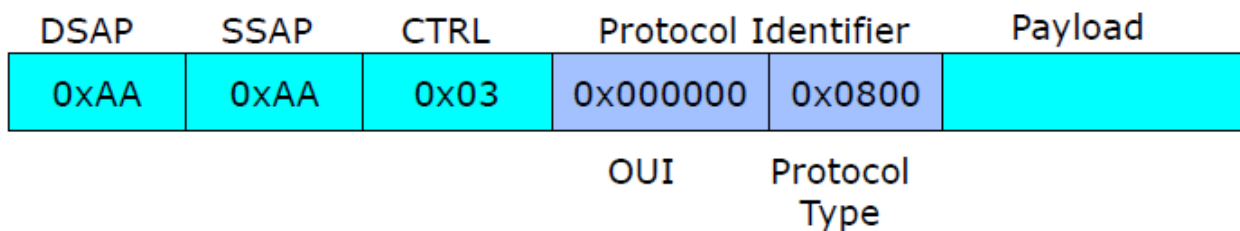


LLC SNAP



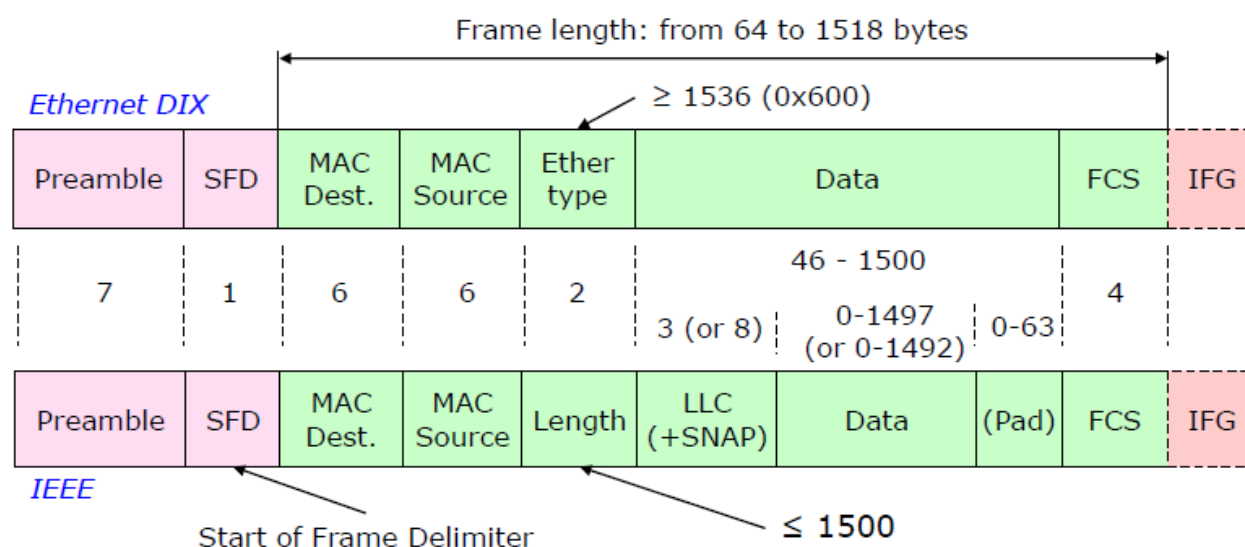
DSAP e SSAP: Protocollo destinazione e sorgente

CTRL: Solitamente 0x03

OUI: Assegnato da IEEE, equivalente ai 3 byte alti del MAC

Protocol Type: Campo più importante, individua il protocollo imbustato.

Ethernet DIX(2.0) e IEEE 802.3



L'IFG nella trama IEEE non è necessario per la presenza del campo length, ma è presente per compatibilità con Ethernet DIX.

Lunghezza minima è 64 byte, massima 1518 (in entrambi i casi bisogna aggiungere Preambolo, SFD, IFG).

Massima distanza teorica di circa 5 Km, in pratica 2 Km.

$$\frac{F_{min} - 1bit}{B} = 2 \cdot \frac{D_{max}}{S_{sig}} \rightarrow D_{max} = \frac{(F_{min} - 1bit) \cdot S_{sig}}{2 \cdot B}$$

Bit time 0.1us. Slot time (tempo per inviare una trama di dimensione minima) di 51.2us.

Fast Ethernet (IEEE 802.3u)

Stessa trama, stesso algoritmo CSMA/CD, ma 10 volte più veloce. Il tempo di bit passa da 0.1us a 0.01 us. Cambia la codifica dei bit (da Manchester a MLT-3).

Massima distanza teorica di 500m, in pratica 200m. Bit time 0.01us, slot time 5.12us, IFG 0.96us.

Viene introdotto il Full-Duplex.

Caratteristiche avanzate delle reti Ethernet

Auto negoziazione

È un meccanismo che permette i due dispositivi ai capi di un link di mettersi d'accordo sulla velocità di trasmissione, sulla modalità Half o Full duplex ed è possibile soltanto se si è connessi ad un host o ad un bridge o switch (non un hub, che lavora a velocità fissa).

Se durante tale procedura la controparte non risponde, la stazione assume essere direttamente connessa ad un hub. Si può fissare la velocità del collegamento anche da management, ma bisogna fare attenzione perché ciò può causare gravi problemi.

Assumiamo che da un lato abbiamo uno switch configurato manualmente in modalità 100Mbps Full-Duplex: ciò fa sì che tale stazione non prenderà parte alle eventuali negoziazioni, essendo già stato configurato.

Un altro switch connesso al primo comincia la procedura di auto negoziazione, e non ricevendo risposta assume essere connesso ad un hub ed imposta la sua modalità in Half Duplex.

Quando questi riceve delle trame in Full Duplex sul link "doppio", poiché egli è in Half Duplex, assume ci sia stata una collisione e scarta la trama, anche se è in grado di ricevere trame in Full-Duplex.

Bisogna quindi prestare attenzione nella configurazione manuale della velocità e della modalità operativa degli apparati poiché uno sbaglio può portare a molte false collisioni.

Dimensione massima frame (oltre 1518 byte)

- 802.1Q aumenta la trama di 4 byte per il tagging VLAN (Baby Jumbo).
- 802.1ad (802QinQ, VLAN Tag Stacking, Provider Bridge) (Baby Jumbo).
- 802.3as propone come nuova dimensione 2000 byte.
- FCoE ha una MTU di 2500 byte (Mini Jumbo).
- MPLS aumenta la dimensione fino a $1518 + (n * 4)$, con n il numero di label.
- Jumbo (Giant, Giant Frames) fino a 9KB, non-standard.

