ETHERNET SWITCHING

Lorenzo Simionato
lorenzo@simionato.org

Dicembre 2007

Questo documento vuol essere una brevissima introduzione agli switch di rete. Per un approfondimento dettagliato, si rimanda invece ad un testo dedicato all'argomento (si veda la bibliografia) o alle numerose risorse presenti in rete.
Il documento è stato redatto nell'ambito di una presentazione per il corso di “Reti di calcolatori”, all'università Ca' Foscari di Venezia.
Nel testo non si escludono errori od inesattezze, pertanto eventuali segnalazioni (errori, suggerimenti, domande) possono essere inviate a lorenzo@simionato.org.

Lorenzo Simionato
12/2007

Indice

Introduzione

Al giorno d'oggi le reti di calcolatori sono diffusissime e vengono utilizzate quotidianamente da milioni di persone, con scopi ludici, informativi, lavorativi, ecc. Basti pensare all'utilizzo di internet, delle LAN (Local Area Network) aziendali o domestiche, delle reti wireless. Il numero di utenti che utilizzano le reti cresce di giorno in giorno, si ha necessità di banda sempre maggiore, per trasferire contenuti sempre più pesanti e si richiede anche qualità del servizio, per consentire ad esempio l'ottimo funzionamento di applicazioni critiche.

In questo contesto la tecnologia deve fornire reti sempre più veloci, affidabili, con buoni tempi di risposta e moltissime altre caratteristiche. Un ruolo chiave lo ha assunto Ethernet, che è di fatto la tecnologia più utilizzata per le reti LAN (siano esse aziendali o domestiche).

In questa relazione si tratterà brevemente come le reti Ethernet, dal loro albore negli anni settanta, abbiano dovuto evolversi, per far fronte a questa crescente domanda di velocità e qualità del servizio. In particolare si tratterà il ruolo degli switch, che hanno rivoluzionato di molto le reti di soli pochi anni fa. Al di la del semplice aumento di velocità, concetti chiave come la condivisione del mezzo trasmissivo o la gestione delle collisioni hanno assunto significati totalmente nuovi.

Si inizierà facendo un breve riepilogo dei metodi di collegamento delle stazioni ad una rete Ethernet (con i relativi problemi), andando quindi a descrivere il funzionamento e le tipologie di switch. Si tratteranno inoltre quali nuove possibilità e benefici questi dispositivi permettano, non dimenticando però anche i problemi da risolvere e gli aspetti negativi.

Si tenga conte che molte sezioni sono solo introduttive, per esempio la descrizione delle VLAN o del funzionamento dell'algoritmo base per determinare lo spanning tree, potrebbero richiedere decine di pagine. L'obiettivo è invece quello di fornire tante argomentazioni, consentendo al lettore di approfondire gli aspetti che più lo interessano o che non gli sono molto chiari.

Mezzi trasmissivi

Le reti Ethernet sono di tipo wired, ovvero i collegamenti tra le varie stazioni sono realizzati tramite dei cavi. Esistono molte varianti, verranno esaminate in ordine cronologico quelle più significative.

Le prime reti Ethernet utilizzavano un grosso cavo coassiale e per questo erano anche chiamate thick Ethernet. In particolare lo standard 10Base5 definiva una velocità di 10Mbps, in banda base, con segmenti lunghi fino a 500 metri. Le stazioni erano collegate al cavo tramite un cablaggio detto "a vampiro".

Successivamente fecero la loro comparsa le reti 10Base2, che utilizzavano questa volta un cavo coassiale più fino (thin Ethernet), che consentiva un migliore maneggevolezza durante la stesura. Le stazioni venivano collegate tramite delle giunzioni a "T" e dei connettori BNC.

In entrambe queste due tipologie vi era quindi un unico cavo (con dei terminatori ai lati), in cui ogni stazione si collegava nel punto desiderato. Si trattava in sostanza di un mezzo completamente condiviso, in altre parole una tensione posta in qualunque punto del cavo veniva trasmessa su tutto il mezzo e quindi rilevata da tutte le stazioni.

Le cose cambiarono con l'introduzione di 10Base-T, dove si utilizzava (e si utilizza) un cavo con dei fili in rame intrecciati a coppie (twisted pair). Vi era inoltre un concentratore (chiamato hub) che collegava tra loro le stazioni. Ogni stazione aveva un proprio cavo che arrivava fino all'hub, da dove partivano tutti i cavi. Era anche possibile collegare tra loro più hub, per creare una rete più estesa (in termini di lunghezza) o per motivi logistici.

Anche se visivamente quella che appare è una topologia "a stella", in realtà il mezzo è sempre condiviso tra tutte le stazioni in quanto l'hub è un semplice dispositivo che ritrasmette i bit in arriva su una porta a tutte le altre. Le comunicazioni avvengono sempre in maniera Half-Duplex, ovvero si possono avere comunicazioni in entrambe le direzioni, ma non contemporaneamente.

Uno dei problemi da gestire in queste tipologie di reti sono le collisioni. Se due stazioni trasmettono in contemporanea, il segnale si "sommerà" ed in sostanza non sarà più possibile far arrivare i dati in maniera corretta. E' stato quindi elaborato il protocollo CSMA/CD (Carrier Sense Multiple Access Collision Detection), che consente alle stazioni di evitare la trasmissione se qualche altra lo sta già facendo. Nel caso ci sia una collisione sarà necessario bloccare la comunicazione e ricominciarla dopo un tempo casuale, sperano che non si verifichi di nuovo.

Le reti twisted-pair aumentarono di velocità, formulando gli standards 100Base-T e 1000Base-T, rispettivamente a 100 e 1000 Mbps. (ci furono anche altri standards, ma non saranno trattati).

