Tecniche di Internetworking

Silvano GAI

Silvano@polito.it

http://www.polito.it/~silvano

Copyright: si veda nota a pag. 2

Nota di Copyright

- Questo insieme di trasparenze (detto nel seguito slides) è protetto dalle leggi sul copyright e dalle disposizioni dei trattati internazionali. Il titolo ed i copyright relativi alle slides (ivi inclusi, ma non limitatamente, ogni immagine, fotografia, animazione, video, audio, musica e testo) sono di proprietà degli autori indicati a pag. 1.
- Le slides possono essere riprodotte ed utilizzate liberamente dagli istituti di ricerca, scolastici ed universitari afferenti al Ministero della Pubblica Istruzione e al Ministero dell'Università e Ricerca Scientifica e Tecnologica, per scopi istituzionali, non a fine di lucro. In tal caso non è richiesta alcuna autorizzazione.
- Ogni altra utilizzazione o riproduzione (ivi incluse, ma non limitatamente, le riproduzioni su supporti magnetici, su reti di calcolatori e stampate) in toto o in parte è vietata, se non esplicitamente autorizzata per iscritto, a priori, da parte degli autori.
- L'informazione contenuta in queste slides è ritenuta essere accurata alla data della pubblicazione. Essa è fornita per scopi meramente didattici e non per essere utilizzata in progetti di impianti, prodotti, reti, ecc. In ogni caso essa è soggetta a cambiamenti senza preavviso. Gli autori non assumono alcuna responsabilità per il contenuto di queste slides (ivi incluse, ma non limitatamente, la correttezza, completezza, applicabilità, aggiornamento dell'informazione).
- In ogni caso non può essere dichiarata conformità all'informazione contenuta in queste slides.
- In ogni caso questa nota di copyright non deve mai essere rimossa e deve essere riportata anche in utilizzi parziali.

Copyright: si veda nota a pag. 2

Terminologia

- Un dispositivo contenente al suo interno almeno i livelli fisico, data link e network è detto
  - nodo
  - system (terminologia corretta OSI)
- I nodi che agiscono come mittente e destinatario finale dei dati sono detti:
  - ES (End System)
  - End Node

Copyright: si veda nota a pag. 2

ES e IS

- Tra due ES possono esserci diversi altri nodi che agiscono da intermediari e svolgono funzioni di instradamento. Questi nodi sono detti:
  - IS (Intermediate System)
  - router
- Tipicamente gli ES implementano tutti e sette i livelli OSI, mentre gli IS sono limitati ai primi tre.

Copyright: si veda nota a pag. 2

Altri Termini

- Router
  - Intermediate System (IS) usato dall'ISO
  - Interface Message Processor (ICMP) usato da Arpanet
  - Gateway usato impropriamente in IP
- End Node
  - End System (ES) usato dall'ISO
  - Data Terminal Equipment (DTE) usato da X.25
  - Host usato in IP

Copyright: si veda nota a pag. 2

Indirizzi e Nomi

- Un nodo è identificato da uno o più indirizzi di livello 3 e da un nome
- La finalità del nome è quella di identificare il calcolatore in modo mnemonico per gli umani
- La finalità dell'indirizzo è quella di determinare l'instradamento dei pacchetti all'interno della rete
- La corrispondenza tra nomi e indirizzi può essere mantenuta tramite:
  - un file locale ad ogni elaboratore (ormai raro)
  - una base di dati distribuita

Copyright: si veda nota a pag. 2

Il livello Network

- Tratta le problematiche di Internetworking
- Instrada i messaggi sulla rete
- Provvedere a trovare instradamenti alternativi in caso di guasti
- Può fornire sia servizi non connessi sia servizi connessi
- È molto sviluppato sugli IS, meno sugli ES.

Copyright: si veda nota a pag. 2

Tecniche di instradamento

- Routing by Network Address
- Label Swapping
- Source Routing

Copyright: si veda nota a pag. 2

Routing by Network Address

- Usato tipicamente nei protocolli non connessi
  - bridge trasparenti
  - Decnet
  - IPX
  - IPv4 e IPv6
  - OSI CLNP
- Ogni pacchetto contiene l'indirizzo del nodo destinatario
- L'indirizzo viene usato come chiave di accesso alle tabelle di instradamento

Copyright: si veda nota a pag. 2

Label Swapping

- È generalmente usato nei protocolli connessi
- Ogni pacchetto è marcato con una label che identifica la connessione
- La label viene usata come chiave per determinare l'instradamento e sostituita in ogni nodo
- È usato in:
  - X.25
  - APPN
  - ATM

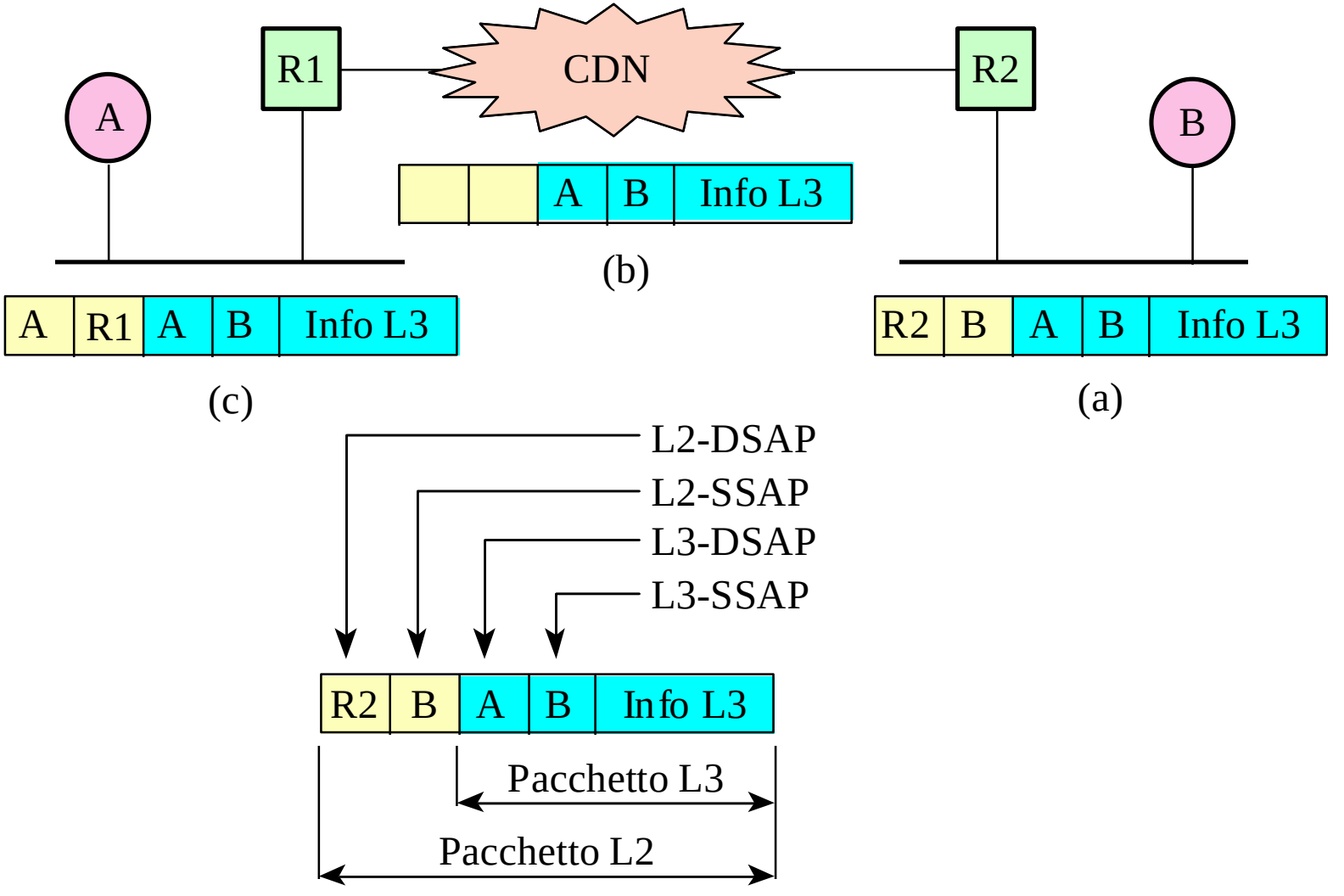
Copyright: si veda nota a pag. 2

Source Routing

- Il nodo mittente scrive nel pacchetto l'elenco completo dei nodi da attraversare
- È usato:
  - nei bridge source-routing
  - in APPN+/HPR

