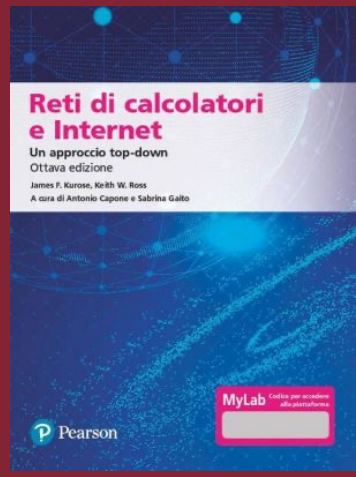
Reti di Elaboratori

Livello di Rete: Dentro i router



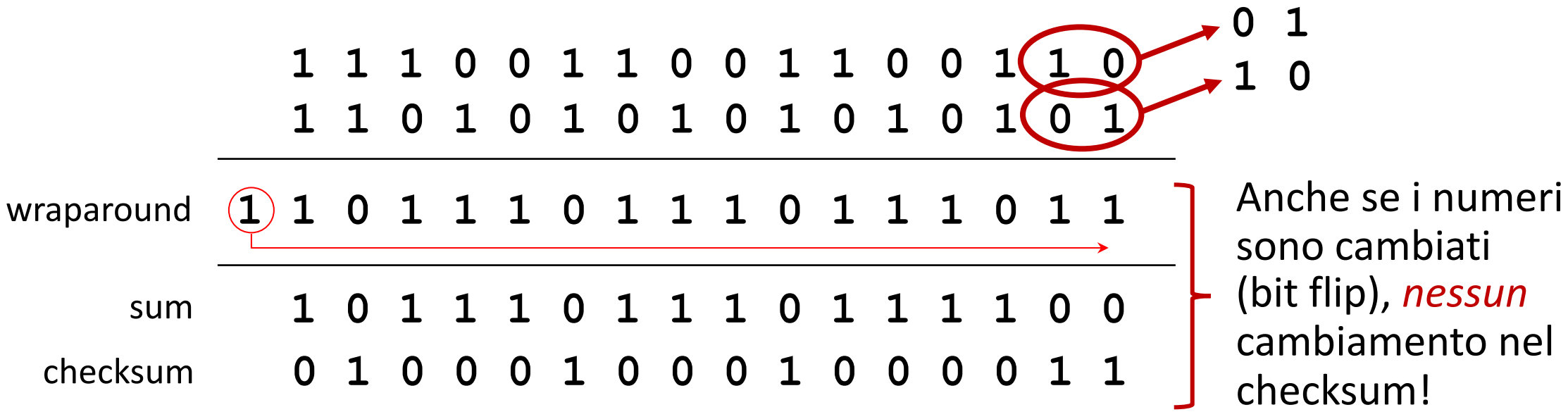
SAPIENZA
UNIVERSITÀ DI ROMA

Alessandro Checco
alessandro.checco@uniroma1.it

Capitolo 4

Checksum Internet: chiarimento (2)

esempio: somma di due numeri interi a 16 bit



Somma 16 bit alla volta dell'header UDP (senza checksum) + pseudoheader IP + data -> complemento a 1 -> checksum

perché?

Se il risultato è 0 va invertito di nuovo (tutti 1)
Se non viene usato checksum = 0

Perché usare complemento a 1 nel checksum

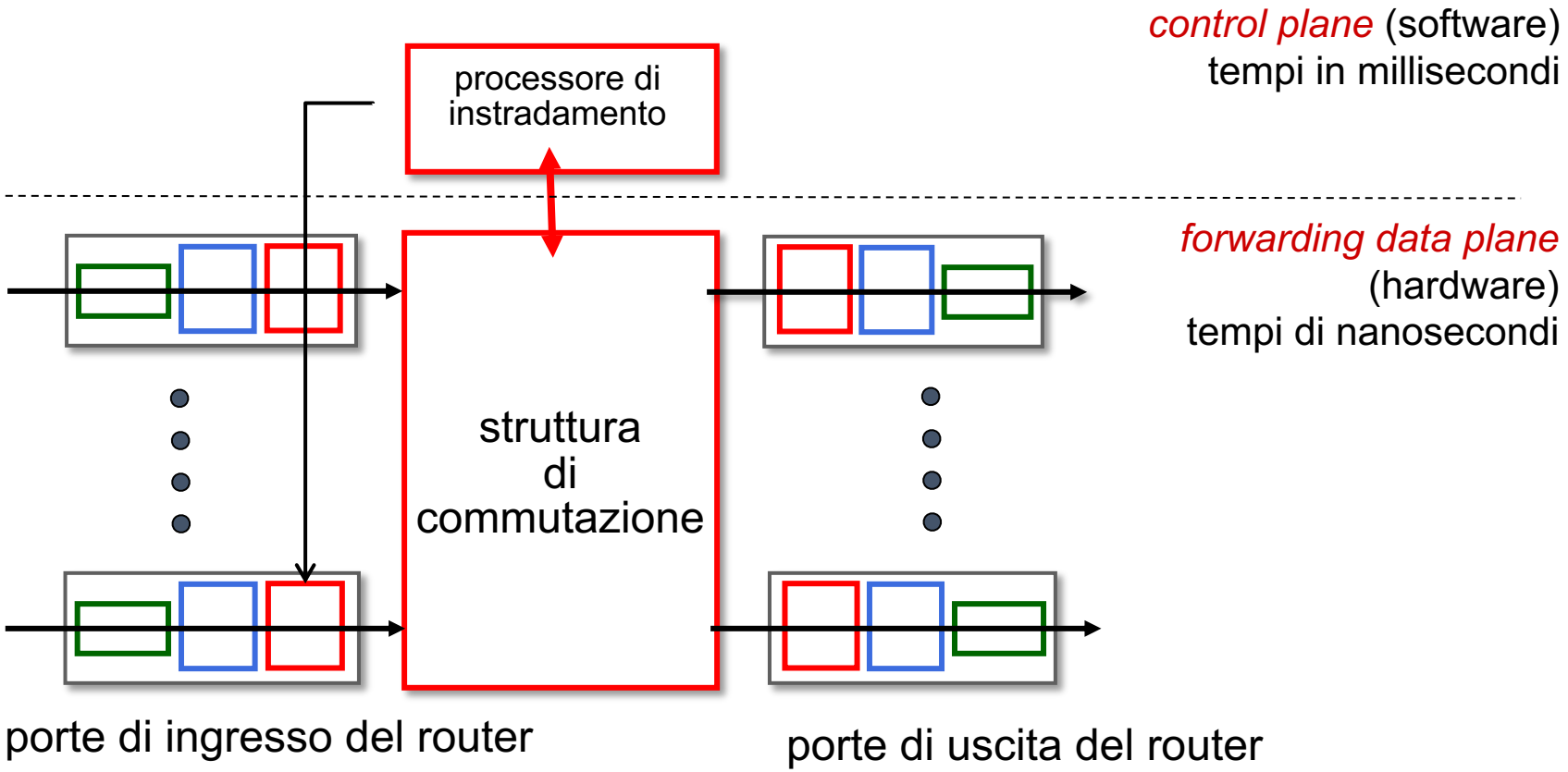
- In pratica non compariamo i due checksum per verificare che siano uguali
- Sommando il campo checksum a quello calcolato in ricezione se non ci sono errori viene 0XFFFF (tutti 1). Il complemento a uno è 0x0000
- Le CPU possono fare il complemento a 1 + controllare che il risultato sia 0 molto velocemente rispetto a comparare due numeri
- Il motivo di invertire di nuovo quando il checksum è zero serve a distinguire un messaggio con checksum 0 da un memory erasure o dal caso in cui il checksum non viene usato

