Hoạt động của Switch

I. Layer 2 Switch  
  
Trong các hệ thống mạng dùng shared Ethernet, thiết bị hub thường được dùng. Nhiều host sẽ được kết nối như là một miền broadcast và miền xung đột (collision). Nói cách khác, các thiết bị shared Ethernet hoạt động ở L1.   
  
Mỗi host lúc này phải chia sẽ băng thông sẵn có cho tất cả các host khác đang kết nối vào hub. Khi có một hoặc nhiều host cố gắng truyền ở một thời điểm, xung đột sẽ xảy ra; lúc này tất cả các host phải lui về và chờ một khoảng thời gian để truyền lại. Cơ chế này áp đặt kiểu hoạt động half-duplex cho các host, nghĩa là các host hoặc là truyền, hoặc là nhận ở một thời điểm. Thêm vào đó, khi một host gửi ra một frame, tất cả các host sẽ nghe frame đó.   
  
Ở mức cơ bản nhất, một Ethernet switch sẽ tách các host kết nối vào nó theo những cách sau:  
  
Mỗi collision domain sẽ bị giới hạn lại. Trên từng switchport, mỗi collision domain bao gồm chính port của switch đó và bao gồm các thiết bị kết nối vào port switch. Thiết bị kết nối này có thể là một host hoặc có thể là một hub khác.  
  
Các host có thể hoạt động ở chế độ fullduplex bởi vì không có sự cạnh tranh trên đường truyền. Các host có thể truyền và nhận ở cùng một thời điểm.  
  
Băng thông không còn chia sẻ, thay vào đó, mỗi switchport cung cấp một phần băng thông dành riêng trên switch fabric từ port này đến port kia. Các kết nối này luôn biến động.  
  
Lỗi trong các frame sẽ không được truyền. Thay vào đó, các frame nhận đươc trên từng port sẽ được kiểm tra lỗi. Các frame tốt sẽ được tái tạo khi nó tiếp tục được chuyển đi. Cơ chế này còn gọi là store-and-forward.  
  
Bạn có thể giới hạn broadcast traffic đến một mức cho trước.  
Switch có thể hỗ trợ các kiểu lọc traffic thông minh.  
  
Khái niệm transparent bridging (TB)  
  
Một layer 2 switch là một transparent bridge có nhiều cổng, trong đó mỗi switchport là một Ethernet segment, tách biệt với những segment khác. Quá trình đẩy frame đi chỉ dựa hoàn toàn vào địa chỉ MAC chứa bên trong từng frame. Một switch sẽ không chuyển một frame cho đến khi nào nó biết địa chỉ đích của frame.   
  
Toàn bộ quá trình đẩy các Ethernet frame đi trở thành quá trình tìm ra những địa chỉ MAC address nào kết hợp với switchport nào. Một switch phải được chỉ dẫn từong minh các host nằm ở đâu (cấu hình MAC tĩnh) hoặc phải tự học các thông tin này. Nếu cấu hình MAC address tĩnh, quá trình này sẽ nhanh chóng quá tải khi các host thay đổi port.  
  
Để học vị trí của một máy, một switch sẽ lắng nghe các frame đi vào và lưu giữ một bảng các thông tin địa chỉ. Khi một frame đến một switchport, switch sẽ kiểm tra MAC nguồn. Nếu địa chỉ MAC nguồn này chưa có trong bảng MAC, địa chỉ MAC, vị trí port và cả thông tin VLAN sẽ được lưu trong bảng. Như vậy, quá trình học vị trí của một host thì dễ dàng và nhanh chóng.   
  
Các frame đi vào cũng có chứa địa chỉ MAC. Một lần nữa, switch sẽ tìm kiếm địa chỉ này trong bảng MAC với hy vọng tìm thấy cổng ra của switch. Nếu tìm thấy, frame có thể được chuyển đi. Nếu địa chỉ không tìm thấy, switch sẽ phát tán frame ra tất cả các port nằm trong cùng một vlan. Động thái này gọi là unknow unicast flooding. Xem hình bên dưới.   
  
Một switch sẽ liên tục lắng nghe các frame đi vào trên các switchport của nó, học các địa chỉ MAC. Tuy nhiên, quá trình này chỉ được phép chỉ khi STP đã quyết định là một port có ổn định cho quá trình sử dụng bình thường hay không. Thuật toán STP sẽ quan tâm đến việc duy trì một mạng không bị loop, khi mà frame không bị đẩy vào vòng bất tận. Đối với các frame chứa địa chỉ broadcast, frame cũng sẽ bị phát tán.   
  
