Fondamenti di Telecomunicazioni

Maurizio Casoni

C.L. Ingegneria Informatica
Università di Modena e Reggio Emilia
Anno Accademico 2011/12

Slide 1

Introduzione

Slide 2

Un pó di storia ...

Nei secoli e in ogni cultura si sviluppano vari metodi per la comunicazione a distanza. Esempi sono i segnali di fumo, i messaggi a ottici, quelli sonori ecc.

Con la scoperta dell'elettricitá nascono vari sistemi basati sulla trasmissione di cariche elettriche. Con l'invenzione della pila (Volta 1799) si ha per la prima volta uno strumento in grado di fornire tensione e corrente.

Telegrafo Introdotto da Morse 1835. É basato sull'azioni di una bobina attraversata da corrente, che aziona un martelletto. Permette di comunicare con simboli e per questo é necessario un codice. Con l'introduzione della telescrivente (1846) diventa la prima applicazione delle comunicazioni elettriche di interesse industriale e di larga diffusione (il servizio telegrafico é tuttora operante, con piccole modifiche, come telex).

Telefono Brevettato da Bell nel 1876. Non utilizza codici. É basato sulla trasformazione del segnale fonico in elettrico e viceversa. L'assenza di codice lo rende accessibile a tutti senza bisogno di passare attraverso un operatore. É il primo servizio di tipo **"home"** ossia accessibile direttamento dalla casa dell'utente. La sua importanza non fu inizialmente compresa dalle compagnie telegrafiche che finirono poi per essere assorbite

Slide 3

dalle neonate compagnie telefoniche. La BELL SYSTEM é oggi uno dei maggiori gruppi mondiali come fatturato.

Radio Nasce nel 1895 ad opera di Marconi, separatamente da telegrafo e telefono. Inizilamente usa segnali in codice (Morse). In seguito diviene possibile la trasmissione di segnali fonici (anche di elevata qualitá con l'FM) e di segnali video (televisione). Sue caratteristiche fondamentali sono la possibilitá di avere collegamenti punto-multipunto che permettono **diffusione radio** o broadcasting e la possibilitá di avere collegamenti fra mezzi in movimento. Il 12 Dicembre 1901 Marconi effettuò la prima tramissione transatlantica dalla Gran Bretagna al Nord America.

Reti di calcolatori I primi esempi risalgono agli anni '50 per il collegamento di terminali periferici a calcolatori. Lo sviluppo dei calcolatori elettronici aumenta le necessitá di collegamento fra elaboratori e terminali, cosí come fra elaboratori. Negli anni '70 negli USA nascono le prime reti per trasmissione dati di tipo "general purpose" (ARPANET, TYMENET).

1991: esplosione della rete Internet a livello planetario con la definizione dello standard WWW.

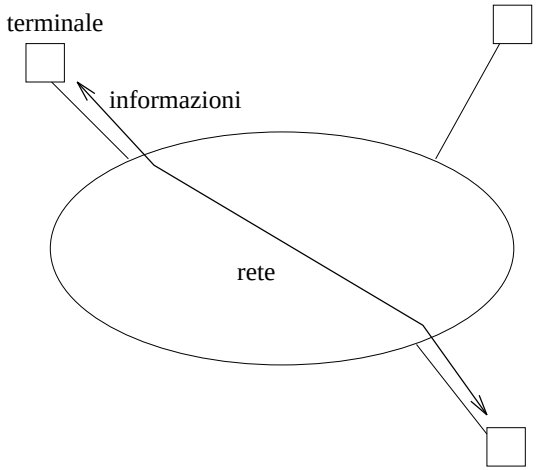
Slide 4

Elementi comuni

Esiste un **servizio** fornito agli **utenti**, che sono tutti coloro che hanno accesso al servizio stesso. L'accesso al servizio avviene attraverso un opportuno apparato che viene detto *terminale di utente*.

Il servizio é basato sulla trasmissione di un certo tipo di informazioni che, solitamente, vengono prima trasformate in segnali elettrici e quindi trasferite.

Per il trasferimento ci si basa sull'uso di una **rete** ossia su di un sistema che permette loa scambio di informazioni (comunicazione) a distanza (tele) fra terminali.



Slide 5

Funzioni di rete

Trasmissione trasferimento dei segnali da punto a punto o da un punto a molti punti utilizzando risorse opportune; *linee di trasmissione* quali collegamenti in doppino telefonico, cavo coassiale, fibra ottica ecc. oppure *canali radio* sia terrestri che via satellite ecc.)

Commutazione reperimento delle risorse all'interno della rete necessarie per realizzare una connessione.

Segnalazione scambio di informazioni fra utente e rete oppure internamente alla rete necessario per la gestione delle connessioni e della rete stessa.

Gestione tutto cio' che concerne il mantenimento delle funzioni della rete; riconfigurazione di fronti a guasti o cambiamenti strutturali, allacciamento di unovi utenti ecc.

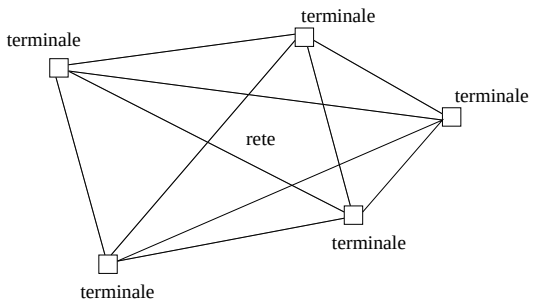
Slide 6

Topologie di rete: rete a maglia completa

Una soluzione apparentemente semplice per realizzare una rete potrebbe essere quella di collegare con una linea di trasmissione dedicata qualunque coppia di terminali. Questa soluzione é detta *rete completamente magliata*.

Se N e' il numero dei terminali, il numero dei collegamenti risulta $C = \frac{N(N-1)}{2}$.

É facile comprendere che in caso di elevato numero di terminali il numero delle linee richieste diventa elevatissimo. Inoltre qualora i terminali siano distribuiti sul territorio sono necessarie linee molto lunghe.



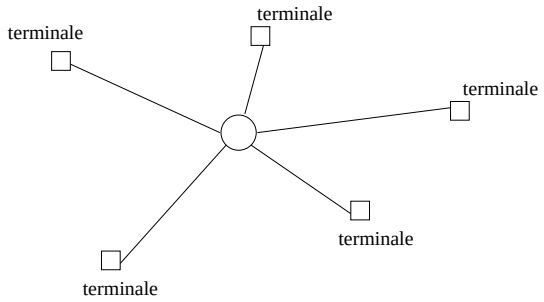
Slide 7

Topologie di rete: rete a stella

Il numero dei collegamenti risulta in questo caso $C = N$.

Il centro stella deve svolgere la funzione di redirezione delle informazioni provenienti da un terminali verso il (o i) destinatari. E' possibile avere un centro stella:

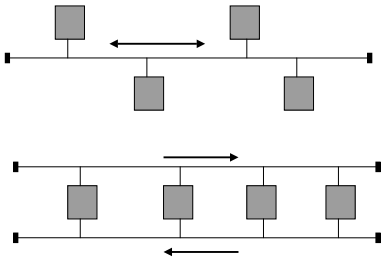
- attivo,
- passivo.



Slide 8

Topologie di rete: rete a bus

Può essere di tipo unidirezionale o bidirezionale.

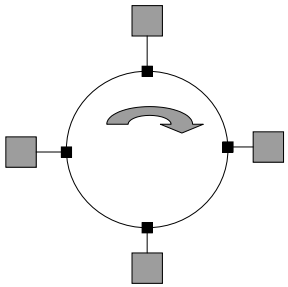


Bidirezionale:

- tutte le stazioni sono collegate ad un mezzo lineare di trasmissione
- dati trasmessi e ricevuti con unica intefaccia hardware
- la trasmissione di una stazione si propaga in entrambe le direzioni ed è ricevuta da tutte le altre stazioni
- ad ogni estremità del bus vi è un terminatore assorbente di segnale

Slide 9

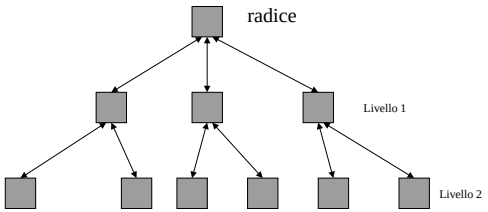
Topologie di rete: rete ad anello



- serie di ripetitori uniti da collegamenti punto-punto unidirezionali
- ogni stazione è collegata alla rete mediante ripetitore e trasmette e riceve dati tramite esso

Slide 10

Topologie di rete: rete ad albero



- assenza di percorsi circolari chiusi
- univocità di percorso sorgente-destinazione

Slide 11

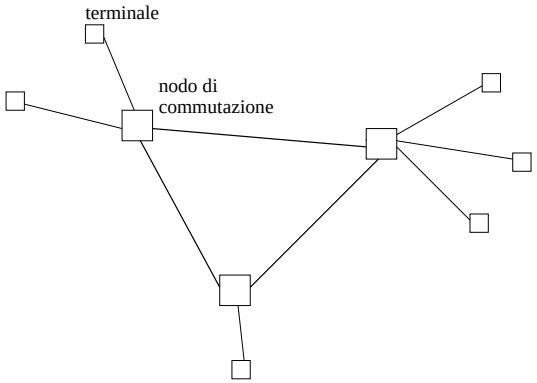
Rete commutata

I terminali sono connessi tramite linee dedicate a dei **nodi di commutazione**, che sono nodi interni alla rete.

I nodi di commutazione sono a loro volta interconnessi. Il loro compito é quello di stabilire un collegamento fra linee entranti e linee uscenti. Le regole con cui viene realizzato questo collegamento sono in generale diverse in reti diverse o in commutatori realizzati con tecnologie diverse.

Nella rete commutata il collegamento fra due terminali viene effettuato solo quando necessario e su richiesta dei terminali stessi.

In reti di grandi dimensioni l'uso della tecnica commutata é piú economico e di piú facile gestione.



Slide 12

Tecniche di commutazione

Esistono due tecniche di commutazione:

- commutazione di circuito
- commutazione di pacchetto

Slide 13

Commutazione di circuito

- La rete instaura un *canale di collegamento dedicato* fra due terminali che vogliono colloquiare.
- Una volta instaurato tale canale é riservato ad **uso esclusivo** dei terminali, in modo analogo ad un collegamento diretto.
- In fase di dialogo la rete é **del tutto trasparente** per gli utenti.

Slide 14

Commutazione di circuito

La comunicazione a commutazione di circuito coinvolge le seguenti fasi:

1. *Instaurazione del circuito.* Prima che i segnali di utente possano essere trasmessi la rete deve instaurare un circuito di collegamento fra terminale chiamante e terminale chiamato (circuito *end-to-end*). Questo viene fatto sulla base di una opportuna segnalazione.
2. *Dialogo.* I due terminali si scambiano informazioni utilizzando il circuito dedicato.
3. *Disconnessione del circuito.* Al termine del dialogo il circuito deve essere rilasciato, al fine di poter essere utilizzato per altre chiamate.

Esiste quindi un ritardo iniziale dovuto al tempo necessario per instaurare il circuito, ma dopo di ció la rete é *trasparente per l'utente* (equivale ad un collegamento fisico diretto).

Slide 15

Commutazione di circuito: pro e contro

Pro:

- il collegamento é dedicato e garantisce sicurezza ed affidabilitá,
- le procedure di controllo sono limitate ad inizio e fine chiamata.

Contro:

- se le sorgenti hanno un basso tasso di attivitá il collegamento é sottoutilizzato,
- la capacitá del canale é fissata dalla capacitá del collegamento e non si puó variare.

Slide 16

Esempio: rete telefonica

La rete telefonica é il piú noto esempio di rete a commutazione di circuito.