Livello di rete: sommario

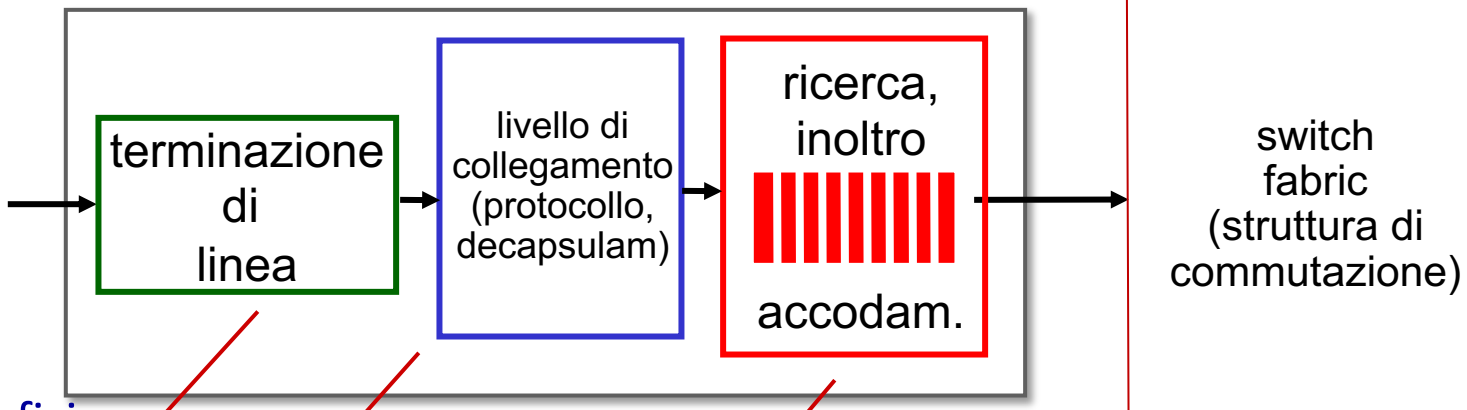
- Livello di rete: panoramica
  - piano dati
  - piano di controllo
- Dentro i router
  - porte di ingresso, commutazione, porte di uscita
  - gestione del buffer, scheduling
- IP: il protocollo Internet
  - formato datagramma
  - indirizzamento
  - traduzione di indirizzi di rete
  - IPv6
- Forwarding generalizzato, SDN
  - Match+action
  - OpenFlow: incontro+azione in azione
- Middleboxes

Panoramica dell'architettura del router

vista di alto livello dell'architettura generica del router:



Funzioni della porta di ingresso



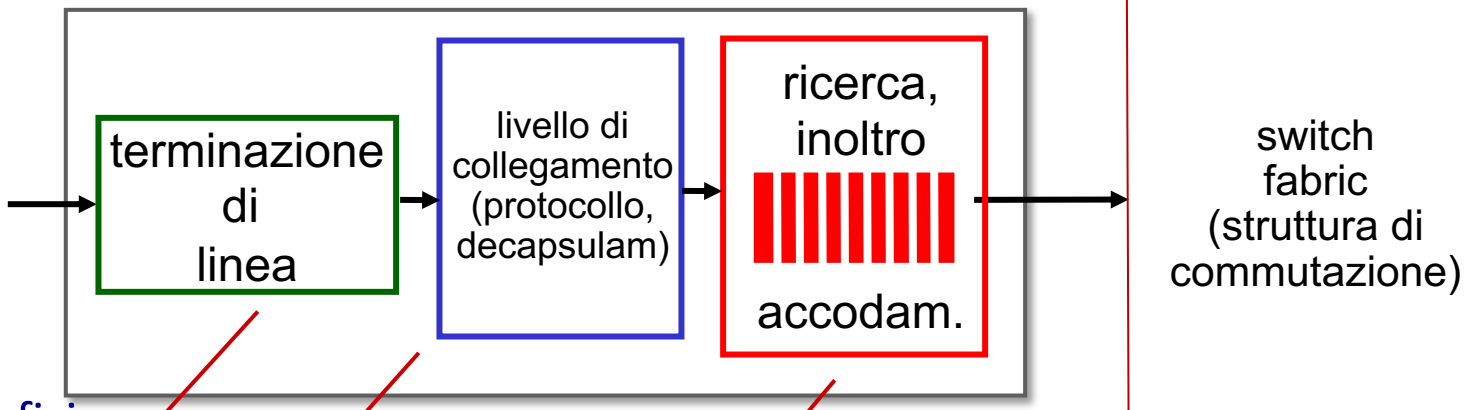
strato fisico:
ricezione a livello di bit

livello di collegamento:
ad esempio, Ethernet (Capitolo 6)

commutazione decentralizzata*:*

- utilizzando i valori del campo di intestazione, cerca la porta di output utilizzando la tabella di inoltro nella memoria della porta di input *("match plus action")*
- obiettivo: completare l'elaborazione della porta di input alla "velocità di linea" (senza diventare bottleneck)
- accodamento della porta di input: se i datagrammi arrivano più velocemente della velocità di inoltro nello switch fabric

Funzioni della porta di ingresso



strato fisico:
ricezione a livello di bit

livello di collegamento:
ad esempio, Ethernet (Capitolo 6)

commutazione decentralizzata*:*

- utilizzando i valori del campo di intestazione, cerca la porta di output utilizzando la tabella di inoltro nella memoria della porta di input *("match plus action")*
- destination-based forwarding: inoltro basato solo sull'indirizzo IP di destinazione (tradizionale)
- generalized forwarding: inoltro basato su qualsiasi insieme di valori del campo di intestazione

Destination-based forwarding

forwarding table

Destination Address Range	Link Interface
11001000 00010111 00010000 00000000 through 11001000 00010111 00010000 00000100	0
11001000 00010111 00010000 00000100 through 11001000 00010111 00010000 00000111	3
11001000 00010111 00011000 11111111	
11001000 00010111 00011001 00000000 through 11001000 00010111 00011111 11111111	2
otherwise	3

D: ma cosa succede se gli intervalli non si dividono così bene?