Una importante rivoluzione, che è poi l'argomento centrale di questa relazione, fu l'avvento degli switch. Questi dispositivi hanno la stessa funzione degli hub, ma operano in maniera diversa come si vedrà successivamente. Da 10Base-T a 1000Base-T è possibile utilizzare alternativamente hub o switch, tuttavia gli hub stanno progressivamente scomparendo.

Un ruolo importante lo hanno anche le reti in fibra ottica, dal loro albore a 10Mbps, fino a superare la soglia dei 10Gbps. Esistono switch in grado di collegare tra loro reti twisted-pair con reti in fibra ottica, tuttavia le fibre ottiche non saranno trattate in questa relazione.

Confronto tra switch e hub

Come si è detto nel paragrafo precedente, dal punto di vista visivo una semplice LAN basata su hub o su switch non presenta differenze: tutte le stazioni sono connesse con un proprio cavo al concentratore. La differenza chiave invece è nel funzionamento.

L'hub è infatti anche noto come multiport-repeater, ovvero si tratta di un dispositivo che opera a livello fisico (ci si riferisce al modello ISO/OSI). Un hub non ha idea di cosa sia un pacchetto o un frame, un hub si occupa solo di bit. Quando arrivano dei bit su una delle sue porte, l'hub li ritrasmette a tutte le altre. Un chiaro vantaggio di questo approccio è la bassissima latenza, non c'è infatti alcun tempo di elaborazione. Man mano che i bit arrivano, vengono in automatico ritrasmessi.

Lo switch è invece noto anche come multiport-bridge (in questa relazione si useranno i termini bridge e switch in maniera intercambiabile) e si tratta di un dispositivo di livello data-link. Uno switch non si ferma alla semplice conoscenza dei bit, ma è in grado di decifrare le informazioni di livello 2, che sono i frame.

Come lo switch utilizzi queste informazioni aggiuntive sarà discusso nel capitolo successivo, quello che è importante capire è quali conseguenze porti questo fatto. Il risultato è quello che i dati inviati ad una porta non vengono più inviati a tutte le altre, ma solo a quelle dove si trova il destinatario (non è del tutto vero come sarà chiaro più avanti). Quindi non si può più parlare di mezzo condiviso, inoltre le collisioni vengono molto ridimensionate.

Si supponga ad esempio di avere 4 stazioni numerate da 1 a 4. 1 sta comunicando con 2 e 3 sta comunicando con 4. Nel caso dell'hub si avrebbero molte collisioni, mentre con lo switch se ne avrebbero pochissime. Infatti è come se 1 e 2 fossero connesse tra loro con un cavo mentre 3 e 4 ne utilizzassero un altro. Questo è un evidente vantaggio in quanto riduce drasticamente il dominio di collisione, consentendo quindi velocità più elevate.

Inoltre il throughput complessivo di uno switch è molto più elevato di quello di un hub. Nell'esempio citato prima, supponendo una rete a 100Mbps, si potranno scambiare dati complessivamente a 200Mbps. Infatti 1 e 2 comunicano a 100Mbps e lo stesso faranno 3 e 4. Nel caso di molte stazioni è ancora più evidente il miglioramento rispetto all'hub, dove la banda dovrà essere invece contesa tra tutte le stazioni che vogliono comunicare.

Gli switch permettono inoltre di operare in modalità Full-Duplex, con evidenti vantaggi. Ci sono inoltre delle differenze per quanto riguarda la sicurezza ed i servizi offerti (es. VLAN), tutti questi aspetti saranno discussi nei prossimi capitoli.

Un aspetto negativo degli switch è invece la latenza, che è superiore a quella degli hub. Uno switch deve infatti analizzare alcune informazioni prima di inviare il pacchetto, naturalmente questo provoca un ritardo superiore dell'hub (dove i bit venivano inviati subito).

Funzionamento degli switch

Uno switch riesce ad inviare i dati solo alla porta dove è connesso il destinatario, con gli evidenti vantaggi che sono stati descritti. E' lecito domandarsi come questo possa avvenire.

In primo luogo è necessario conoscere l'indirizzo del destinatario, visto però che gli switch operano a livello data-link sono in grado di estrarre tutte le informazioni presenti in un frame. Ogni frame infatti contiene l'indirizzo MAC del destinatario del messaggio (ci si riferisce naturalmente ad Ethernet), ed è proprio questa l'informazione che interessa.

Questo dato non è però sufficiente, manca un modo per associare ciascun destinatario alla porta dove è collegato. Viene utilizzata un tabella (CAM table) che mantiene queste corrispondenze. Quando arriva un frame viene estratto l'indirizzo MAC del destinatario, cercato nella tabella, ed inviato alla porta indicata.

Il problema abbastanza semplice da intuire è come costruire questa tabella. Un host potrebbe infatti essere connesso allo switch, sebbene non abbia ancora mai comunicato. Inoltre nulla vieta di scollegare una stazione da una porta e connetterla su di un'altra. Naturalmente ci si aspetta che tutto continui a funzionare, che non sia necessario riavviare lo switch o che ci voglia del tempo prima che la rete inizi a funzionare. Si faccia anche attenzione al fatto che è anche possibile collegare ad una porta dello switch un hub. In questo modo si potrebbero avere, ad esempio, 20 host connessi tutti alla porta numero 1.

Appena si accende lo switch, la sua CAM table è vuota, in quando è impossibile sapere dove sia connessa ciascuna stazione. Supponiamo, per esempio, l'arrivo del primo frame di dati dalla stazione con indirizzo A connessa alla porta 1, destinato alla stazione B connessa alla porta 2. Lo switch cerca quindi l'indirizzo B nella sua tabella ma non lo trova (come abbiamo detto la tabella è vuota), quindi invia il frame a tutte le porte (tranne la porta d'arrivo, 1). Contemporaneamente però salva nella tabella il fatto che l'host A è collegato alla porta 1. Nelle prime fasi, quindi, lo switch opererà in maniera del tutto simile ad un hub.

