Livello Network e protocollo IP

Davide Quaglia

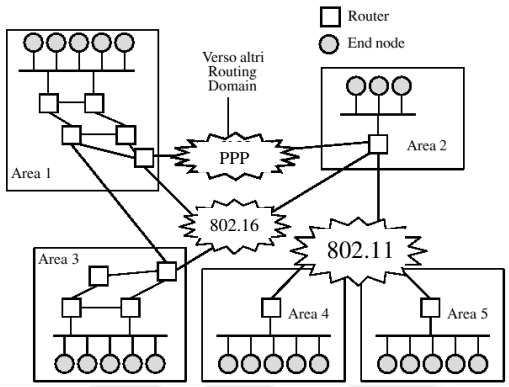
Motivazioni

- Necessità di far comunicare diversi tipi di reti di livello 2
  - Diversi mezzi trasmissivi
  - Diversi formati di Datalink PDU
  - Diverse dimensioni max di frame (Max Transfer Unit – MTU)
  - Diversi formati di indirizzi (o assenti)
  - Presenza di percorsi multipli per aumentare l'affidabilità

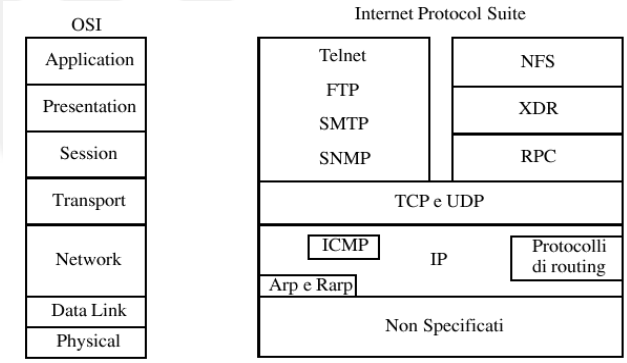
Funzionalità del livello network

- Aggregazione di reti di livello 2 per creare reti molto grosse (fino a Internet mondiale)
- Indirizzamento delle stazioni indipendente dallo standard di livello 2
- Routing: trovare la strada (migliore) tra due nodi qualsiasi della rete globale
- Gestione delle diverse MTU mediante frammentazione e riassemblaggio

Aggregazione di reti di liv. 2



Architettura TCP/IP



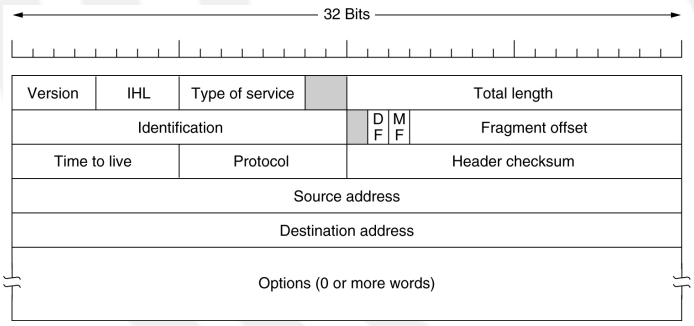
Internet Protocol (IP)

- Negli anni 70' DARPA progetta ARPANET
- Internet Engineering Task Force (IETF)
- Request for Comment (RFC)
  - RFC 791 descrive IP
- Internet = Inter Networking = come far parlare due nodi che non si vedono a livello 2

Internet Protocol (2)

- Indirizzamento dei nodi
- Servizio un-acknowledged connectionless: ogni PDU è indipendente dalle altre, deve contenere l'indirizzo di destinazione e non ne viene confermata la ricezione
- Routing
- Altre funzionalità:
  - Frammentazione
  - Rilevazione debole degli errori (solo checksum dell'header della Network PDU)
  - Forwarding (host+router+reti di livello 2)