Dòng chảy của frame trong switch  
  
Phần này sẽ khảo sát tiến trình của một frame khi nó đi qua một L2 switch. Khi một frame đến trên một port, frame sẽ được đặt vào hàng đợi inbound. Mỗi hàng đợi có thể chứa các frame các mức ưu tiên khác nhau. Switchport có thể được hiệu chỉnh sao cho các frame quan trọng được xử lý trước. Chức năng này cho phép các dữ liệu quan trọng không bị loại bỏ khi có nghẽn xảy ra.   
  
Khi các hàng đợi được phục vụ và frame được giải phóng ra khỏi hàng đợi, switch phải xác định không chỉ port đích mà còn phải xác định là có nên đẩy các frame đó không (whether) và bằng cách nào (how). Ba quyết định cơ bản cần phải được thực hiện: một quyết định liên quan đến tìm ra cổng ra, hai quyết định còn lại là tìm ra chính sách để đẩy frame đi. Cả ba quyết định này được thực hiện đồng thời bởi các thành phần phần cứng độc lập của switch. Các thành phần này là:  
  
Bảng L2 forwarding: Địa chỉ đích chứa trong frame sẽ đựoc dùng như là thông số để so sánh vào bảng CAM. Nếu địa chỉ là tìm thấy, cổng ra của switch và thông tin vlan tương ứng sẽ được đọc và sử dụng. Nếu không tìm thấy, frame sẽ được đánh dấu để phát tán.  
Các ACL bảo mật có thể đựoc dùng để lọc các frame theo địa chỉ MAC, kiểu giao thức, thông tin L4. Bảng TCAM sẽ chứa các ACL trong một dạng đã được biên dịch sao cho quyết định forward một frame hay không sẽ được thực hiện chỉ trong một động tác tìm kiếm bảng TCAM.   
  
Các QoS ACL có thể phân loại các frame đi vào theo các thông số QoS hoặc để định hình hay kiểm soát tốc độ của dòng traffic. Bảng TCAM cũng được dùng trong quá trình quyết định này.  
  
Sau khi quá trình tìm kiếm trong bảng CAM hay TCAM đã diễn ra, frame sẽ được đặt bên trong hàng đợI của outbound switchport. Các hàng đợi outbound được xác định bằng các thông số QoS chứa trong frame hay thông số được truyền cùng với frame.   
  
Hoạt động của Multilayer switch  
  
Các Catalyst switch, chẳng hạn như 3560, 4500 và 6500 có thể đẩy các frame dựa trên thông tin L3 và L4 chứa trong gói tin. Tiến trình này gọi là chuyển mạch đa tầng (multilayer switching – MLS). Một cách tự nhiên, tiến trình L2 switch cũng phải được thực hiện vì suy cho cùng, các giao thức lớp cao hơn vẫn phải chứa trong các Ethernet frame.   
  
Các kiểu MLS  
  
Có hai thế hệ MLS: route-caching (thế hệ đầu) và topology-based (thế hệ thứ 2). Hiện nay các dòng switch như 3560, 4500 và 6500 chỉ hỗ trợ thế hệ thứ hai của MLS.   
  
Route-caching: Thế hệ đầu. Kiểu công nghệ này đòi hỏi về mặt phần cứng phải trang bị thêm một route processor RP và một switch engine SE. RP phải xử lý gói tin đầu tiên của một dòng các traffic để tìm ra địa chỉ đích. SE sau đó sẽ lắng nghe cả gói tin đầu tiên và địa chỉ đích cần đích, sau đó tạo ra một đường đi tắt trong cache. SE sau đó sẽ đẩy các gói tin kế tiếp trong cùng một dòng traffic dựa trên thông tin trong cache. Kiểu hoạt động MLS này còn được gọi là netflow LAN Switching, flow-based hoặc “route once, switch many”. Ngay cả khi ngày nay kiểu chuyển mạch này không được dùng trong các Catalyst switch, kỹ thuật này vẫn được dùng để tạo ra các thông tin về dòng lưu lượng và các thông tin thống kê.  
  
Topology-based: Thế hệ thứ hai của MLS sử dụng các phần cứng chuyên dụng. Các thông tin định tuyến lớp 3 sẽ được xây dựng và đưa vào một cơ sở dữ liệu về toàn bộ sơ đồ mạng. Cơ sở dữ liệu này, bản chất sẽ được kèm theo một cơ chế tìm kiếm bằng phần cứng rất hiệu quả, sẽ được tham khảo sao cho các gói tin có thể được đẩy đi ở tốc độ rất cao. Khi có một so trùng dài nhất được tìm thấy (longest match), kết quả này sẽ được dùng. Khi cấu trúc mạng thay đổi, database chứa trong phần cứng này cũng sẽ được cập nhật động trong thời gian rất ngắn. Kiểu MLS này được gọi là Cisco Express Forwarding CEF. Một tiến trình định tuyến chạy trên phần cứng của switch sẽ download bảng định tuyến thông thường vào trong bảng FIB.  
  