Supponiamo che successivamente B invii un frame ad A (ad esempio l'acknowledgement di quello ricevuto prima), la situazione sarà ora diversa. Lo switch questa volta troverà A nella sua tabella e manderà quindi il frame solamente alla porta numero 1. Inoltre salverà anche l'associazione di B con la porta 2.

In sostanza, con la circolazione di un po' di traffico la tabella sarà progressivamente riempita dagli indirizzi dei mettenti con la relativa porta. In questo modo, dopo un po', si troverà quasi sempre una corrispondenza nella tabella. Il risultato sarà quello atteso, ovvero la comunicazione avverrà sempre tra due sole porte distinte.

Come è stato introdotto prima però, bisogna considerare anche la possibilità che un host sia spostato da una porta all'altra. Quando l'host, dopo lo spostamento, comunicherà di nuovo lo switch aggiornerà l'entry nella tabella con la nuova porta e tutto continuerà a funzionare. Nel caso però in cui, dopo lo spostamento, l'host non comunichi mai (caso raro), devono comunque essere presenti dei meccanismi per rilevare questo spostamento. Quello che si fa è quindi rinfrescare frequentemente la tabella CAM. Ovvero quando una entry non viene aggiornata da molto, la si rimuove dalla tabella. Naturalmente si potrebbe avere anche la perdita di qualche frame durante uno spostamento. Il tempo infatti in cui nella tabella si avrà la vecchia porta sarà pari al minino fra il tempo di “invecchiamento” delle entry e il tempo della prima trasmissione della stazione spostata.

Al lato pratico questi problemi sono di piccola entità, in quanto si hanno sempre delle comunicazioni (es. richieste ARP o acknowledge) tra gli host. Inoltre non è così frequente spostare un host da una porta all'altra e il caso della perdita di qualche frame è comunque gestito dai livelli superiori.

Tipologie di switch

Esistono moltissimi modelli di switch presenti sul mercato, con caratteristiche e funzionalità molto diverse e prezzi per tutte le tasche. Sebbene sia possibile distinguerli per moltissimi fattori, in questo contesto faremo solo due classificazioni: una riguarda i metodi di instradamento, l'altra la configurazione.

Per quanto riguarda la prima classificazione, possiamo distinguere quattro tipologie:

1. **Store and forward**
   Con questo tipo di tecnica lo switch legge l'intero frame in arrivo. Una volta che è stato interamente ricevuto, ne viene verificato il CRC. Solamente se il CRC è corretto, il frame viene inviato alla porta di destinazione. Questa tecnica ha due importanti vantaggi: non si fanno circolare frame corrotti: se un frame non è valido viene scartato immediatamente; è possibile collegare allo switch host con velocità differenti (ad esempio 10 e 100Mbps). Lo svantaggio è invece la maggior latenza, in quando è necessario ricevere tutto il frame prima di inviarlo. E' in genere la tecnica più utilizzata.
2. **Cut Through**
   In questo caso viene letta solamente la prima parte del frame, fino all'indirizzo MAC del destinatario. A questo punto si inizia a spedire il frame alla porta trovata nella tabella, mentre la coda del frame sta ancora arrivando. Il grande vantaggio di questa tecnica è la bassa latenza, come si può facilmente intuire. Non si possono naturalmente connettere host con differenti velocità e i frame corrotti vengono propagati nelle rete.
3. **Fragment Free**
   In questo caso vengono letti i primi 64bytes del frame. Come dice il nome l'obiettivo è quello di eliminare i frammenti in circolo per la rete. Un frammento si genera tipicamente quando si hanno collisioni. Una stazione inizia a trasmettere e si ha quindi la collisione, una parte del frame (il frammento appunto) però è già in circolo per la rete. Questa tecnica cerca quindi di evitare che i frammenti più piccoli di 64bytes non vengano diffusi nella rete. Naturalmente anche in questo caso non si possono collegare host con velocità differenti, la latenza è un compromesso tra i due metodi elencati precedentemente.
4. **Adaptive switching**
   Questa tecnica è presente tipicamente solo sugli switch managed (vedi sotto). L'idea è quella di operare inizialmente in una delle modalità descritte prima e di passare poi ad un'altra a seconda di alcuni eventi. Si potrebbe per esempio iniziare con una tipologia cut through, se si notano però un gran numero di pacchetti con CRC errato, si passa a store and forward.

Per quanto concerne invece la seconda classificazione, abbiamo:

1. **Unmanaged switches**
   Si tratta dei dispositivi più economici, adatti in ambito domestico o piccole azienda (SOHO: Small Office Home Office). Sono periferiche *plug and play*, ovvero è sufficiente connetterle e basta. Non è possibile configurare alcun tipo di parametro.
2. **Managed switches**
   Sono i dispositivi più costosi, sono dotati di una interfaccia web, telnet, o protocolli proprietari che consentono di configurare molti aspetti. E' tipicamente possibile scegliere il metodo di instradamento descritto prima, configurare le VLANs, esaminare le statistiche del traffico e moltissime altre cose. Naturalmente in questo tipo di switch le funzionalità offerte variano da modello a modello.

Sicurezza

Gli switch introducono delle novità anche per quanto concerne, sebbene in lieve misura, aspetti relativi la sicurezza. Come è stato descritto sulle reti in cui il mezzo è completamente condiviso (es. Ethernet con l'hub), i frame vengono inviati a tutte le stazioni. Quando un host riceve un frame controlla se l'indirizzo MAC del destinatario corrisponde al suo e solo in quel caso lo accetta, altrimenti lo scarta. Tuttavia questo avviene in condizioni normali, è però anche possibile configurare la NIC (Network interface card, ovvero la scheda di rete) in una modalità detta promiscua. In questa modalità vengono ricevuti tutti i frame ed è quindi possibile esaminarli con un qualsiasi programma di sniffing.