802.3af – 802.3at (Power Over Ethernet, PoE)

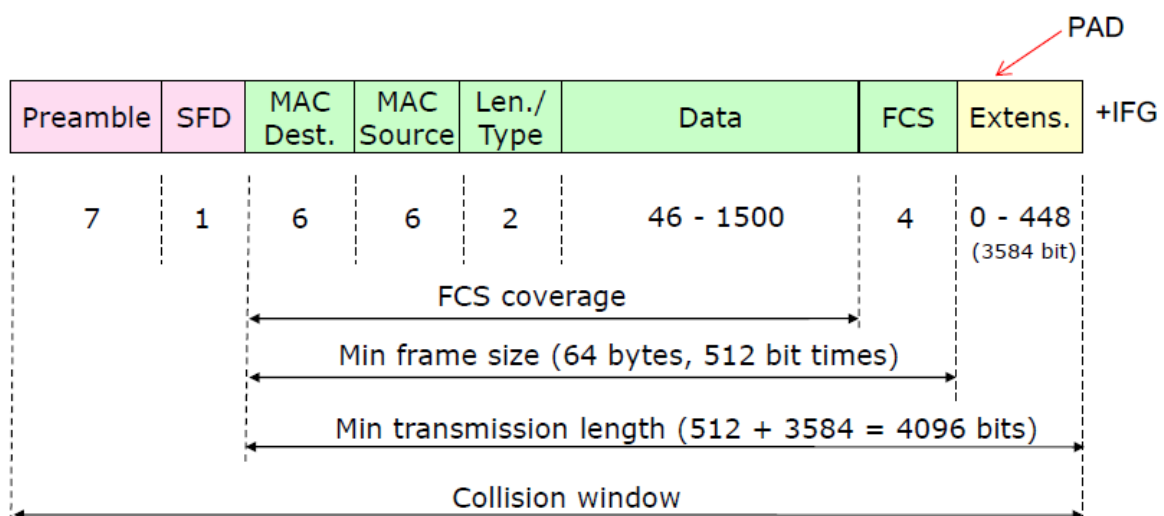
Distribuzione di corrente elettrica sul cavo Ethernet. Adatto per dispositivi con moderate necessità energetica (telefoni VoIP, access point Wi-Fi, videocamere di sorveglianza, ...).

Evita l'installazione di ulteriori cavi ed è compatibile con stazioni non PoE, ma si ha un maggiore consumo energetico da parte degli switch.

Gigabit Ethernet (IEEE 802.3z)

Stessa trama per mantenere la compatibilità, stesso algoritmo CSMA/CD, anziché se non implementato. Funziona soltanto in Full-Duplex. Introduce un aumento dello slot time.

Non si può aumentare la lunghezza della trama minima per motivi di compatibilità, quindi viene aumentata la lunghezza minima della trasmissione che è di 512 bytes. Poiché la trama minima è sempre di 64 bytes, uso il padding.



I frame più grandi di 1512 bytes non vengono estesi. Questa tecnica è detta Carrier Extension.

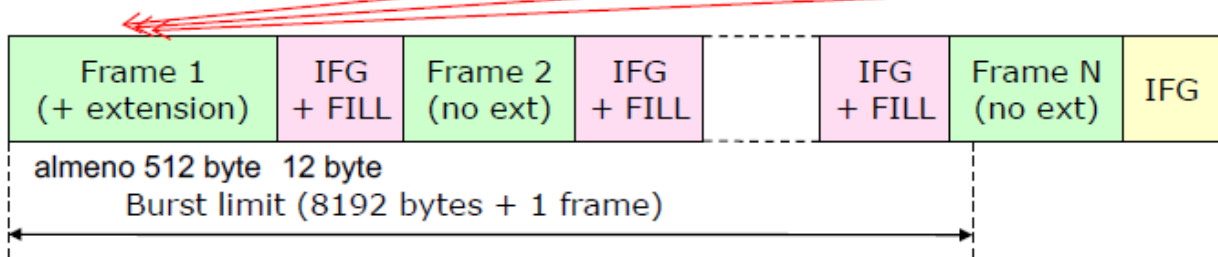
La dimensione massima della trama è ancora 1536 bytes.

Viene introdotto anche il Frame Bursting, cioè l'invio in successione di diverse trame una dopo l'altra.

Il limite massimo è detto Burst-limit ed equivale a 8192 bytes + 1 frame.

Ad ogni trama inviata vedo se ho raggiunto il burst limit, in caso negativo invio un'altra trama.

■ All frames include SFD, Preamble and the actual frame



Ho 2 IFG diversi, ma entrambi di 12 byte:

- IFG+FILL: invio un'altra trama senza rilasciare il canale.
- IFG: fine trama complessiva, rilascio il canale. IFG normale.

10 Gigabit Ethernet (IEEE 802.3ae)

Stessa trama di 802.3, Full-Duplex (no repeater, no CSMA/CD, no Carrier Extension). Adatto per Backbone, DataCenter, MAN.

Dispositivi LAN

- L1: Repeater (Hub multiporta)
Separano i domini fisici, stesso dominio di collisione. Ricevono e propagano sequenze di bit e possono essere usati per connettere due segmenti della stessa LAN Ethernet con livelli fisici diversi (es. cavo coassiale e fibra ottica). Il segnale ritrasmesso viene amplificato se necessario. È un dispositivo attivo.
- L2: Bridge/Switch
Separano i domini di collisione, stesso dominio broadcast. Il dominio di collisione (LAN fisica) è quella porzione di rete in cui opera un'istanza di CSMA/CD. Il dominio broadcast (LAN logica) è quella porzione di rete in cui i frame possono essere direttamente propagati e può includere diversi domini di collisione. Vengono divisi i domini di collisione e viene esteso il dominio broadcast.
Permettono l'interconnessione a livello Data-Link (tra Ethernet e WiFi, Ethernet e Fast Ethernet)
- L3: Router (L3 Switch)
Separano domini broadcast.

Full Duplex

Indica la possibilità per una NIC di trasmettere e ricevere data contemporaneamente. Le due ragioni principali che hanno portato alla nascita e allo sviluppo del Full Duplex sono la modalità operativa "store-and-forward" dei bridge e le nuove tecnologie (il Full Duplex non è disponibile per Hub e Repeater e si può avere solo in link punto-punto) in ambito del cablaggio.

Vantaggi:

- Raddoppio teorico del throughput.
- Vantaggio minimo per client e server (i primi sfruttano solo il downlink, gli altri l'uplink).
- Grosso vantaggio per bridge/switch di backbone, favorendo maggiore simmetria nell'ampiezza di banda.
- Il vantaggio principale è però che il CSMA/CD non è più necessario, dato che le collisioni non sono più possibili. Non c'è più bisogno di mantenere una trama minima, mantenuta invece per compatibilità, ed il diametro massimo della rete dipende soltanto dall'attenuazione del segnale sul cavo.

Attacchi ai Filtering Database

- MAC Flooding Attack: generazione di frame con indirizzi MAC sorgenti casuali.
Il filtering database si riempiono, cominciando ad inoltrare tutte le trame in flooding, per cui funzioneranno come degli Hub, cosicché le altre stazioni possono sniffare il traffico di rete.
Viene rallentata anche la rete. Si può controllare limitando il numero di indirizzi MAC nel FD che si possono associare ad ogni porta.
- Packet Storm: generazione di frame con MAC destinazione inesistenti.
Ciò porta gli switch ad inoltrare tutti i frame in flooding su tutta la rete rallentandola.

QoS nelle LAN

Alcuni scenari di congestione nelle reti locali:

- Congestione quando ho N link in ingresso ed uno in uscita, tutti quanti della stessa velocità (micro-congestioni se agli N link sono connessi degli host il cui traffico è di tipo bursty).
- Congestioni temporanee nei client se connessi con un host più veloce (più veloce in termini di velocità del link, CPU, ...). Poco valido al giorno d'oggi.
- Congestione permanente sui server edge (es. link in ingresso ad uno switch a 10Gb, ma collegato al server con un collegamento ad 1Gb).

Le conseguenze sono aumento del jitter (variazione del tempo di inter-arrivo dei pacchetti), non adatto per applicazioni real-time (VoIP, streaming, videoconferenza) e storage (file system distribuiti).