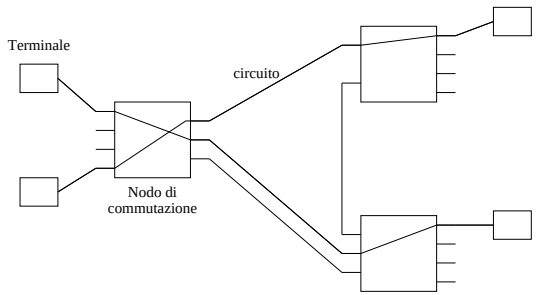
Qui il circuito trasporta una comunicazione vocale full-duplex.

Dal punto di vista dell'utente le fasi suddette si possono qui riconoscere come segue:

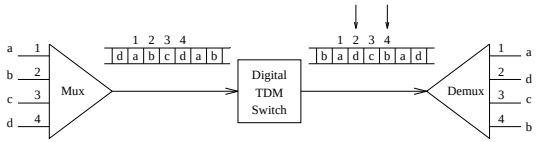
- sgancio della cornetta e composizione del numero ⇒ il circuito viene instaurato dalla rete,
- conversazione = dialogo trasparente,
- riaggancio della cornetta ⇒ richiesta di disconnesione del circuito.

Slide 17

Commutazione di circuito analogica



Commutazione di circuito numerica



Slide 18

Commutazione di pacchetto

- Trasporta informazioni in forma numerica.
- Le informazioni di utente vengono strutturate in blocchi, detti **pacchetti**, unitamente ad oppurtune informazioni di segnalazione (di indirizzamento, per la verifica della correttezza delle informazioni di utente, ecc.).
- I pacchetti vengono trasmessi da un nodo di commutazione all'altro fino a raggiungere la destinazione **utilizzando in tempi diversi risorse comuni**.

Slide 19

Commutazione di pacchetto: pro e contro

Pro:

- L'efficienza nell'utilizzazione delle linee é maggiore, poiché la stessa linea é condivisa in modo dinamico da piú chiamate.
- La rete puó supportare diverse velocitá ed effettuare anche conversioni tramite memorizzazione.
- É possibile implementare prioritá per favorire traffici particolarmente importanti in caso di congestione.

Contro:

- é difficile garantire un predeterminato tempo di transito, quindi non é adatta per servizi di tipo *real time*.

Slide 20

Tecniche di commutazione di pacchetto

Circuito virtuale :

- lo scambio delle informazioni é preceduta da una procedura di apertura della connessione, durante la quale viene stabilito il percorso dei pacchetti da origine a destinazione,
- a tale percorso viene associato un numero di **circuito virtuale** che verrá usato per l'instradamento dei pacchetti di informazione,
- tutti i pacchetti di dati seguono lo stesso percorso,
- non contengono l'indirizzo della destinazione finale ma solamente il numero di circuito virtuale.

Datagramma :

- ogni pacchetto viene gestito e instradato in modo independente, senza relazione con pacchetti precedenti o successivi, anche appartenenti alla stessa connessione,
- ogni pacchetto porta tutte le informazioni di indirizzamento necessarie per raggiungere la destinazione finale,
- pacchetti diversi di una stessa connessione possono seguire percorsi diversi e quindi avere tempi di percorrenza diversi.

Slide 21

Esempi

La reti a commutazione di pacchetto piú comune:

- **Internet** : rete mondiale a commutazione di pacchetto secondo lo standard **TCP-IP** (che vedremo),

Slide 22

Modello ISO-OSI

Slide 23

Organismi di standardizzazione

- **ITU**: International Telecommunications Union é la piú antica associazione di stati del mondo (1865) ed è un organo specializzato delle Nazioni Unite. Si compone di vari gruppi di lavoro dei quali i piú importanti sono:
  1. **ITU-T**: Comitato Consultivo Internazionale Telegrafico e Telefonico
  2. **ITU-R**: Comitato Consultivo Internazionale di Radiocomunicazione
- **ISO**: International Standard Organization (1946) si occupa di standardizzazione un pó in tutti i campi ed in particolare anche nel mondo delle reti, in Italia la **UNI** lavora per ISO.

Altri importanti enti sono:

1. Institute of Electrical and Electronic Engineering (IEEE)
2. Internet Society (con IETF)
3. Electrical Industry Association (EIA)
4. Telecommunications Industry Association (TIA)
5. American National Standardization Institute (ANSI)

Slide 24

Sistemi chiusi

Tutte le reti di calcolatori della prima generazione nascono e si evolvono come **sistemi chiusi**:

- nel mondo dell'Informatica: tutti i componenti della rete devono essere dello stesso costruttore (captivity),
- nel mondo delle Telecomunicazioni: una rete specializzata per ogni servizio, con i suoi propri terminali, nodi, mezzi trasmissivi e protocolli.

Questo crea *incompatibilita' e disordine*, ponendo ostacoli alla comunicazione a vari livelli:

- a causa della diversita' delle reti gli apparati non riescono ad interpretare i segnali degli altri,
- anche se i calcolatori riescono a connettersi non riescono a colloquiare perche' parlano linguaggi diversi,
- i programmi applicativi non riescono ad operare in ambiente distribuito.

Nota: Tutte le soluzioni proposte hanno in comune una architettura a *strati*.

Slide 25

Reti aperte

Obiettivo: **sistemi aperti**, in cui *qualunque* terminale comunica con *qualunque* fornitore di servizi mediante *qualunque* rete.

Per realizzare un sistema aperto é necessario stabilire delle regole comuni per lo scambio delle informazioni: sono necessari degli **standard**.

Slide 26

Modello euristico

- Un professore servendosi del servizio postale contatta un editore in Giappone per chiedere informazioni sul suo materiale didattico e decide di ordinare alcuni articoli.
  - Tutta la trattativa puó essere vista come una comunicazione (scambio di lettere e materiale) fra Professore ed Editore durante la quale ognuno fa il suo mestiere (problemi di **Applicazione**).
  - Le lettere ed i libri debbono avere una forma comprensibile: a questo fine possono intervenire delle Segretarie (problemi di **Presentazione**).
  - Il movimento delle merci fa nascere questioni amministrative: ordini, fatture, ricevute. Anche per questo si possono usare impiegati appositi (problemi di **sessione**).
  - Occorre scegliere il servizio postale a cui affidarsi, inoltre le lettere ed i pacchi vanno preparati in modo conforme alle regole di questo servizio (problemi di **Trasporto**).
  - La posta viene consegnata ad un servizio pubblico. In primo luogo un ufficio smistamento, la **instrada** verso la sua destinazione (problemi di **Rete**).

Slide 27

  - La posta in generale deve percorrere diverse tratte. Viene pertanto messa in contenitori (sacchi ad es.) che vengono confezionati e controllati da appositi incaricati (problemi di **Controllo di linea**).
  - Infine un **mezzo di trasporto** (treno, aereo, nave) porta materialmente a destinazione (**mezzo fisico**).
- Si possono evidenziare due tipi di interazione fra le persone coinvolte:
  - fra corrispondenti remoti, ad es. Professore-Editore, Segretaria-Segretaria, ecc. (**protocolli**)
  - fra superiore e subordinato locali, ad es. Professore-Segretaria, ecc. (**interfaccie**)

Slide 28

ISO-OSI

A partire dai primi anni '70 l'ISO ha dato il via a lavori per giungere ad una serie di standard unificati per la realizzazione di reti di calcolatori aperte.

L'ISO ha quindi proposto un modello di riferimento detto **"Open System Interconnection (OSI)"** che e' diventato standard internazionale nel 1983 (ISO 7498).

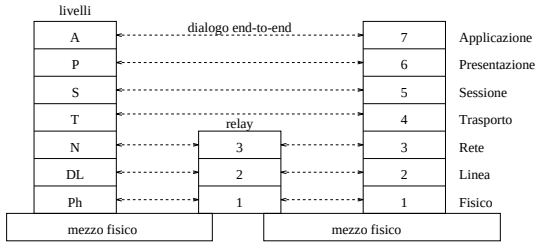
L'OSI e' basato sul concetto centrale di una *architettura a strati*. Lo standard definisce solamente i servizi che ogni strato deve fornire e le realtive interfaccie.

L'architettura a strati ha alcuni grandi vantaggi:

- scompone il problema in sottoproblemi piu' semplici da trattare,
- rende i vari livelli indipendenti,
- definendo solamente servizi e interfaccie, livelli diversi possono essere sviluppati da enti diversi.

Slide 29

I 7 strati del modello OSI



- si possono avere funzione di **relay** ai livelli 1, 2, 3 (lower layers)
- i protocolli dal 4 in su sono sempre **end-to-end**
- gli strati 5, 6, 7 sono detti upper layers.

Slide 30

Descrizione OSI

Coinvolge tre livelli di astrazione

Modello di riferimento : schema concettuale e definizione dei confini e relazioni dei 7 strati.

Definizione dei servizi : definizione astratta di quali servizi sono offerti.

Specifiche dei protocolli : descrizione di come uno strato fornisce i servizi.

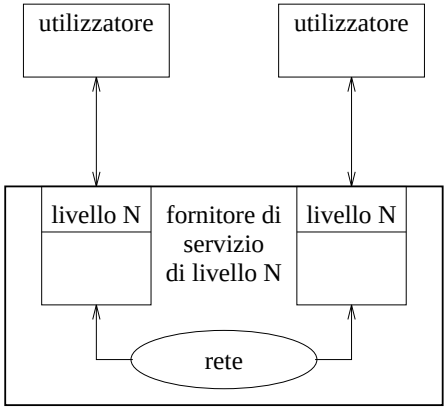
Slide 31

Service User - Service Provider

Il termine *servizio (service)* si riferisce all'interazione verticale fra i livelli. Un servizio viene definito sulla base delle relative interazioni e dei parametri scambiati.

Il *fornitore di servizio a livello N (N-service provider)* comprende il livello N e tutti i livelli inferiori di cui il livello N fa uso.

L'*utilizzatore di servizio di livello N (N-service user)* é l'entitá di livello N+1 che fa uso dei servizi del livello N.



Slide 32

Modalitá di Servizio

Il modello ISO-OSI specifica due modalitá di trasferimento delle informazioni:

Connection oriented. Si compone di tre fasi come segue

1. instaurazione della connessione, tramite lo scambio di opportune informazioni iniziali,
2. trasferimento dei dati veri e propri,
3. chiusura della connessione.

Connectionless. Tutti le informazioni di indirizzo e i parametri necessari per il trasferimento dei dati sono forniti, insieme ai dati di utente, per ogni accesso al servizio. Ogni unitá di dati viene traferita in modo indipendente dalle altre.

Slide 33

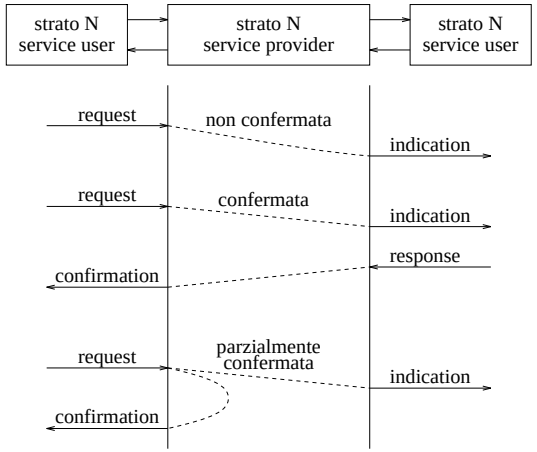
Primitive di servizio

Sono la *rappresentazione astratta* dell'interazione fra service-user e service-provider.

Sono indipendenti dall'implementazione e seguono un uguale schema per ogni interfaccia.

Sono 4:

1. request
2. indication
3. response
4. confirme



Slide 34

SAP

Definizione: Si definisce *entitá* in un livello ogni elemento attivo del livello stesso

Ciascun livello comprende una o piú entitá.