In questo modo un qualsiasi host connesso ad una LAN è in grado di esaminare completamente tutto il traffico in transito. Sarà quindi necessario porre dei rimedi ai livelli superiori, per esempio utilizzando per i dati sensibili opportuni protocolli cifrati (https per il web, pop3ssl per la posta, ecc).

Con l'uso degli switch la situazione è però molto diversa. Gli host potranno ancora settare la propria NIC in modalità promiscua, ma non ne avranno alcun vantaggio. Come detto infatti lo switch recapiterà loro solamente i frame che li riguardano. Si può dire dunque che gli switch siano in un certo senso più sicuri.

L'impossibilità di “sniffare” la rete, può però portare anche a dei problemi. In alcuni casi infatti la modalità promiscua ha uno scopo ben preciso. Si considerino ad esempio i vari software di NIDS (Network Intrusion Detection System). Questi programmi tentano di rilevare intrusioni nella rete, ma per svolgere il loro compito hanno bisogno di raccogliere tutto il traffico in transito, cosa che con gli switch non risulta più possibile.

Negli switch di fascia alta tuttavia è stato posto rimedio a questo problema. In questi dispositivi esiste infatti il concetto Monitoring Port, Switch Port Analyzer (SPA) o Roving Analysis Port(RAP). Si tratta però in sostanza dello stesso concetto: esiste una porta sulla quale viene copiato tutto il traffico. In questo modo è sufficiente connettere lo sniffer (il NIDS ad esempio) su questa porta e il dispositivo sarà in grado di ricevere tutto il traffico della rete. In alcuni dispositivi esiste proprio una porta speciale per questa funzionalità, su quelli più evoluti è invece tutto configurabile via software. Si può quindi collegare lo sniffer su una qualsiasi porta e configurare quale traffico far arrivare. Con questo sistema è possibile, ad esempio, ricevere il traffico di una particolare VLAN o di particolari porte. Si evita così un enorme traffico e si riescono a monitorare le sole cose che interessano (es. i server).

Come è stato evidenziato sembra non ci sia modo di “sniffare” il traffico per un normale host, che è anche il motivo per cui si ricorre a soluzioni come la monitoring port descritta poco fa, quando questa è una necessità. In realtà esistono delle tecniche per riuscire comunque a portare a termine questo compito.

Una delle più usate è nota come ARP Poisoning (o ARP Spoofing), una soluzione abbastanza semplice che la maggior parte delle volte è efficace. Senza entrare troppo nei dettagli, se ne vedrà in breve il funzionamento. Generalmente un host comunica con un altro utilizzando il protocollo IP (si parla delle reti utilizzate più di frequente), tuttavia gli switch sono periferiche di livello 2 e quindi ignorano cosa sia un IP, comunicano invece con indirizzi MAC. Chi è interessato dunque a comunicare su una LAN non deve limitarsi a conoscere l'indirizzo IP del destinatario, ma deve sapere anche l'indirizzo MAC (si parla sempre di Ethernet). Il protocollo ARP si occupa proprio di mantenere una associazione fra questi due indirizzi. Quando un host vuole comunicare con un altro chiede in broadcast (a tutti) l'indirizzo MAC dell'host con l'indirizzo IP specificato (ARP Request). Il pacchetto è ricevuto da tutti e solo la macchina con l'IP indicato risponderà (inviando una ARP Reply) comunicando il proprio indirizzo MAC. A questo punto il mittente conosce anche l'indirizzo

MAC del destinatario (che viene inserito in una tabella simile alla CAM table descritta prima) e può inviare i frame senza problemi.

ARP spoofing tenta di ingannare direttamente il mittente, facendoli credere che l'indirizzo MAC del destinatario sia quello dell'attaccante. In questo modo il mittente invierà il pacchetto direttamente all'attaccante, che avrà poi cura di rispedirlo al reale destinatario, per consentire la continuazione della comunicazione. Tuttavia l'attaccante potrebbe anche modificare il pacchetto prima di rispedirlo, non limitandosi al semplice sniffing. Si parla in questo caso di attacchi MITM (Man in the Middle).

Per ingannare il mittente, l'attaccante invia una serie di pacchetti di ARP Reply, senza che nessuno abbia fatto alcuna ARP Request. In questi pacchetti indicherà sempre il suo indirizzo MAC, ma specificando gli indirizzi IP dei destinatari di cui intende carpire il traffico in arrivo. Questi pacchetti di ARP Reply saranno ricevuti da tutti, che provvederanno ad aggiornare le proprie tabelle arp, in maniera erronea.

Una soluzione per difendersi da questo problema è quella di utilizzare delle tabella ARP, almeno parzialmente, statiche. In questo modo un host non dovrà fare alcuna ARP request, ma conoscerà direttamente l'indirizzo MAC del suo destinatario, non incorrendo nel problema. Naturalmente con questa soluzione diventa scomodo sostituire schede di rete, host, cambiare indirizzi IP, ecc. Esistono anche altre tecniche per risolvere il problema, che non verranno però trattate. Quello che è importante aver appreso è il fatto che, in linea teorica, lo switch è più sicuro di un hub. Tuttavia non bisogna assolutamente pensare che il traffico in una LAN switched sia sempre ricevuto dal mittente. Pertanto tutte le precauzioni da prendere per i dati sensibili (in particolare la cifratura), sono ancora valide.

Simplex, Half-duplex e Full-Duplex

E' utile a questo punto fare un breve riepilogo delle tipologie dei canali di comunicazione.

Si parla di canali di tipo Simplex, quando il mezzo trasmissivo consente di trasmettere i dati in una sola direzione. Per esempio in una fibra ottica i dati viaggiano in un solo verso. Canali simplex sono utilizzati in situazioni molto particolari oppure, nel caso ad esempio delle fibre, ne vengono utilizzati due. In questo modo si ottiene una vera e propria trasmissione full-duplex (vedere sotto).