Il buffer delle NIC è piccolo e non può tamponare il problema, TCP va in timeout e dimezza la finestra di trasmissione, e il throughput cala. Aggiornare soltanto la rete d'accesso e non il backbone porta ad un declino delle prestazioni.

IEEE 802.1p

Non risolve il problema delle congestioni, ma lo limita.

È basato su 8 livelli di priorità, o meglio di 8 diverse classi di servizio, in quanto lo standard non indica una relazione di importanza tra i diversi livelli, né come questi vanno gestiti: ciò dipende dalla particolare implementazione del costruttore. Queste 8 classi di servizio sono definibili con i 3 bit del capo user priority di un frame Ethernet che adotta la codifica prevista da IEEE 802.1Q, quindi 802.1p richiede che il link sia di tipo trunk.

Si possono usare diversi algoritmi di scheduling delle trame: round robin, weighted round robin, weighted fair queueing. La priorità può essere settata dalla scheda di rete dell'host (nessun software lo fa, necessità di link trunk), dallo switch d'accesso (che dovrebbe capire anche i livelli 3, 4, 5; il pacchetto può essere cifrato) in base all'indirizzo MAC.

IEEE 802.3x

Implementa il controllo di flusso a livello Ethernet, grazie ai pacchetti PAUSE, che dicono al ricevente di interrompere la trasmissione per una certa quantità di tempo specificata all'interno del pacchetto stesso.

Ci sono due modalità di controllo di flusso:

- Asimmetrico: solo una macchina sul link può inviare pacchetti PAUSE.
- Simmetrico: entrambe le macchine sul link possono inviare e ricevere i pacchetti PAUSE.

Alcuni problemi: su quale porta in ingresso ad uno switch bisogna agire affinché smetta di trasmettere quando dalla porta in uscita arriva un pacchetto PAUSE?

Se l'invio di PAUSE è configurato in maniera simmetrica ed un host invia un PAUSE allo switch a lui direttamente connesso, questi manderà il PAUSE ad un altro switch, che smettendo di trasmettere, rallenterà il traffico anche di quegli host che congestionati non erano.

Quindi, in pratica, si usa un controllo di flusso asimmetrico sulla rete d'accesso (gli host possono inviare i PAUSE, gli switch no) e nessun controllo di flusso nel backbone.

Link Aggregation (IEEE 802.3ad)

Se volessimo aumentare l'ampiezza di banda tra due switch, e mettessimo un secondo link che li colleghi, questi verrebbe bloccato dallo Spanning Tree, in quanto formerebbe un loop con il primo. Abbiamo più affidabilità, ma non aumentiamo di fatto la banda.

Il link aggregation associa più link fisici (solitamente 2 o 4) ad un link logico, che è visto come unico dallo Spanning Tree. 802.3ad è valido solo per Ethernet e usa solo link full-duplex e tutti i link fisici dello stesso aggregato devono avere la stessa velocità.

Per la creazione di tali link logici vengono scambiate periodicamente le LACPDU (alla connessione della porta, ogni 3/90s, quando un link cade).

Un link fisico è un canale affidabile e le trame non debbono arrivare fuori ordine: con il link aggregation potrebbe accadere che una trama più piccola su un link fisico arrivi prima di un frame più grosso su un altro link fisico dello stesso aggregato, anche se il frame più piccolo è stato inviato dopo di quello più grande.

Ciò non deve accadere, che le trame appartenenti alla stessa conversazione arrivino fuori ordine. ma lo standard non definisce un algoritmo per la distribuzione dei pacchetti, ma suggerisce alcune soluzioni.

Es, in base al MAC destinazione (ciò non va bene, in quanto un server connesso ad uno switch con due link riceverebbe traffico da un solo link), oppure quello sorgente, o criteri più sofisticati che renderebbe gli switch più complessi e meno performanti.

Inoltre, bisogna fare attenzione nell'assegnazione del giusto costo per il link aggregato affinché rifletta il valore di banda aggregata cosicché lo (Rapid) Spanning Tree funzioni correttamente.

IGMP Snooping

È una tecnica che permette il controllo del traffico multicast su reti di livello 2, in quanto normalmente esso viene inoltrato in flooding dagli switch, penalizzando scalabilità e banda.

Esistono delle tecniche che permettono di conoscere la posizione dei membri di un gruppo multicast.

Una di queste è il protocollo GMRP, che permette di registrare la propria membership ad un gruppo multicast presso uno switch, il quale inoltrerà tale informazione anche agli altri switch.

GMRP è però poco usato perché non implementato da molte applicazioni.

Si è scelto di usare un protocollo di livello 3, l'IGMP, che non è uno standard, ma molto diffuso: esso si basa sul presupposto che tutto il traffico sia IPv4.

Lo switch quindi deve essere in grado di comprendere tali pacchetti e da questi ricavare la posizione degli host appartenenti ad uno specifico gruppo multicast.

Ciò viene realizzato sfruttando il fatto che c'è un mapping diretto tra un indirizzo IP multicast ed il suo indirizzo MAC corrispondente (01:00:5E:0[23 bit bassi dell'IP]).

Non tutti gli indirizzi multicast IP vengono però gestiti da IGMP.

Gli indirizzi nel range 224.0.0.0 – 224.0.0.244 non richiedono IGMP in quanto debbono essere comunque inoltrati in flooding senza alcuna segnalazione IGMP (multicast “well-known”).

Richiedono IGMP, invece, gli indirizzi nel range 224.0.1.0 – 238.255.255.255 e 239.0.0.0 – 239.255.255.255.

Come funziona IGMP.

Il router invia un Host Membership Query per sapere se ci sono host in ascolto su un qualche gruppo multicast (al router basta sapere che ci sia almeno un membro per gruppo).

Ogni host, come risposta, invia un Host Membership Report per richiedere il flusso multicast. Tale pacchetto è ricevuto da tutti gli host della LAN e se uno di questi rileva che qualcun altro è in ascolto sul suo stesso gruppo, non invia a sua volta un Host Membership Report.

Gli switch devono quindi essere in grado di distinguere i multicast “well-known” da quelli dinamici in quanto i primi debbono essere trasmessi in flooding su tutta la rete, i secondi soltanto agli host che hanno segnalato di volere ricevere tali pacchetti.

Per far ciò gli switch intercettano i pacchetti di Host Membership Report e riempiono delle tabelle con le corrispondenze tra host membri di un certo gruppo multicast e una porta dello switch.

IGMP, poiché si basa sul fatto che tutto il traffico sia IPv4, fallisce con IPv6.

IPv6 fa largo uso dei pacchetti multicast, ed un MAC costruito a partire da un indirizzo IPv6 multicast viene interpretato dallo switch come un multicast dinamico, da inoltrare, quindi, soltanto agli host in ascolto su quel gruppo.

Poiché nessuno avrà fatto tale richiesta di membership, tale pacchetto viene erroneamente scartato dallo switch, e quindi non arriverà mai a destinazione.

Una soluzione potrebbe essere aggiornare lo switch o inoltrare tutti i pacchetti multicast in flooding.

Spanning Tree (IEEE 802.1D)

Se la rete è magliata, le trame possono entrare in loop ed il backward learning fallisce perché crede che un host sia raggiungibile da porte sempre diverse ed i filtering database sono continuamente in uno stato inconsistente.

Le trame che generano loop sono quelle broadcast e quelle dirette ad host inesistenti.