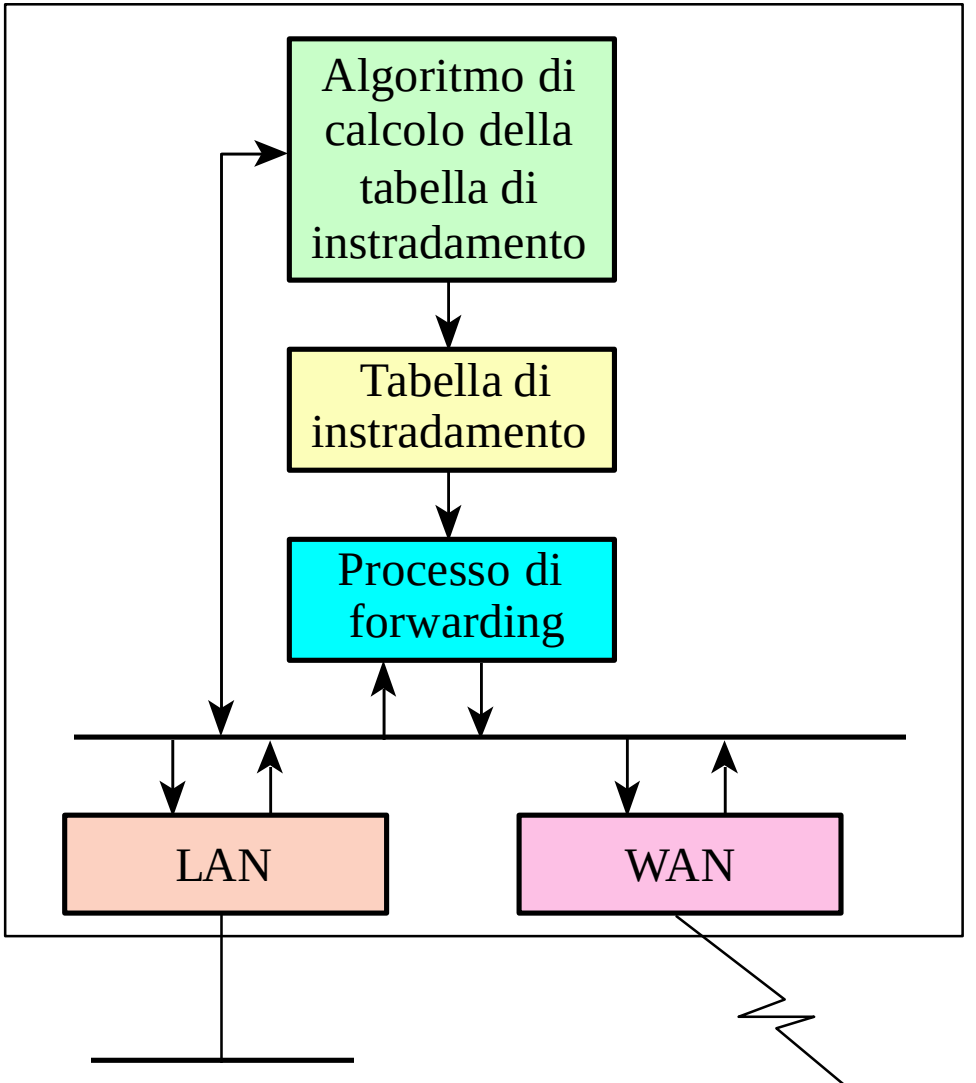
Copyright: si veda nota a pag. 2

Indirizzi di livello 2 e di livello 3



Copyright: si veda nota a pag. 2

Architettura di un Router



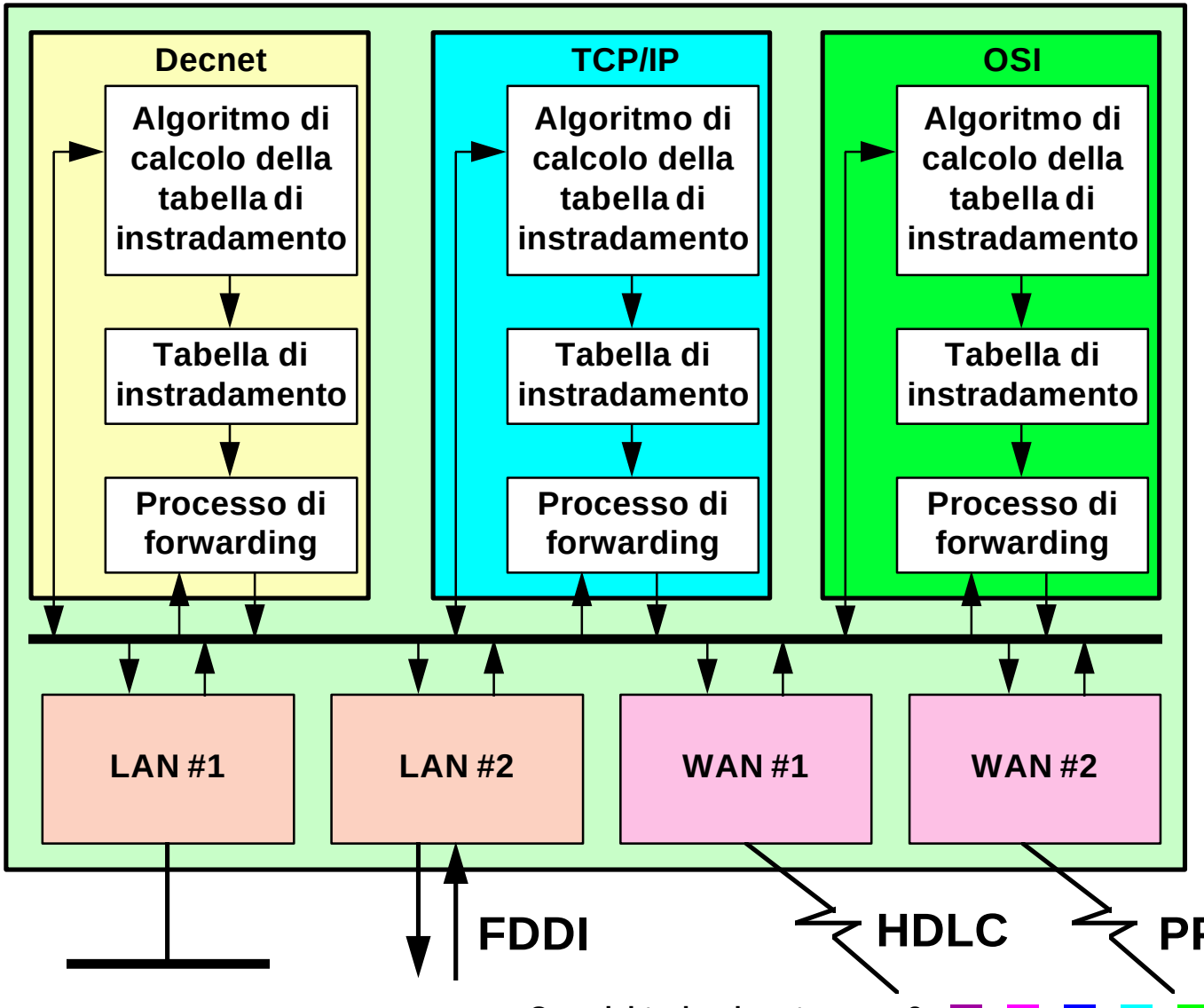
Copyright: si veda nota a pag. 2

Tabelle di instradamento

- Il livello network per instradare i pacchetti si basa:
  - sull'indirizzo del destinatario finale
  - su tabelle di instradamento presenti negli IS
- Le tabelle di instradamento possono essere:
  - scritte manualmente
  - calcolate da algoritmi che imparano la topologia della rete e si adattano ai suoi cambiamenti
- Normalmente le tabelle di instradamento sono contenute solo negli IS

Copyright: si veda nota a pag. 2

Multiprotocol Router



Copyright: si veda nota a pag. 2

Servizi Offerti

- **Servizi connessi:**
  - **CONS: Connection Oriented Network Services**
  - **ITU-T e le PTT li implementano in reti dati a pacchetto quali quelle conformi al protocollo X.25 (ISO 8208)**
- **Servizi non connessi:**
  - **CLNS: ConnectionLess Network Services o datagram**
  - **sono adottati nelle reti proprietarie quali DECNET e TCP/IP e anche nelle reti OSI (ISO 8473)**

Copyright: si veda nota a pag. 2

Servizi Offerti

ISSUE	CONS	CLNS
Initial Setup	Required	Not Possible
Addresses	During Setup	Every Packet
Packet Sequence	Guaranteed	Not Guaranteed
Error Control	Network Layer	Transport Layer
Flow Control	Provided	Not Provided
Option Neg.	Yes	No
Connection ID	Yes	No

Copyright: si veda nota a pag. 2

Scelta dell'algoritmo

- Non semplice: più criteri di ottimalità spesso contrastanti. Ad esempio:
  - minimizzare il ritardo medio di ogni pacchetto
  - massimizzare l'utilizzo delle linee
- Complicata dalla presenza di un elevato numero di nodi collegati con una topologia qualsiasi
- Algoritmi troppo complessi, operanti su reti molto grandi, potrebbero richiedere tempi di calcolo inaccettabili

Copyright: si veda nota a pag. 2

Esempio

- Si supponga di voler inviare un messaggio a tutti i nodi di una rete carica evitando l'utilizzo di tecniche di broadcast per contenere il carico:
  - occorre individuare un algoritmo di routing che garantisca che il pacchetto tocchi sequenzialmente tutti i nodi
- Per ottimizzare il cammino di questo pacchetto:
  - occorre risolvere il tipico problema del Commesso Viaggiatore che è NP-Hard

Copyright: si veda nota a pag. 2

Caratteristiche di un algoritmo

- Semplicità: i router hanno CPU e memoria finite
- Robustezza e adattabilità alle variazioni di topologia
- Stabilità: l'algoritmo deve convergere
- Equità: evitare di danneggiare particolari nodi

Copyright: si veda nota a pag. 2

Ottimalità

- Ottimalità: i criteri adottati sono tipicamente i due esposti precedentemente.
- Occorre definire una metrica. Gli unici due parametri universalmente accettati sono:
  - HOPS*:* numero di salti effettuati, cioè il numero di IS attraversati lungo il cammino
  - COSTO*:* somma dei costi di tutte le linee attraversate (il costo di una linea è inversamente proporzionale alla sua velocità)