I canali di tipo Half-duplex invece consentono la trasmissione in entrambe le direzioni, ma non contemporaneamente. Un tipico esempio sono le reti 10Base5, 10Base2 o le reti basate su hub. In questi contesti, come si è descritto nei precedenti capitoli, è spesso necessario un meccanismo di rilevamento delle collisioni.

I canali di tipo Full-duplex, infine, consentono la comunicazione in entrambe le direzione, anche contemporaneamente. Questa è naturalmente la situazione desiderabile, sia in termini di collisioni, che di performance.

Praticamente tutti gli switch in commercio supportano questa ultima modalità. Quando si utilizzano dunque solo comunicazioni full-duplex le collisioni non ci sono più. Sebbene possa sembrare sorprendente, non c'è poi molto di strano. Ogni host infatti idealmente comunica in maniera diretta con un altro e la comunicazione viaggia in entrambe le direzioni. Pertanto non è possibile alcun tipo di collisione, cioè significa anche che il protocollo CSMA/CD non ha più senso e non viene utilizzato.

I vantaggi in termini di performance sono evidenti. Tanto per cominciare il tempo di contesa del mezzo non c'è più, non si deve poi ritrasmettere alcun frame (tranne in caso di errori). La velocità di trasmissione è dunque limitata semplicemente dalla capacità degli host e dal mezzo trasmissivo. Inoltre se si considera, ad esempio, una rete a 100Mbps essendo la comunicazione full-duplex la banda completamente utilizzabile sale a 200Mbps. Anche tutte le considerazioni riguardanti le distanze massime dei collegamenti vanno riviste. Infatti non ci sono più problemi riguardanti il tempo necessario perché tutti possono “sentire” la collisione. I limiti di distanza saranno dunque unicamente dettati dai problemi dell'attenuazione del segnale.

Tutti questi motivi hanno determinato un grande successo dei link full-duplex e anche la diffusione degli switch.

Controllo di flusso

Si consideri ora una moderna rete LAN Ethernet in cui le stazioni sono collegate da switch e i collegamenti siano di tipo full-duplex. Come si è accennato nel capitolo precedente, non ci sono più collisioni. Pertanto la velocità di invio del mittente non sarà più ostacolata dal esse, ma l'host potrà limitarsi ad inviare i frame il più veloce possibile. Se si pensa poi ad una Ethernet gigabit, il rischio di "inondare" il destinatario di frame, riempiendo il suo buffer in queste condizioni è davvero molto alto. Naturalmente esistono varie soluzioni per problemi relativi al controllo di flusso, ad esempio a livello trasporto. In questa sede però esamineremo la soluzione adottata da Ethernet.

Nel caso di collegamenti full-duplex, è stato definito lo standard IEEE 802.3x, che è obbligatorio per gigabit Ethernet, mentre facoltativo per le velocità più basse. L'idea è quella di inviare speciali frame chiamati PAUSE. Quando il destinatario rileva che sta ricevendo troppi frame (es. il suo buffer è in esaurimento) invia un frame PAUSE specificando il tempo per il quale non intende ricevere altri dati.

Chi sta spedendo è quindi tenuto a sospendere la trasmissione per il tempo indicato. Il destinatario potrà anche inviare altri frame PAUSE, allo scopo di allungare o diminuire il tempo di attesa. Specificando un tempo di lunghezza 0, sarà anche possibile far riprendere subito la trasmissione.

Un problema di questo approccio è il fatto che i frame PAUSE vengono spediti allo speciale indirizzo MAC 01:80:C2:00:00:01. Quindi una stazione non può chiedere una sospensione solamente ad un particolare mittente, ma devo farlo per tutti. Questo fatto crea importanti implicazioni riguardanti questo tipo di controllo di flusso negli switch.

Praticamente tutti gli switch supportano 802.3x, lo standard però specifica che è necessario sospendere la trasmissione alla ricezione dei frame PAUSE, ma non obbliga assolutamente a mandare frame PAUSE. Il risultato è che la maggior parte degli switch, pur rispettando lo standard, inviano frame PAUSE solo in situazioni critiche o addirittura mai. Un esempio di situazione critica potrebbe essere quella in cui tutte le porte e tutti i buffer siano quasi saturi.

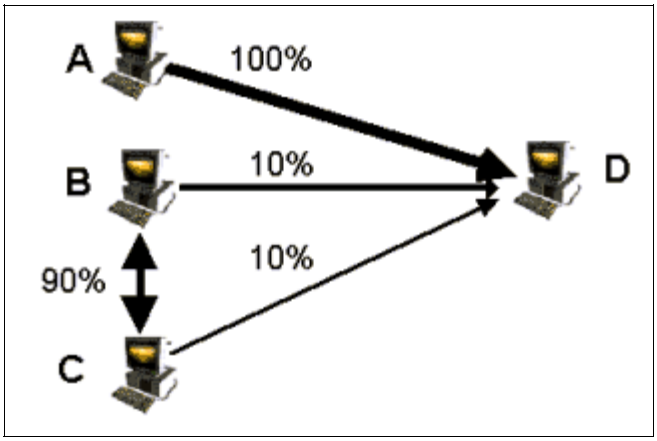

Figura 1: External Head of Line Blocking example

Si consideri invece l'esempio riportato in Figura 1, chiamato spesso *external head of line blocking*. Si supponga che le 4 stazioni siano tutte collegate a differenti porte dello stesso switch. Come si vede D dovrebbe ricevere dati per una capacità del collegamento del 120% (es. 120Mbps su Fast Ethernet). Chiaramente questo non è possibile, si potrebbe dunque pensare che lo switch debba inviare dei frame PAUSE agli host A,B,C. Purtroppo come è stato detto prima i frame PAUSE hanno l'effetto di sospendere tutte le trasmissioni verso l'host che li manda. Si avrebbe dunque un pensate rallentamento fra la comunicazione fra B e C, cosa assolutamente non voluta.