L'unica soluzione al problema è creare rete senza maglie (ma la rete sarebbe poco robusta) o disabilitare (logicamente) un link, in quanto a livello 2 non abbiamo l'equivalente del TTL a livello 3.

È quest'ultima l'idea che sta alla base dello Spanning Tree, che trasforma la rete in un albero, in cui c'è un solo percorso che unisce due host e tale albero ha la radice nel root bridge.

Parametri di STP:

- BridgeID: BridgePriority(2 byte, meglio multiplo di 4096)+MAC Bridge(MAC di una porta del bridge).
- Root Path Cost: costo per raggiungere il root bridge (il costo dipende dalla velocità del link).
- PortID: identificatore di ogni porta di un bridge. PortPriority(1 byte)+PortNumber(1 byte).

Algoritmo:

- Scelta del root bridge:
 - BridgePriority
 - Bridge con MAC più basso
- Scelta root port:
 - porta con root path cost minore
 - porta connessa al bridge remoto con BridgeID più basso
 - porta connessa alla porta remota con PortID più basso
 - porta locale con PortID più basso
- Scelta designated port:
 - porta con root path costo minore
 - porta connessa al bridge con BridgeID più basso
 - porta con PortID più basso

Configuration BPDU (C-BPDU): possono essere generate soltanto dal root bridge, e vengono inviate ad ogni Hello Time. Gli altri bridge si limitano a propagarle solo quando le ricevono dalla root port. Le root port sono quelle che ricevono la BPDU migliore. Anche le porte bloccate ricevono e processano le BPDU. Se non vengono ricevute BPDU per MaxAge le porte bloccate, nel rispetto dei timer, diventano designate.

Funzionamento:

- Ogni bridge all'avvio assume essere root e tutte le sue porte sono designate, ed invia in flooding su tutte le porte le BPDU da egli stesso generate con il suo ID nel Root Identifier e nel BridgeID e root path cost 0.
- Se un bridge riceve una BPDU con RootID minore del suo BridgeID smette di generare BPDU e propaga tali BPDU aggiornando alcuni campi (mette il suo BridgeID, PortID della porta attraverso la quale propaga, aggiorna il root path cost, incrementa MessageAge).
- Se un bridge riceve una BPDU con RootID maggiore del suo la ignora.
- Se un bridge riceve dalla porta X una BPDU con root path cost minore di quello corrente, la porta X diventa root port (in caso di più valori uguali si sceglie come root port quella che ha ricevuto la BPDU con il campo BridgeID più basso, o ancora con PortID più basso, o dalla porta locale con PortID più basso).
- Se un bridge riceve una BPDU con root path cost maggiore dell'attuale la ignora.
- Una volta scelto il root bridge e la root port, il bridge deve decidere se sulle altre porte lui è il bridge designato (cioè quella porta è designata). Ciò viene fatto comparando i messaggi inviati da lui su quella porta e quelli che egli riceve da quella porta. Se vede arrivare una BPDU il cui root path cost è più basso, o è più basso il BridgeID o ancora il PortID, significa che chi sta dall'altra parte è "più bello", e quella porta è bloccata (viceversa, dall'altro lato sarà designata).

Quando una porta deve diventare nello stato forwarding (designata o root), non inoltra subito il traffico dati, ma per sicurezza attende prima un tempo pari a $2 * \text{ForwardDelay}$ (circa 30s, 15 secondi nello stato listening e 15 secondi in quello learning). Quando una porta deve diventare bloccata, invece, per sicurezza, lo diventa immediatamente.

In generale l'algoritmo converge in circa 50s con i parametri di default $20 (\text{MaxAge}) + 30 (2 * \text{ForwardDelay})$.

Quando un link cade, il bridge che vede il cambiamento invalida il suo FD, ma gli altri non invalideranno i loro, che faranno fare alle trame un percorso sbagliato. Per questo quando un bridge vede un cambiamento nella topologia (caduta/aggiunta/cambio costo di un link, cambia la Bridge Priority) invia una Topology Change Notification (TCN BBPDU) dalla root port ad intervalli di Hello Time fino a quando non riceve l'Acknowledgment (BPDU + TCA) dall'upstream bridge. Quando tale BPDU raggiunge il root bridge, questi invia per $\text{MaxAge} + \text{ForwardDelay}$ una BPDU + TC che raggiungerà tutti i bridge in cascata, indicando loro di impostare le loro entry nel FD a ForwardDelay, così da velocizzarne l'invecchiamento, e non cancellare quelli ancora attivi e raggiungibili, riducendo il flooding.

I bridge connessi direttamente ad un host, vedranno continuamente un cambiamento nella topologia quando il PC si spegne e accende: ciò porterà ad un continuo flush dei FD, ad un aumento del flooding sulla rete e all'invio delle TCN BPDU e conseguenti BPDU+TCA e BPDU+TC. Per questo si decide di disabilitare lo Spanning Tree sulle porte edge (per i router Cisco vuol dire impostare la modalità PortFast per cui si va nello stato forwarding senza aspettare quei 30 secondi di listening e learning).

IEEE 802.1t

Introduce nuovi costi per i link per le nuove velocità ed aggiorna il campo BridgePriority del BridgeID (4 bit di priorità + 8 bit per la specifica istanza dello Spanning Tree).

Rapid Spanning Tree (IEEE 802.1w)

Lo Spanning Tree era pensato per reti a bus commune, in cui accorgersi se un host è down, un link è caduto, era più lento. Nelle moderni reti switched punto-punto full-duplex ciò avviene molto più rapidamente, ed è stato necessario sviluppare una versione migliorata dello Spanning Tree classico, di cui, però, condivide i principi fondamentali.

La convergenza di tali algoritmo è di circa 10ms.

Lo Spanning Tree distingueva in stato delle porte (disabled, blocking, listening, learning, forwarding) e ruoli delle porte (root, designata, bloccata). Nel RSTP lo stato delle porte è Discarding, Learning e Forwarding, ed il ruolo è root, designata, alternate, backup, edge.

La porta alternate non fa passare le trame dati, ed è la sostituta della root port in caso questa fallisce (è quella che riceve una BPDU migliore da un altro bridge).

Anche la porta backup non fa passare trame dati ed è la porta di "riserva" per quella LAN (è la porta che riceve una BPDU dallo stesso bridge).

Le porte edge sono quelle collegate agli host finali e non ad un bridge.

Algoritmo:

- Scelta del root bridge
- Scelta delle root port
- Scelta delle porte designata
- Le altre sono alternate (se connesse ad una porta di un altro bridge), backup (se connesse ad una porta dello stesso bridge) o edge.