Definizione: L'interfaccia logica fra una entitá di livello N+1 e una di livello N, attraverso la quale viene fornito un servizio, e' detta **punto di accesso al servizio (Service Access Point - SAP)**

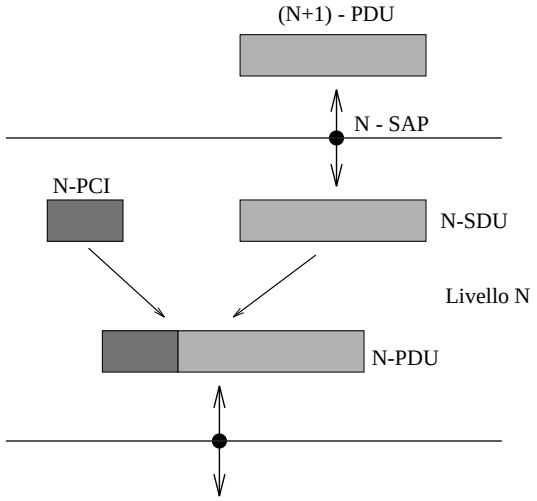
Ogni N-SAP ha un indirizzo unico che serve per identificarla al confine fra i due livelli.

- Un entitá di livello N puó servire piú N-SAP contemporaneamente.
- Un utilizzatore di livello N puó servirsi di piú N-SAP contemporaneamente.

Slide 35

Trasferimento dati

- I dati di livello N+1 sono passati al livello N attraverso un N-SAP sotto forma di **Service Data Unit (SDU)**.
- I dati sono trasferiti sulla rete fra entitá dello stesso livello sotto forma di **Protocol Data Unit (PDU)**.
- Una PDU contiene sia dati di livello N+1 che informazioni di controllo di livello N dette **Protocol Control Information (PCI)**.



Slide 36

Strato Fisico

Scopo dello strato fisico é quello di attivare, mantenere e deattivare la la connessione fisica fra due entitá di strato 2.

Per fare questo deve specificare le caratteristiche:

meccaniche : forma di prese e spine, numero di contatti,

elettriche : voltaggio e caratteristiche elettriche dei segnali associati all'interfaccia,

funzionali : significato dei vari segnali,

procedurali : combinazioni e sequenze dei segnali all'intefaccia necessarie al fine di regolarne il funzionamento.

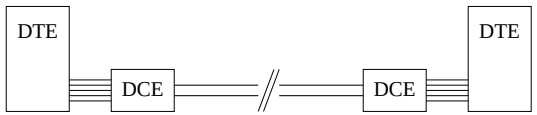
Slide 37

DTE e DCE

Tradizionalmente per gli apparati in cui implementare l'interfaccia di strato 1 si fa uso della seguente terminologia.

DTE - Data Terminal Equipment Tipicamente il terminale di utente (terminale, calcolatore, mainframe ecc.)

DCE - Data Circuit terminating Equipment Tipicamente l'apparato di accesso alla rete (modem o terminazione di rete) che realizza il reale accesso al mezzo fisico.



Nota: Solamente l'interfaccia DTE-DCE é oggetto dello strato 1, mentre non lo sono le modalitá precise con cui il DCE accede al reale mezzo fisico utilizzato per la trasmissione.

Slide 38

Principali standard di strato 1

Collegamenti punto punto e per reti pubbliche per dati.

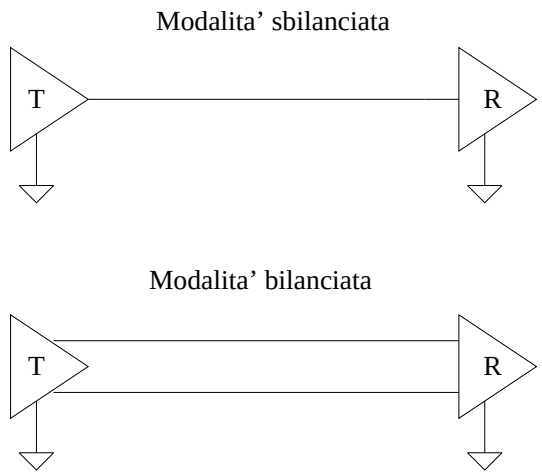
- Per reti analogiche: X.21 bis e famiglia delle raccomandazioni V.
- Per reti digitali: X.21.

Collegamenti per reti locali.

- A bus: ISO 8802.3 8802.4 8802.6
- Ad anello: ISO 8802.5 8802.7

Slide 39

Modalitá di connessione elettrica



T: trasmettitore
R: ricevitore

Nota: La modalitá bilanciata é meno sensibile ai disturbi sulla massa e per questo permette migliori prestazioni in termini di rapporto segnale/rumore. Questo risulta in maggiori velocitá e distanze di collegamento, rispetto alla configurazione sbilanciata.

Slide 40

EIA RS-232-E

Nasce come standard EIA. É incluso nelle raccomandazioni CCITT nella X.21 bis che a sua volta fa riferimento ad alcune raccomandazioni della serie V (V.24, V.28).

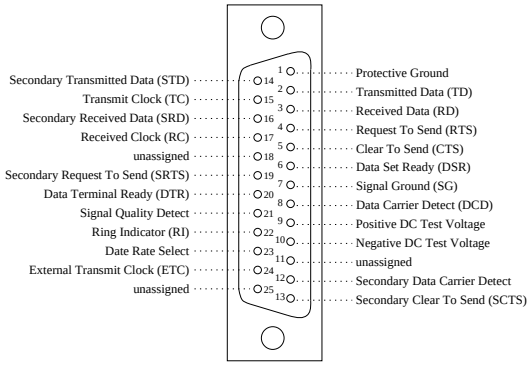
Caratteristiche meccaniche Connettore Canon (ISO 2110) a 25 pin. Viene permesso l'uso di connettore a 9 pin nella versione per PC.

Caratteristiche elettriche Modalitá sbilanciata,
tensione massima < 25 Volt
velocitá massima ≤ 20 Kbit/s
lunghezza del collegamento $\simeq 10 \div 15$ metri

Valore Logico	Stato	Tensione
0	ON	V > +3 Volt
1	OFF	V < -3 Volt

Slide 41

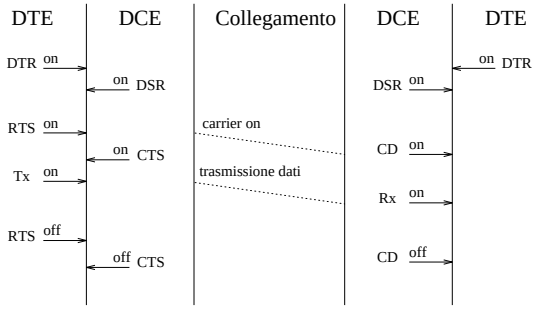
Caratteristiche funzionali



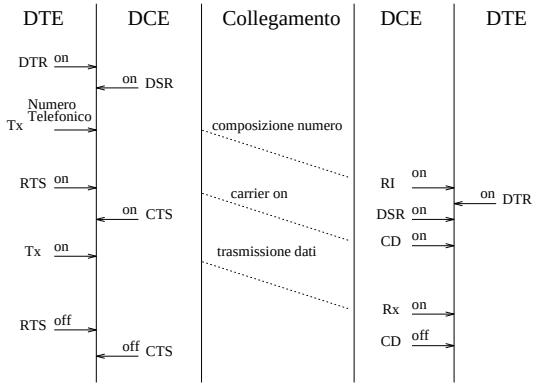
Slide 42

Caratteristiche procedurali

Esempio di scambio dati monodirezionale utilizzando un collegamento diretto.



Esempio di scambio dati monodirezionale su linea pubblica commutata.

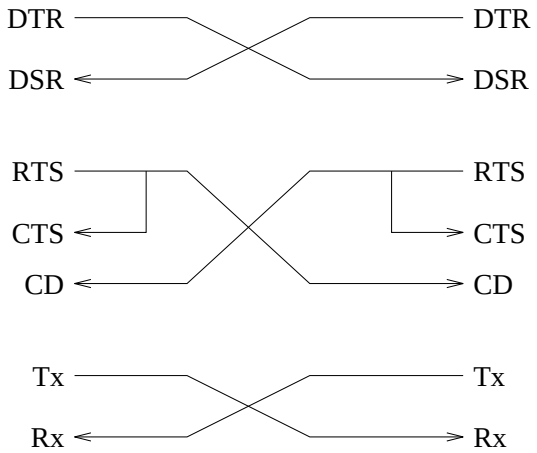


Slide 43

Collegamento DTE-DTE

Standard ISO TR 7477.

É necessario connettere i pin in modo tale da soddisfare le modalitá procedurali. Questa interconnessione é stata studiata in modo tale da simulare la presenza del DCE per ognuno dei due DTE.



Slide 44

EIA-449 ed EIA-422-A, EIA-423-A

EIA-449 é stat sviluppata per superare i limiti in velocitá e distanza della RS-232.

Utilizza un connettore Canon a 37 pin (ISO 4902).

Puó usare a livello elettrico due diverse modalitá:

EIA-423-A • generatori sbilanciati,

- livelli elettrici: 4/6 Volts, ON, -4/6 Volts, OFF,
- velocitá 1Kbit/s fino a 1Km, 100 Kbit/s fino a 10 m.

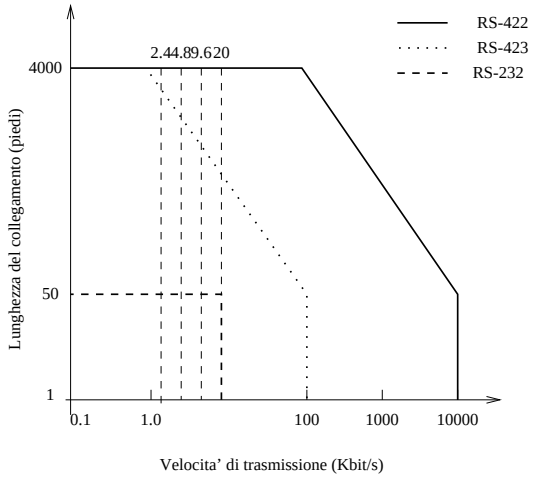
EIA-422-A • generatori bilanciati,

- livelli elettrici da 2 a 6 Volts differenziali,
- velocitá 100Kbit/s fino a 1Km, 10 Mbit/s fino a 10 m.

A livello funzionale e procedurale é analogo ad RS-232.

Slide 45

Confronto fra le varie interfacce



Slide 46

Universal Serial Bus (USB) (cenni)

Lo standard USB consente una trasmissione seriale su una coppia di cavi in rame.

Un' altra coppia di cavi in rame può essere utilizzata per alimentare periferiche poste in cascata.

USB definisce due tipi di cavi:

- cavi per comunicazioni ad alta velocità (12 Mbps) di maggiore qualità con migliore schermatura, di lunghezza fino a 5m.
- cavi per comunicazioni a 1.5 Mbps di lunghezza fino a 3 m.

e due tipi di connettori, meccanicamente diversi:

- di tipo "A"
- di tipo "B"

Lo standard USB consente la realizzazione di piccole reti di dispositivi in cui un apparato master (PC) controlla diversi dispositivi slave.

Slide 47

Strato 2 - Data Link Layer

Scopo dello strato di linea é quello di **attivare, mantenere e deattivare la connessione fisica fra due entitá di strato 3**, con lo scopo di trasformare la linea tra due nodi adiacenti in un collegamento affidabile.

Le funzioni tipicamente svolte dallo strato 2 sono le seguenti:

- strutturazione del flusso di dati in segmenti (denominati **frame** o **trame**).
- controllo di flusso,
- controllo di sequenza,
- controllo e gestione degli errori di trasmissione.

Slide 48

Full Duplex e Half Duplex

Lo scambio di dati tra una sorgente ed una destinazione collegate, per esempio, da una linea di trasmissione, può essere:

- **Half duplex**: solo una delle due stazioni può trasmettere e l' altra deve tacere; vi alternanza di trasmissioni;
- **Full duplex**: le due stazioni possono contemporaneamente trasmettere e ricevere informazioni.