Longest prefix matching

match del prefisso più lungo

quando si cerca la voce della tabella di inoltro per un determinato indirizzo di destinazione, usa il prefisso dell'indirizzo *più lungo che è un match con l'indirizzo di destinazione*

Intervallo di indirizzi di destinazione	link interface
11001000 00010111 00010 *** ********	0
11001000 00010111 00011000 ********	1
11001000 00010111 00011 *** ********	2
Altrimenti	3

esempi :

11001000 00010111 00010110 10100001 quale interfaccia?

11001000 00010111 00011000 10101010 quale interfaccia?

Longest prefix matching

match del prefisso più lungo

quando si cerca la voce della tabella di inoltro per un determinato indirizzo di destinazione, usail prefisso dell'indirizzo *più lungo che è un match con l'indirizzo di destinazione*

esempi :

Longest prefix matching

match del prefisso più lungo

quando si cerca la voce della tabella di inoltro per un determinato indirizzo di destinazione, usail prefisso dell'indirizzo *più lungo che è un match con l'indirizzo di destinazione*

Intervallo di indirizzi di destinazione	link interface
11001000 00010111 00010 *** ********	0
11001000 00010111 00011000 ********	1
11001000 00010111 00011 *** ********	2
Altrimenti	3

match

esempi :

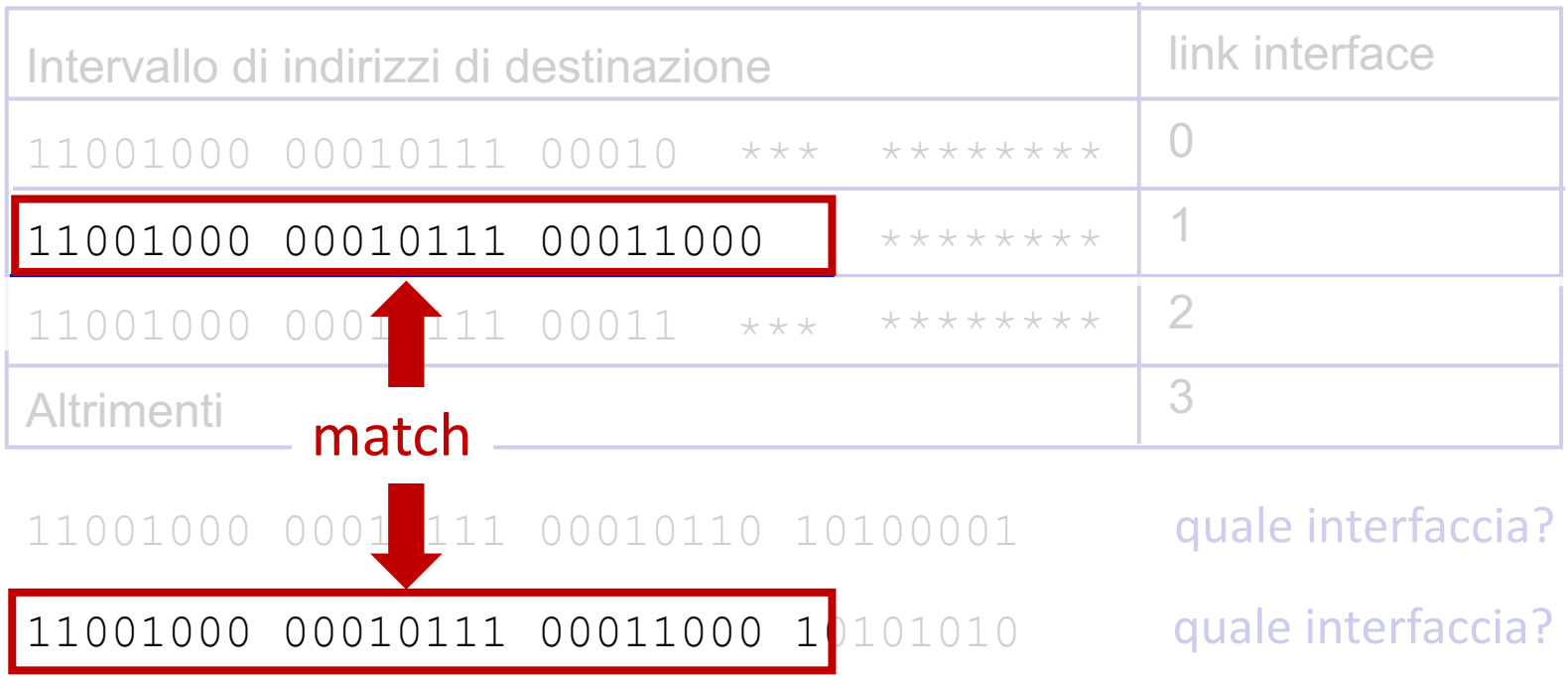
11001000 00010110 10100001 quale interfaccia?

11001000 00010111 00011000 10101010 quale interfaccia?

Longest prefix matching

match del prefisso più lungo

quando si cerca la voce della tabella di inoltro per un determinato indirizzo di destinazione, usail prefisso dell'indirizzo *più lungo che è un match con l'indirizzo di destinazione*



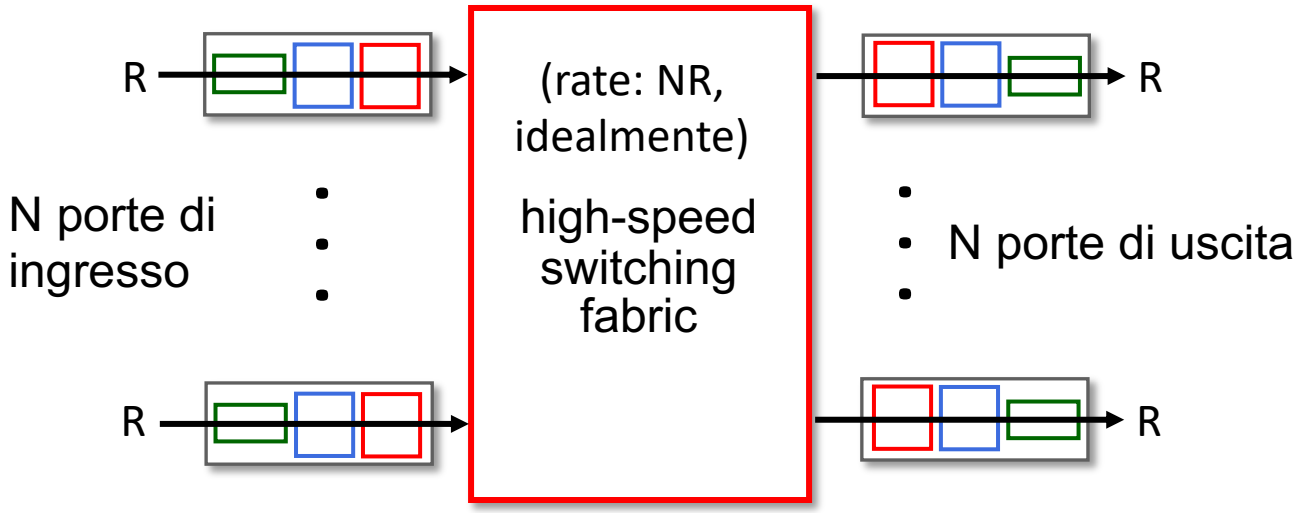
esempi :