Miglioramenti rispetto a STP:

- Tutti i bridge generano BPDU ad ogni Hello Time (default 2s), così non si attende per MaxAge.
- Invecchiamento veloce delle informazioni. Se non ricevo BPDU per 3 Hello Time consecutivi, la BPDU del root bridge scade (per i link punto-punto full-duplex me ne accorgo subito).
- Accettazione di BPDU inferiori. Se un bridge riceve una BPDU "più brutta" dalla root port, questa viene accettata e l'algoritmo RSTP riparte subito senza aspettare MaxAge, per cui ogni bridge si ricandida come root.
- Transizione rapida allo stato forwarding. Le porte edge non fanno parte della topologia e diventano immediatamente forwarding, senza attendere $2 * \text{ForwardDelay}$ (30s). Le porte alternate diventano subito root se la root port fallisce. Se non ci sono porte alternate il bridge si propone come root (tutte le porte diventano designate).
- Proposal/Agreement. Tale meccanismo è valido per link, ed è disponibile solo sui link full-duplex (che sono sempre punto-punto). Quando un link va up, la porta ha il ruolo di designata, ma il suo stato è discarding. Tale algoritmo permette ai due bridge di mettersi d'accordo velocemente su chi deve essere il bridge designato sul link punto-punto. Il più veloce si propone come designato, se all'altro bridge sta bene (la BPDU che invierebbe ha root path cost maggiore, bridgeID più alto, portID più alto), porta in discarding le porte root e designate, ed invia l'agreement, altrimenti invia a sua volta un proposal che l'altro accetterà inviando lui l'agreement. Ciò viene ripetuto in cascata partendo dal root bridge a livelli verso il basso, tagliando la rete ad ogni livello mentre è in corso il proposal/agreement. Se non si riceve un agreement dopo un proposal, la porta diventa di tipo 802.1D e attraversa quindi gli stati listening e learning.

Topology Change

A differenza di STP, le Topology Change vengono mandate solo quando un link va in up, non quando va in down perché in questo caso le trame prima o poi verranno scartate e le entry del FD corrispondenti a quella porta vengono fatte scadere subito.

Se una porta non edge va up, viene fatto partire il while timer ($2 * \text{HelloTime}$) su tutte le porte designate e root, e su di esse viene fatto il flush degli indirizzi MAC corrispondenti (quelle porte potrebbe più non essere attive dopo e/o il percorso per quegli host potrebbe più non essere lo stesso di prima) ed invio BPDU+TC su quelle porte fino alla scadenza del timer (nello STP invece veniva prima informato il root bridge che a cascata notificava tutti gli altri). Quando un bridge riceve una BPDU+TC può voler dire che in quella direzione c'è qualcuno che prima era raggiungibile da un'altra direzione. Su tutte le altre porte, quindi, azzerò le entry del FD per evitare che una trama vada nella direzione sbagliata, attivo il while timer, ed invio anch'io le BPDU+TC.

Se una porta edge va up, non vengono generate Topology Change.

RSTP e STP

Quando una porta va up, sono attive entrambi le modalità 802.1w e 802.1D per un tempo pari a Migration Delay (3s). Alla scadenza si adatta alla modalità della prossima BPDU ricevuta; comunque, si passa in modalità 802.1D se viene ricevuta una BPDU 802.1D.

Un bridge che comincia in modalità 802.1D, può non essere in grado di funzionare in modalità 802.1w a meno di configurazione da management.

Alta reattività di RSTP

Non sempre la veloce reattività del RSTP ci è in aiuto. Se un collegamento è sporco, la porta farà continuamente up e down, scatta il while timer e i FD vengono purgati continuamente e la rete resterà continuamente in uno stato transitorio. Una soluzione proprietaria Cisco è l'anti-flapping, che pone in questi casi la porta in stato error-disable.

VLAN

Una VLAN è una partizione di gruppi omogenei di host, che permette di gestire reti IP differenti a livello 2.

Ciò facilita la gestione delle rete, poiché non sposto cavi, ma faccio tutto via software, management, ottimizzo l'utilizzo dell'infrastruttura (se devo isolare una rete non metto uno switch o un router, ma cambio la configurazione delle porte degli switch), guadagno in scalabilità (la riassegnazione delle porte è immediata).

Assegnazione delle VLAN:

- Port-based: viene settata da management ogni porta dello switch (access o trunk). È trasparente all'host che però cambia VLAN a seconda della porta dello switch a cui è collegata, per cui non si ha mobilità a livello 3.
- Assegnazione trasparente: una tabella che mappa gli indirizzi MAC alla VLAN di appartenenza. Permette la mobilità.
- Per-user assignment (802.1x): l'host si autentica presso lo switch che lo assegna alla VLAN di appartenenza.
- Cooperative assignment: le trame vengono taggate dalla stazione di partenza. Fatto solitamente su apparati controllati dal management. Ci sono due modalità:
 - Singola VLAN per NIC
 - VLAN multiple per NIC attraverso interfacce virtuali, create dal driver della scheda di rete.

Anche se le VLAN sono indipendenti tra loro, l'isolamento non è però completo, in quanto il traffico di una VLAN può comunque saturare la banda disponibile alle altre VLAN: si potrebbe usare un Per-VLAN QoS.

Differenza tra link access e link trunk:

- Link Access: quando una porta sullo switch è di tipo access, viene configurata una sola VLAN per quella porta, cioè quella che al dispositivo è permesso di accedere. Tale dispositivo è però ignaro di appartenere a quella VLAN, in quando esso è in grado di capire soltanto i normali frame Ethernet (cioè non taggati): lo switch, infatti, rimuove le informazioni sulla VLAN dalla trama prima di inoltrarla al dispositivo sul link d'accesso.
- Link Trunk: un link trunk è in grado di trasportare traffico proveniente da diverse VLAN ed è usato per connettere gli switch ad altri switch o ad un router. Per differenziare le diverse VLAN ci si basa sul VLAN ID secondo lo standard 802.1Q.

VLAN e Spanning Tree

Sono teoricamente indipendenti tra loro.

Prima si calcola lo Spanning Tree sulla rete, eliminando i loop, poi si creano le VLAN sulla topologia risultante, con un unico albero di instradamento per tutte le VLAN. Quasi tutti i costruttori implementano anche il PVST, che permette la coesistenza di più istanze di Spanning Tree sulla rete, una per VLAN.

Ciò aumenta il carico di lavoro per gli switch, e se cade un link, la topologia va ricalcolata per tutte le istanze dello Spanning Tree, e di conseguenza si avranno tante BPDU in giro per la rete.

Il PVST ottimizza il carico sulla rete, permettendo l'utilizzo di tutti i link fisici disponibili, ma non ottimizza i percorsi di rete all'interno della VLAN: pur avendo un link diretto tra due switch, tale link può essere stato tagliato dallo Spanning Tree, e le trame faranno un percorso più lungo per arrivare a destinazione.

HSRP

Se volessimo ridondare il default gateway, ed aggiungessimo un router, sbaglieremmo, perché se cade il primo, gli host dovrebbero essere esplicitamente configurati affinché contattino l'altro.

L'HSRP è un protocollo proprietario di Cisco che permette la ridondanza del default gateway e il bilanciamento del carico.

Tra i diversi DG viene eletto uno come active, e si fa uso dei alcuni messaggi di Hello per capire se l'active cade ed eleggere un altro active.

Ai router vengono assegnati 2 indirizzi IP, uno fisico ed uno virtuale, e 2 indirizzi MAC, uno fisico ed uno virtuale.

L'indirizzo IP virtuale è quello che gli host usano come IP del DG, il MAC virtuale è generato automaticamente da HSRP (per Cisco 00:00:0C:07:AC:xx, con xx numero del gruppo HSRP).

Ci sono 3 tipi di router: active, stand-by e listen.

Il router che diventa active è quello a priorità più alta, oppure quello con IP fisico più alto. Il router active e quello stand-by inviano continuamente messaggi di Hello come keep-alive.

I messaggi di Hello vengono inviati ad intervalli di Hello Time (default 3s); l'Hold Time è il tempo di validità dell'ultimo messaggio di Hello, quando scade, il router stand-by si propone come active.