Slide 49

Sincronismo

É possibile parlare di sincronizzazione:

- a livello di bit,
- a livello di trama.

Un protocollo é detto sincrono a livello di bit quando i bit iniziano e finiscono in istanti predeterminati.

I principali protocolli di strato 2 si basano su **trasmissioni sincrone a livello di bit**.

A livello di trama un protocollo puó essere:

asincrono per cui é necessario introdurre delle sequenze di bit o di caratteri particolari che permettano al ricevitore di rivelare inizio e fine di una trama,

sincrono ossia tale che le trame possono iniziare e finire solamente in istanti predefiniti.

Slide 50

Controllo di flusso

É necessario per accordare le velocitá di ricevitore e trasmettitore.

Se il trasmettitore é piú veloce di quanto lo sia il ricevitore esiste il pericolo di saturare di dati il ricevitore, deve quindi esistere un meccanismo per comunicare al trasmettetitore di fermare la trasmissione e attendere.

Controllo di sequenza

- Durante la trasmissione puó accadere che:
  - alcune trame vadano perdute,
  - alcune trame vengano ricevute piú di una volte e quindi replicate
  - alcune le trame vengano ricevute in ordine diverso da quello di trasmissione.
- É pertanto necessario controllare che le trame ricevute siano nella giusta sequenza.
- Avviene tramite numerazione delle trame.
- Occorre porre attenzione a non confondere trame diverse con uguale numero di sequenza.

Slide 51

Controllo dell'errore

La trasmissione é in generale affetta da errori dovuti a fenomeni quali disturbi, interferenza, perdita di sincronismo ecc., ed é perció caratterizzata da un certo tasso d'errore.

É possibile rivelare o correggere questi errori tramite una opportuna codifica dei messaggi:

correzione di errore : richiede molti bit aggiuntivi di controllo, permette di identificare e correggere gli errori,

rivelazione d'errore : richiede pochi bit aggiuntivi di controllo, ma rivela solamente l'errore, **per cui é necessario ritrasmettere i dati errati per avere una comunicazione corretta**.

Nota: La decisione se usare una tecnica o l'altra viene fatta valutando la convenienza in termini di bit totali da trasmettere.

Slide 52

Esempio: Supponiamo di avere un canale con messaggi lunghi 1000 bit.

Supponiamo che un codice a correzione di errore richieda 10 bit di controllo per ogni trama.

Per rivelare un singolo errore basta un solo bit di paritá. Alla rivelazione dell'errore fa poi seguito la ritrasmissione della trama.

Supponiamo di avere un canale con tasso di errore per bit pari a 10^{-6}. É necessaria una ritrasmissione mediamente una volta ogni 1000 trame.

Se si vuole trasferire 1 Mbit di dati, il totale di bit aggiuntivi é di:

- 10000 bit nel caso a correzione,
- 1000+1001=2001 nel caso a rivelazione.

Supponiamo di avere un canale con tasso di errore per bit pari a 10^{-5}. É necessaria una ritrasmissione mediamente una volta ogni 100 trame.

Se si vuole trasferire 1 Mbit di dati, il totale di bit aggiuntivi é di:

- 10000 bit nel caso a correzione,
- 1000+10*1001=11010 nel caso a rivelazione.

Supponiamo di avere un canale con tasso di errore per bit pari a 10^{-4}. É necessaria una ritrasmissione mediamente una volta ogni 10 trame.

Slide 53

Se si vuole trasferire 1 Mbit di dati, il totale di bit aggiuntivi é di:

- 10000 bit nel caso a correzione,
- 1000+111*1001=112111 nel caso a rivelazione.

Si vede come la convenienza di un metodo rispetto all'altro dipenda dal tasso d'errore e quindi dalla qualitá del collegamento.

In generale:

- conviene la rilevazione se il canale é affidabile per cui ci sono pochi errori,
- conviene la correzione se il canale produce molti errori di trasmissione.
- nelle reti di telecomunicazione si mira a rilevare il maggior numero possibile di errori piuttosto che a correggerli.

Slide 54

Codici a blocchi

Si abbiano k bits di informazione che si vogliono proteggere dagli errori mediante r bits di controllo aggiuntivi opportunamente elaborati chiamati di *ridondanza*. L'insieme degli $n = k + r$ bits costituisce il blocco che viene trasmesso.

Codici a paritá

Codici molto semplici che a k bits di informazione aggiungono un solo bit in modo tale che il numero totale di "1" sia *pari* o *dispari*. Tali semplici codici hanno il difetto di non rilevare un numero pari di errori.

Esempio: Usato per proteggere il set di caratteri ASCII a 7 bits

- 7 bits di informazione;
- 1 bit di ridondanza

Slide 55

Codici lineari

L'algebra binaria utilizza l'operazione *ex-or* o modulo 2 ($\oplus$) che origina una struttura di gruppo e con l'aggiunta della moltiplicazione una struttura di anello. Perció, l'addizione si intenderá da ora in poi quella modulo 2 ($\oplus$) che é l'unica, assieme alla sua complementare, ad avere elemento neutro ed opposto.

Si abbiano due messaggi di k bits di informazione $m_k^{(1)}$ e $m_k^{(2)}$ a cui corrispondono le parole di codice $c_n^{(1)}$ e $c_n^{(2)}$. Un codice si dice *lineare* se al messaggio di informazione $m_k^{(3)} = m_k^{(1)} \oplus m_k^{(2)}$ corrisponde la parola di codice $c_n^{(3)} = c_n^{(1)} + c_n^{(2)}$.

Codici sistematici

Un codice si dice *sistematico* quando nella sequenza di n bits sono mantenuti distinti i k bits di informazione dagli r bits di ridondanza.

Slide 56

Codici polinomiali

Con i k bits di informazione viene definito il seguente polinomio di grado $k-1$ nella variabile aleatoria binaria x:

$$P_{k-1}(x) = b_0 + b_1 x + b_2 x^2 + ... + b_{k-1} x^{k-1}$$

I codici polinomiali si basano sulla definizione, a livello di normativa internazionale, di un particolare polinomio $G_r(x)$ detto *generatore* e pertanto noto a tutti.

Slide 57

Codici polinomiali - polinomio trasmesso

Per calcolare gli n bit che verranno trasmessi e perció il corrispondente polinomio $T_{n-1}(x)$ si procede come segue:

1. Si moltiplica il polinomio $P_{k-1}(x)$, rappresentativo dei k bits di informazione, per x^r venendo cosi a porgli in coda r zeri;
2. Si esegue la divisione polinomiale fra $P_{k-1}(x)x^r$ e $G_r(x)$:

$$P_{k-1}(x)x^r = G_r(x)Q_{k-1}(x) + R_{r-1}(x)$$

3. Determinazione di $T_{n-1}(x)$ notando che nell'algebra adottata somma e sottrazione coincidono:

$$T_{n-1}(x) = P_{k-1}(x)x^r + R_{r-1}(x) = G_r(x)Q_{k-1}(x)$$

Nota: La moltiplicazione di $P_{k-1}(x)$ per x^r consente al codice di ottenere la proprietá di sistematicitá. Inoltre $T_{n-1}(x)$ risulta multiplo, per come é stato calcolato, di $G_r(x)$.

Slide 58

Codici polinomiali - polinomio ricevuto

Il polinomio ricevuto verrá indicato con $T'_{n-1}(x)$ per evidenziare la possibilitá che si siano verificati errori nel corso della trasmissione e propagazione nel polinomio $T_{n-1}(x)$.

Perció si avrá:

$$T'_{n-1}(x) = T_{n-1}(x) + E(x)$$

dove $E(x)$ rappresenta il polinomio errore che ha i coefficienti non nulli in corrispondenza dei bit di $T_{n-1}(x)$ che vengono corrotti.

Il ricevitore per rilevare eventuali errori esegue la divisione cercandone il resto:

$$Resto\left[\frac{T'_{n-1}(x)}{G_r(x)}\right] = Resto\left[\frac{T_{n-1}(x)+E(x)}{G_r(x)}\right] = Resto\left[\frac{E(x)}{G_r(x)}\right]$$

In assenza di errori la divisione dá resto nullo. Puó peró accadere che:

$$Resto\left[\frac{E(x)}{G_r(x)}\right] = 0$$

anche se $E(x) \neq 0$ e quindi occorre scegliere accuratamente $G_r(x)$.

Slide 59

Scelta del polinomio generatore

Il polinomio generatore deve essere scelto in modo tale che $E(x)$ non sia divisibile per $G_r(x)$. In realtá, ció non é semplice e quindi tale scelta sará fatta in modo da rendere bassa la possibilitá che $E(x)$ sia multiplo di $G_r(x)$. In generale si osserva che affinché due polinomi siano divisibili occorre che il gr(Num) > gr(Den), altrimenti tale divisione fornisce necessariamente un resto.

Procediamo ora per gradi:

1. Un singolo errore. $E(x)$ puó scriversi come

   $$E(x) = x^i$$

   e pertanto é sufficiente che in $G_r(x)$ vi siano almeno due "1";

2. Numero dispari di errori. Si prenda $G_r(x)$ multiplo di $(1+x)$:

   $$G_r(x) = g(x)(1+x)$$

   Se $E(x)$ é divisibile per $G_r(x)$ lo sará pure per $(1+x)$:

   $$E(x) = e(x)g(x)(1+x)$$

   Ora peró, per ipotesi, $E(1) = 1$ che é in contrasto con l'essere multiplo di $(1+x)$ poiché $(1+1) = 0$ che origina invece $E(1) = 0$. Pertanto si é dimostrato per assurdo che nessun polinomio con un numero dispari di termini é divisibile per $(1+x)$;

Slide 60

3. Numero pari di errori. In questo caso non esiste una regola altrettanto rigorosa ma comunque ITU ha proposto il seguente polinomio che risulta ottimale pure in questo caso:

$$x^{16} + x^{12} + x^5 + 1$$

4. Errori a burst. Nelle reti di telecomunicazione é molto piú frequente rispetto ad una distribuzione uniforme degli errori una concentrazione degli stessi in certi intervalli rispetto ad altri e questo fenomeno si puó vedere come un filotto di bit lungo ℓ i cui bits intermedi sono inaffidabili. Tale filotto é quindi rappresentato da un polinomio di grado $\ell - 1$ per cui si possono avere i seguenti casi:
   - $\ell - 1 < r$: errore viene sempre rilevato;
   - $\ell - 1 = r$: si ha resto nullo $\Leftrightarrow E(x) \equiv G_r(x)$ e questo evento puó verificarsi con prob $= 1/2^{r-1}$;
   - $\ell - 1 > r$: l'errore ora puó sfuggire con prob $= 1/2^r$.

Slide 61

Trasmissione asincrona a carattere

- Protocollo pre-OSI che si colloca a metá strada fra strato 1 e 2.
- Usato nella comunicazione fra terminali e calcolatori tramite linea seriale.
- Viene inviato un carattere per volta, iniziando in un qualunque istante di tempo, per cui l'intervallo che intercorre fra due caratteri non é multiplo del tempo di bit.

Slide 62

Il riconoscimento del carattere avviene come segue:

1. quando non c'é nulla da trasmettere si invia sulla linea sempre lo stesso livello logico (solitamente 1) detto STOP,
2. quando vi é un carattere da trasmettere si ha una transizione di livello da STOP a START, il livello logico START (solitamente 0) viene mantenuto per un tempo di bit,
3. vengono inviati i bit del carattere (utilizzando un opportuno set di codifica dei caratteri, ad esempio ASCII) piú l'eventuale bit di paritá,
4. si conclude la trasmissione con un carattere di STOP che puó avere la durata di 1, 1.5 o 2 tempi di bit, che serve per separare caratteri successivi.