Longest prefix matching

- vedremo *perché la* corrispondenza del prefisso più lungo viene utilizzata a breve, quando studieremo l'indirizzamento
- longest prefix matching: spesso eseguita utilizzando memorie indirizzabili a contenuto ternario (TCAM)
  - *contenuto indirizzabile:* presenta l'indirizzo a TCAM: recupera l'indirizzo in un ciclo di clock, indipendentemente dalle dimensioni della tabella
  - Cisco Catalyst: ~1 milione di voci della tabella di routing in TCAM

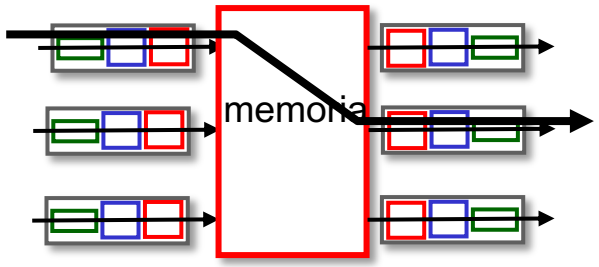
Switching fabrics

- trasferisce il pacchetto dal collegamento di input al collegamento di output appropriato
- switching rate: velocità alla quale i pacchetti possono essere trasferiti dagli ingressi alle uscite
  - spesso misurato come multiplo della velocità della linea di ingresso/uscita
  - N ingressi: desiderabile velocità di commutazione N volte la velocità di linea

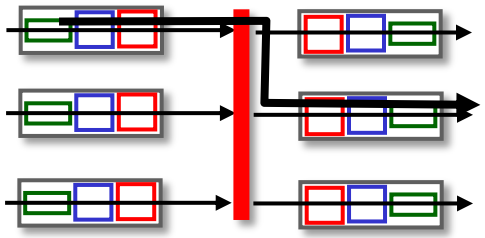


Switching fabrics

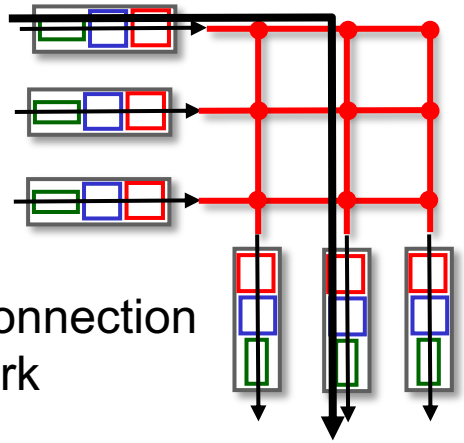
- trasferisce il pacchetto dal collegamento di input al collegamento di output appropriato
- switching rate: velocità alla quale i pacchetti possono essere trasferiti dagli ingressi alle uscite
  - spesso misurato come multiplo della velocità della linea di ingresso/uscita
  - N ingressi: desiderabile velocità di commutazione N volte la velocità di linea
- tre tipi principali di switching fabric:



memoria



bus

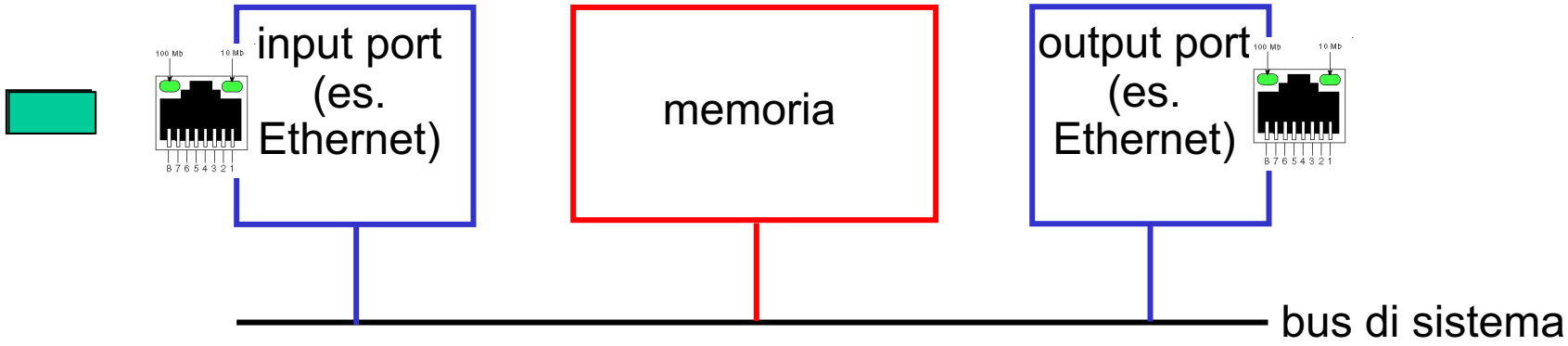


interconnection network

Commutazione tramite memoria

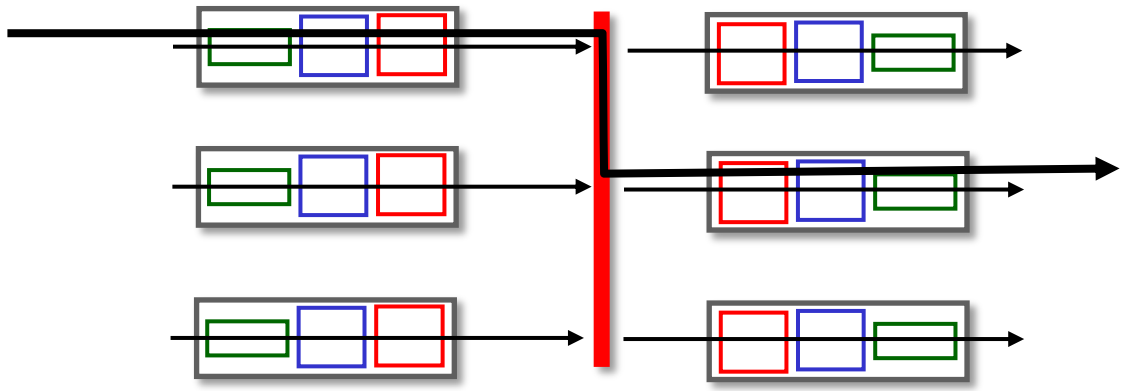
router di prima generazione (1970-1980):

- computer tradizionali con switching sotto il controllo diretto della CPU
- pacchetto copiato nella memoria del sistema
- velocità limitata dalla larghezza di banda della memoria (2 attraversamenti di bus per datagramma)