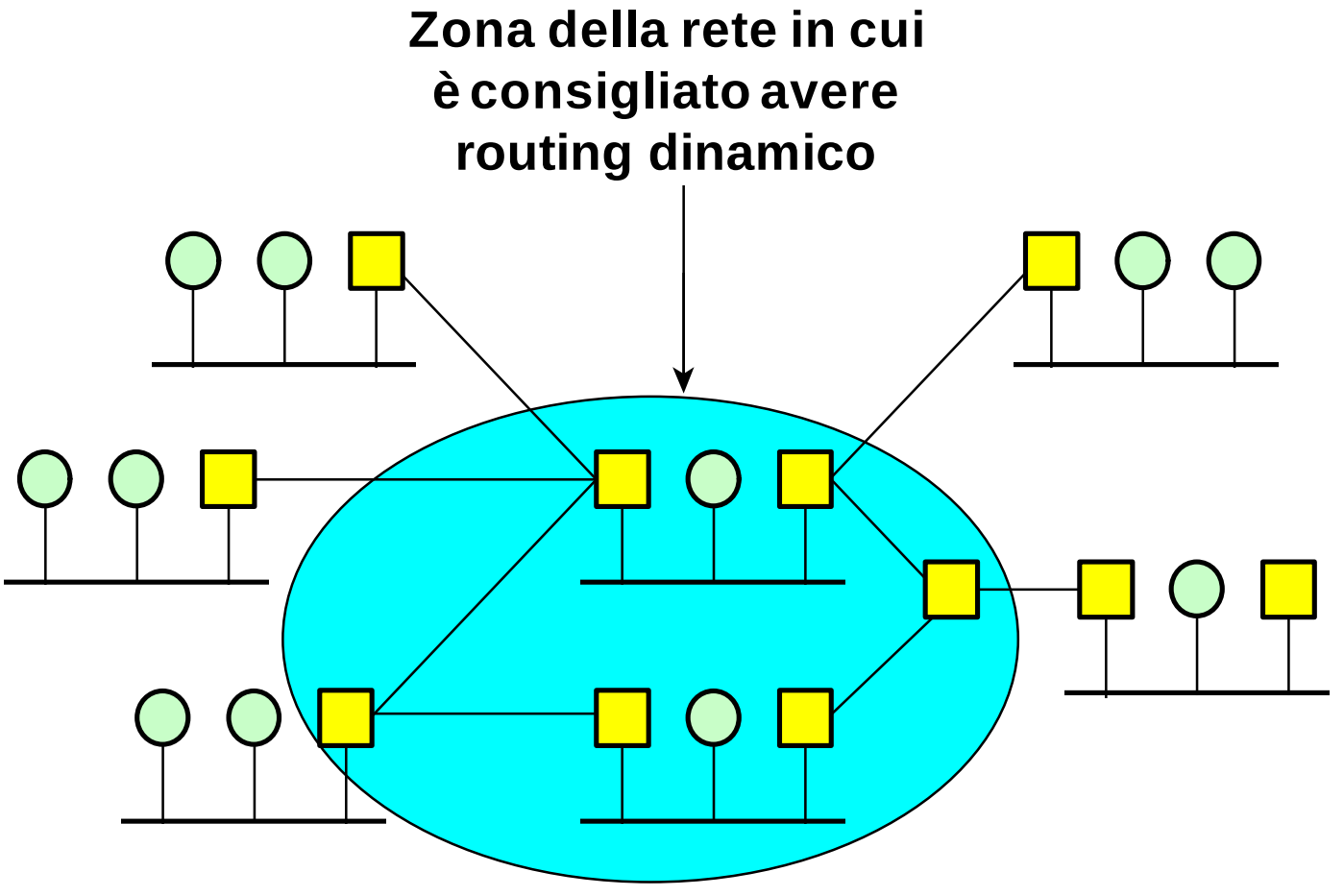
Copyright: si veda nota a pag. 2

Algoritmi di Routing

- Non adattativi (statici, deterministici): criteri fissi di instradamento
  - Fixed Directory Routing
  - Flooding
- Adattativi (dinamici, non deterministici): calcolano le tabelle di instradamento in funzione della topologia della rete, dello stato dei link e del carico
  - Routing Centralizzato
  - Routing Isolato
  - Routing Distribuito

Copyright: si veda nota a pag. 2

Routing Statico e Dinamico



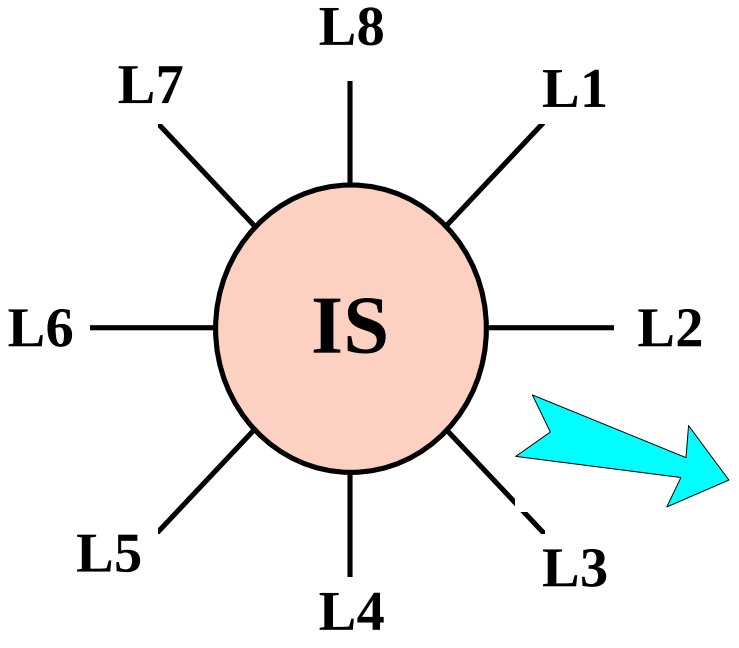
Copyright: si veda nota a pag. 2

Fixed Directory Routing

- Ogni nodo ha una tabella di instradamento scritta manualmente (gestione complessa)
  - Indirizzo di destinazione --> linea da usarsi
- Il gestore della rete ha il totale controllo dei flussi di traffico
  - interventi manuali per ridirigere il traffico in presenza di guasti
  - Questo secondo punto può essere parzialmente evitato adottando tabelle di priorità che diano più scelte in caso di guasto
- Usato con successo in reti TCP/IP e SNA

Copyright: si veda nota a pag. 2

Fixed Directory Routing



Indirizzo	Prima scelta	Seconda scelta
pol88a	L2	L4
vaxrom	L4	L3
vaxto	L2	L5
infngw	L1	L7
poldidL6	-	
vaxlnfL2	L4	

Copyright: si veda nota a pag. 2

Flooding

- Algoritmo non adattativo
- Ciascun pacchetto in arrivo viene ritrasmesso su tutte le linee eccetto quella su cui è stato ricevuto
- Adatto a carichi bassi
- Si può migliorare:
  - introducendo un age-counter
  - scartando i pacchetti la seconda volta che passano in un nodo (richiede memoria nei nodi)

Copyright: si veda nota a pag. 2

Selective Flooding

- I pacchetti vengono ritrasmessi solo su linee selezionate:
  - **Random Walk**. Il pacchetto in arrivo su un nodo viene trasmesso in modo casuale su una delle linee disponibili
  - **Hot Potato**. Ogni nodo ritrasmette il pacchetto sulla linea con la coda di trasmissione più breve

Copyright: si veda nota a pag. 2

Routing Centralizzato

- Esiste un Routing Control Center (RCC) che calcola e distribuisce le tabelle
- Il RCC riceve informazioni sullo stato della rete da tutti i nodi e le usa per calcolare le nuove tabelle
- È usato nella rete TYMNET
- Ottimizza le prestazioni, ma è poco robusto (aggiornamenti parziali delle tabelle dovuti a guasti possono generare loop)
- Induce un notevole carico sulla rete specialmente in prossimità del RCC

Copyright: si veda nota a pag. 2

Routing Isolato

- È l'opposto del Routing Centralizzato in quanto non esiste un RCC e ogni nodo decide l'instradamento in modo autonomo senza scambiare informazioni con gli altri IS
- Ad esempio: Backward Learning
  - Usato dai bridge 802.1D (integrato dall'algoritmo di spanning tree)

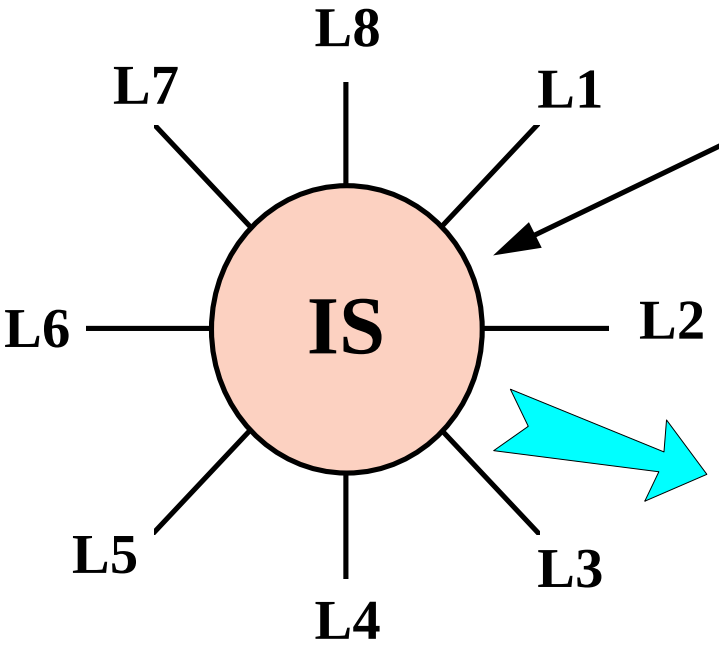
Copyright: si veda nota a pag. 2

Backward Learning

- Ogni pacchetto ha un campo in cui vengono sommati i costi di tutte le linee attraversate
- Gli IS osservano questo campo nei pacchetti ricevuti e imparano quanto dista il mittente tramite la linea su cui è avvenuta la ricezione
- Occorre limitare la validità temporale delle annotazioni nella tabella in quanto questa tecnica non è in grado di rilevare se un cammino non è più disponibile