Quando un router stand-by diventa active, invia un ARP Reply gratuito in broadcast con i suoi IP e MAC virtuali, anche se ciò non è strettamente necessario (il mapping delle ARP Cache non cambia ed i filtering database si aggiornano non appena il nuovo active comincia ad inviare trame).

Solo il router attivo può usare l'indirizzi MAC virtuale (se così non fosse, i FD degli switch sarebbe sempre in una condizioni instabile), gli altri possono usare soltanto i MAC fisici.

HSRP è incapsulato in UDP (porta sorgente e destinazione 1985). L'indirizzo IP destinazione è il multicast 224.0.0.2 (tutti i router, tale indirizzo è un multicast well-known ed inoltrato sempre, anche senza segnalazione IGMP). L'indirizzo IP sorgente è quello fisico del router (necessario per vedere chi ha l'IP più alto ed eleggere così l'active) e TTL è pari a 1. Il MAC destinazione è quello generato automaticamente da HSRP, quello sorgente è quello virtuale per l'active, quello fisico per gli altri.

La preemption va configurata su ogni router e permette ad un router con una priorità più alta dell'active router di prendere il suo posto.

La funziona di track permette di monitorare anche il link verso la WAN e se cade, abbassa la priorità del router di 10.

Poiché ogni VLAN è una LAN separata con il suo default gateway, sono necessarie più istanze di HSRP, da configurare per ogni sotto-interfaccia virtuale.

Posso creare più gruppi HSRP sulla stessa LAN per fare del bilanciamento del carico, ciò però comporta una suddivisione statica degli host che non sempre è conveniente.

HSRP agisce solo sul traffico in uscita dalla LAN, quello in entrata è gestito dai protocolli di routing a livello 3.

L'algoritmo converge in circa 10 secondi.

VRRP

È una versione leggermente modificata di HSRP sviluppata con lo scopo di non infrangere alcun brevetto di Cisco.

Queste alcune differenze:

- Pacchetto direttamente imbustato in IP e non più UDP.
- Trasmesso all'indirizzo multicast di destinazione 224.0.0.18 e non 224.0.0.2.
- Ci sono diversi indirizzi MAC associati ad ogni gruppo, anziché uno virtuale unico.
- TTL = 255, non più pari ad 1 (anche se un router che riceve un pacchetto con TTL diverso da 255 deve scartare il pacchetto)
- Diversa nomenclatura: Master/Backup, anziché Active/Stand-by, messaggi di Advertisement anziché di Hello, Virtual Router ID anziché gruppo HSRP.
- Ogni master VRRP può controllare più di un indirizzo IP.
- Ci possono essere uno o più router di backup (in HSRP c'era solo un router standby).
- C'è un router master e tutti gli altri sono backup (non esistente l'equivalente listen di HSRP).
- Funzione di track non disponibile
- Preemption attiva di default.
- Converge in circa 4 secondi.

GLBP

Permette una gestione del bilanciamento del carico migliore rispetto ad HSRP.

Un gruppo di router condivide lo stesso IP virtuale ed ha diversi MAC virtuali. Tra tutti questi viene eletto l'Active Virtual Gateway (AVG), che alle richieste ARP degli host risponde in maniera intelligente con gli indirizzi MAC virtuali degli altri router, gli Active Virtual Forwarder (AVF), che prenderanno in carico il traffico del client.

Ci sono quattro possibili algoritmi per il bilanciamento del carico:

1. Nessuno. GLBP in questo caso è equivalente ad HSRP.
2. Pesato. Inoltre il traffico ai diversi AVF in percentuali diverse, a seconda delle diverse capacità dei link in uscita.
3. Dipendente dagli host. Associa un host ad uno specifico AVF. Ciò è utile nel caso l'host sia dietro NAT, altrimenti ad ogni passaggio di consegne bisogna aggiornare anche la tabella di traduzione del NAT del nuovo AVF per quell'host.
4. Round Robin. Utilizzo gli AVF in sequenza uno dopo l'altro.

IEEE 802.1x

È uno standard IEEE basato sul controllo delle porte di accesso alla rete LAN e provvede ad autenticare e autorizzare i dispositivi collegati alle porte della rete (switch e access point) stabilendo un collegamento punto a punto e prevenendo collegamenti non autorizzati alla rete locale. Viene utilizzato dalle reti locali wireless per gestire le connessioni agli access point e si basa sul protocollo EAP.

Funzionamento.

1. Quando un nuovo nodo richiede l'accesso alle risorse di una LAN, l'access point (AP) ne richiede l'identità. Nessun altro tipo di traffico è consentito oltre a EAP, prima che il nodo sia autenticato (la "porta" è chiusa). Il nodo che richiede l'autenticazione è spesso denominato supplicant, tuttavia sarebbe più corretto dire che esso contiene un supplicant, un programma applicativo installato nel computer. Il supplicant ha il compito di fornire risposte all'autenticatore che ne verificherà le credenziali. Lo stesso vale per l'access point; l'autenticatore non è l'access point. Invece, l'access point contiene un autenticatore. L'autenticatore non deve necessariamente trovarsi all'interno dell'access point; può essere un componente esterno. Viene instaurato un tunnel TLS cifrato per lo scambio d'identità tra supplicant ed access point.
2. Dopo l'invio dell'identità, comincia il processo di autenticazione. Il protocollo utilizzato tra il supplicant e l'autenticatore è EAP, o, più correttamente, EAP incapsulato su LAN (EAPOL). L'autenticatore re-incapsula i messaggi EAP in formato RADIUS (protocollo AAA usato per lo scambio di messaggi di autenticazione tra autenticatore e server), e li passa al server di autenticazione (EAP over RADIUS). Durante l'autenticazione, l'autenticatore semplicemente ritarda i pacchetti tra il supplicant e il server di autenticazione. Quando il processo di autenticazione si conclude, il server di autenticazione invia un messaggio di successo (o di fallimento, se l'autenticazione fallisce). L'autenticatore quindi apre la "porta" al supplicant.
3. Dopo un'autenticazione andata a buon fine, viene garantito al supplicant l'accesso alle altre risorse della LAN e/o ad Internet.

Viene chiamata autenticazione basata sulle porte perché l'autenticatore ha a che fare con due tipi di porte: controllata e non controllata. La porta controllata è quella a cui si connette il supplicant, ed è quella che permette o nega il traffico di rete, quella non controllata è quella usata per l'invio e la ricezione delle trame EAPOL con il server RADIUS.

Un'altra tecnica molto usata nelle reti wireless è quella del Captive Portal.

Il Captive Portal forza un client http a visitare una speciale pagina web usata per l'autenticazione prima di poter accedere alla navigazione.

Ciò si ottiene intercettando tutti i pacchetti fin dal momento in cui l'utente apre il proprio browser e tenta l'accesso a Internet. In quel momento il browser viene rediretto verso una pagina web la quale può richiedere l'autenticazione oppure semplicemente l'accettazione delle condizioni d'uso del servizio. Anche in questo caso può essere necessario lo scambio di messaggi di autenticazione con un server RADIUS.

