAN. Multiple promiscuous ports can be defined within a single Private VLAN. Within the previously mentioned hosting environment, the Layer 3 switch default gateways are commonly connected to promiscuous ports.

Isolated ports are typically used for those endpoints that only require access to a limited number of endpoints. An endpoint connected to an isolated port can only communicate with those endpoints connected to promiscuous ports. Endpoints connected to adjacent

isolated ports cannot communicate. Within a Web-hosting environment, isolated ports are reserved for hosts that only require access to default gateways.

In some cases, communication is required between endpoints that belong to the same customer. "Front-end" content replication, high-availability NICs, and server clustering all require some form of front-end replication. In such cases, the community feature of the Private VLANs is very useful. A Private VLAN community is a grouping of isolated ports that belongs to one customer. Within the community, endpoints can communicate with one another and with any defined promiscuous port. Endpoints that belong to one community cannot communicate with endpoints of a different community.

Regardless of the combination of isolated, community, and promiscuous ports used within a Private VLAN, it is still one Layer 2 structure and therefore requires only one IP subnet. The addressing model now changes, and instead of allocating an individual subnet to each customer, a range of addresses from one or two common large IP networks is assigned. By allocating addressing from one or two common larger IP networks, the address waste is severely reduced.

All restriction functions of the Private VLAN are implemented in hardware and therefore do not rely on a software learning mechanism for setup. A hardware implementation ensures the utmost in port security and does not subject the Private VLAN feature to the effects of potential software bugs. In addition, private VLANs and regular VLANs can co-exist on the same chassis at the same time. You can also define multiple Private VLANs within a switch if required.

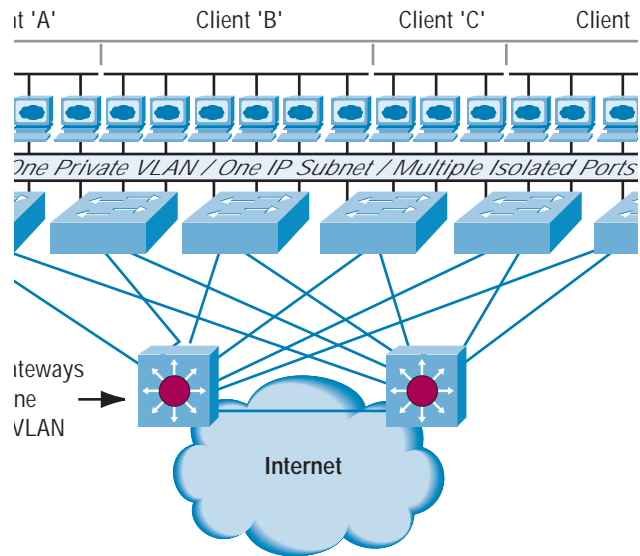
Application of Private VLANs

Private VLANs provide two main benefits, namely optimized IP address management and reduced consumption of VLANs all within a multiclient environment. Because the entire Private VLAN structure is considered to be one Layer 2 domain, it can be addressed with one large address range. The need to estimate future address requirements for each client within the multiclient environment and subsequently assign a variable length subnet to each client has been alleviated. With Private VLANs, as new clients or new servers that belong to existing clients are added to the service, they receive subsequent addresses from the common contiguous network. Although the total address range given to any particular client can be fragmented, all addresses are routed as if they belong to one network. Address space fragmentation is a negligible concern within the Private VLAN system. The primary goal is to provide Layer 2 segregation between adjacent servers and access to a common service such as a default gateway, backup server, and so forth. Address space fragmentation was also common in the old model involving one VLAN per client when a client exceeded the number of servers that could be addressed by the assigned subnet.

In the scenario in Figure 4, communities are not used because separate back-end networks are included for server-to-server communication. Therefore, on the front end of the servers, one Private VLAN has been deployed. Each server is connected to an isolated port of the Private

VLAN. The entire infrastructure consumes only one IP network and two VLANs comprising one primary and one secondary VLAN or the isolated VLAN.

Figure 6 Private VLAN Application



The Private VLAN feature, which is available on the Catalyst 6500 Series of Multilayer Switches, greatly reduces the complexity of deploying a large-scale multiclient switched environment. IP address management is simple because new servers are assigned the next consecutive address of a common address pool. Instead of the client having ownership of a particular IP subnet, the client now owns a pool of IP addresses that may or may not be consecutive. The head-end Layer 3 switches only need to route for one IP network, namely



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

European Headquarters
Cisco Systems Europe
11, Rue Camille Desmoulins
92782 Issy-les-Moulineaux
Cedex 9
France
www.cisco.com
Tel: 33 1 58 04 60 00
Fax: 33 1 58 04 61 00

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

Asia Pacific Headquarters
Cisco Systems Australia, Pty., Ltd
Level 17, 99 Walker Street
North Sydney
NSW 2059 Australia
www.cisco.com
Tel: +61 2 8448 7100
Fax: +61 2 9957 4350

Cisco Systems has more than 200 offices in the following countries and regions. Addresses, phone numbers, and fax numbers are listed on the Cisco.com Web site at www.cisco.com/go/offices.

Argentina • Australia • Austria • Belgium • Brazil • Bulgaria • Canada • Chile • China PRC • Colombia • Costa Rica • Croatia • Czech Republic • Denmark • Dubai, UAE • Finland • France • Germany • Greece • Hong Kong SAR • Hungary • India • Indonesia • Ireland • Israel • Italy • Japan • Korea • Luxembourg • Malaysia • Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland • Portugal • Puerto Rico • Romania • Russia • Saudi Arabia • Scotland • Singapore • Slovakia • Slovenia • South Africa • Spain • Sweden • Switzerland • Taiwan • Thailand • Turkey • Ukraine • United Kingdom • United States • Venezuela • Vietnam •

All contents are Copyright © 1992–2001 Cisco Systems, Inc. All rights reserved. Important Notices and Privacy Statement. Catalyst, Cisco, Cisco IOS, Cisco Systems, the Cisco Systems logo, and EtherChannel are registered trademarks of Cisco Systems, Inc. or its affiliates in the U.S. and certain other countries. All other brands, names, or trademarks mentioned in this document or Web site are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (00010R)