Copyright: si veda nota a pag. 2

Routing Isolato



Su L1 giunge un pacchetto da vaxrom con costo 6

Costi

pol88a	L2	12	L4	14
vaxrom	L4	11	L3	12
vaxto	L2	5	L5	34
infngw	L1	33	L7	34
poldid	L6	26	-	-
vaxlnf	L2	8	L4	12

La riga della tabella relativa a vaxrom diviene:

vaxrom	L1 6	L4 11

Copyright: si veda nota a pag. 2

Routing Distribuito

- Utilizzato in varie reti proprietarie (es. DECNET) e in OSI
- Ogni router calcola le sue tabelle dialogando con gli altri router e con gli end-node
- Tale dialogo avviene tramite dei protocolli ausiliari a livello 3
- Esistono due approcci principali al routing distribuito:
  - Algoritmi Distance Vector
  - Algoritmi Link State

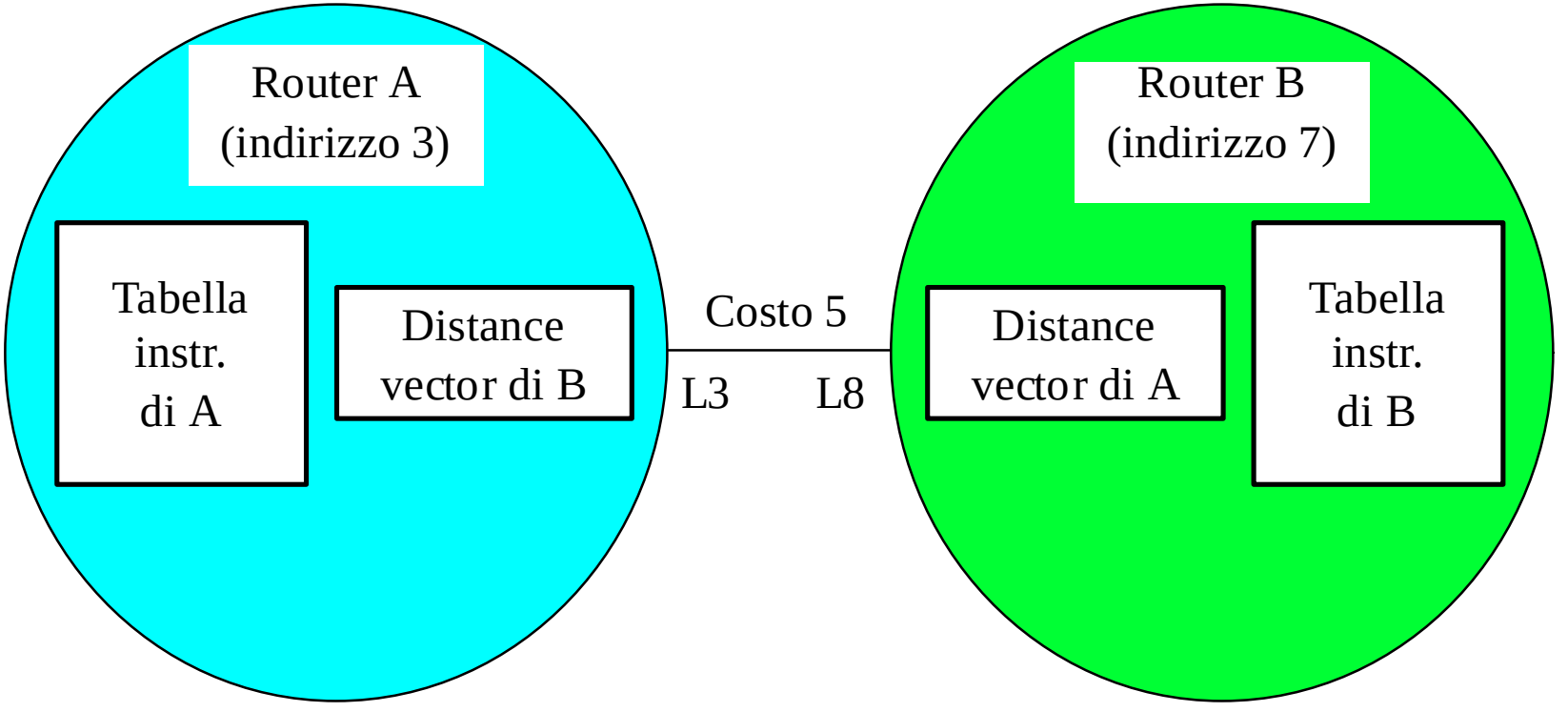
Copyright: si veda nota a pag. 2

Distance Vector

- Noto anche come algoritmo di Bellman-Ford
- Ogni nodo, quando modifica le proprie tabelle di instradamento, invia ai nodi adiacenti un distance vector
- Il distance vector è un insieme di coppie
  - [indirizzo - distanza]
- La distanza è espressa tramite metriche classiche quali numero di hops e costo
- Ogni nodo memorizza per ogni linea l'ultimo distance vector ricevuto

Copyright: si veda nota a pag. 2

Esempio Distance Vector



Copyright: si veda nota a pag. 2

Distance Vector

- Un router ricalcola le sue tabelle se:
  - cade una linea attiva
  - riceve un distance vector da un nodo adiacente diverso da quello memorizzato.
- Il calcolo consiste nella fusione di tutti i distance vector delle linee attive
- Se le tabelle risultano diverse da quelle precedenti, invia ai nodi adiacenti un nuovo distance vector
- L'algoritmo ha un worst case esponenziale e un comportamento normale tra $O(n^2)$ e $O(n^3)$

Copyright: si veda nota a pag. 2

Distance Vector: caratteristiche

- Vantaggi:
  - Molto semplice da implementare
- Svantaggi
  - Possono innescarsi dei loop a causa di particolari variazioni della topologia
  - Converge alla velocità del link più lento e del router più lento
  - Difficile capirne e prevederne il comportamento su reti grandi: nessun nodo ha una mappa della rete!

Copyright: si veda nota a pag. 2

Es.: tabella di instradamento di A

Indirizzo	Hops	Costo	Linea
1	5	25	3
2	3	20	2
3	0	0	0
4	2	15	3
5	7	55	1
6	4	23	1
7	1	5	3
...	...	...	...

Copyright: si veda nota a pag. 2

Es.: Distance Vector di A in B

Indirizzo	Hops	Costo
1	6	30
2	4	25
3	1	5
4	3	20
5	8	60
6	5	28
7	2	10
...	...	...