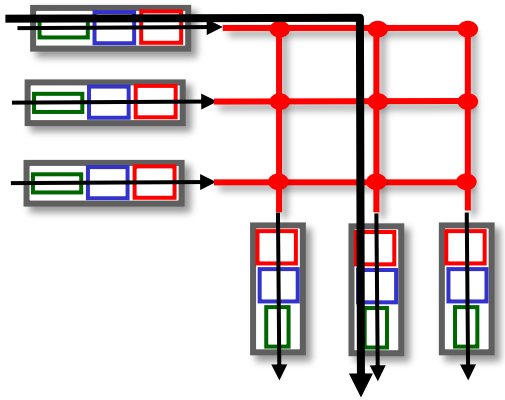
Commutazione tramite bus

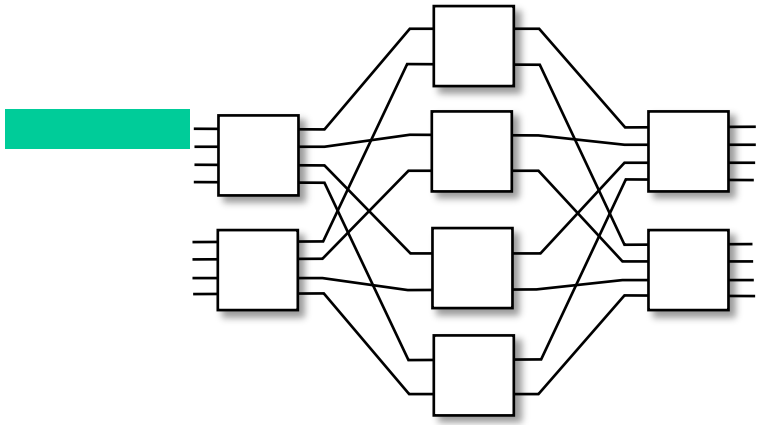
- datagramma dalla memoria della porta di ingresso alla memoria della porta di uscita tramite un bus condiviso
- *contesa bus:* velocità di commutazione limitata dalla larghezza di banda del bus
- Bus 32 Gbps, Cisco 5600: velocità sufficiente per router di accesso



Commutazione tramite reti di interconnessione

- Crossbar, reti Clos, altre reti di interconnessione sviluppate inizialmente per connettere processori
- interruttore multistadio: interruttore *nxn* da più stadi di interruttori più piccoli
- sfruttando il parallelismo:
  - frammentare il datagramma in celle di lunghezza fissa all'ingresso
  - commuta le celle attraverso il la rete di interconnessione, riassembla il datagramma in uscita



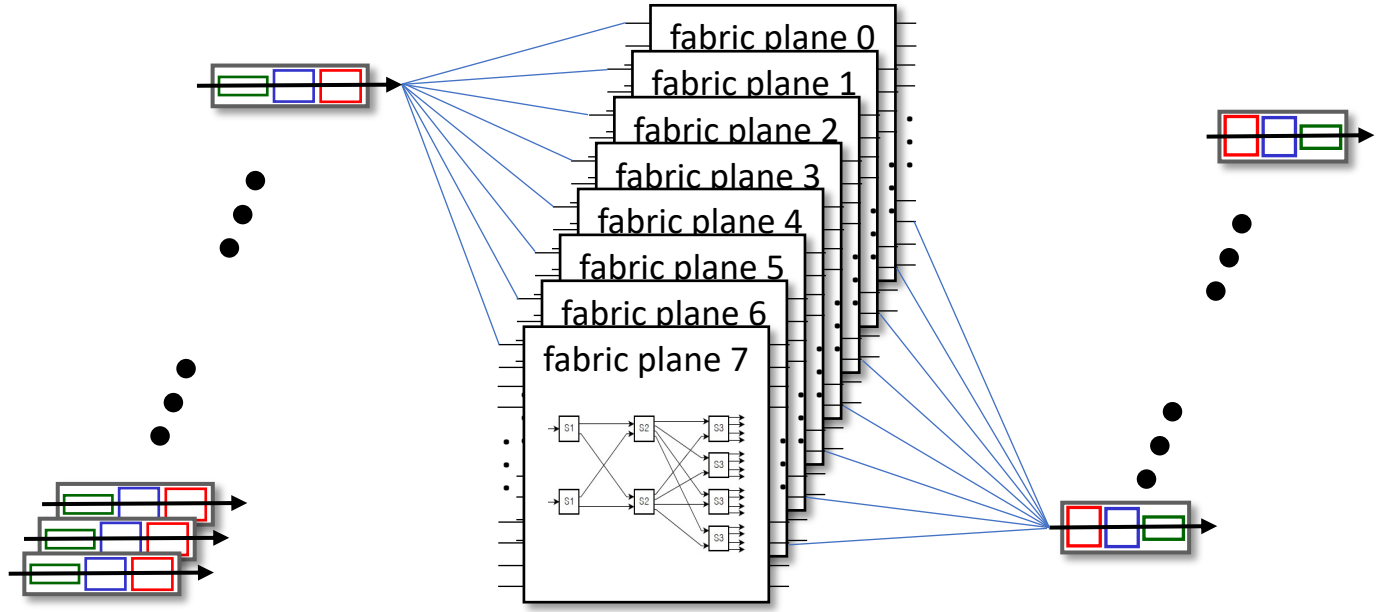


Interruttore multistadio 8x8
costruito da interruttori di dimensioni più piccole

Commutazione tramite reti di interconnessione

- scaling, utilizzando più "piani" di commutazione in parallelo:
  - speedup, scaleup tramite parallelismo

- Router Cisco CRS:
  - unità base: 8 piani di commutazione
  - ogni piano: rete di interconnessione a 3 stadi
  - capacità di commutazione fino a 100 Tbps

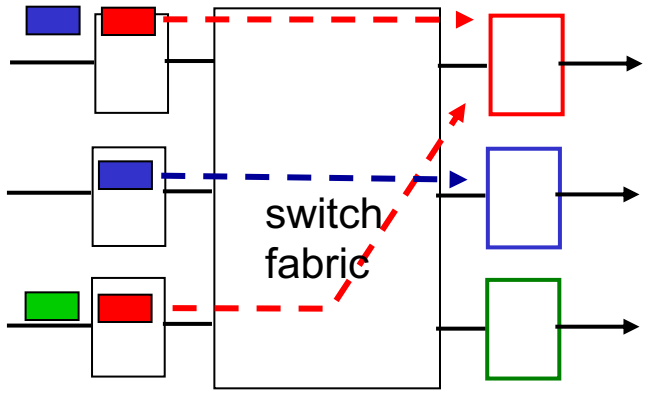


Livello di rete: sommario

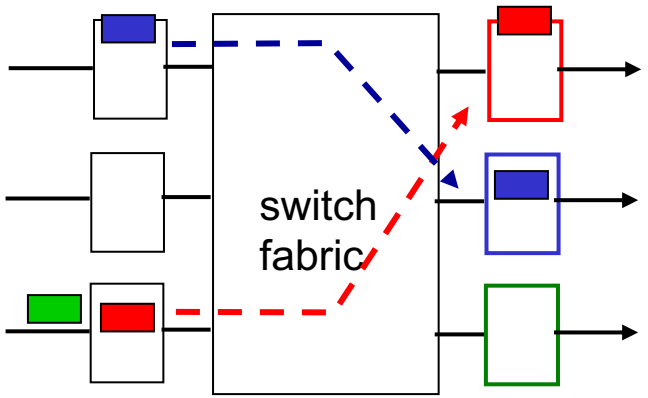
- Livello di rete: panoramica
  - piano dati
  - piano di controllo
- Dentro i router
  - porte di ingresso, commutazione, porte di uscita
  - gestione del buffer, scheduling
- IP: il protocollo Internet
  - formato datagramma
  - indirizzamento
  - traduzione di indirizzi di rete
  - IPv6
- Forwarding generalizzato, SDN
  - Match+action
  - OpenFlow: incontro+azione in azione
- Middleboxes