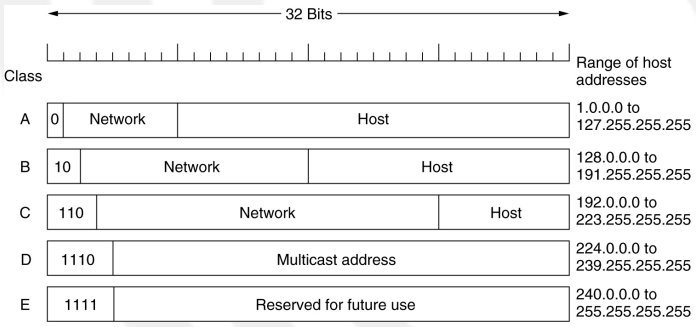
Formato dell'header del pacchetto IP



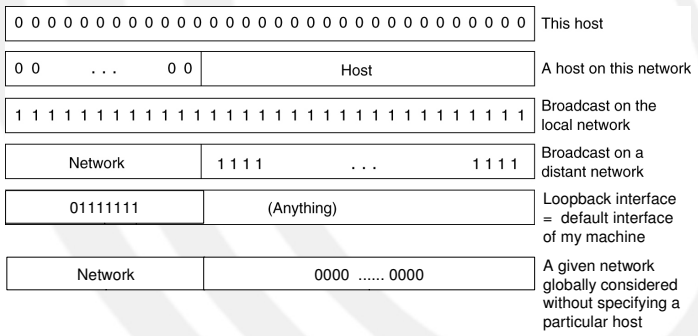
Indirizzi

- Assegnati alle interfacce (e non alle macchine !)
- 32 bit divisi tra Network e InterfaceID
- Determinazione della parte Network
  - Automatica mediante suddivisione in 5 classi di indirizzi
  - Manuale mediante utilizzo di network bitmask (netmask)

Classi degli indirizzi IP



Indirizzi IP speciali



Notazione decimale “dotted”

- Elencare 32 bit può essere scomodo
- I 32 bit vengono raggruppati in 4 numeri da 8 bit (intervallo 0-255) che vengono scritti
  - In base 10
  - Separati da punti
- Esempio:

  01111111 00000000 00000000 00000001 -->

  127.0.0.1 (interfaccia di loopback = l'interfaccia di default della propria macchina)

Netmask

```
130.192.17.15 da solo --> indirizzo di classe B
                  Net: 130.192.0.0 Host: 0.0.17.15

130.192.17.15 + "11111111 11111111 11110000 00000000"
--> Net: 130.192.16.0  Host: 0.0.1.15

Si può anche scrivere 130.192.17.15/20
```

- Se non si specifica la netmask vale la lunghezza del prefisso determinata dalla classe dell'indirizzo
- La netmask viene usata per suddividere grossi lotti di indirizzi in lotti più piccoli (subnetting)

Primo livello di routing

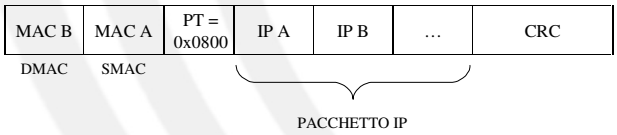
- Ogni interfaccia conosce:
  - Il proprio IP
  - La lunghezza della parte network (ufficiale o netmask)
  - IP di un router (default GW)
- Si confronta la parte dell'IP destinatario corrispondente alla parte network del proprio IP. Si verificano due casi:
  - Uguali: raggiungibilità diretta sulla rete di livello 2
  - Diversi: occorre spedire il pacchetto al default gateway
- Conseguenza: mittente e default gateway hanno lo stesso prefisso

Primo livello di routing (2)

- Subnet IP = insieme di tutte le interfacce con lo stesso prefisso IP (per lunghezza e per valore)
- Subnet IP $\subseteq$ rete di livello 2
- L'indirizzo del default GW impostato sul mio host deve essere nella stessa subnet e quindi avere lo stesso prefisso del mio IP
  - /sbin/ifconfig -a ---> IP e eventuale netmask
  - /sbin/route ---> IP del default GW

Invio di un pacchetto IP su rete Ethernet

Formato di un frame MAC per l'invio di un pacchetto IP tra 2 host (da A a B) della **stessa** sottorete IP:



Per conoscere l'indirizzo MAC di B la stazione A usa l'Address Resolution Protocol (ARP)

ARP

- A invia la richiesta Address Resolution Protocol in un frame MAC broadcast:

FFFFFFFFFFFF	MAC A	PT = 0x0806	Chi ha IP B?	CRC

ARP

- Solo B risponde con:

MAC A	MAC B	PT = 0x0806	Sono io!	CRC

- A può spedire il pacchetto IP
- A mette la terna (MAC B, IP B, timestamp) in una cache per usarla le volte successive
  - le righe più vecchie di 15 minuti vengono eliminate

Invio di un pacchetto IP su rete Ethernet (caso 2)

Formato di un frame MAC per l'invio di un pacchetto IP tra 2 host (da A a B) appartenenti a sottoreti IP **diverse**:

MAC DEFAULT GW	MAC A	PT = 0x0800	IP A	IP B	…	CRC

DMAC SMAC

PACCHETTO IP

Per conoscere l'indirizzo MAC del default GW la stazione A usa l'Address Resolution Protocol (ARP)

ARP (caso 2)

- Richiesta ARP broadcast per l'indirizzo MAC corrispondente all'IP del default GW

FFFFFFFFFF	MAC A	PT = 0x0806	Chi ha IP GW?	CRC

- Il default GW risponde:

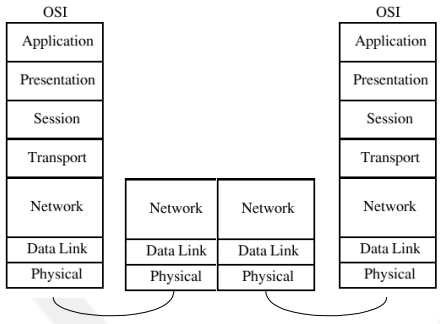
MAC A	MAC GW	PT = 0x0806	Sono io!	CRC

- La stazione memorizza la terna (MAC GW, IP GW, timestamp) nella cache ARP

Il traffico broadcast è notevole in una rete 802.X/Ethernet su cui c'è il protocollo IP !