Copyright: si veda nota a pag. 2

Es.: fusione di Distance Vector

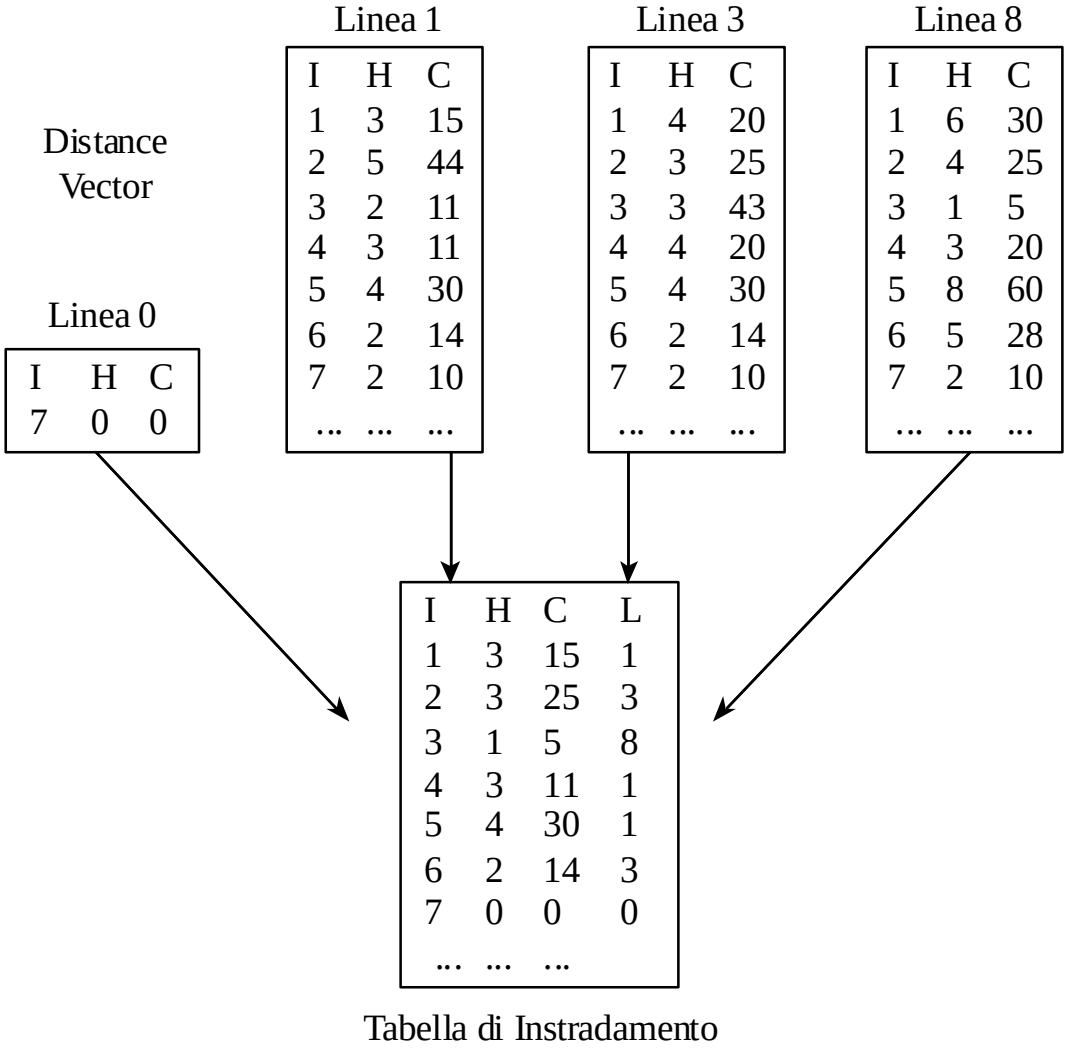


Tabella di Instradamento

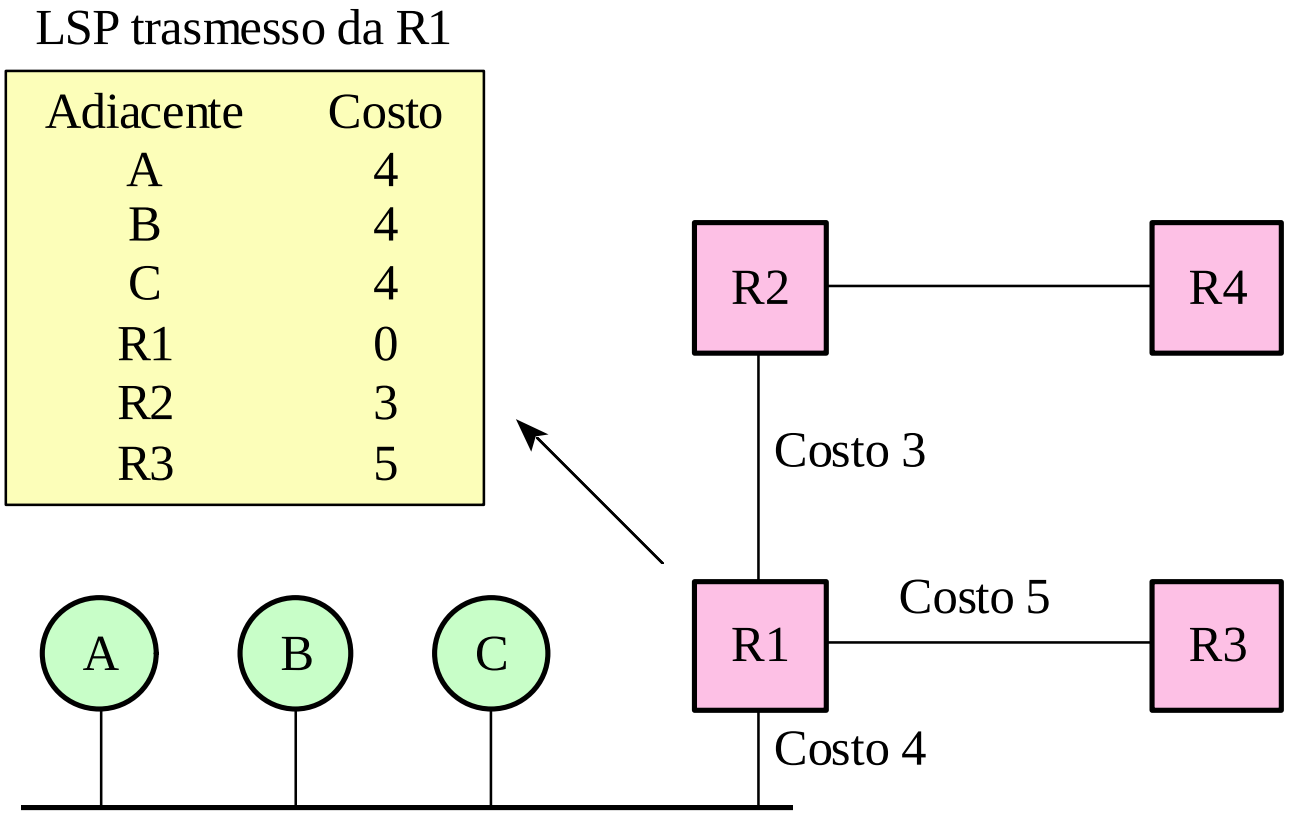
Copyright: si veda nota a pag. 2

Link State

- Ogni router impara il suo ambito locale: linee e nodi adiacenti
- Trasmette queste informazioni a tutti gli altri router della rete tramite un Link State Packet (LSP)
- Tutti i router, memorizzando i LSP trasmessi dagli altri router, si costruiscono una mappa della rete

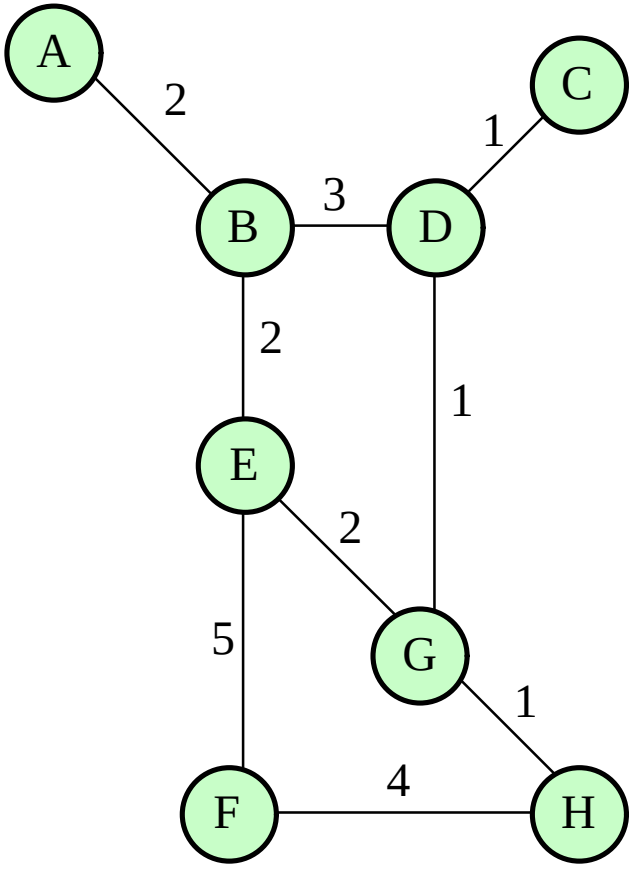
Copyright: si veda nota a pag. 2

Esempio: trasmissione di un LSP



Copyright: si veda nota a pag. 2

Esempio: grafo della rete e LSP-DB



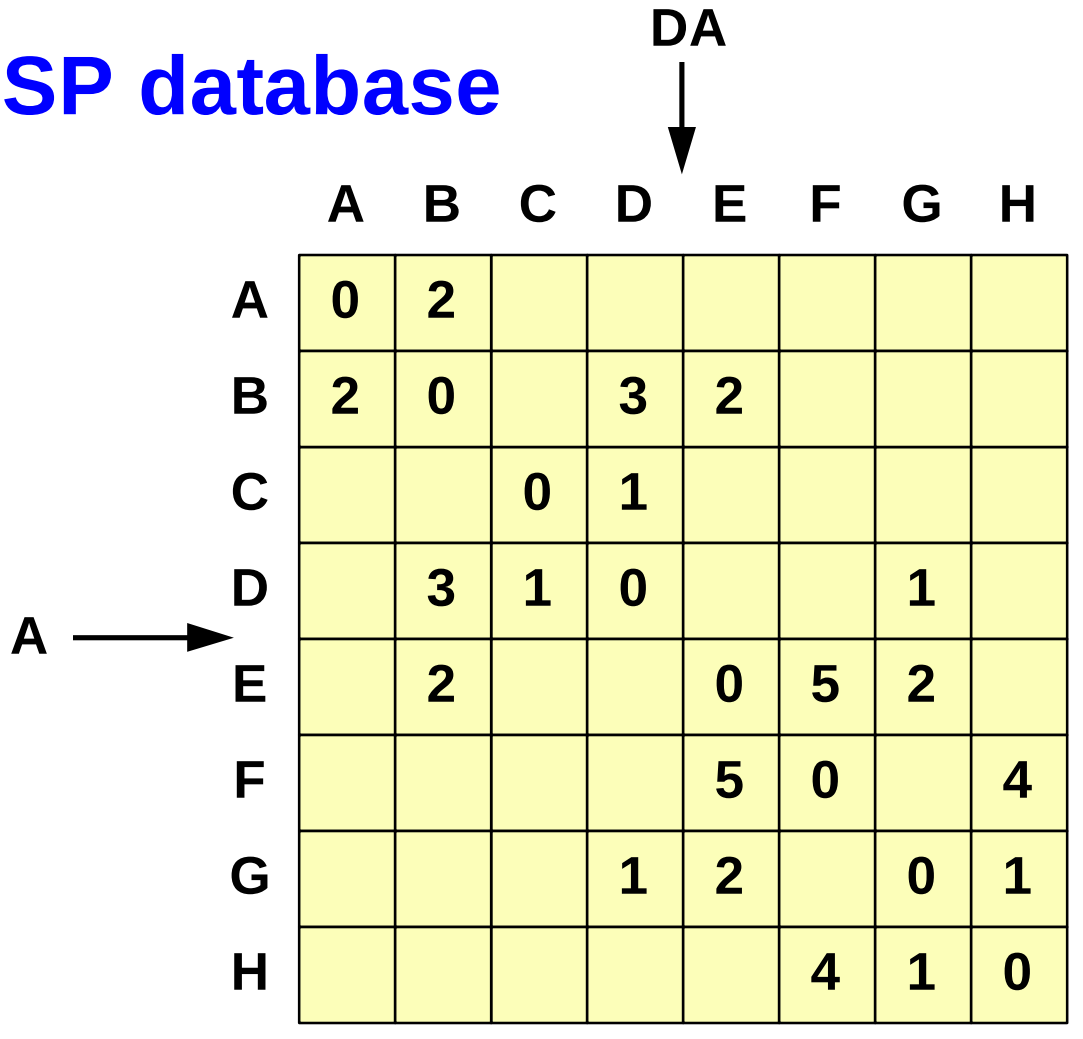
LSP Database

A	B/2		
B	A/2	D/3	E/2
C	D/1		
D	B/3	C/1	G/1
E	B/2	F/5	G/2
F	E/5	H/4	
G	D/1	E/2	H/1
H	F/4	G/1	

(replicato su ogni IS)