Tramite la transizione STOP-START e il successivo START bit, il ricevitore si sincronizza sui bit di ogni nuovo carattere.

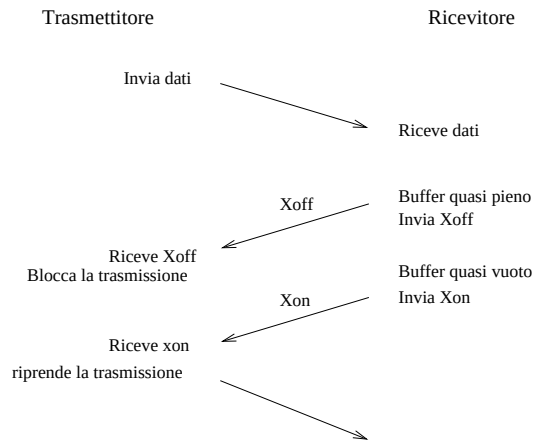
Slide 63

Controllo di flusso, uso di Xon/Xoff

Per effettuare il controllo di flusso in una trasmissione asincrona si usano i due caratteri:

Xoff (codice ASCII 19) con questo carattere il ricevitore dice che non é piú in grado di ricevere e pone quindi il trasmettitore in attesa,

Xon (codice ASCII 17) con questo carattere il ricevitore comunica al trasmettitore che é in grado di ricevere e quindi la comunicazione riprende.



Slide 64

Automatic Repeat Request ARQ

Principio di funzionamento dei protocolli ARQ:

1. il ricevitore controlla le trame ricevute:
   - individua quelle errate, utilizzando codici a rivelazione d'errore,
   - individua quelle replicate o fuori sequenza,
2. le trame errate, replicate o fuori sequenza vengono scartate,
3. viene richiesto al trasmettitore di ritrasmettere le trame errate o fuori sequenza.

Obiettivo: il protocollo deve essere progettato in modo da essere corretto, ossia in modo da garantire un corretto trasferimento dei dati se la rivelazione degli errori di trasmissione é corretta.

Si assume inoltre che ricevitore e trasmettitore siano connessi da un collegamento diretto, per cui:

- le trame arrivano nella stessa sequenza con cui sono inviate a meno degli errori,
- tutte sperimentanoo ritardi di propagazione circa uguali.

Slide 65

Stop and Wait

La connessione é fra A e B.

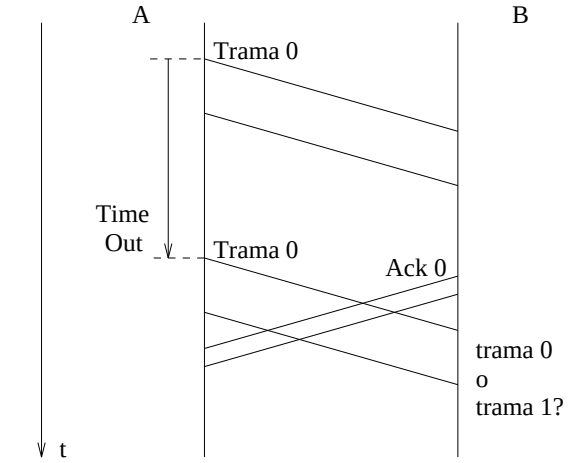
Il protocollo opera come segue:

1. A invia una trama a B
2. B riceve,
   - se la trama é corretta invia una trama di conferma (acknowledgment),
   - se la trama é errata, tace,
3. A trasmette una nuova trama quando:
   - riceve l'acknowledgment da B
   - scade un tempo di time out prefissato, che indica che B non ha ricevuto la trama.

Nota: Il **time out** é necessario per evitare che il protocollo entri in una condizione di stallo se una trama od un acknowlegment vanno perduti (**deadlock**).

Slide 66

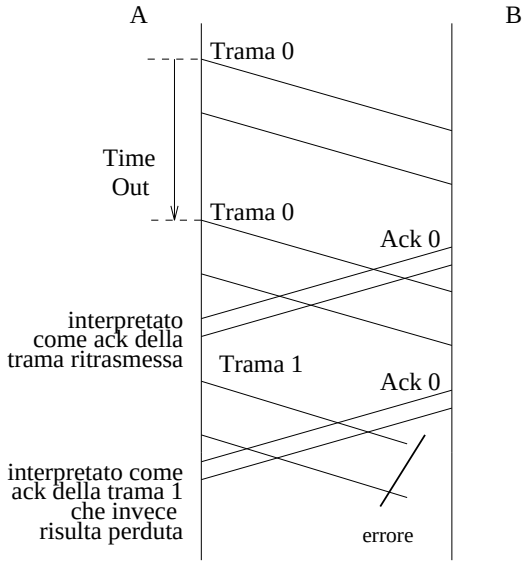
Possibili errori del protocollo e relative soluzioni



- Il time out scatta troppo presto,
- la trama ritrasmessa viene scambiata per la seconda trama del dialogo.

Soluzione: inserire nelle trame un numero sequenziale che serva a distinguerle fra loro.

Slide 67



- due ACK vengono confusi,
- a seguito di un errore si confondono una trama ritrasmessa con una trama nuova e non ci accorge dell'errore sulla trama nuova.

Soluzione: numerare gli ACK in relazione alla trama ricevuta.

Tradizionalmente si usa porre nell'ACK il numero della trama ricevuta + 1 (ossia il numero della prossima trama da ricevere).

Slide 68

Numerazione

Non é praticabile una numerazione delle trame all'infinito, poiché sarebbe necessario un campo infinito nella trama per contenere il numero.

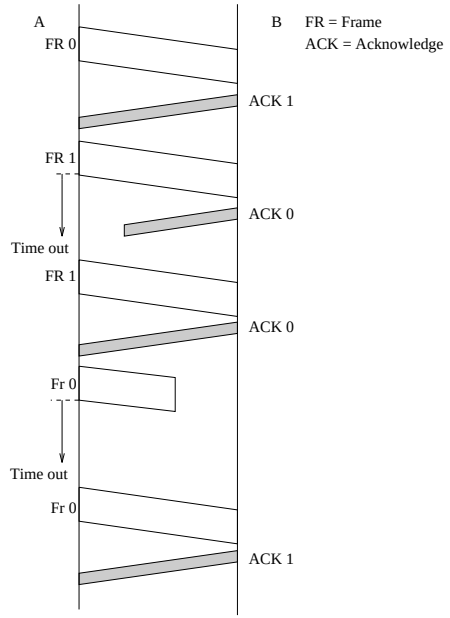
1. Quando A ha inviato la trama i nell'ACK successivo puo' ricevere i seguenti numeri:
   - $i + 1$ se la trama é stata ricevuta correttamente
   - i se la trama non é stata ricevuta correttamente.
2. Quando B ha inviato un ACK con numero i puó ricevere una trama con i seguenti numeri:
   - i se l'ACK é stato ricevuto correttamente,
   - $i - 1$ se l'ACK é andato perduto.

Conclusione

É sufficiente usare una numerazione modulo 2 (cifre 0 ed 1), in quanto sufficiente per non generare ambiguitá.

Slide 69

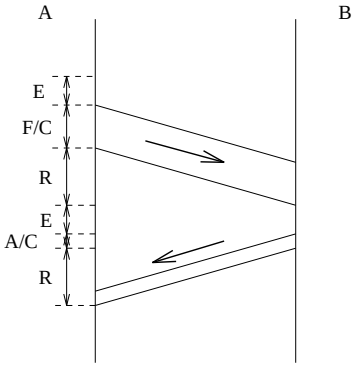
Esempio di funzionamento



Slide 70

Efficienza del protocollo stop-and-wait in assenza di errori

D : numero di bit di dati contenuti in un frame,
H : numero di bit contenuti nell'header (intestazione o PCI) del frame,
$F = D + H$: lunghezza totale del frame,
A : lunghezza dell'ACK,
E : tempo di elaborazione necessario per la composizione del frame in partenza e per il controllo dell'errore sui fame in arrivo,
R : tempo di propagazione del segnale elettrico da un capo all'altro del collegamento,
$I = E + R$,
C : capacitá del canale di trasmissione (in bit al secondo).



Slide 71

Il tempo totale intercorso fra la trasmissione di due frame successivi senza errori é:

$$TT = \frac{F}{C} + I + \frac{A}{C} + I$$

Definizione: Si definisce efficienza η di un protocollo il rapporto $\eta = \frac{D}{TT\ C}$ pari al numero di bit utili trasmessi divisi per il numero totale di bit che si potrebbero trasmettere in TT.

Per cui risulta:

$$\eta = \frac{D}{F + 2IC + A} = \frac{D}{D + H + 2IC + A}$$

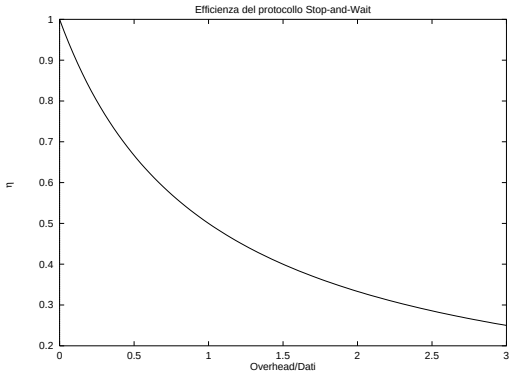
assumendo che l'ACK sia praticamente composto dalla sola PCI di livello 2 risulta $A = H$ e perció:

$$\eta = \frac{D}{D + 2H + 2IC} = \frac{D}{D + 2H + 2IC} = \frac{D}{D + \mathcal{O}}$$

$\mathcal{O}$ viene chiamato **overhead** e rappresenta la quantitá di dati aggiuntivi introdotti dal protocollo (si noti che $\mathcal{O}$ é una grandezza in bit).

η risulta tanto minore quanto maggiore é $\mathcal{O}$.

Slide 72



Nota: Si noti come il valore di $\mathcal{O}$ sia legato non solo alla quantitá di informazioni di controllo introdotte dal protocollo (A ed H), ma anche al tempo di propagazione ed al ritardo di trasmissione.

Un protocollo stop-and-wait é molto inefficiente in caso di:

- elevato I e quindi per collegamenti lunghi,
- grandi C ossia elevate velocitá di trasmissione.

Slide 73

Sliding Window

Il protocollo opera come segue:

1. A puó inviare a B una serie di trama fino ad un numero massimo di W, tutte numerate in successione,
2. quando B riceve una trama opera in modo analogo allo stop-and-wait
   - se una trama é corretta ed in sequenza invia una trama di conferma (acknowledgment),
   - se una trama é errata o fuori sequenza tace oppure invia una trama apposita indicante la ricezione errata (Not ACK o NACK).

La gestione delle trame da inviare avviene come segue:

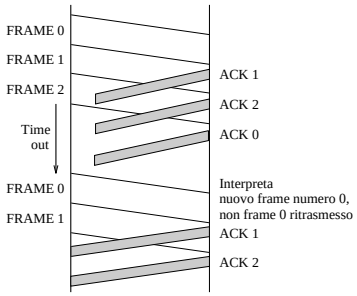
- A invia le trame in sequenza numerandole in modo progressivo,
- l'insieme dei numeri delle W trame inviate viene detto **finestra di trasmissione**, e contiene i numeri delle trame giá inviate e non ancora confermate con ACK dal ricevitore,
- quando il trasmettitore riceve un ACK per una certa trama toglie il relativo numero dalla finestra di trasmissione, lasciando un posto libero per una nuova trama da trasmettere.

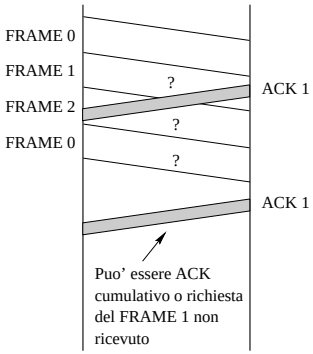
Slide 74

Numerazione

Poiché la finestra di trasmissione puó contenere al piú W trame sembrerebbe sufficiente numerare le trame modulo W ossia con i numeri da 0 a $W - 1$.