L2 vs L3

	L2		L3	
	Pro	Contro	Pro	Contro
Generale	Semplice da produrre, poco costoso e plug-and-play			Hardware complesso (estrarre informazione a livello 3, ricostuire il frame, longest prefix matching, protocolli di routing sofisticati)
	Performante			Non molto veloci
	Mobilità a livello 3			Costosi
	Trasparente agli host			No mobilità No trasparenza, necessità di riconfigurare gli host
Sicurezza		No isolamento di rete (ARP spoofing, MAC flooding)	Isolamento di rete per migliore sicurezza	
		Filtering database vulnerabile	Indirizzi IP configurabili da management	
		ACL da definire per MAC	ACL definibili per rete	
		Poco intelligente per operare a livello di porte TCP/UDP	Capiscono informazioni a livello trasporto e applicativo	
Indirizzamento		Indirizzamento piatto	Indirizzamento gerarchico	
Traffico broadcast		Cresce con il numero degli host	I router non inoltrano il traffico broadcast	
		Broadcast storm		
Dimensione della rete		Diametro di rete limitato	Nessun limite	
Percorsi di rete		Limitate capacità di routing (unico Spanning Tree), percorsi non ottimizzati (può non essere importante), link sotto-utilizzati	Ogni router ha il proprio Spanning Tree	
Transitorio		Recupero guasti lenti, meglio con RSTP		
		Vulnerabile agli attacchi flooding durante il transitorio		

Conviene creare la rete d'accesso ed il backbone a livello 2 e scegliere il livello 3 per il gateway d'uscita e l'interconnessione tra VLAN.

NAS e SAN

I dati sono importanti. Prima si usavano i mainframe per la loro memorizzazione ed elaborazione, poi si è arrivati al modello client-server: i primi elaboravano i dati, i secondi li memorizzavano.

Adesso i nuovi trend sono quello peer-to-peer e datacenter.

Nel primo caso, i dati sono distribuiti su varie sorgenti, è scalabile e la cui importanza non è percepita in ambito business. Il secondo, quello dei datacenter, sposta tutti i server in un'unica ubicazione, il datacenter appunto.

Le due tecniche non sono in antitesi.

Il datacenter all'interno è gestito in maniera peer-to-peer, in quanto il software di gestione dei datacenter è basato su un modello p2p.

I datacenter vengono utilizzati sia per lo storage che per l'elaborazione, in quanto ci si è accorti che la CPU era per molto tempo scarica.

Prima lo storage era all'interno della macchina server, ma ciò non era molto flessibile in quanto se mancava l'alimentazione al server, perdeva accesso anche ai miei dati.

Si è quindi passati ad un modello DAS (Directly Attached Storage), in cui i dati erano staccati dal server, ma direttamente connessi ad esso.

Il protocollo di comunicazione era SCSI, una pila protocollare di rete (dal livello fisico a quello applicativo) che permetteva la comunicazione tra sistema operativo e dispositivi di memorizzazione.

La comunicazione era caratterizzata da bassa latenza e basso tasso di errore, da cui protocolli di recupero errori inefficienti.

La flessibilità non era così alta: se si guastava il server andava direttamente sostituito, altrimenti l'accesso ai dati era impossibile, si preferiva avere l'accesso ai dati da più server contemporaneamente, la distanza tra storage e server è ancora abbastanza ridotta.

Il modello che si è sviluppato è NAS (Network Attached Storage), in cui lo storage ed il server sono fisicamente lontani, non hanno un'interfaccia SCSI, ma Ethernet, e la comunicazione avviene attraverso la rete IP.

Il NAS esporta e virtualizza i dischi, ho cioè accesso al file system, e non ai singoli settori del disco.

Si hanno i vantaggi dell'interfaccia ad alto livello del file system, ma se vogliamo un controllo più fine, non abbiamo accesso ai settori del disco; inoltre si hanno difficoltà di interoperabilità tra protocolli di comunicazione tra macchine di produttori diversi.

Le SAN (Storage Area Network), invece, non esportano dischi, ma i blocchi dei dischi e la comunicazione tra i dischi ed i server non avviene sulla rete pubblica IP, ma una rete dedicata, una rete specificatamente progettata per i data center e lo storage, con caratteristiche di low latency e low error rate.

Questa rete ad hoc necessita di una pila protocollare particolare.

In cima c'è sempre SCSI per motivi di compatibilità, ma sotto di esso non conviene mettere TCP/IP, per i ritardi che esso provoca. Sotto SCSI c'è Fibre Channel, una tecnologia di livello fisico, datalink e trasporto: è equivalente a SCSI (low latency e low error rate) ma tecnologicamente più evoluta per i moderni dispositivi.

Il Fibre Channel include una modalità lossless e implementa anche il controllo di flusso.

Ogni volta che invio un pacchetto consumo un token, quando non ho più token non posso più inviare pacchetti, debbo attendere la ricezione di altri token dal mio interlocutore: questi token si chiamano anche crediti.

Alcune pile protocollari usate nelle SAN sono:

SCSI		
Fibre Channel	Fibre Channel	iSCSI
	FCIP (Fibre Channel over IP)	TCP
	TCP	IP
	IP	Ethernet
	Ethernet	

Con l'avvento di 10GbE, si sviluppa FCoE. Per risolvere il problema dell'affidabilità si applica il per-priority PAUSE, e posso fermare il traffico soltanto di una certa priorità (posso avere il traffico Ethernet a priorità 0 e quello Storage a priorità 1), così da avere una lossless ethernet. 10GbE è molto usato nei DataCenter moderni.

Cavi e Cablaggio

Le caratteristiche principali di un mezzo trasmissivo sono:

- Velocità di propagazione del segnale espresso come frazione della velocità della luce.
- Impedenza della linea ($Z = R + jI$).
- Dimensione dei conduttori, espressa in AWG (American Wire Gage).

Le caratteristiche elettriche dipendono da:

- Caratteristiche meccaniche e geometriche del cavo
 - Numero e diametro dei conduttori
 - Distanza dei conduttori
 - Concentricità tra conduttore e isolante
 - Presenza di schermi
- Materiali usati nella costruzione

Fattori che compromettono l'impedenza di un cavo:

- Centratatura del conduttore rispetto all'isolante
- Variazioni nella geometria del cavo dovuti a difetti di fabbricazione e/o incuria nell'installazione.

L'attenuazione è la riduzione di ampiezza del segnale di uscita di un cavo rispetto al segnale in ingresso.

La diafonia, anche detta crosstalk, è l'interferenza elettromagnetica che si può generare tra due cavi vicini.

Riguarda soltanto i cavi in rame ed aumenta con l'aumentare della frequenza.

La diafonia viene distinta in paradiafonia e telediafonia:

A <-----> B (1)

C <-----> D (2)

Si ha paradiafonia quando il telefono A del circuito (1) genera un disturbo ricevuto dal telefono C del circuito (2).

Si ha telediafonia quando il telefono A del circuito (1) genera un disturbo ricevuto dal telefono D del circuito (2).

Il circuito (1) si definisce disturbante, il circuito (2) si definisce disturbato.

Data l'attenuazione sul canale, la paradiafonia è evidente soltanto per i primi 20-30m di cavo.

Per Delay Skew si intende la variazione del tempo di propagazione dovuto alla differente lunghezza delle coppie in un cavo multicoppia a causa del diverso passo di binatura (ritorsione).

In un cavo twisted vengono utilizzati diversi passi di binatura per ridurre il crosstalk.

Per Return Loss si intende la quantità di energia del segnale che viene riflessa nei punti di variazione di impedenza del conduttore (disadattamento di impedenza).

Vantaggi della schermatura:

- Maggiore immunità ai disturbi elettromagnetici.
- Maggior costanza dell'impedenza.
- Riduzione della diafonia se applicata alle singole coppie.