Accodamento nella porta di ingresso

- Se la switch fabric è più lenta delle porte di input combinate -> potrebbe verificarsi l'accodamento nelle code di input
  - ritardo di accodamento e perdita a causa dell'overflow del buffer di input!
- Blocco HOL (Head-of-the-Line): il datagramma in coda nella parte anteriore della coda impedisce agli altri in coda di andare avanti

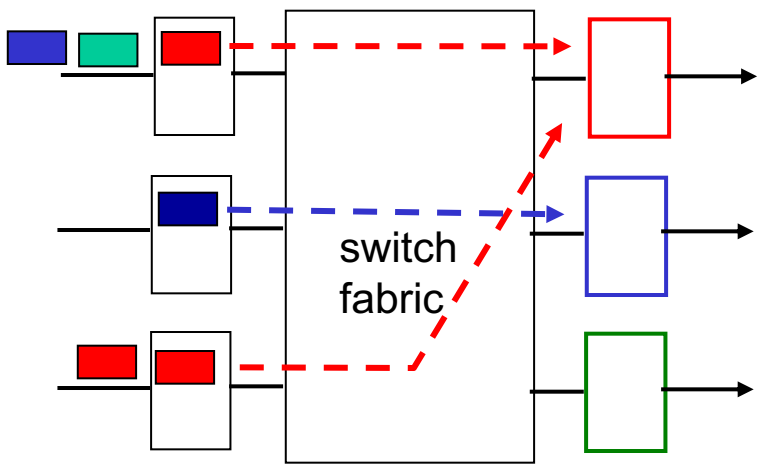


contesa della porta di uscita: può essere trasferito solo un datagramma rosso. il pacchetto rosso inferiore è *bloccato*

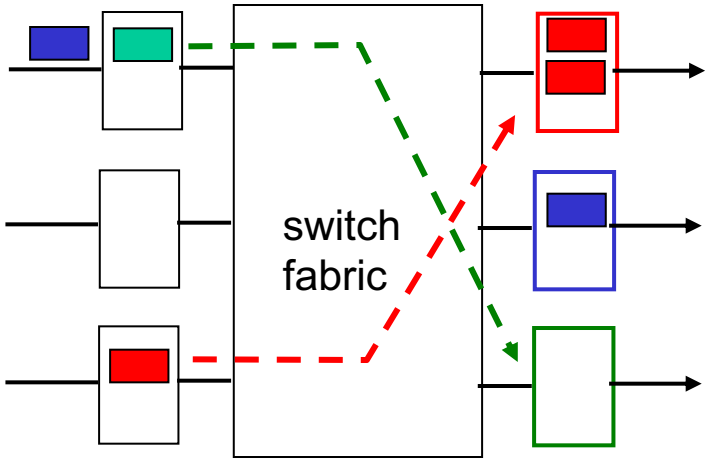


un pacchetto dopo: il pacchetto verde subisce un blocco HOL

Accodamento nella porta di uscita



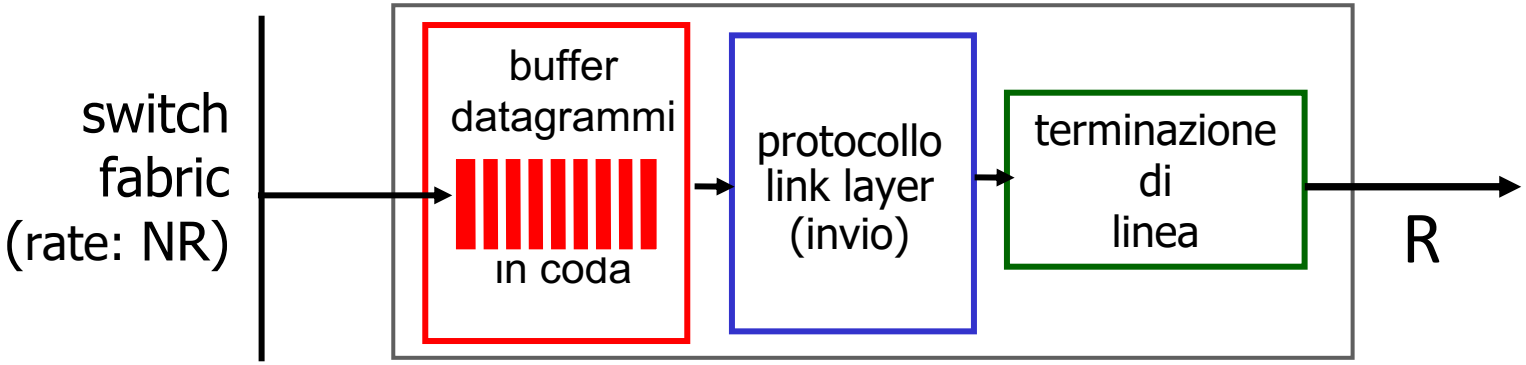
pacchetti in arrivo
a rate maggiore di
quello in uscita



un pacchetto più tardi

- buffering quando la velocità di arrivo tramite switch supera la velocità della linea di uscita
- *accodamento (ritardo) e perdita a causa dell'overflow del buffer della porta di uscita!*

Accodamento nella porta di uscita



- *Buffering* richiesto quando i datagrammi arrivano dal fabric più velocemente della velocità di trasmissione del collegamento.
  Drop policy: quali datagrammi eliminare se non ci sono buffer liberi?

→ I datagrammi possono andare persi a causa di congestione, buffer troppo piccoli

- *La disciplina di scheduling*
  sceglie tra i datagrammi in coda quali trasmettere

→ Priority scheduling: chi ottiene le migliori prestazioni, neutralità della rete

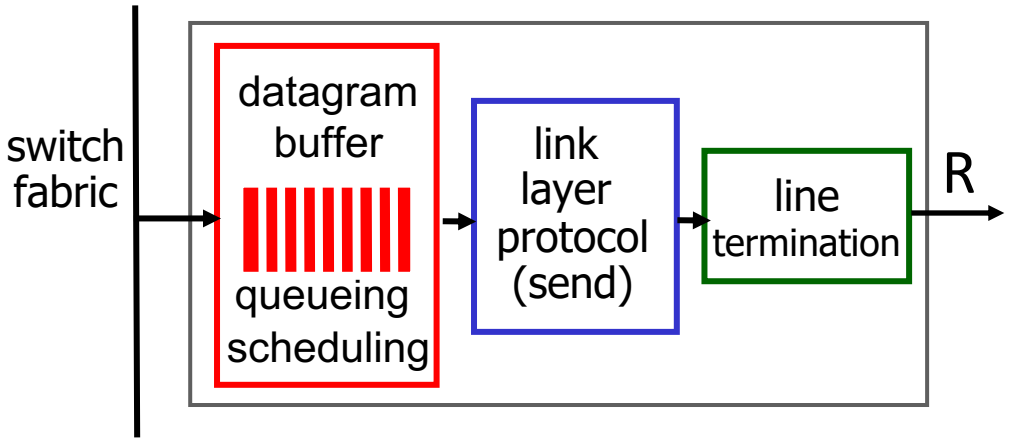
Quanto buffering?