In realtá cosí facendo possono aversi i seguenti errori:



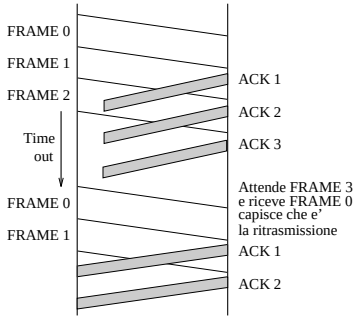


Slide 75

Numerazione

Si usano i numeri da 0 a W invece che a $W - 1$.

Con k bits si ha una finestra di ampiezza massima pari a $W = 2^k - 1$.



Slide 76

ACK implicito

Per risparmiare in trame da trasmettere é possibile effettuare ACK implicito, ossia:

- l'ACK di numero n serve non solo per confermare la ricezione corretta della trama $n-1$ ma anche implicitamente quella di tutte le trame precedenti non ancora confermate.

Slide 77

Controllo di flusso

Se il ricevitore vuole fermare la trasmissione dei dati per un certo tempo, deve semplicemente sopsendere l'invio degli ACK (come nello stop-and-wait).

Controllo dell'errore e della sequenza

Il trasmettitore si accorge se c'é stato un errore di trasmissione sulla trama n perché:

- non riceve l'ACK delle trame successive a $n-1$,

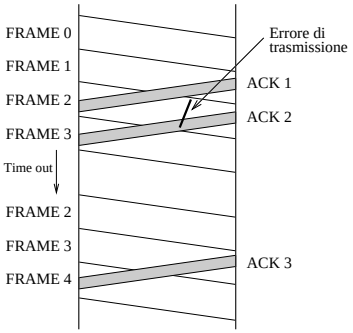
Dopo aver trasmesso tutta la finestra scatta un time-out come nel caso dello stop-and-wait. Scaduto il time out per la ritrasmissione esistono due possibili politiche:

- go back n
- selective repeat

Slide 78

Go back n ARQ

Se una trama risulta errata, si ritrasmettono tutte le trame giá trasmesse a partire da quella errata.

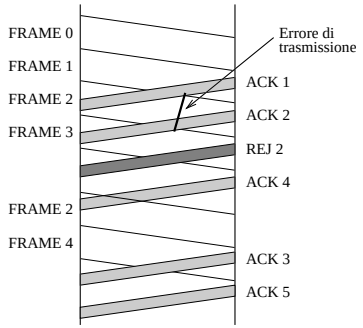


- Perde in efficienza in quanto vengono ritrasmesse informazioni giá inviate correttamente.
- Non richiede elaborazione o memorizzazione aggiuntiva nel ricevitore.

Slide 79

Selective repeat ARQ

Se una trama risulta errata viene ritrasmessa solamente quella.



- É piú efficiente del “go back N” poiché cerca di non ritrasmettere trame giá trasmesse correttamente.
- Le trame devono essere memorizzate nel ricevitore al fine di ripristinare la giusta sequenza.
- Richiede una maggiore capacitá di memorizzazione ed elaborazione in ricevitore e trasmettitore.

Slide 80

Piggybacking

Se il dialogo é full duplex, per inviare il numero di ACK si possono utilizzare direttamente le trame.

In questo caso ogni trama porta due numeri, quello di trama e quello di ACK.

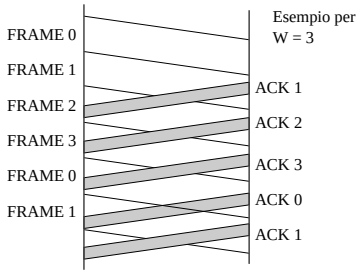
Slide 81

Efficienza dei protocolli sliding-window in assenza di errori

Le grandezze in gioco sono le stesse del caso sotp-and-wait, piú l'ampiezza della finestra W.

Per calcolare l'efficienza di dei protocolli sliding-window é necessario distingure due casi.

$W\ F > C\ TT$

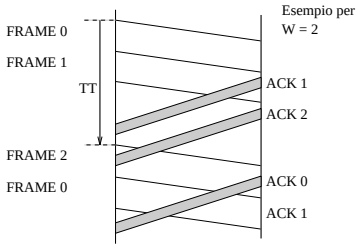


in questo caso il trasmettitore non interrompe mai la trasmissione delle trame per cui risulta:

$$\eta = \frac{D}{D + H}$$

Slide 82

$W\ F \leq C\ TT$



In questo caso si trasmettono W trame in untempo pari a TT calcolato come per lo stop-and-wait, per cui

$$\eta = \frac{W\ D}{D + 2H + 2I}$$

Slide 83

HDLC - High-level Data Link Control (ISO 4335)

Sviluppato per fornire i seguenti servizi:

- connessioni punto-punto e punto-multipunto
- connessioni a breve e lunga distanza,
- connessioni full-duplex e half-duplex
- connessioni basate su comunicazione di tipo primario-secondario (master-slave) o con stazioni di pari importanza (peer-to-peer).

Questi servizi sono tutti previsti e selezionabili all'atto di instaurazione della connessione.

Slide 84

Caratteristiche di base

Tipi di stazioni:

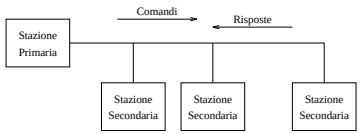
primarie : sono stazioni aventi la capacitá di comandare le operazione sul collegamento (inizio e fine della comunicazione ecc.),

secondarie : sono stazioni che operano solamente a fronte di un comando proveniente da una stazione primaria,

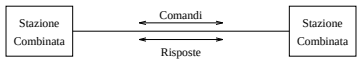
combinate : sono stazioni in grado di funzionare sia come primarie sia come secondarie,

Configurazioni del collegamento:

sbilanciato : viene utilizzato nei collegamenti punto-punto e punto-multipunto e consiste di una stazione primaria ed una o piú stazioni secondarie,



bilanciato : viene usato solamente in comunicazioni punto-punto e consiste di due stazioni combinate,



Slide 85

Modalitá di trasferimento delle informazioni:

Normal response mode (NRM) : detta anche configurazione sbilanciata, la stazione primaria (master) inizia la comunicazione e le stazioni secondarie (slave) possono solamente rispondere a fronte di una richiesta esplicita,

Asyncronous balanced mode (ABM) : detta configurazione bilanciata, le due stazioni sono combinate e possono iniziare entrambe in qualunque momento la comunicazione,

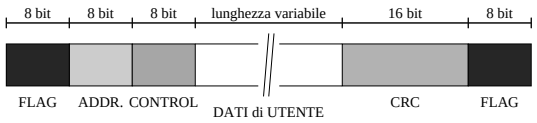
Asyncronous response mode (ARM) : é simile alla precedente ma solamente una delle due stazioni ha la responsabilitá dell'inizializzazione del collegamento.

Slide 86

Struttura della trama

L'unitá informativa é il bit (non il carattere), per cui si dice che HDLC é orientato al bit.

Il formato della trama é il seguente.



FLAG 8 bit, usato per delimitare la trama,

ADDRESS 8 bit (estenbile a 16), contiene l'indirizzo, del destinatario o della sorgente a seconda dei casi,

CONTROL 8 bit (estendibile a 16) serve per definire il tipo di trama,

DATI contiene i dati di utente (livello 3) e non ha una lunghezza prefissata,

CRC 16 bit, contiene i bit per il controllo d'errore calcolati utilizzando un codice polinomiale.

Slide 87

Delimitazione della trama

Per riconoscere l'inizio e la fine della trama si utilizza una speciale sequenza di 8 bit detta **flag** (bandiera):

`01111110`

- Poiché la trama HDLC deve essere in grado di trasportare qualunque sequenza di bit come dati di utente, esiste il pericolo che una sequenza di bit uguale al flag sia presente nei dati.
- Questo provocherebbe una errata delimitazione della trama.
- Per evitarlo si usa il meccanismo del **bit stuffing**.

Dopo aver inserito il flag di inizio, quando durante la trasmissione dei dati viene vista una sequenza di 5 bit con valore `1` gli viene *automaticamente* inserito uno 0 in coda.

Esempio:

`1111111111110011111000111000000111111`

`11111011111011001111100001110000001111101`

Slide 88

Tipi di trame

Il campo **control** della trama viene utilizzato per distinguere tre tipi di trame, come segue:

bit	1	2	3	4	5	6	7	8	
	0	N(S)			P/F	N(R)			I informative
	1	0	S		P/F	N(R)			S supervisione
	1	1	M		P/F	M			U non numerate

N(S) : numero di trama inviata
N(R) : numero di trama ricevuta (+1)
S : tipo di trama di supervisione (2 bit per un massimo di 4 trame)
M : tipo di trama non numerata (5 bit per un massimo di 32 trame)
P/F : bit poll/final

Slide 89

trame informative : sono utilizzate per il trasporto delle informazioni nella fase di dialogo normale, il campo control contiene i numeri di trama inviata e trama ricevuta, necessari per l'ARQ,

trame di supervisione : servono per regolare il dialogo, in alternativa al piggybacking, contengono il solo numero di trama ricevuta e sono di quattro tipi:

- Receiver Ready (RR) : serve come acknowledge se non si hanno trame dati da trasmettere,
- Receiver not Ready (RNR) : come la precedente, ma sospende l'invio dei delle trame fino all'invio di un RR,
- Reject (REJ) : serve come "not acknowledge", quando si riceve una trama errata,
- Selective reject (SRJ) : serve per richiere la ritrasmissione quando si usa un protocollo selective repeat.

trame non numerate : sono utilizzate per tutte la varie funzioni di controllo, non contengono numerazione e non alterano la sequenza dei numeri di finestra.

Slide 90

Uso del campo di indirizzo

- Non é necessario nelle comunicazioni punto-punto, ma viene trasmesso ugualmente per uniformitá.
- Nelle comunicazioni punto-multipunto contiene sempre l'indirizzo della stazione slave a cui il master vuole mandare informazioni o dalla quale le riceve.

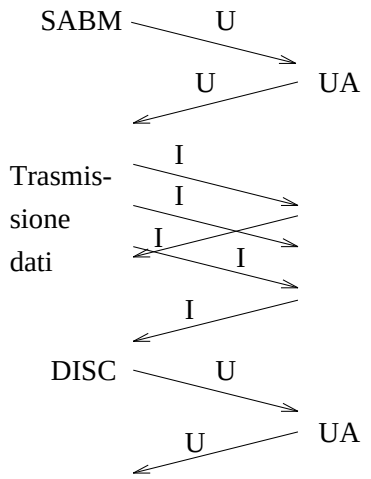
Uso del bit P/F

- nel caso di dialogo NRM, con collegamento punto multipunto, il bit P/F viene posto ad 1 quando il master intende dare il permesso di parlare ad una stazione slave, questa risponde con il bit P/F a 0 e lo pone ad 1 quando invia l'ultima trama, cosicché il master possa riprendere il controllo della linea,
- nel dialogo ABM punto-punto viene usato per la gestione delle ritrasmissioni e delle pause nelle trame S.