Tipi di schermi:

- Foglio (in mylar alluminato che avvolge il cavo sotto la guaina esterna).
- Calza (treccia di fili di rame che avvolge il cavo).
- Foglio + Calza (schermatura migliore, ma aumentano dimensioni e costi del cavo).

Fibre Ottiche

Le fibre ottiche da utilizzare vanno scelte in base alla distanza da coprire e alle disponibilità economiche.

Se la distanza è inferiore ai 3 Km, si possono usare le fibre multimodali; per distanze maggiori, bisogna optare per le monomodali (che sono anche più costose).

Differenza tra fibre monomodali e multimodali.

Le fibre multimodali hanno un nucleo di luce relativamente largo (62.5 micron di diametro), vengono usate per brevi distanze e sono basate su una tecnologia LED.

Le fibre monomodali hanno un nucleo di luce più piccolo (8-10 micron di diametro). Sono usate per lunghe distanze e sono basate su una tecnologia a diodo laser.

Content Delivery Network

Le CDN riprendono un concetto risalente agli inizi del Web quando si voleva aumentare la velocità di accesso alle pagine risparmiando sulla banda, che era al tempo costosa: furono introdotte quindi le Web Cache, dispositivi che memorizzano una copia in locale delle pagine HTTP più recenti visitate dagli utenti e fungevano da Proxy Server.

Le Web Cache poi sono cadute in disuso, dato l'aumento della banda e la riduzione del suo costo, il grosso aumento di contenuti sul Web che diventava difficile e costoso mantenere in una cache.

Lato server, i motivi per lo sviluppo delle CDN sono dovuti al volume di traffico da gestire, per cui un singolo Datacenter non era più sufficiente, e vengono costruiti server distinti, la cui delocalizzazione permette anche di poter diminuire l'RTT verso l'utente finale (migliore user experience nella navigazione), di migliorare l'affidabilità (un terremoto in un sito non distrugge tutti i miei server) e il load balancing.

Le Content Delivery Network sono reti di server overlay che gestiscono la delocalizzazione dei contenuti sulla rete.

Il concetto è simile a quello delle Web Cache, e su di esse mettono le pagine web che necessitano di maggiore banda (pagine con video, immagini, live streaming). Un altro motivo è avere del contenuto dinamico, cioè personalizzato per l'utente che visita quella pagina (es. luogo in cui si trova).

Dal punto di vista architetturale, il fornitore di contenuti (es. CNN) fa l'upload su un server, ed un software provvede a distribuire tali informazioni su tante repliche distribuite su Internet da cui i client prenderanno il contenuto.

Il metodo più usato per fare in modo che i client accedano alla replica corretta in modo trasparente (loro digitano sempre `www.cnn.com`, e non l'indirizzo del server replica) è una tecnica basata sul DNS-based routing, per cui l'indirizzo IP ritornato non dipende soltanto dal nome dell'host da risolvere, ma anche dall'IP da cui proviene la richiesta di risoluzione.

Normalmente una richiesta ad un sito, es. `cnn.com`, comincia con una richiesta al nostro DNS autoritativo (che eventualmente farà una richiesta al DNS di CNN) che ci fornirà l'IP cercato e il nostro browser farà partire una serie di richieste HTTP verso il server CNN delle pagine e i contenuti di esse.

Nel caso di un DNS-based routing, la situazione è un po' diversa, e richiede una modifica nel DNS di CNN, che diventa un CDN DNS.

L'utente fa una richiesta al suo DNS autoritativo, che chiede al CDN DNS di CNN, il quale, conoscendo l'IP sorgente della richiesta e quindi il suo dominio, cerca la replica a lui più vicina e risponde con il suo indirizzo IP. A questo punto l'utente finale, in maniera del tutto trasparente a lui, non contatterà più il server della CNN, ma la replica.

Vantaggi:

- Trasparenza verso l'utente finale.

Svantaggi:

- Per le aziende il cui core business non è l'informatica, modificare il DNS, creare e mantenere le repliche, non è un lavoro da poco, o una spesa che conviene sostenere.

L'approccio di Akamai.

Prendiamo come esempio la pagina internet della home page del sito della CNN.

Essa è formata da HTML, e da link a video e immagini: le URL di quest'ultime vengono modificate, vengono akamaizzate. La prima parte dell'URL identifica le rete Akamai, la seconda identifica lo User-ID (codice client di Akamai), la terza parte indirizza il dato fisico a cui accedere (l'immagine, il video, ...).

La pagina principale è sempre ospitata da CNN, sono i suoi contenuti ad essere memorizzati sulle repliche di Akamai.

Così facendo, CNN non deve neanche modificare il suo DNS, ma solo le URL dei contenuti che vuole delocalizzare e far gestire dal Content Provider (Akamai, per esempio).

L'utente, dopo aver scaricato la pagina principale del sito dal server di CNN, fa una richiesta al suo DNS per le URL modificate dei suoi contenuti, che a sua volta contatterà un server dell'overlay di Akamai (il CDN ISP) che risponderà con l'indirizzo IP della replica più vicina all'utente che ha fatto la richiesta. L'utente richiederà quindi i contenuti a tali repliche.

Vantaggi:

- Trasparenza sia verso l'utente finale, sia verso il fornitore di contenuti che non modifica il suo DNS (es. CNN).

Server Load Balancing

La duplicazione e la delocalizzazione dei contenuti, non risolve però completamente i problemi di scalabilità e bilanciamento del carico dei Datacenter.

Il classico SLB (Server Load Balancer), riceve una richiesta in ingresso, e la gira in maniera intelligente (es. Round Robin) a diversi server fisici. Ci sono 2 tipi di SLB: Content-unaware e Content-aware.

Il primo lavora fino a livello 4, distribuendo il traffico in base a IP e Porte sorgente (IP e Porta destinazione non cambiano): a questo livello viene ridiretta l'intera connessione e non un singolo pacchetto (cioè un client una volta che è stato associato ad un server, contatterà quest'ultimo fino alla chiusura della connessione TCP).

Il secondo tipo lavora fino a livello 7, facendo sì che ogni server risponda a richieste di una specifica sezione del sito.

Per far ciò bisogna poter accedere al payload http, cioè l'URL.

In questo caso, il SLB non può ridirigere il client immediatamente verso un server, perché non sa a quale sezione del sito accederà (la connessione inizia con il 3-way handshake, la GET http viene dopo).

Le connessioni dei client debbono quindi necessariamente terminare sul SLB, e non sul server fisico finale: si ha un 3-way handshake tra client e SLB, e un 3-way handshake tra SLB e server finale.

Il SLB deve essere anche in grado di gestire i certificati SSL nel caso in cui le connessioni siano cifrate.

Alcune applicazioni necessitano che le connessioni TCP provenienti dallo stesso client vengano ridirette allo stesso server (Sticky Connections), per esempio carrelli della spesa virtuali, transazioni economiche.

Ciò può essere gestito con i cookies, così il server può mantenere lo stato della nostra transazione.

Con un SLB di tipo L4, anche se siamo ridiretti a server diversi, se ci è stato dato un cookie, ogni volta che ci presentiamo ad un nuovo server, questi, interrogando dei server di backend (es. DB Server) usando come chiave il cookie, è in grado di riconoscerci, recuperare lo stato della nostra transazione e servirci in maniera opportuna.

Se ho un SLB di tipo L7, questi, essendo in grado di capire il cookie, può ridirigerci sempre sullo stesso server fisico, il quale ha le informazioni sul mio stato in memoria, e non deve chiederle al backend.