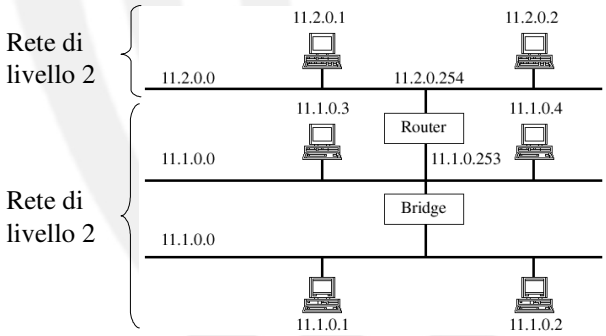
Router

Application layer	Application gateway
Transport layer	Transport gateway
Network layer	Router
Data link layer	Bridge, switch
Physical layer	Repeater, hub



Un router lavora a livello Network

Router e switch/bridge



Architettura di un router

- Il routing è diverso dal forwarding
- Marche: Cisco, Juniper, HP, ...

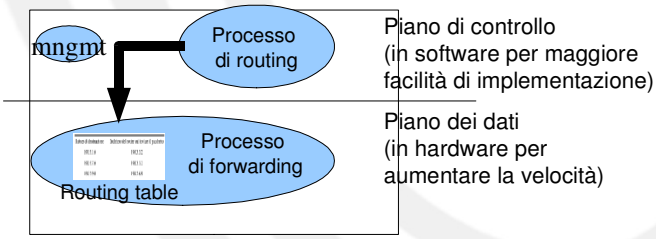
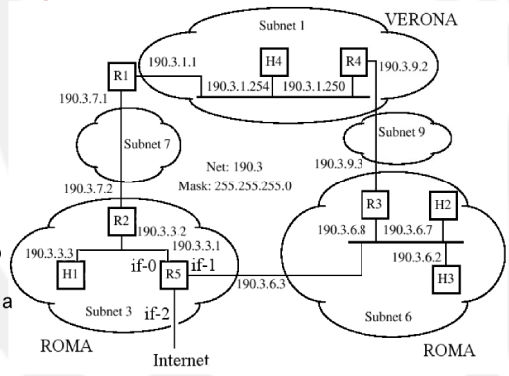


Tabella di instradamento

- Due colonne:
  - rete di destinazione
  - interfaccia di uscita
- Una riga per ogni sotto-rete a cui il router non è collegato direttamente
  - possibilità di accorpare più destinazioni che escono dalla stessa interfaccia
- Possibilità di impostare una default route (= “tutte le altre direzioni”)
- Non si specificano le sotto-reti di destinazione in cui il router ha delle interfacce (conoscenza diretta)

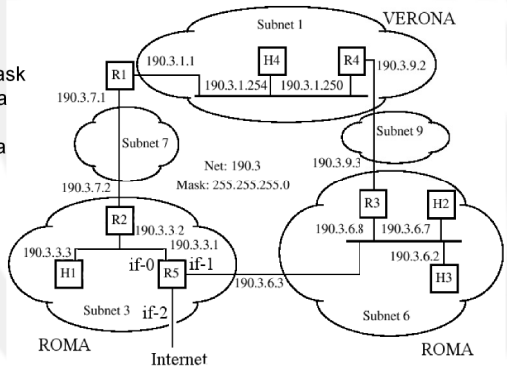
Esempio di rete IP

- Azienda con 2 sedi lontane
- Acquisizione di lotto classe B 190.3.0.0
- Ulteriore subnetting con netmask 255.255.255.0 per creare delle sottoreti:
  - Verona sede
  - Roma produzione
  - Roma commerciale
- Affitto di 2 canali punto-punto
  - 1 canale 5 Mb/s flat
  - 1 canale 5 Mb/s tariffato a tempo da usare solo in caso di backup
- L'uscita su Internet avviene a Roma



Esempio di rete IP

- Tutte le interfacce devono avere IP con la stessa netmask
- I router non appartengono a nessuna sotto-rete
- Tra 2 router c'è sempre una sotto-rete con almeno due indirizzi
- Ogni interfaccia di ciascun router ha associato
  - Un IP
  - Una netmask
  - Un default GW