Slide 91

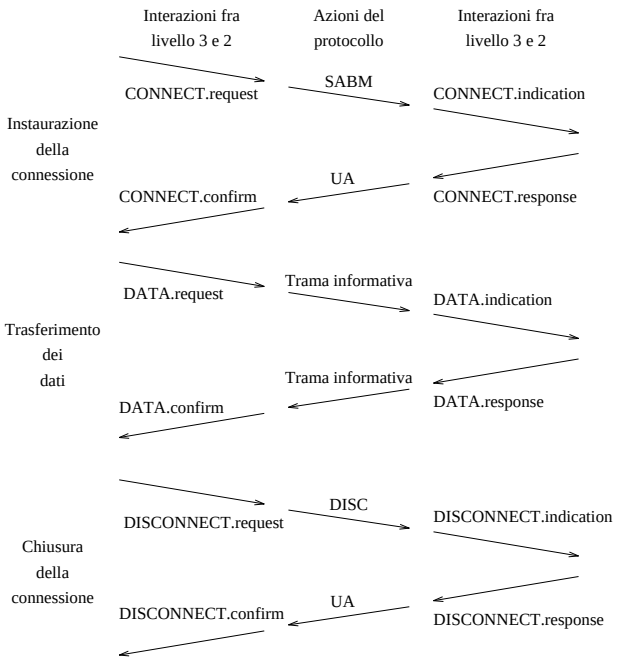
Esempio di comunicazione bilanciata

1. Chi attiva la chiamata invia una trama di inizializzazione (Set Asyncronous Balanced Mode - SABM),
2. l'accettazione viene notificata tramite un Unnumbered Acknowledgment - UA,
3. la chiamata é attiva ed é possibile inviare trame informative nelle due direzioni,
4. per concludere viene inviata una trama non numerata di disconnessione (DISC).



Slide 92

Inquadramento della procedura bilanciata nel modello OSI



Slide 93

Teoria del traffico

Slide 94

Processi stocastici

Sia $X(t)$ una funzione aleatoria di cui non é nota a priori l'evoluzione nel tempo.

All'atto di un esperimento si manifesta un certo andamento appartenente all'insieme di tutti (solitamente infiniti) i possibili andamenti di $X(t)$, che indichiamo con:

$$X_1(t), X_2(t), \cdots, X_N(t), \cdots$$

Definizione: Si definisce *processo stocastico*, l'insieme delle funzioni $X_i(t)$.

Definizione: Ogni funzione $X_i(t)$ si dice *realizzazione* o *funzione campione* del processo.

Nel seguito verranno considerati processi stocastici:

- discreti nei valori e tempo continui,
- discreti nei valori e tempo discreti.

Nota: Nel caso di processi stocastici a valori discreti l'insieme dei valori che puó assumere la variabile aleatoria $X(t)$ é anche detto *spazio degli stati*.

Slide 95

Esempi

- Il numero delle chiamate telefoniche giunte ad una centrale ad un certo istante, a partire dall'inizio della giornata.
- La lunghezza totale in bit dei file su di un hard disk.
- Il numero di volte che un utente telefonico trova occupato quando chiama.

Slide 96

Funzione probabilitá di massa del primo ordine

Fissato un certo istante di osservazione t_0, il valore $X(t_0)$ assunto dalla funzione aleatoria é una variabile aleatoria discreta.

Definizione: La funzione probabilitá di massa

$$p_1 = \mathrm{P}(X(t_0), t_0)$$

é detta *funzione probabilitá di massa del primo ordine* del processo stocastico.

Definizione: La probabilitá congiunta

$$p_2 = \mathrm{P}((X(t_0), t_0), (X(t_1), t_1))$$

é detta *funzione probabilitá di massa del secondo ordine* del processo stocastico.

In modo analogo si possono definire le p_i per valori di i superiori a 2.

Slide 97

Valori medi

Dato un processo stocastico discreto nei valori e tempo continuo, possiamo definire due tipi di medie:

- medie statistiche o di insieme,
- medie temporali.

Valor medio statistico é funzione del tempo ed é definita come:

$$E[X(t)] = \sum_{\forall k} X_k p_1(X_k(t), t)$$

ed é determinabile a priori nota la p_1,

Valor medio temporale non é funzione del tempo ed é definita come:

$$< X(t) >= \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t) dt$$

ed é determinabile solamente a posteriori noto l'andamento della funzione campione.

Slide 98

Processi stazionari ed ergodici

Definizione: Un processo stocastico si dice *stazionario* quando i processi $X(t)$ e $X(t+t_0)$ hanno la medesima descrizione statistica e quindi sono il medesimo processo, qualunque sia t_0.

Teorema 1 *Se un processo é stazionario valgono le seguenti proprietá:*

1. $p_1(X(t),t) = p_1(X)$
2. $E[X(t)] = E[X] \ \forall t$

Definizione: Un processo stocastico si dice *ergodico* se ogni funzione campione é caratteristica con probabilitá 1 del processo stesso.

Ció significa che il risultato di un qualunque rilievo statistico, effettuato su di una generica funzione campione, coincide con probabilitá 1, con quello dell'analogo rilievo effettuato sull'insieme delle funzioni campione ad un tempo qualunque. In particolare per il valor medio:

$$E[X(t)] = < X(t) > \quad \forall t$$

Teorema 2 *Un processo ergodico é necessariamente stazionario.*

Slide 99

Processo di Bernoulli

- É un processo tempo discreto a valori discreti. Essendo tempo discreto la variabile temporale $t_i = 1, 2 \ldots \infty$ é un numero naturale e nel seguito la funzione $X(t_i)$ si indicherá con X_i
- I possibili valori di X_i sono:

  0 insuccesso,

  1 successo.
- La funzione probabilitá di massa per X_i e':

  $p_1(0) = q \ \forall t_i$

  $p_1(1) = p = 1 - q \ \forall t_i$
- I risultati in intervalli successivi sono fra loro indipendenti.

Il processo é stazionario per definizione ($p_1(X)$ non dipende dal tempo).

Il processo é ergodico.

Slide 100

La probabilitá di avere k successi su N intervalli di tempo é data dalla formula:

$$P(N,k) = \binom{N}{k} p^k \, q^{N-k}$$

$P(N,k)$ é detta **formula binomiale**.

Teorema 3

$$E[k] = \sum_{i=0}^{N} i \, P(N,i) = Np$$

Teorema 4

$$\mathrm{var}(k) = N \, p \, (1-p)$$

Slide 101

La probabilitá di avere un intervallo lungo n fra due arrivi successivi é data dalla sequente probabilitá congiunta:

$$P(n) = \mathrm{P}(0 \text{ arrivi nei primi } n-1 \text{ intervalli,}$$

$$1 \text{ arrivo nell}'\text{intervallo n} - \text{esimo}) =$$

$$= q^{n-1} \, p$$

$P(n)$ é detta funzione probabilitá di massa **geometrica**.

Teorema 5

$$E[n] = \sum_{i=1}^{\infty} i \, q^{i-1} \, p = \frac{1}{p}$$

Teorema 6

$$\mathrm{var}(n) = \frac{1-p}{p^2}$$

Slide 102

Processo di Poisson

Fu introdotto come approssimazione di quello di Bernoulli quando il valore di N é molto elevato.

- Consideriamo l'asse dei tempi suddiviso in intervalli di lunghezza Δt.
- La probabilitá p di avere un arrivo in Δt é $p = \lambda\,\Delta t$.
- $N = \frac{t}{\Delta t}$ é il numero di intervalli da 0 a t.

La probabilitá di avere k arrivi nell'intervallo $[0,\, t]$ é data dalla formula binomiale $P(N, k)$.

Se $\Delta t \to 0$ anche $N \to \infty$, $p \to 0$ e:

$$\lim_{N\to\infty} \binom{N}{k} p^k (1-p)^{N-k} = \lim_{N\to\infty} \binom{N}{k} \left(\frac{\lambda t}{N}\right)^k \left(1 - \frac{\lambda t}{N}\right)^{N-k}$$

$$= \lim_{N\to\infty} \frac{N!}{k!(N-k)!} (\lambda t)^k \left(\frac{1}{N^k}\right) \left(1 - \frac{\lambda t}{N}\right)^{-k} \left(1 - \frac{\lambda t}{N}\right)^{N}$$

$$= \frac{(\lambda t)^k}{k!} \lim_{N\to\infty} \frac{N!}{(N-k)!} \frac{1}{(N-\lambda t)^k} \left(1 - \frac{\lambda t}{N}\right)^{N}$$

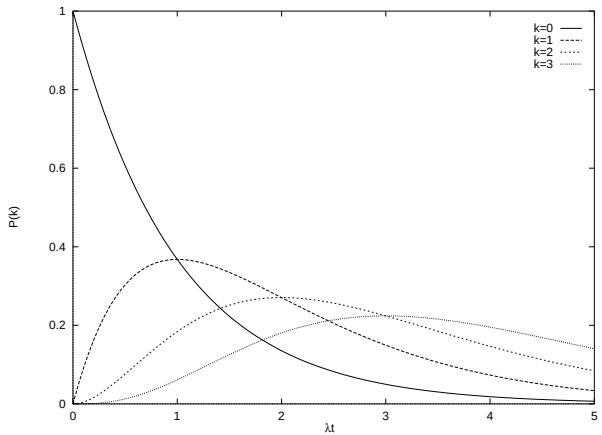
$$= \frac{(\lambda t)^k}{k!} \lim_{N\to\infty} \frac{N}{N-\lambda t} \frac{N-1}{N-\lambda t} \cdots \frac{N-k+1}{N-\lambda t} \left(1 - \frac{\lambda t}{N}\right)^{N}$$

$$= \frac{(\lambda t)^k}{k!} e^{-\lambda t} \Rightarrow \mathrm{P}(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Slide 103

Andamento della distribuzione di Poisson

$$\mathrm{P}(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

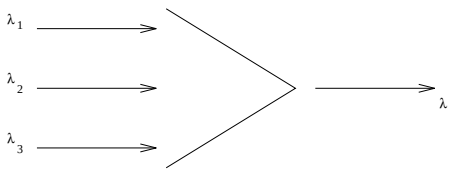


Slide 104

Proprietá

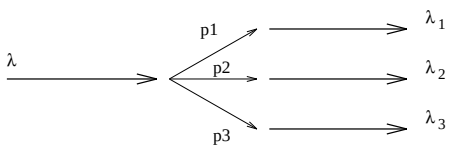
La sovrapposizione di h processi di Poisson aventi frequenze di arrivo λ_i $i = 1 \dots h$ é ancora un processo di Poisson con

$$\lambda = \sum_{i=1}^{h} \lambda_i$$



La scomposizione casuale di un processo di Poisson in h processi diversi da origine ancora ad h processi di Poisson aventi frequenze di arrivo:

$$\lambda_i = p_i \, \lambda \;\; \forall i$$



Slide 105

Teorema 7 *Il numero medio di arrivi in un tempo t é:*

$$E[k] = \bar{k} = \sum_{i=1}^{\infty} i \, \frac{(\lambda t)^i}{i!} e^{-\lambda t} = \lambda t$$

Teorema 8 *La varianza del numero di arrivi in un tempo t é:*

$$\mathrm{var}[k] = \sigma_n(k)^2 = \lambda t$$

Slide 106

Distribuzione del tempo interarrivo

Si calcola la distribuzione di probabilitá del tempo che intercorre fra due arrivi successivi di un processo di Poisson.

$$F(t) = \mathrm{P}(\text{tempo interarrivo} \leq t) =$$

$$= 1 - p(0,t) = 1 - e^{-\lambda t}$$

$F(t)$ é detta distribuzione *esponenziale negativa*. La corrispondente densitá di probabilitá é:

$$f(t) = \frac{dF(t)}{dt} = \lambda\, e^{-\lambda t}$$

Teorema 9

$$E[t] = \frac{1}{\lambda}$$

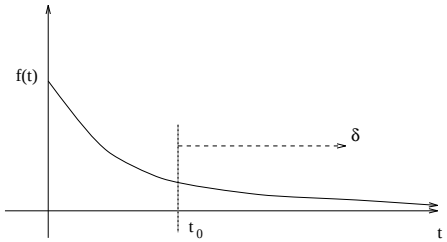
Teorema 10

$$\mathrm{var}(t) = \frac{1}{\lambda^2}$$

Slide 107

Distribuzione del tempo residuo

Calcoliamo la distribuzione del tempo di attesa *residuo* (δ) per il prossimo arrivo dato che si é giá atteso un tempo t_0.



$$\begin{aligned} F(\delta) &= \mathrm{P}(t \leq t_0 + \delta \mid t > t_0) \\ &= \frac{\mathrm{P}(t_0 < t \leq t_0+\delta)}{\mathrm{P}(t > t_0)} \\ &= \frac{\mathrm{P}(t \leq t_0+\delta) - \mathrm{P}(t \leq t_0)}{\mathrm{P}(t > t_0)} \\ &= \frac{1-e^{-\lambda(\delta+t_0)} - (1-e^{-\lambda t_0})}{1-(1-e^{-\lambda t_0})} = 1 - e^{-\lambda\delta} \end{aligned}$$

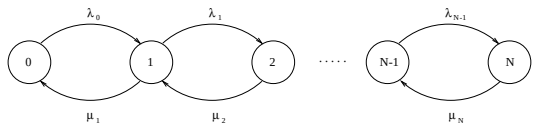
- La distribuzione esponenziale é *priva di memoria*.
- La distribuzione esponenziale é l'unica distribuzione continua avente questa proprietá.