- RFC 3439 regola empirica: buffering medio pari a RTT "tipico" (diciamo 250 ms) moltiplicato per la capacità del collegamento C
  - ad esempio, collegamento C = 10 Gbps: buffer da 2,5 Gbit
- raccomandazione più recente: con *N* flussi, buffering pari a

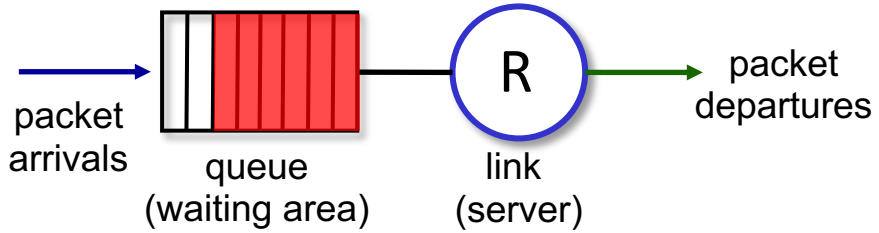
$$\frac{RTT \cdot C}{\sqrt{N}}$$

- ma un buffering *eccessivo può aumentare i ritardi (in particolare nei router domestici)*
  - RTT lunghi: scarse prestazioni per le app in tempo reale, risposta TCP lenta
  - "mantieni il collegamento del collo di bottiglia abbastanza pieno (occupato) ma non troppo". I buffer dovrebbero solo assorbire le **fluttuazioni statistiche di occupazione in mancanza di congestione**

Gestione del buffer



Astrazione: coda



Gestione del buffer:

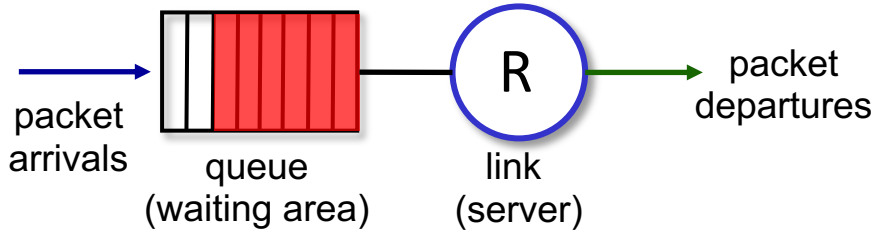
- drop: quale pacchetto inserire nella coda e quale scartare quando il buffer è pieno
  - tail drop: scartare il paccheto in arrivo
  - priority: scartare o rimuovere selettivamente
- marking: quali pacchetti marcare per indicare congestione (ECN, RED)

Scheduling dei pacchetti: FCFS (first come first served)

scheduling dei pacchetti:
decidere qual è il prossimo pacchetto da inviare sul link

- first come, first served
- priorità
- round robin
- weighted fair queueing

Astrazione: coda



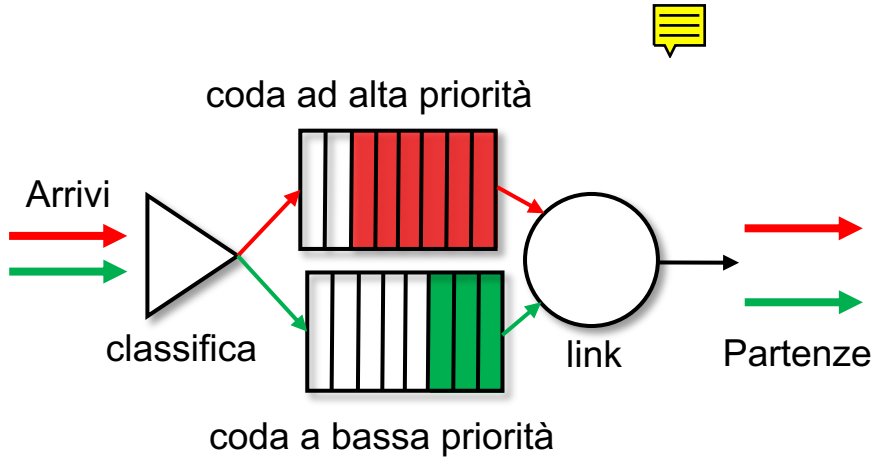
FCFS: pacchetti trasmessi in ordine di arrivo alla porta di uscita

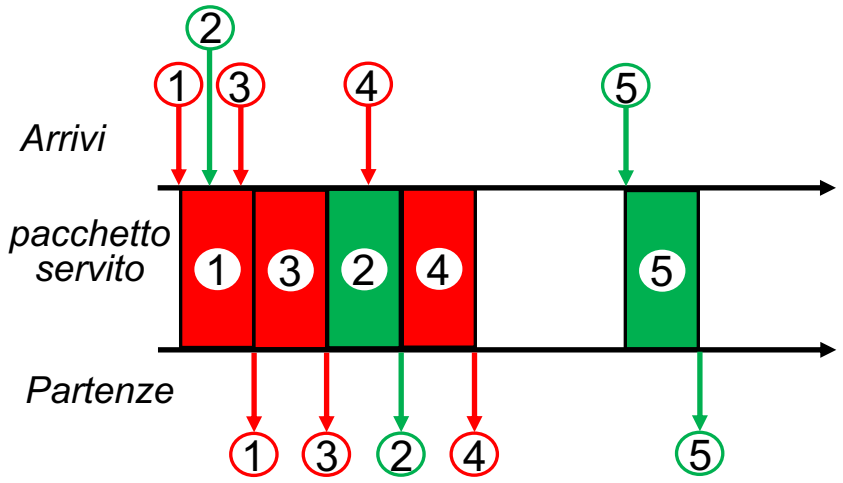
- noto anche come: First-in-first-out (FIFO)
- esempi del mondo reale?