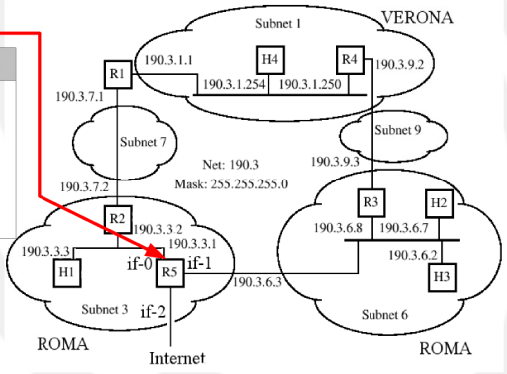


Esempio di rete IP

Tabella di instradamento

Destinazione	IFacc
190.3.6.0/24	if-1
190.3.3.0/24	if-0
190.3.1.0/24	if-0
190.3.7.0/24	if-0
190.3.9.0/24	if-1
default	if-2

Inserite automaticamente dal router per conoscenza diretta



Routing dinamico e distribuito

- Non esiste un punto della rete privilegiato che gestisce il calcolo dei percorsi
- Tutti i router si scambiano periodicamente dei pacchetti con le informazioni sulla loro raggiungibilità e il relativo costo (ad es. numero di router da attraversare)
- Ciascun router, in base alle info ricevute, si calcola la propria tabella di routing
- La periodicità garantisce l'adattamento a cambiamenti della topologia
  - Quando i router sono numerosi la probabilità di disservizio non è trascurabile

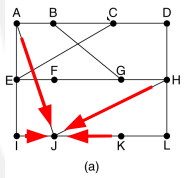
Modello di rete mediante grafi e albero dei cammini minimi

- Non si considerano gli host (end system)
- Si crea un grafo dove ogni nodo rappresenta un router e gli archi sono i collegamenti tra router
- Si può assegnare un peso positivo ad ogni arco che rappresenta il costo di attraversamento del collegamento
  - ritardo
  - inverso della capacità
  - tradizionalmente si mette peso unitario
- Si usa l'algoritmo *Shortest Path First* (SPF) di Dijkstra per calcolare i percorsi minimi tra ciascun router e tutti gli altri
  - in caso di pesi unitari si minimizza il numero di router attraversati

Tecnica del Distance Vector

- Ciascun router deve:
  1. Compilare una tabella dove indica le distanze verso **tutti i router della rete**
  2. Trasmettere ogni T secondi tale tabella ai **router adiacenti**
  3. Utilizzare le tabelle ricevute dai router adiacenti per aggiornare la propria tabella
  4. Ripetere i passi 2 e 3 finché la tabella non è stabile (uguale rispetto al passo precedente)

Esempio di Distance Vector



(a)

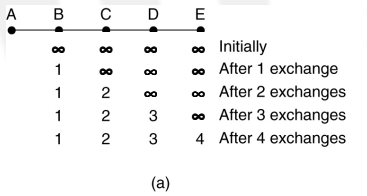
To	A	I	H	K	New estimated delay from J	Line
A	0	24	20	21	8	A
B	12	36	31	28	20	A
C	25	18	19	36	28	I
D	40	27	8	24	20	H
E	14	7	30	22	17	I
F	23	20	19	40	30	I
G	18	31	6	31	18	H
H	17	20	0	19	12	H
I	21	0	14	22	10	I
J	9	11	7	10	0	–
K	24	22	22	0	6	K
L	29	33	9	9	15	K
	JA delay is 8	JI delay is 10	JH delay is 12	JK delay is 6	New routing table for J	

Vectors received from J's four neighbors

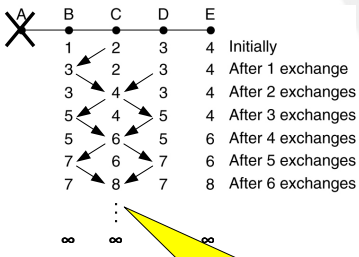
(b)

(a) Una rete (b) Input da A, I, H, K e la nuova tabella di routing di J

Reazione del Distance Vector ai cambi di topologia



(a)



(b)

Per ogni router la colonna corrispondente riporta la distanza da A

Caso a: il router A da non attivo diventa attivo (buona notizia)

Caso b: il router A da attivo diventa non attivo (cattiva notizia)

Occorre impostare un limite per terminare il conteggio e concludere che A non è raggiungibile (per il RIP tale limite è 16)

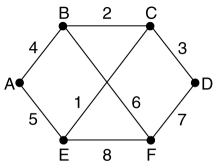
Pro e contro del Distance Vector

- Facile implementazione (algoritmo di Bellman-Ford)
- Veloce adattamento a “buone notizie”
- Lento adattamento a “cattive notizie”
- Difficile diagnostica perché nessun router ha la mappa della rete

Tecnica Link State

- Ciascun router deve:
  1. Compilare una tabella dove indica le distanze verso i **router adiacenti**
  2. Trasmettere ogni T secondi tale tabella a **tutti i router della rete**
  3. Utilizzare le tabelle ricevute da tutti i router per aggiornare la propria tabella
  4. Costruire l'albero dei cammini minimi a partire da tali informazioni