Copyright: si veda nota a pag. 2

LSP database



Questa rappresentazione è quella più appropriata per applicare l'agoritmo di Dijkstra

Copyright: si veda nota a pag. 2

Link State

- Ogni router calcola indipendentemente le sue tabelle di instradamento applicando alla mappa della rete l'algoritmo di Dijkstra o SPF (Shortest Path First)
- La complessità è ElogN
  - E è il numero di link
  - N è il numero di nodi
- È utilizzato nello standard ISO 10589 (IS-IS) e nel protocollo OSPF (adottato in alcune reti TCP/IP)

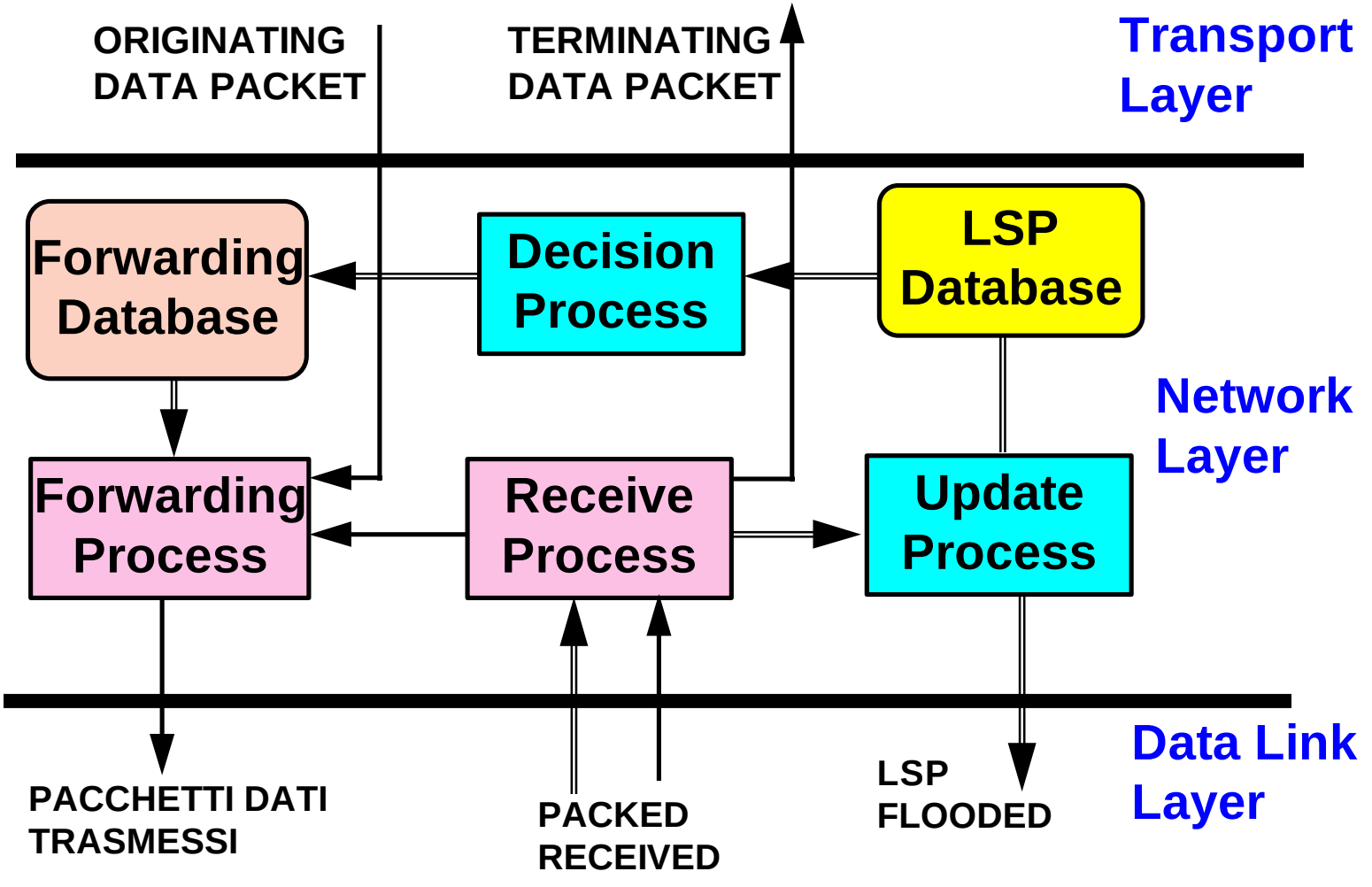
Copyright: si veda nota a pag. 2

Link State: caratteristiche

- Vantaggi:
  - Può gestire reti di grandi dimensioni
  - Ha una convergenza rapida
  - Difficilmente genera loop, e comunque è in grado di identificarli e interromperli facilmente
  - Facile da capire: ogni nodo ha la mappa della rete
- Svantaggi:
  - Molto complesso da realizzare (la prima implementazione ha richiesto a Digital 5 anni)
  - Necessita di meccanismi speciali per le LAN

Copyright: si veda nota a pag. 2

Architettura di un router Link State



Copyright: si veda nota a pag. 2

Routing: Update Process

- Ogni router genera un Link State Packet (LSP) che contiene:
  - Stato di ogni link connesso al router
  - Identità di ogni vicino connesso all'altro estremo del link (sulle LAN possono esserci migliaia di vicini)
  - Costo del link
  - Numero di sequenza per il LSP
  - Checksum
  - Lifetime

Copyright: si veda nota a pag. 2

LSP flooding

- Il LSP è trasmesso in flooding su tutti i link del router
- I pacchetti LSP memorizzati nei router formano una mappa completa e aggiornata della rete:
  - Link State Database

Copyright: si veda nota a pag. 2

Ricezione di un LSP

- All'atto del ricevimento di un LSP il router compie le seguenti azioni:
  - Se non ha mai ricevuto LSP da quel router o se il LSP è più recente di quello precedentemente memorizzato, memorizza il pacchetto e lo ritrasmette in flooding su tutte le linee eccetto quella da cui l'ha ricevuto
  - Se il LSP ha lo stesso numero di sequenza di quello posseduto non fa nulla
  - Se il LSP è più vecchio di quello posseduto trasmette al mittente il pacchetto più recente

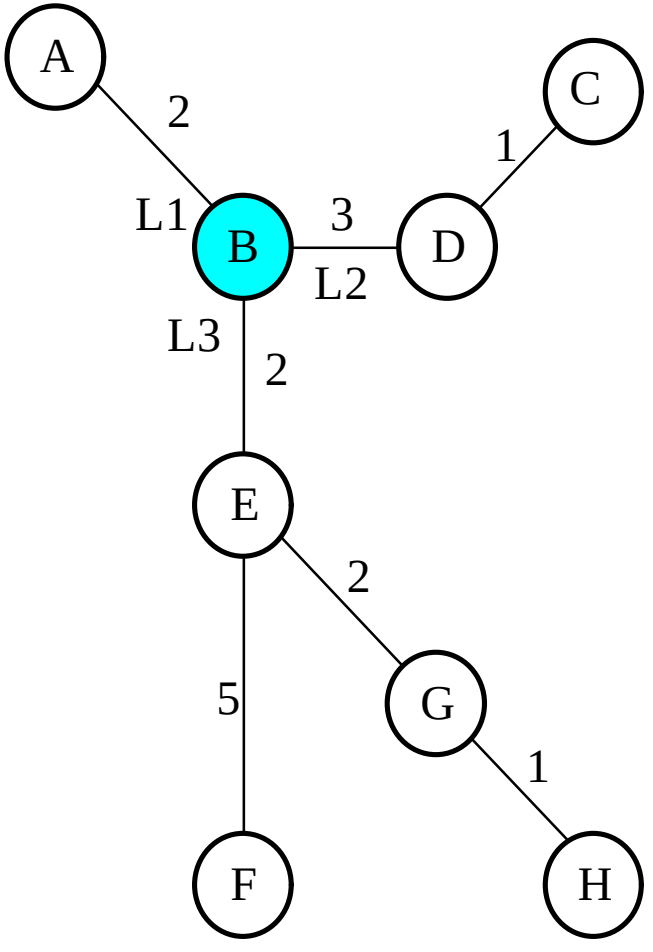
Copyright: si veda nota a pag. 2

Routing: decisioni

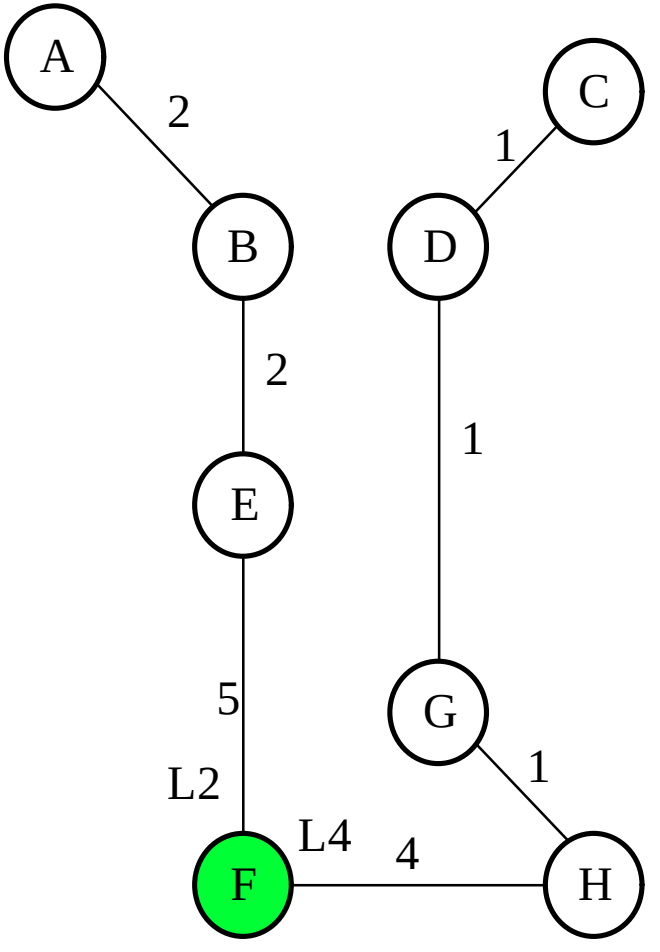
- Il router elabora il Link State Database per produrre il Forwarding Database
  - Si pone come radice dello shortest-path tree
  - Cerca lo shortest path per ogni destination node
  - Memorizza il vicino (i vicini) che sono sullo shortest path di ogni destination node
- Il forwarding Database contiene per ogni nodo destinazione:
  - insieme delle coppie {circuito, vicino}
  - dimensione dell'insieme determinata dal parametro MaxSplit