Inoltre quando una linea o una risorsa è sovraccarica, bisogna operare delle scelte su quali pacchetti scartare e quali no. Evidentemente, utilizzando i frame PAUSE, non si potrà fare alcuna scelta di questo tipo, non arrivando più dati. Questo è particolarmente importante per gestire bene il QoS (Quality of Service), che deve consentire il passaggio dei pacchetti con priorità maggiore anche in caso di problemi, invece di rallentare tutto il traffico indistintamente.

Il risultato dunque è che in molti switch non si utilizzano di fatto i frame PAUSE (si sospende però la trasmissione al loro arrivo). Tuttavia nel caso del collegamento diretto di due host (tramite cavo cross-over) 802.3x funziona molto bene.

Spanning Tree

Quando si inizia ad avere a che fare con reti più estese, l'uso di un solo switch può essere molto limitativo. Nel caso di stazioni molto lontane, si dovrebbero stendere lunghi cavi (senza contare le distanze massime di 100m) che di fatto farebbero strade simili, con gli evidenti svantaggi in termini di costi, cablaggio, ecc.

In molte situazioni è quindi utile collegare tra loro più switch, mantenendo comunque la stessa rete (niente routing o indirizzi IP di classi differenti). Naturalmente però se un cavo che collega tra loro due switch si rompe oppure le porte dove è connesso hanno dei problemi, la comunicazione tra host sui differenti switch viene a mancare. In alcune situazioni è dunque utile avere dei link ridondanti.

Dei collegamenti ridondati creano però dei cicli all'interno della rete. I cicli sono molto pericolosi e vanno assolutamente evitati. In primo luogo si potrebbero avere dei frame duplicati, in quanto un frame potrebbe seguire più strade che poi portino alla stessa destinazione. Situazione ben peggiore è invece quella del broadcast storm. Se viene inviato un frame in modalità broadcast, tutti gli switch lo invieranno a tutte le loro porte. Quello che succede è che il frame inizierà letteralmente a "girare" nei cicli all'infinito, arrivando senza troppi problemi a saturare l'intera rete.

Per risolvere tutti questi problemi è necessario che non ci siano cicli e che tutti gli switch siano raggiungibili, bisogna quindi creare un albero disabilitando alcuni collegamenti ridondanti. L'algoritmo STP(Spanning Tree Protocol) definito in IEEE 802.1D si occupa proprio di risolvere questo problema. Segue una breve spiegazione del suo funzionamento, senza entrare troppo nei dettagli. Si tenga presente che lo stesso algoritmo deve essere eseguito su tutti gli switch per funzionare.

Per prima cosa viene designato uno switch come root, ovvero radice dello spanning tree. Per realizzare questa operazione vengono scambiati dei particolari frame detti BPDU (Bride Protocol Data Unit). Questi frame contengono varie informazioni: il root bridge corrente, il costo del cammino, il bridge che ha generato il BPDU, ecc. Il root bridge viene identificato da un indirizzo formato da priorità e indirizzo MAC.

Ogni switch genera un BPDU contenente il proprio identificatore (indirizzo MAC e priorità) come root bridge e lo inoltra a tutte le porte. Quando un altro switch riceve un BPDU confronta l'identificatore contenuto nel frame con il proprio. Se identificatore BPDU è minore del proprio allora lui non è root bridge quindi si limita ad inviare in broadcast le configurazioni che riceve. Se invece è maggiore allora il root bridge potrebbe essere lui, quindi continua ad inviare i proprio pacchetti BPDU e non diffonde invece quello che ha ricevuto.

Quello che succede dopo un po' di tempo è che tutti gli switch sanno chi è il root switch. Come si dovrebbe aver intuito sarà quello con priorità minore. A parità di priorità, quello con indirizzo MAC più piccolo. Una volta eletto il root bridge, ogni switch calcola il percorso più breve tra se e root. La porta che "passa" per quel percorso prende il nome di *root port* e funzionerà come in uno switch tradizionale. Le altre porte invece vanno in strato *blocked*, ovvero sono solo in grado di ricevere BPDU ma per il resto appaiono disabilitate (in realtà esistono anche le porte *designated*, ma non verranno trattate). Nel caso della caduta di qualche link sarà necessario ricominciare l'algoritmo, quindi è necessario che anche le porte *blocked* possano ricevere BPDU.

A questo punto si è praticamente generato un albero, quindi si è raggiunto lo scopo prefisso e non si hanno più problemi con i cicli. Nel caso di modifiche topologiche (es. caduta di collegamenti), lo spanning tree viene ricalcolato.

Esistono poi varie evoluzioni di questo algoritmo, come il Rapid Spanning Tree Protocol oppure il Multiple Spanning Tree Protocol. Tuttavia questi algoritmi non saranno trattati.

VLAN

Con l'ingrandimento delle reti locali è utile in vari casi segmentare la rete in più parti. Uno dei motivi potrebbe essere la sicurezza, per esempio si intende tenere separati due diversi reparti in maniera netta. Un altro motivo potrebbe essere legato alla dimensione della rete. Infatti in una rete switched si ha un solo dominio di broadcast, cosa che implica la trasmissione a tutti dei pacchetti broadcast, cosa che diventa costosa se ci sono tanti host. Tuttavia non si vuole nemmeno stendere nuovi cavi nell'edificio, non si vogliono comprare nuovi apparati (servirebbe almeno uno switch per ogni parte separata della rete), si vuole permettere agli utenti di spostarsi magari da una postazione all'altra senza troppi problemi.