Esempio di Link State



(a)

Link A		Link B		Link C		State D		Packets E		Packets F	
Seq.		Seq.		Seq.		Seq.		Seq.		Seq.	
Age		Age		Age		Age		Age		Age	
B	4	A	4	B	2	C	3	A	5	B	6
E	5	C	2	D	3	F	7	C	1	D	7
		F	6	E	1			F	8	E	8

(b)

(a) Una rete

(b) Pacchetti link state

Pro e contro del Link State

- Veloce convergenza sia in caso di buone notizie sia in caso di cattive notizie
- Migliore diagnostica (ogni router ha un grafo dell'intera rete con annesso albero dei cammini minimi)
- Maggior complessità a causa del maggior numero di messaggi circolanti rispetto al Distance Vector

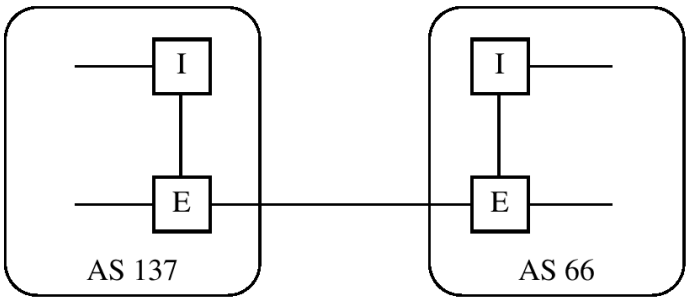
Distance Vector e Link State sono approcci duali

- DV = scambio tra **router adiacenti** di una tabella di raggiungibilità di **tutti i router**
- LS = scambio tra **tutti i router** di una tabella di raggiungibilità dei **router adiacenti**

Autonomous System

- Insieme di reti IP sotto un'unica autorità
- Numerazione intera e univoca a livello mondiale
- Interior Gateway Protocol (IGP): usato all'interno di un AS
  - All'interno di un AS tutti i router usano lo stesso tipo di protocollo IGP
- Exterior Gateway Protocol (EGP): usato tra reti di diversi AS

Autonomous System



E: Exterior router
I: Interior router

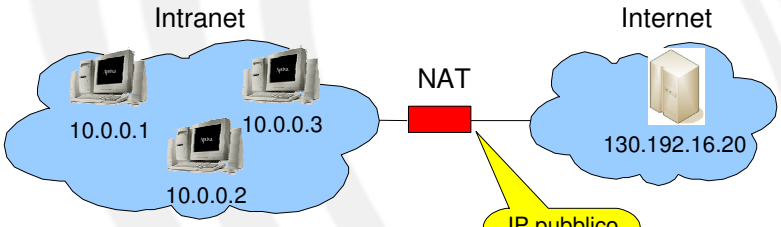
Protocolli di routing in Internet

- Interior Gateway Protocols (IGP):
  - Distance vector: RIP, RIP2, IGRP
  - Link state: Open Shortest Path First (OSPF)
- Exterior Gateway Protocol (EGP):
  - Border Gateway Protocol (BGP)
- Criterio di ottimalità di cammino:
  - Numero di hop (numero di router attraversati)

Indirizzi privati

- Definiti in RFC 1918 - Address Allocation for Private Internets
- Tre lotti di indirizzi
  - 10.0.0.0 - 10.255.255.255 (10.0.0.0/8)
  - 172.16.0.0 - 172.31.255.255 (172.16.0.0/12)
  - 192.168.0.0 - 192.168.255.255 (192.168.0.0/16)
- Non vengono mai “annunciati” dai protocolli di routing distribuito
- Le reti fatte di indirizzi privati non si possono collegare a Internet se non tramite un Network Address Translator (NAT)

NAT: un esempio



Il NAT sostituisce gli indirizzi 10.0.0.X con l'unico indirizzo 157.27.242.5. Per capire il funzionamento occorre aver spiegato il livello trasporto

Internet Control Message Protocol (ICMP)

- Imbustato dentro il pacchetto IP
- Messaggi di supporto al funzionamento IP:
  - Anomalie
  - Raggiungibilità
  - Miglioramento routing (redirect)
- Funzionamento di PING
- Funzionamento di TRACE ROUTE

Trace Route

- 3 messaggi ICMP
  - Echo Request
  - Echo Reply
  - Host Unreacheable
- Time to Live (TTL)

Qualità del Servizio in Internet

- Il progetto originale di IP non prevede controllo della Qualità del Servizio
  - Modello *Best effort*: i router fanno del loro meglio per consegnare i pacchetti a destinazione ma nulla viene garantito all'utente
- L'utilizzo della rete da parte degli utenti varia continuamente nel tempo e da punto a punto della rete
  - Se in un router la quantità di byte in entrata eccede la capacità di smaltimento si ha una *congestione*

Comportamento nei router

- Il servizio di trasmissione dati fornito da IP è non-connesso e non-confermato
- Un pacchetto che arriva ad un router, può trovare la porta su cui dovrebbe essere trasmesso impegnata da un altro pacchetto in trasmissione.
- In questo caso, viene memorizzato in una coda, e subisce per questo un ritardo.
- Nel caso la coda sia piena, il pacchetto viene scartato.