Dòng chảy của gói tin trong switch L3  
  
Đường đi mà một gói tin lớp 3 đi vào một MLS thì cũng tương tự như của L2 switch. Rõ ràng, một vài cách thức để xử lý thông tin lớp 3 cần phải được thêm vào. Hình dưới đây mô tả một MLS switch tiêu biểu và các tiến trình quyết định bên trong.   
  
Các gói tin đến một switchport sẽ được đặt trong hàng đợi phù hợp giống như trong L2 switch.   
  
Mỗi gói tin sẽ được lấy ra khỏi hàng đợi và kiểm tra cả thông tin L2 và L3. Ở thời điểm này, switch quyết định đẩy gói tin về đâu sẽ được dựa trên hai bảng địa chỉ. Cũng giống như trong L2 switching, tất cả các quyết định chuyển mạch MLS sẽ được thực hiện đồng thời bằng phần cứng.   
  
Bảng L2 forwarding: Địa chỉ MAC được dùng như một thông số trong bảng CAM. Nếu frame chứa một gói tin L3 cần phải được chuyển đi, địa chỉ MAC đích là địa chỉ MAC của port L3 trên switch. Trong trường hợp này, kết quả của bảng CAM chỉ được dùng để quyết định rằng frame nên được xử lý ở L3.  
  
Bảng L3 forwarding: Khi switch tham khảo đến bảng FIB, địa chỉ đích của gói tin sẽ được dùng. Nếu tìm thấy một hàng trong bảng FIB theo nguyên tắc longest match (trùng cả phần địa chỉ và phần mask), địa chỉ next-hop L3 sẽ được ghi nhận. Bảng FIB cũng chứa địa chỉ L2 MAC và cổng ra của switch sao cho quá trình tìm kiếm trong bảng về sau là không cần thiết.   
  
Các ACL bảo mật sẽ được biên dịch vào thành các hàng của bảng TCAM sao cho các quyết định đẩy gói tin sẽ chỉ cần xác định thông qua một động tác tìm kiếm trong bảng.   
  
Các chức năng QoS khác như phân loại gói tin, định hình và đánh dấu có thể được thực hiện như là một quá trình tìm kiếm duy nhất trong bảng TCAM. Cũng giống như trong L2 switch, cuối cùng, gói tin cũng phải được đặt trong các hàng đợi phù hợp trên cổng ra của switch.   
  
Tuy nhiên cũng cần nhớ rằng trong quá trình chuyển mạch MLS, địa chỉ next-hop sẽ nhận được từ bảng FIB cũng giống như một router nào đó. Sau khi đã có địa chỉ L3, thông thường router/L3 switch sẽ xác định giá trị nexthop và tìm địa chỉ L2 của nó. Sau đó, chỉ có địa chỉ L2 được dùng, sao cho L2 frame có thể được gửi. Tiến trình này chính là quá trình encapsulation. Địa chỉ L2 nexthop phải được đặt vào frame thay cho địa chỉ đích ban đầu (chính là địa chỉ L2 của MLS switch). Địa chỉ nguồn L2 của frame cũng sẽ được đổi lại thành địa chỉ L2 của MLS switch trước khi nó được gửi đến thiết bị nexthop. Nghĩa là, cả địa chỉ nguồn MAC và địa chỉ đích MAC của một frame khi đi qua một L3 switch sẽ phải thay đổi.   
  
Ngoài ra, cũng giống như trong routers, giá trị TTL trong gói tin L3 phải được trừ đi 1. Bởi vì nội dung của gói tin L3 (giá trị TTL) đã thay đổi, giá trị L3 header checksum phải được tính toán lại. Và bởi vì cả nội dung L2 và L3 cũng đã thay đổi, giá trị L2 checksum cũng phải được tính toán lại. Nói cách khác, toàn bộ frame phải được viết lại trước khi nó đi ra hàng đợi bên ngoài. Toàn bộ quá trình này hoàn tất bằng phần cứng.   
  
Các ngoại lệ đối với quá trình MLS  
  
Để đẩy gói tin đi dùng các quyết định đồng thời được mô tả ở trên, gói tin phải là dạng “MLS-ready”. Ví dụ, CEF có thể đẩy gói tin IP đi trực tiếp giữa các host. Điều này diễn ra khi cả địa chỉ nguồn và địa chỉ đích là đã biết và không có một thông số IP nào cần phải thao tác. Các gói tin dạng khác không thể chuyển mạch theo kiểu CEF thì phải được xử lý chi tiết hơn. Các gói tin/ traffic dạng như sau sẽ bị đánh dấu và gửi về CPU của switch để xử lý theo kiểu process switching:  
  
ARP requests and replies.  
IP packets requiring a response from a router (TTL has expired, MTU is exceeded, fragmentation is needed, and so on)  
IP broadcasts that will be relayed as unicast (DHCP requests, IP helper-address functions).  
Routing protocol updates.  
Cisco Discovery Protocol packets.  
IPX routing protocol and service advertisements.  
Packets needing encryption.  
Packets triggering Network Address Translation (NAT)  
Other non-IP and non-IPX protocol packets (AppleTalk, DECnet, and so on)

Các bảng dùng trong quá trình switching  
  
Các Catalyst switch duy trì vài bảng trong quá trình hoạt động. Các bảng này đuợc dùng trong quá trình chuyển mạch L2 hoặc MLS và được lưu trong các bộ nhớ rất nhanh sao cho nhiều trường trong gói tin hay frame có thể được so sánh song song.   
  