Per tutti questi e molti altri motivi nascono le Virtrual LAN (VLAN). Come suggerisce il nome si tratta di LAN "virtuali", nel senso che non sono separate e costruite in maniera fisica. Si occupa infatti di tutto lo switch (tipicamente di fascia alta) tramite una interfaccia di configurazione. Dal punto di vista fisico infatti le VLAN non si distinguono dalle LAN tradizionali. Tutti gli host sono infatti collegati allo switch in maniera tradizionale.

Nella configurazione dello switch si dovranno definire le varie VLAN che si vogliono creare e dire da quali elementi saranno formate. La configurazione tipica è quella per porte. Si decide che un insieme di porte farà parte della VLAN 1, un altro insieme della VLAN 2, e così via.

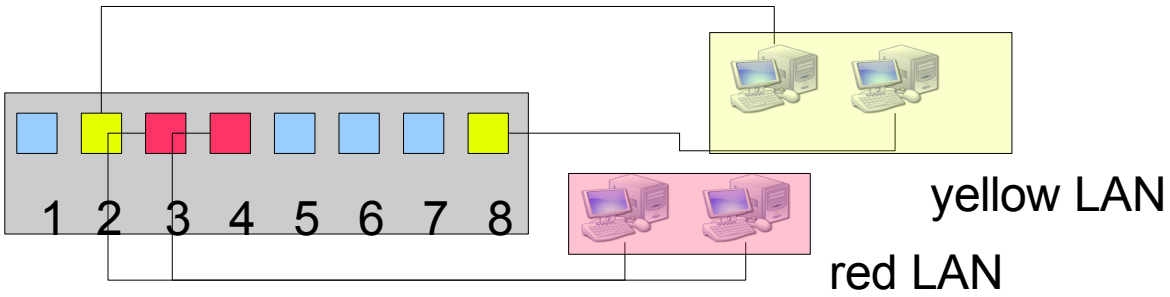

Figura 2: Esempio di VLAN

Nell'esempio riportato in Figura 2 sono state create 3 VLAN: azzurra, gialla e rossa. Come si vede la VLAN gialla utilizza le porte 2 e 8, quella rossa 3 e 4 e quella azzurra le restanti. Sarebbe anche possibile associare ad una porta più VLAN, ma non tratteremo questo aspetto.

VLAN diverse si comportano proprio come se si trattasse di switch differenti. Il dominio di broadcast è limitato a ciascuna, bisogna occuparsi del problema del routing dei pacchetti da una VLAN all'altra. Si potrebbe utilizzare un router o far gestire il tutto dallo stesso switch Layer 3, tuttavia non esamineremo in dettaglio questo problema.

Quando la rete aumenta di dimensioni come è stato descritto precedentemente, si possono collegare tra loro più switch. Nel caso di VLAN però la soluzione si complica, si parla in generale di Inter-Switching VLAN. Il problema è che una stessa VLAN potrebbe estendersi tra due switch, a questo punto però uno switch che riceva i frame dall'altro, non può sapere di che VLAN ciascun frame faccia parte. Per risolvere il problema si possono utilizzare soluzioni (proprietarie) in cui i frame sono incapsulate in altri frame che aggiungano le opportune informazioni. Trattandosi di soluzione proprietarie però si potrebbero avere dei problemi con dispositivi diversi, inoltre ciascun frame in transito tra due switch va incapsulato opportunamente (maggior latenza).

Per questi motivi è stato sviluppato lo standard IEEE 802.1Q, implementato da praticamente tutti gli switch di fascia alta che supportano le VLAN.

Sebbene possa sembrare strano, con questo standard si è deciso di apporre delle modifiche al frame

Ethernet. Viene infatti aggiunto un nuovo campo chiamato VLAN ID e anche un campo di priorità e altri dettagli. L'idea è quindi quella del tagging diretto dei frame, ovvero associare ad ogni frame Ethernet un identificativo della VLAN. Con queste informazioni si risolve il problema in maniera abbastanza semplice. Quando il frame arriva allo switch, basterà che esso estragga il VLAN ID per sapere a quali porte dovrà inoltrarlo.

Naturalmente a prima vista una modifica del frame Ethernet fa sorgere molte perplessità, non è chiaro infatti come si comporteranno le NIC un po' datate. Sembra infatti che un aspetto importantissimo come quello della retro-compatibilità sia stato violato per consentire alle VLAN di funzionare.

In realtà si è tenuto conto di questi problemi, ed infatti gli switch distinguono tra due differenti tipi di porte:

- Access Port: E' una porta tradizionale, su cui andranno connessi gli host. In questa porta i frame Ethernet sono quelli tradizionali, che non contengono il campo VLAN ID e le altre modifiche.
- Trunk Port: Questa porta invece utilizza i frame del nuovo standard (con il campo VLAN ID). Si utilizza per connettere un altro switch o agli host che supportano 802.1Q.

L'idea quindi è quella che l'host tradizionale, senza alcuna modifica, dialoghi con lo swicth (chiamiamolo A) tramite una access port. Se il frame dovrà passare ad un altro switch (chiamiamolo B), allora A creerà un nuovo frame con VLAN ID secondo lo standard e lo farà passare per la trunk port che è collegata con B. A questo punto B saprà a che porte inoltrare il frame (leggendo VLAN ID). Naturalmente se lo dovrà inviare a delle access port, manderà un frame tradizionale, rimuovendo VLAN ID e le altre informazioni.

In questo modo si risolvono tutti i problemi, a condizione che gli switch supportino lo standard, ma d'altronde è una condizione indispensabile.