Copyright: si veda nota a pag. 2

Es.: alberi di instradamento



(a)



(b)

Copyright: si veda nota a pag. 2

Es.: tabelle di instradamento

Tabella di B

A	L1
C	L2
D	L2
E	L3
F	L3
G	L3
H	L3

Tabella di F

A	L2
B	L2
C	L4
D	L4
E	L2
G	L4
H	L4

Copyright: si veda nota a pag. 2

Algoritmo di Dijkstra

- Ogni nodo ha a disposizione il grafo della rete (dove i nodi sono i router e gli archi le linee) e agli archi è associata una metrica
- Ogni nodo usa l'algoritmo di Dijkstra per costruire lo Spanning Tree del grafo (albero dei cammini di costo minimo)
- Ad ogni nodo si assegna una etichetta che rappresenta il costo massimo per la raggiungibilità di quel nodo
- L'algoritmo modifica le etichette cercando di minimizzarle e di renderle permanenti

Copyright: si veda nota a pag. 2

Strutture dati per l'algoritmo di Dijkstra

- Sia V l'insieme dei vertici del grafo
- Sia C[i,j] il costo della connessione diretta da i a j, assunto infinito se tale connessione non esiste
- Sia D[i] il costo del cammino dal vertice 1 al vertice i
- L'algoritmo inizializza D[i] = C[1,i] e poi iterativamente cerca di minimizzare D[i]
- L'algoritmo usa un insieme S in cui memorizza quali vertici hanno già un valore di D definitivo

Copyright: si veda nota a pag. 2

Procedura

```
procedure Dijkstra; (* calcola il costo del cammino più
   corto dal nodo 1 a tutti gli altri nodi *)
begin
    S := {1};
    for i := 2 to n do D[i] := C[1,i]; (* inizializza D *)
    for i := 1 to n-1 do
    begin
        si scelga un nodo w in V-S tale che D[w] sia minimo;
        si aggiunga w a S;
        per ogni vertice v in V-S do D[v] := min(D[v], D[w] + C[w,v])
    end
end; (* Dijkstra *)
```


Copyright: si veda nota a pag. 2

Esempio di LSP database

	A	B	C	D	E	F	G	H
A	0	2						
B	2	0		3	2			
C			0	1				
D		3	1	0			1	
E		2			0	5	2	
F					5	0		4
G				1	2		0	1
H						4	1	0

Questa è la matrice C[i,j] dove 1=A, 2=B, 3=C, ecc.

Copyright: si veda nota a pag. 2

Routing: complessità

- La complessità del calcolo dello spanning tree è *ElogN* dove:
  - E è il numero di links
  - N è il numero di nodi
- Poichè i costi dei link sono numeri interi piccoli si riescono a realizzare strutture dati sofisticate che fanno tendere questo valore a N
- Su un processore da 1 MIPS inserito in una rete con 600 nodi e 300 link il tempo di calcolo è di circa 150 millisecondi

Copyright: si veda nota a pag. 2

Routing: le LAN

- L'algoritmo Link State presuppone di lavorare su canali punto-punto (gli archi del grafo)
  - Le reti locali si comportano canali broadcast!
- Occorre modellarle le LAN tramite una struttura equivalente punto-punto
  - La più semplice struttura equivalente è una maglia completa
  - Non accettabile per la crescita quadratica del numero di link al crescere del numero di nodi!

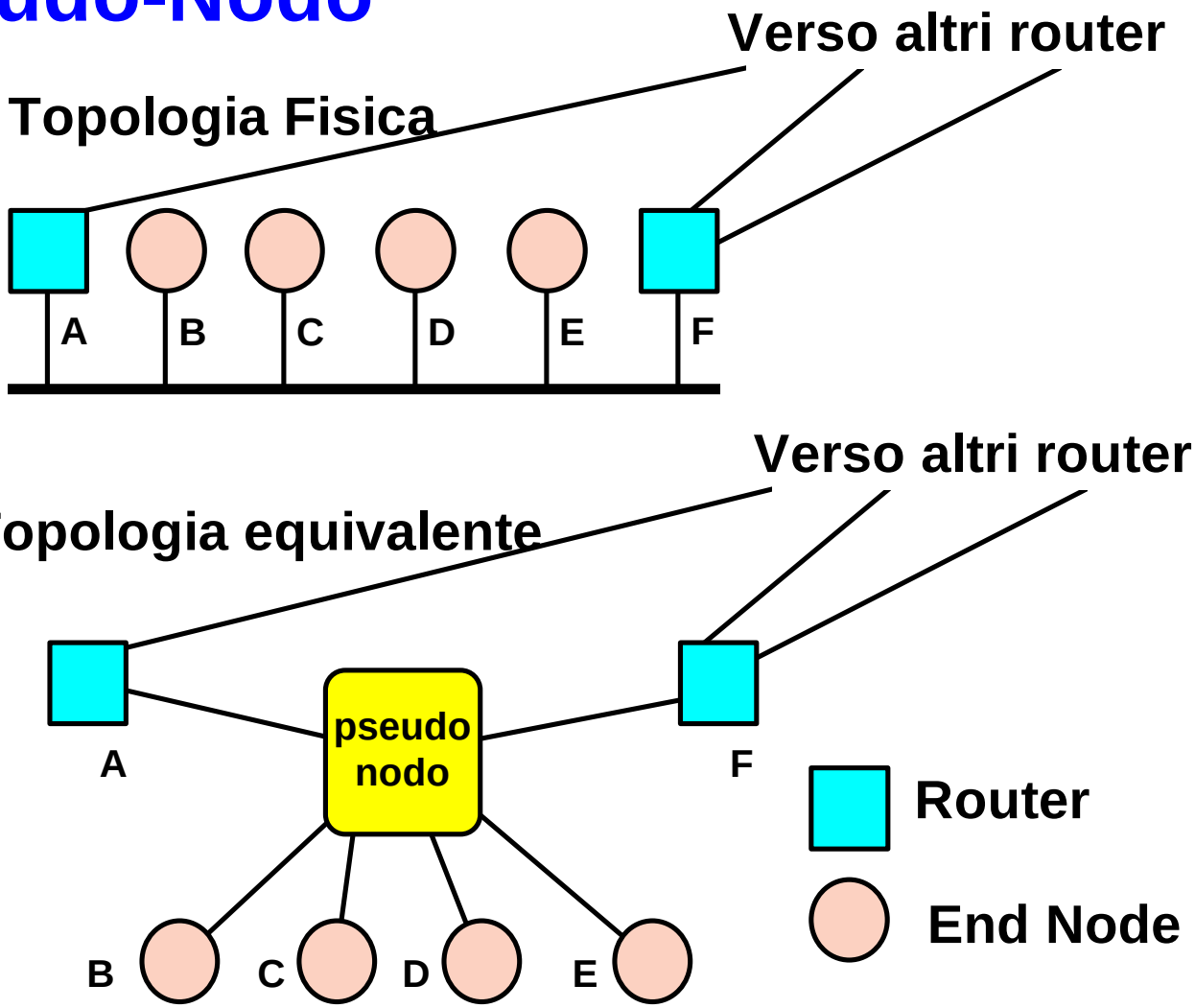
Copyright: si veda nota a pag. 2

Lo Pseudo-Nodo

- Si introduce il concetto di pseudo-nodo. È un nodo fittizio non esistente sulla rete che viene realizzato da uno dei router presenti sulla LAN (designated router)
- La topologia equivalente è una stella con al centro lo pseudo-nodo