Politiche di scheduling: priorità

Priority scheduling:

- traffico in arrivo classificato, accodato per classe di priorità
  - qualsiasi campo di intestazione può essere utilizzato per la classificazione
- invia il pacchetto dalla coda con la priorità più alta che contiene pacchetti nel buffer
  - FCFS all'interno della classe di priorità

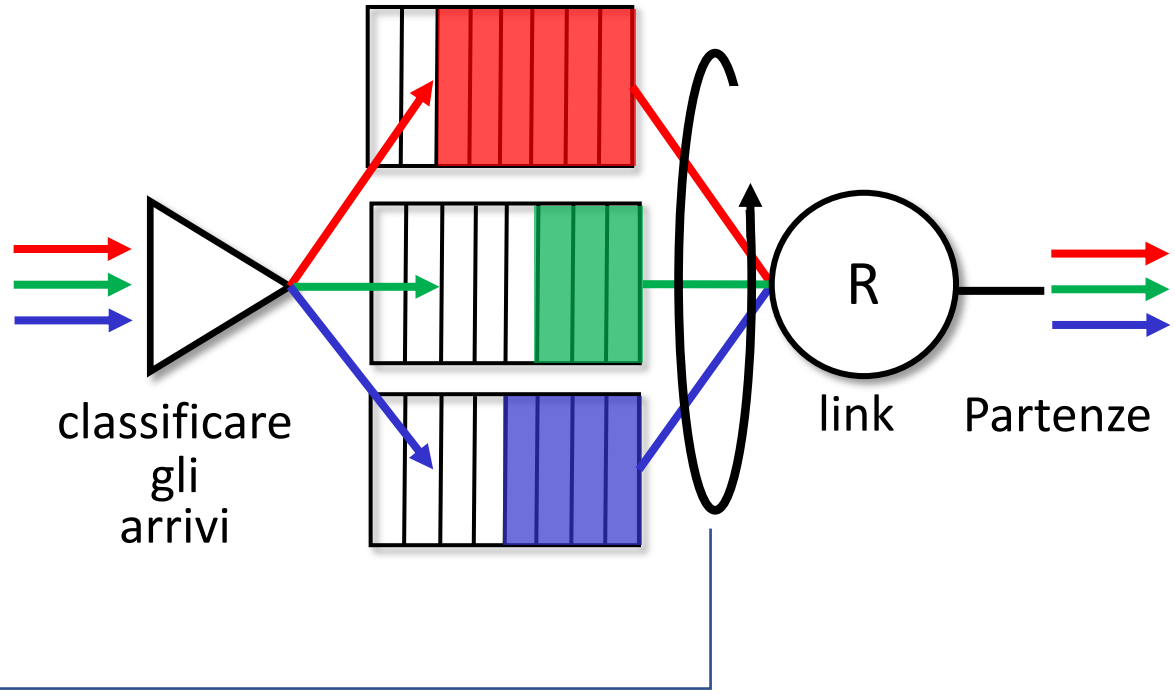




Politiche di scheduling: round robin

Scheduling Round Robin (RR):

- traffico in arrivo classificato, in coda per classe
  - qualsiasi campo di intestazione può essere utilizzato per la classificazione
- server ciclicamente, esegue ripetutamente la scansione delle code di classe, inviando a turno un pacchetto completo da ciascuna classe (se disponibile)



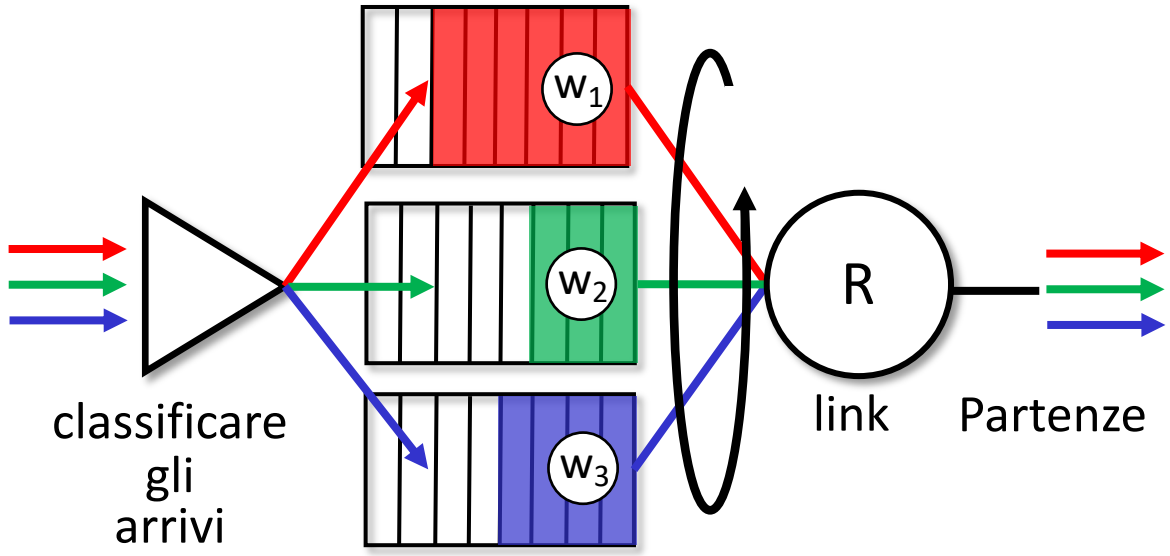
Politiche di schedulazione: weighted fair queuing

Weighted fair queuing (WFQ):

- Round Robin generalizzato
- ogni classe *i* ha peso w_i, e riceve una quantità ponderata di servizio a ogni ciclo:

$$\frac{w_i}{\sum_j w_j}$$

- garanzia di larghezza di banda minima (per classe di traffico)



Neutralità della rete

Cos'è la neutralità della rete?

- *tecnico:* come un ISP dovrebbe condividere/allocare le proprie risorse
  - scheduling dei pacchetti e gestione del buffer sono i *meccanismi*
- principi *sociali, economici*
  - proteggere la libertà di parola
  - incoraggiare l'innovazione, la concorrenza
- Provvedimenti *legali* per stabilire regole e politiche di gestione

Diversi paesi hanno diversi approcci sulla neutralità della rete

Neutralità della rete in USA

2015 US FCC *Order on Protecting and Promoting an Open Internet:* tre princìpi:

- no blocking ... "non si debbono bloccare contenuti, applicazioni, **servizi** o **dispositivi** *leciti* e *non dannosi*, soggetti a una *ragionevole* gestione della rete"
- no throttling ... "non deve compromettere o degradare il traffico Internet legittimo sulla base di contenuti, applicazioni o servizi Internet o l'uso di un dispositivo non dannoso, soggetto a una ragionevole gestione della rete."
- no priorità a pagamento. … "non deve prevedere la prioritizzazione retribuita"

ISP: servizi di telecomunicazioni o servizi di informazione?

Un ISP è un fornitore di "servizi di telecomunicazioni" o di "servizi di informazione"?

- la risposta conta *davvero* dal punto di vista normativo!

Legge sulle telecomunicazioni degli Stati Uniti del 1934 e del 1996:

- *Titolo II: impone* "doveri di servizio pubblico" sui *servizi di telecomunicazione*: tariffe ragionevoli, non discriminazione e richede *regolamentazione*
- *Titolo I:* si applica ai *servizi di informazione:*
  - nessun dovere di servizio pubblico (*non regolamentato*)
  - ma concede autorità alla FCC in alcuni casi

Europa: Legge 2015 che può avere delle scappatoie legali