Bảng CAM  
  
Tất cả các kiểu Catalyst switch dùng bảng CAM cho quá trình L2 switching. Khi frame đến trên switchport, địa chỉ nguồn MAC sẽ được học và lưu trong bảng CAM. Port đầu vào và thông tin VLAN tương ứng sẽ được học. Nếu một địa chỉ MAC học được trên một switch port sau đó được chuyển sang một port khác, địa chỉ MAC và các nhãn thời gian tương ứng sẽ được lưu lại trên port mới nhất. Sau đó, thông tin cũ trong bảng MAC sẽ bị xóa. Nếu một địa chỉ MAC được nhận ra đã có sẵn trên chính port đó, chỉ có nhãn thời gian (timestamp) là được cập nhật.   
  
Các switch thông thường có bảng CAM lớn sao cho nhiều địa chỉ có thể tìm kiếm. Tuy nhiên, sẽ không có đủ chổ cho tất cả các địa chỉ có thể trên một hệ thống mạng lớn. Để quản lý không gian của bảng CAM, các entry cũ (không được cập nhật) sẽ được xóa ra khỏi bảng CAM. Mặc định, các hàng trong bảng CAM này có thời gian aged-out là 300 giây. Để thay đổi thời gian mặc định này, ta có thể dùng lệnh  
  
Switch(config)# mac address-table aging-time seconds  
  
Mặc định, các địa chỉ MAC có thể được học động khi có frame đi vào. Bạn cũng có thể cấu hình MAC tĩnh. Lúc này, hãy dùng lệnh:  
  
Switch(config)# mac address-table static mac-address vlan vlan-id interface type mod/num

Bảng TCAM  
  
Trong quá trình định tuyến truyền thống, ACL có thể lọc hay kiểm soát traffic. Các ACL có thể được tạo ra bởi một hoặc nhiều đối tượng hoặc các phát biểu match có thể được tính toán theo trình tự. Việc tính toán một ACL có thể tốn thêm thời gian, làm tăng độ trễ của gói tin. Trong MLS, tất cả các tiến trình so sánh của ACL đều hiện thực bằng phần cứng. TCAM cho phép một gói tin được kiểm tra với toàn bộ ACL chỉ thông qua một động tác tìm kiếm đơn giản. Phần lớn các switch có nhiều bảng TCAM để các ACL về bảo mật và QoS có thể được kiểm nghiệm đồng thời và xử lý song song với các quyết định đẩy gói tin ở L2 và L3.   
  
Có hai thành phần trong bảng TCAM:  
  
Feature Manager: sau khi một ACL được tạo ra hoặc cấu hình, FM sẽ biên dịch và trộn các hàng của ACL vào bảng TCAM. Bảng TCAM sau đó sẽ được tham chiếu ở tốc độ chuyển frame.   
Switching Database Manager SDM: bạn có thể chia bảng TCAM trên vài Catalyst switch ra thành các vùng có chức năng khác nhau.   
  
Cấu trúc bảng TCAM:  
  
TCAM là một mở rộng của khái niệm bảng CAM. Hãy nhớ rằng một bảng CAM sẽ dùng một index hoặc một giá trị khóa (thường là địa chỉ MAC)  
  
Các hàng trong bảng TCAM thường bao gồm các giá trị Value, Mask và result. Các trường từ gói frame hoặc gói tin sẽ được nạp vào bảng TCAM, trong đó các trường này sẽ so sánh với các cặp value/match.  
  
Cột giá trị (value) luôn là 134bit, có thể chưa địa chỉ nguồn và địa chỉ đích và các thông tin liên quan khác. Thông tin kết hợp để hình thành nên cột value này phụ thuộc vào kiểu của ACL. Trường mask cũng có chiều dài 134bit. Mask giúp chỉ ra các bit đang quan tâm. Cột kết quả là các con số chỉ ra hành động cần phải thực hiện sau khi bảng TCAM đã được tìm kiếm. Cần lưu ý là so với ACL truyền thống, bảng TCAM cho phép một số result có thể. Ví dụ result có thể là permit/deny hoặc một giá trị index đến một chính sách QoS hoặc một pointer đến giá trị nexthop khác.