Effetto sui pacchetti

- **Perdita di pacchetto:** quando esso arriva ad una coda già piena
- **Ritardo (media e variazione):** la presenza di code introduce ritardo nel percorso; se la lunghezza delle code varia, varia anche il ritardo; il ritardo dipende dal percorso e dall'istante di tempo

Effetto sui pacchetti (2)

- **Consegna fuori ordine:** siccome ogni pacchetto è indipendente dagli altri, può fare un percorso diverso dagli altri diretti alla stessa destinazione
- **Errori sui bit:** il protocollo IP non ha meccanismi sofisticati di controllo degli errori che viene delegato ai protocolli di livello datalink

Superamento del Best Effort

- I problemi creati dal Best effort sono finora stati recuperati a livello trasporto (col TCP) ma questa soluzione è ottima quando si ha essenzialmente a che fare con applicazioni non troppo esigenti
- Oggi si vuol far convivere su IP applicazioni molto diverse: dal web alla telefonia
- Occorrono nuove soluzioni:
  - Over-provisioning
  - Amministrazione della capacità dando diverse priorità ai pacchetti delle varie applicazioni

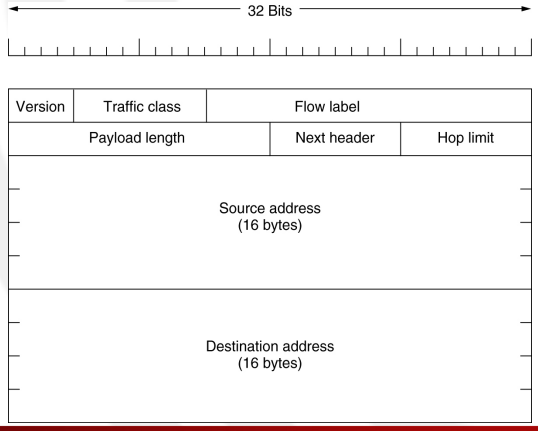
Priorità

- Occorre inserire un valore di priorità ai flussi di byte
- I byte appartenenti a priorità diverse devono essere gestiti in code diverse (come le corsie preferenziali dei taxi/autobus)
- Si può fare a diversi livelli
  - Datalink (Virtual LAN, 802.11 PCF, 802.11e)
  - IP (campo ToS) --> modello Differentiated Services
  - Trasporto --> modello Integrated Services

IP versione 6

- Nuovo formato di indirizzi per aumentarne la disponibilità
- Nuovo formato dell'header IP che elimina (rende opzionali) i campi meno usati
- Tutti i protocolli che viaggiano su IPv4 possono viaggiare anche su IPv6
- Regole di coesistenza tra IPv4 e Ipv6
- Introdotta la sicurezza a livello network

Header minimo di IPv6



Header minimo di IPv6

- Version (4 bit): ha valore 6
- Traffic class (8 bit): eredita le funzioni del ToS
- Flow label (20 bit): assieme a Src Addr e Dest Addr può servire a creare una label per funzioni “tipo MPLS”
- Payload length (16 bit): lunghezza del payload
- Next header (bit): codice della prossima intestazione o del tipo di PDU di livello 4
- Hop limit (8 bit): eredita le funzioni del TTL
- Source address (128 bit): vedere slide sul formato
- Destination address (128 bit): vedere slide sul formato

Differenze rispetto a header IPv4

- L'header ha lunghezza fissa (40 byte)
- Eliminazione campo *IP Header Length*
- Eliminazione campo *Protocol* (esiste *next header*)
- Eliminazione dei campi per gestire la frammentazione
  - E' stata ridotta la necessità di frammentare (minore lavoro nei router → maggiore throughput)
  - E' stata introdotta un'estensione opzionale di header se proprio serve la frammentazione
- Eliminazione del campo *Checksum* che doveva essere ricalcolata in ogni router
  - minore lavoro nei router → maggiore throughput

Formato degli indirizzi

- 128 bit (16 byte) → più degli atomi dell'universo
- Notazione: 8 gruppi (separati da ":") di 4 cifre esadecimali
  - 8000:0000:0000:0000:0023:0567:0000:CDEF
- Abbreviazioni:
  - Si possono omettere gli zeri iniziali di ciascun gruppo
    - 8000:0000:0000:0000:23:567:0:CDEF
  - Gruppi contenenti solo zeri possono essere sostituiti da una coppia di ":" (vale per UNA SOLA sequenza di gruppi a zero)
    - 8000::123:4567:0:CDEF