Prima di concludere questo capitolo, si fa notare che sono possibili altre modalità per caratterizzare una VLAN, oltre a quella basata sulle porte che è stata descritta. Se ne elencheranno altre due che sono utilizzate in alcune situazioni:

- Associazione tramite indirizzo MAC:

  Si configurano sullo switch gli indirizzi MAC di ogni host della VLAN, in questo modo sarà possibile connettere le stazioni su qualunque porta. Lo svantaggio principale è che sostituendo ad esempio la NIC dell'host sarà necessario riconfigurare la VLAN. Inoltre bisogna porre attenzione ai problemi di sicurezza, è possibile infatti modificare l'indirizzo MAC.
- Associazione tramite indirizzo IP:

  Si configurano sullo switch gli indirizzi IP di ogni host della VLAN. Ci si potrà ancora collegare su qualunque porta, tuttavia eventuali sostituzioni della NIC o dell'host non richiederanno riconfigurazioni. Va posta sempre attenzione alla sicurezza. La cosa che invece dovrebbe saltare all'occhio è che IP è un protocollo di livello 3, mentre gli switch sono dispositivi di livello 2! Una spiegazione più approfondita di questo fatto è trattata nel prossimo capitolo.

Switch Layer 3 e oltre

E' stato messo in luce più volte nei capitoli precedenti il fatto che gli switch siano periferiche di livello 2, con le dovute conseguenze di questo fatto (per esempio la questione che essi non sappiano cosa sia un indirizzo IP). Tuttavia, specie negli ultimi anni, il ruolo degli switch sta progressivamente cambiando e si vedono sul mercato sempre di più dispositivi che non si limitano ad operare a livello 2. Si parla in particolare di switch layer 3. Questi dispositivi sono a tutti gli effetti dei router, in quanto sono in grado di instradare pacchetti formulando opportune scelte ed utilizzando vari protocolli di routing. Inoltre uno switch Layer 3 è tipicamente più veloce di un router in quanto le funzioni di inoltro (look-up nelle tabelle) sono realizzate in hardware, diversamente da molti router che lo fanno via software.

Gli switch layer 3 sono molto utilizzati per effettuare il VLAN routing, descritto nel capitolo precedente. Supponendo infatti di creare diverse VLAN appartenenti a reti IP diverse, sarà necessario un routing tra le stesse. Uno switch layer 3 può quindi assolvere a questo compito, in maniera del tutto trasparente. Un amministratore di rete “distratto” infatti non dovrà fare quasi nulla, collegherà semplicemente gli host allo switch ignorando probabilmente il fatto che venga effettuato del routing tra VLAN diverse. Inoltre in questo modo si avrà un sistema centralizzato, occupandosi di tutto lo switch. Questa situazione è una delle applicazioni principali degli switch layer 3.

Esistono però anche switch che non si fermano a livello 3, si parla di switch layer 4. Questi dispositivi sono quindi in grado di esaminare anche i dati dei protocolli TCP e UDP comunemente utilizzati in internet. E' quindi possibile effettuare operazioni di filtraggio molto avanzate, si potrebbe per esempio bloccare tutto il traffico TCP sulla porta 3000. In situazioni tipiche (dipende molto dal modello) è possibile anche dare priorità a determinato traffico, per esempio utilizzando più code a priorità diverse. Si potrebbe avere una coda per il traffico normale, ed un'altra per pacchetti UDP su una particolare porta (per esempio utilizzata dalle applicazioni VOIP).

Infine in situazioni molto particolari, fanno la loro comparsa anche switch layer 7, anche se forse la parola switch non è più appropriata in questo caso. Si tratta infatti di dispositivi in grado di analizzare dati fino a livello applicazione. In questo caso la possibilità delle operazioni effettuabili cresce enormemente. Un utilizzo tipico è quello dei load-balancer per i siti web. Se si hanno infatti molti utenti che accedono ad un sito web, potrebbe essere necessario utilizzare più server. Serve però un sistema per instradare le varie connessioni in maniera bilanciata ai vari server. Uno switch layer 7 potrebbe essere in grado di compiere operazioni di questo tipo.

In alcuni casi, nella comunicazione HTTP, si ha però la necessità che un client comunichi sempre con il server web con in quale ha iniziato la comunicazione. E' il caso tipico di una procedura di login o dei carrelli della spesa utilizzati nei siti di e-commerce. In queste situazioni lo switch devi quindi riconoscere l'utente e trasmettere il suo traffico sempre allo stesso server. Alcune soluzioni vanno quindi addirittura a catturare informazioni presenti nei cookie, in maniera da mantenere la “sessione” creata dall'applicazione web. La questione quindi è molto ampia e gli scenari possibili sono davvero moltissimi, tuttavia non è lo scopo di questa relazioni analizzarli.

Sebbene quanto visto possa sembrare molto interessante, è bene prenderlo con le dovute cautele. Innanzitutto si tratta di un modifica totale al ruolo degli switch, come è stato accennato infatti parlare ancora di switch in questi contesti non ha quasi più senso. Inoltre le soluzioni layer 4 e layer 7 non sono assolutamente standard, inoltre si tratta per lo più di implementazioni proprietarie. L'applicazione di queste tipologie di apparati, inoltre, ha senso solo in realtà molto grandi dove si hanno decine e decina di migliaia di utenti. Si tratta infine di dispositivi di costo molto elevato, dove le soluzioni proposte cambiano di giorno in giorno e da produttore a produttore, anche in relazione al servizio che si vuole ottenere.

Bibliografia

- Andrew S. Tanenbaum. *Reti di calcolatori*, 2003, Pearson Education
- Cisco Networking Academy, CCNA online course materials (cisco.netacad.net)
- http://support.intel.com/support/express/switches/sb/cs-014410.htm
- http://www.cisco.com/warp/public/473/41.html
- http://www.networkworld.com/netresources/0913flow2.html
- IEEE 802 Part 3 standard, www.ieee.org
- Slides del libro *Switched LAN*, M.Baldi, P.Nicoletti, 2002, McGraw-Hill