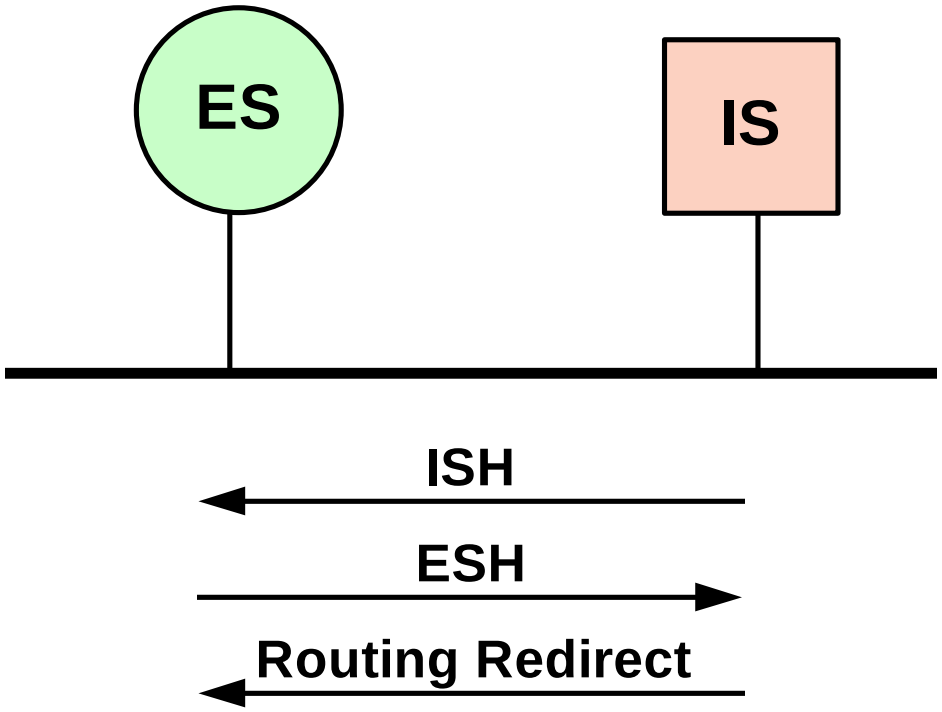
Copyright: si veda nota a pag. 2

Lo Pseudo-Nodo



Copyright: si veda nota a pag. 2

Neighbor Greetings



Copyright: si veda nota a pag. 2

ES connessi ad una LAN

- Inviano periodicamente un End System Hello (ESH) per informare gli IS della loro esistenza
- Gli IS inviano periodicamente un Intermediate System Hello (ISH) per lo stesso scopo
- Un IS “R” invia un Redirect Message ad un ES “E”, tutte le volte che E invia un pacchetto a R destinato a “D” ed R stabilisce che D è raggiungibile direttamente tramite la LAN o tramite un router collegato alla LAN

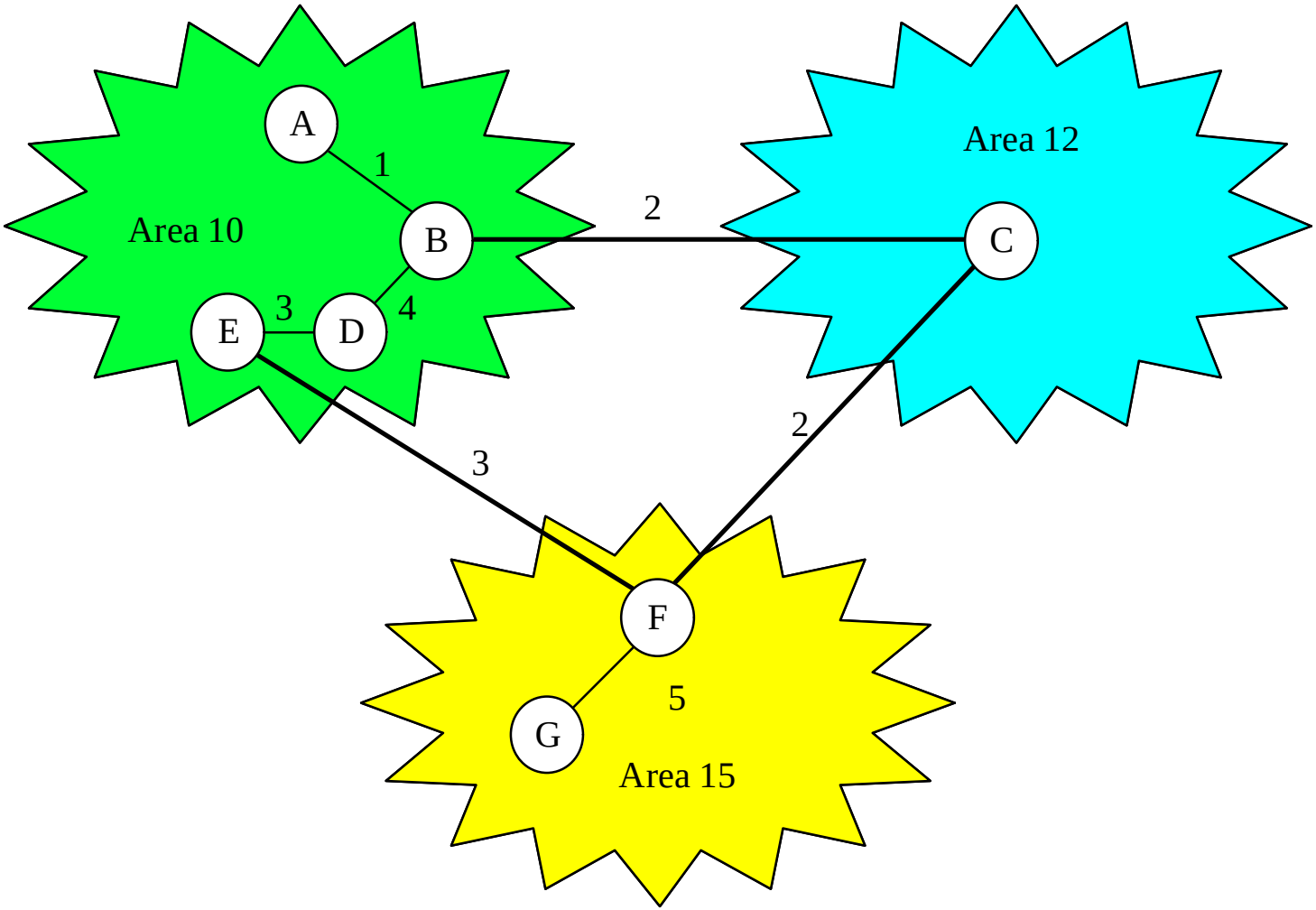
Copyright: si veda nota a pag. 2

Routing gerarchico

- Gli algoritmi precedenti spesso non possono o devono operare su tutta la rete
- La rete viene divisa gerarchicamente in domini (SNA, OSI), aree (OSI, DECNET) e/o network e subnetwork (TCP/IP)
- Il routing opera allora a più livelli, ad esempio all'interno dell'area, tra aree diverse, tra domini diversi
- Un routing gerarchico migliora le prestazioni della rete, seppure in taluni casi può portare ad instradamenti non ottimi, ma corretti

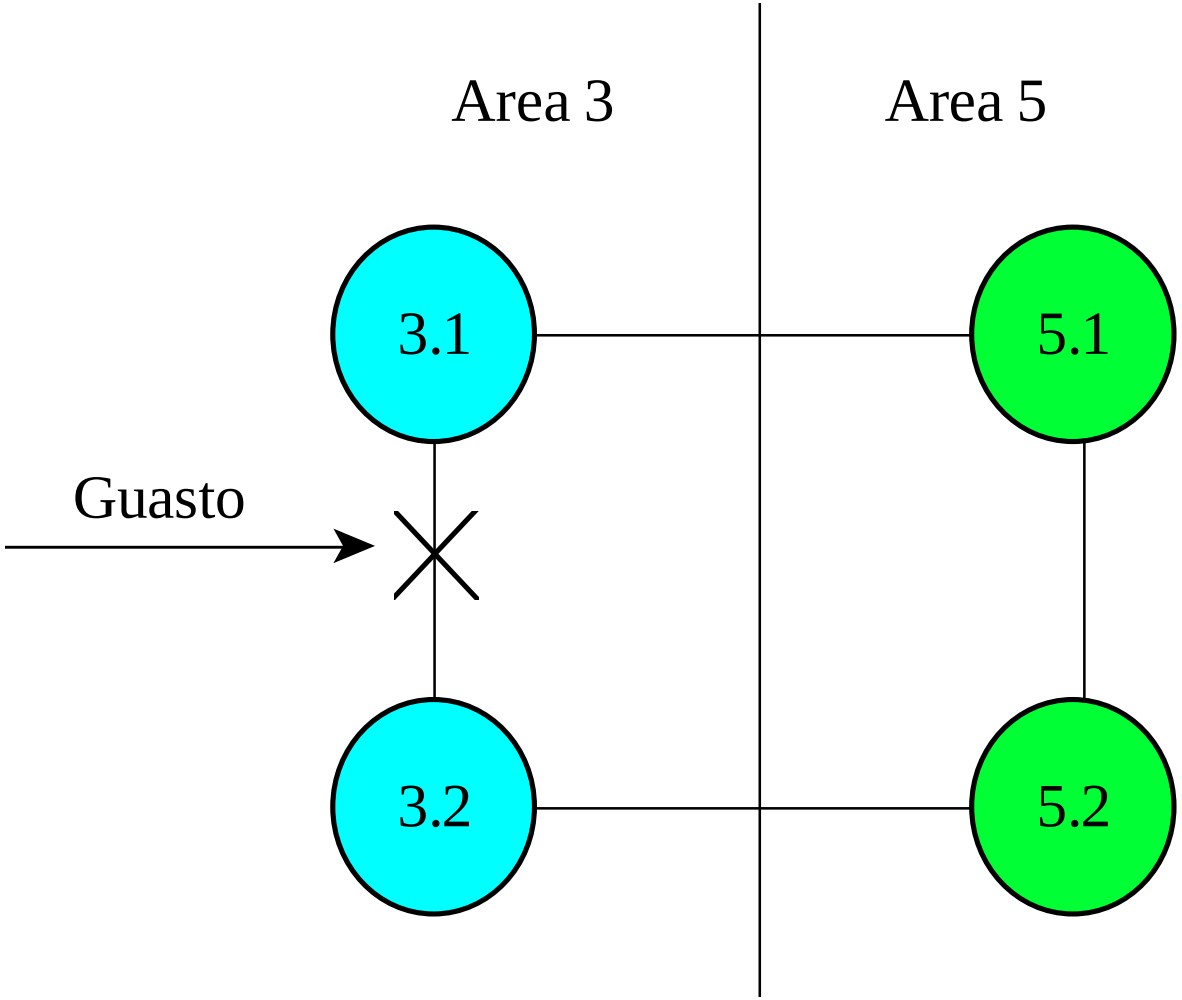
Copyright: si veda nota a pag. 2

Es.: Routing gerarchico



Copyright: si veda nota a pag. 2

Area partizionata



Copyright: si veda nota a pag. 2