Formato degli indirizzi (2)

- Gli indirizzi IPv4 possono essere scritti in due modi:
  - Padding con (128 - 32) bit a “0”
    - notazione IPv6
    - in notazione decimale puntata (o “dotted”) portata a 128 bit con degli zeri (rappresentati da una coppia di “:”)
      - ::157.27.242.10
  - Anteponendo 16 bit a “1” ai 32 bit dell'indirizzo e i restanti (128-16-32) messi a “0”
    - Notazione IPv6
    - in notazione decimale puntata portata a 128 bit con il prefisso “::FFFF:”)
      - ::FFFF:157.27.242.10

Tipi di indirizzi

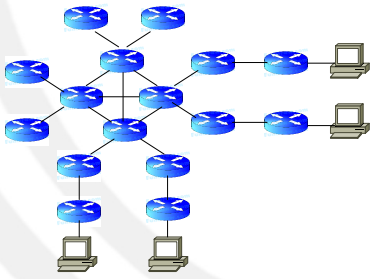
- Non esistono più le classi
- Indirizzo non assegnato: 00..0 (tutti 128 bit a 0)
- Indirizzo di loopback: 00..1 (127 zeri e un “1”)
- **Unicast global addresses**
  - indirizzi pubblici
  - attualmente sono assegnati quelli che iniziano con i bit 001
- Indirizzi privati
  - Link local unicast: iniziano con i bit 1111 1110 10
    - da usare nella procedura di *autoconfigurazione* (vedi lucido apposito)
  - Site local unicast: iniziano con i bit 1111 1110 11
    - da usare in intranet (indirizzi privati mai propagati su Internet)
- Multicast: FFxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx

Netmask

- La maschera di bit indicava quale parte dell'indirizzo era da considerarsi come prefisso della rete
  - Es: 157.27.242.10/255.255.255.0
- Negli ultimi anni si era affermata la notazione “/XX”
  - Es: 157.27.242.10/24
- In IPv6 rimane solo quest'ultima notazione
  - Es: 8000:0000:0000:0000:0123:4567:89AB:CDEF/112

Problema dell'assegnamento degli indirizzi

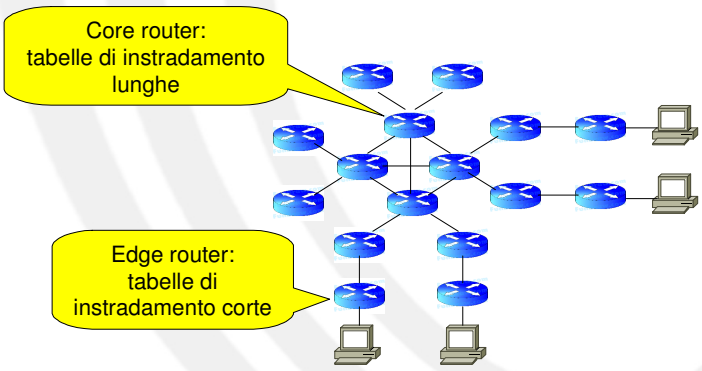
- La rete Internet, come anche la rete telefonica, ha una organizzazione gerarchica



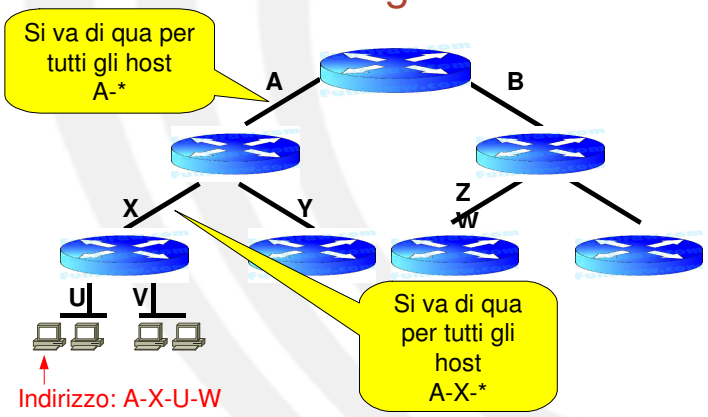
Problema dell'assegnamento degli indirizzi (2)

- I numeri del telefono riflettono la gerarchia della rete telefonica (prefisso della nazione, codice d'area, numero dell'abbonato)
  - Questo semplifica il lavoro nelle centrali di commutazione
- Gli indirizzi IPv4 invece non riflettono la gerarchia della rete Internet (in realtà tengono conto di un solo livello)
  - Conseguenza: le tabelle di routing crescono man mano che si sale verso i core router

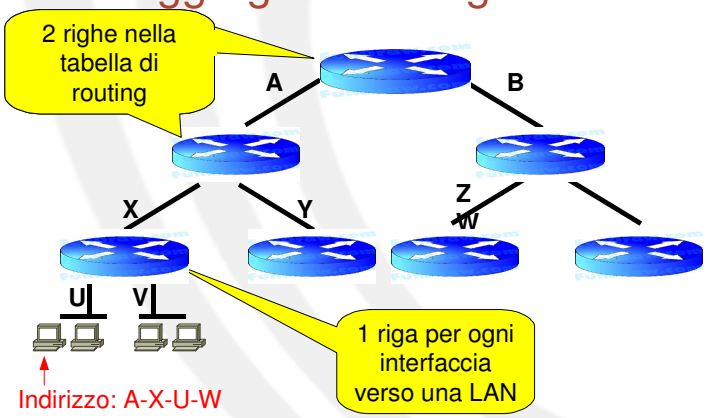
Problema dell'assegnamento degli indirizzi (3)



Indirizzi gerarchici



Aggregazione degli indirizzi

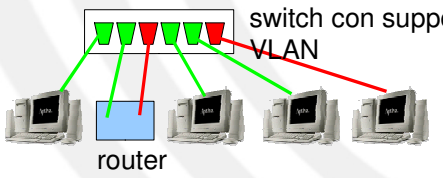


Virtual LAN

- Gli switch separano domini di collisione ma non di multicast/broadcast:
  - Protocollo ARP e malfunzionamenti generano traffico broadcast che occupa inutilmente banda
- Problemi di sicurezza:
  - Selective flooding nel transitorio
  - Possibilità di poisoning
- Soluzione: partizionamento di una LAN in tante LAN da collegare tramite router IP (creando corrispondenti sottoreti IP)

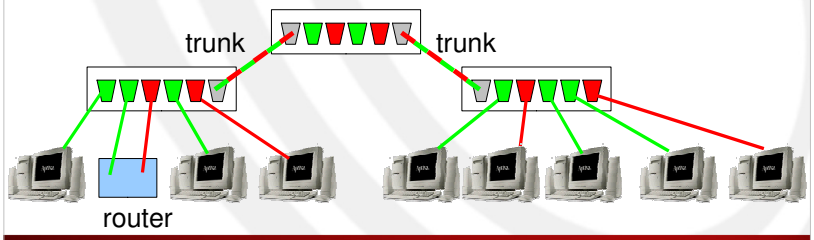
Virtual LAN (2)

- Separazione di stazioni tra LAN diverse anche se collegati allo stesso switch
  - L'amministratore decide l'assegnazione delle porte tramite SW di net-management
  - Assegnazione facile da cambiare senza spostare cavi

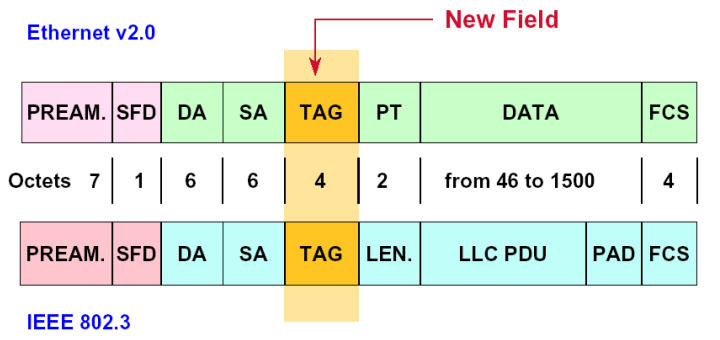


Virtual LAN (3)

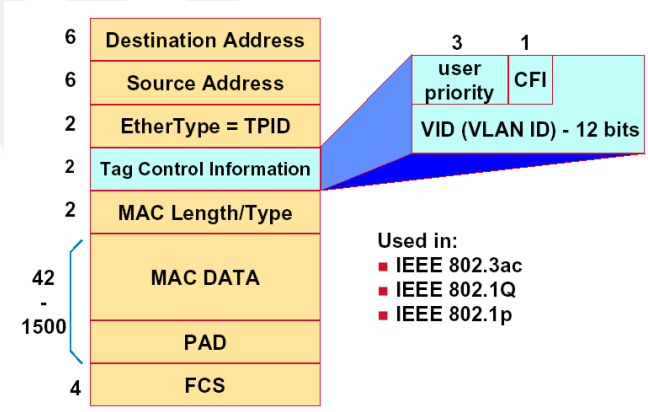
- Come distribuire VLAN su più switch?
  - Occorre scrivere un ID della VLAN nella trama ethernet (standard VLAN 802.1Q)
    - Può essere aggiunto o rimosso dagli switch LAN
    - Non crea problemi di compatibilità con le stazioni
    - Utile per dare priorità



VLAN 802.1Q



VLAN 802.1Q



Used in:

- IEEE 802.3ac
- IEEE 802.1Q
- IEEE 802.1p