Slide 108

Processi di nascita e morte

Definizione: Un processo di *nascita e morte* é un processo stocatico tempo continuo a valori discreti tale che:

1. la probabilitá di trovarsi in un certo stato nel futuro dipende solamente dallo stato attuale del sistema,
2. il suo stato puó variare di una sola unitá alla volta.



- λ_k : frequenza di nascita dallo stato k
- μ_k : frequenza di morte dallo stato k

Slide 109

Probabilitá di stato

Preso un intervallo Δt supponiamo di *trascurare* le probabilitá che avvengano piú di una nascita o di una morte per intervallo.

Allora possiamo scrivere le seguenti equazioni che legano fra loro le probabilitá di stato.

$$P_k(t+\Delta t) = P_{k-1}(t)p_{k-1,k}(\Delta t) + P_{k+1}(t)p_{k+1,k}(\Delta t) +$$
$$+P_k(t)p_{k,k}(\Delta t)$$

$$P_0(t+\Delta t) = P_1(t)p_{1,0}(\Delta t) + P_0(t)p_{0,0}(\Delta t)$$

Inoltre:

$$p_{k,k}(\Delta t) = 1 - (\lambda_k + \mu_k)\Delta t$$

$$p_{k-1,k}(\Delta t) = \lambda_{k-1}\,\Delta t$$

$$p_{k+1,k}(\Delta t) = \mu_{k+1}\,\Delta t$$

Da queste facendo il limite per $\Delta t \to 0$ otteniamo:

$$\frac{dP_k(t)}{dt} = -(\lambda_k + \mu_k)P_k(t) + \lambda_{k-1}P_{k-1}(t) + \mu_{k+1}P_{k+1}(t)$$

$$\frac{dP_0(t)}{dt} = -\lambda_0 P_0(t) + \mu_1 P_1(t)$$

Slide 110

Se il processo é stazionario, una volta raggiunto l'equilibrio, le P_k non dipendono dal tempo, perció otteniamo:

$$\frac{dP_k(t)}{dt} = 0 \ \ \forall t$$

e quindi:

$$\begin{cases} -\lambda_0 P_0 + \mu_1 P_1 = 0 \\ -(\lambda_k + \mu_k) P_k + \lambda_{k-1} P_{k-1} + \mu_{k+1} P_{k+1} = 0 \\ \sum_{i=0}^{N} P_i = 1 \end{cases}$$

Risolvendo questo sistema otteniamo:

$$P_k = P_0 \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}$$

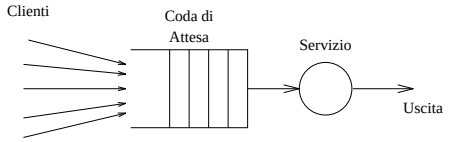
$$P_0 = \frac{1}{1 + \sum_{k=1}^{N} \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}}$$

Slide 111

Sistemi a coda

Si definisce *sistema a coda* un sistema nel quale un flusso di entitá, chiamate **clienti**, richiede un certo tipo di **servizio**, fornito da opportuni servitori.

I clienti possono accedere immediatamenti al servizio qualora vi siano servitori disponibili altrimenti possono essere posti in attesa, oppure rifiutati.



Gli arrivi dei clienti, cosí come i tempi di servizio sono in genere grandezze aleatorie ed in particolare processi stocastici, per cui il comportamento del sistema a coda é in generale un processo stocastico.

Nota: In seguito assumeremo che questo processo sia stazionario, per cui raggiunga l'equilibrio a seguito di un periodo transitorio.

Slide 112

Per specificare un sistema a coda é necessario specificarne le modalitá di funzionamento che nelle linee generali si possono riassumere nei seguenti punti:

- tipo di popolazione dei clienti, in particolare se la popolazione é in numero finito o infinito,
- il tempo fra arrivi successivi, questa grandezza é solitamente un processo stocastico,
- il numero di servitori esistenti,
- il tempo di servizio richiesto da ogni cliente, anche questa grandezza é solitamente rappresentata da un processo stocastico,
- la politica con cui i servitori scelgono i clienti da servire nell'insieme di quelli in attesa (scheduling),
- il numero di posti di attesa presenti all'interno del sistema.

Slide 113

Scheduling

I seguenti sono alcuni dei metodi di scheduling piú frequentemente utilizzati:

FCFS First Come, First Served : il primo entrato é anche il primo servito,

LCFS Last Come, First Served : l'ultimo entrato é anche il primo servito,

PS Processor Sharing : i servitori vengono condivisi fra i clienti mano a mano che arrivano.

PRIO : scheduling con prioritá.

Slide 114

Notazione di Kendall

Per descrivere in modo compatto queste caratteristiche é stata coniata una notazione abbreviata (notazione di Kendall) A/B/m/n/p/S, con i seguenti significati:

A descrive il processo degli arrivi,

B descrive il processo di servizio,

m indica il numero di servitori,

n indica il numero di posti in coda,

p indica il numero di clienti,

S indica la modalitá di scheduling.

Molto spesso nella notazione di Kendall si omettono n, p, S. In questo caso si sottintende un numero infinito di posti di attesa, una popolazione infinita di clienti e una politica di scheduling FCFS.

Alcuni processi utilizzati sovente per lo studio di sistemi a coda sono:

M processo di Poisson,

G processo generico (Generale),

D processo deterministico.

Slide 115

Risultato di Little

Definiamo le grandezze:

1. $\bar{N}$: numero di utenti nel sistema al tempo t
2. λ_s : frequenza media di arrivo dei clienti nel sistema,
3. $\bar{\delta}$: tempo medio speso nel sistema da ogni cliente,
4. $\bar{\eta}$: tempo medio speso nella coda di attesa da ogni cliente,
5. $\bar{\theta}$: tempo medio di servizio richiesto da un cliente.

Il *risultato di Little* dice che esiste il seguente legame fra queste grandezze:

$$\bar{N} = \lambda_s \bar{\delta}$$

ossia:
il numero medio di utenti presenti in sistema a coda é dato dal prodotto fra la frequenza media di arrivo dei clienti per il tempo medio che ogni cliente spende nel sistema.

Nota: Questo risultato é valido per qualunque sistema a coda, qualunque sia il processo degli arrivi e qualunque sia il processo di servizio.

Slide 116

In particolare il risultato di Little vale anche per le varie parti del sistema a coda.

Per il solo servizio

$$\bar{N}_s = \lambda_s \bar{\theta}$$

Per la sola coda di attesa

$$\bar{N}_c = \lambda_s \bar{\eta}$$

Slide 117

Utilizzazione

Definizione: Si definisce *utilizzazione* U del sistema a coda il rapporto:

$$U = \frac{\lambda_s \bar{\theta}}{m}$$

Osservazioni:

- Affinché il sistema sia stabile deve essere

  $$0 \leq U < 1$$

  altrimenti il numero degli utenti nel sistema cresce all'infinito.

- U fornisce la percentuale media di tempo per cui un servitore risulta occupato (nell'ipotesi che tutti i servitori siano uguali).

Slide 118

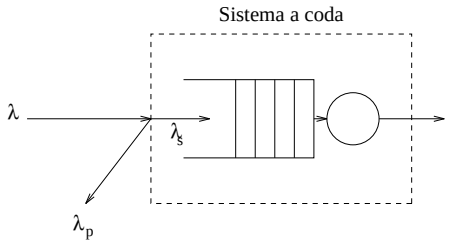
Sistemi con perdita

Se il numero di posti all'interno del sistema a coda é **finito** allora e' possibile che un cliente si presenti al sistema e gli venga rifiutato l'ingresso. In questo caso si dice che: **il cliente e' bloccato o perduto**.

Definiamo λ la frequenza con cui gli utenti richiedono di entrare nel sistema, una parte di questi verrá accettata e una parte rifiutata. Vale la relazione:

$$\lambda = \lambda_s + \lambda_p$$

dove λ_p é la frequenza con cui i clienti vengono rifiutati.



Slide 119

Traffico Offerto

Definizione: Si definisce *intensitá media di traffico offerto* ad un sistema la quantità:

$$A_0 = \lambda\bar{\theta}$$

Definizione: Si definisce *intensitá media di traffico smaltito* dal sistema a coda la quantitá:

$$A_s = \lambda_s\bar{\theta} = mU$$

Definizione: Se il sistema é a perdita si definisce *intensitá media di traffico perduto* la quantitá:

$$A_p = \lambda_p\bar{\theta} = A_0 - A_s$$

É ovvio che per un sistema senza perdita valgono le relazioni:

$$A_s = A_0 \quad A_p = 0$$

Queste quantitá sono numeri puri, ma convenzionalmente si misurano in **Erlangs (Erl)**.

Esempio: Se si dice che $A_s = 1$Erl cio' significa, che se il sistema ha un solo servitore, questo é occupato per il 100% del tempo, se ha m servitori questi sono impegnati per $\frac{100}{m}\%$ del tempo.

Slide 120

Coda M/M/1

- Arrivi poissoniani con frequenza di arrivo λ
- Durata del servizio esponenziale con durata media $\bar{\theta} = \frac{1}{\mu}$
- Un solo servitore
- Coda infinita

Probabilitá di stato

Le frequenze di nascita e di morte risultano:

$$\begin{cases} \lambda_n = \lambda \ \ \forall n \\ \\ \mu_n = \mu \ \ \forall n \end{cases}$$

da cui

$$P_n = P_0 \left(\frac{\lambda}{\mu}\right)^n = P_0 \, \rho^n$$

$$\rho = A_0 = A_s = \frac{\lambda}{\mu}$$

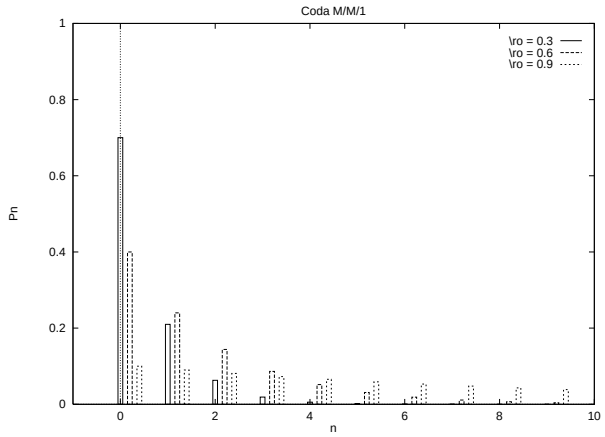
Se $\rho < 1$ il sistema é stabile ed il valore di P_0 risulta:

$$P_0 \sum_{i=0}^{\infty} \rho^n = 1 \Rightarrow \sum_{i=0}^{\infty} \rho^n = \frac{1}{1-\rho} \Rightarrow P_0 = 1 - \rho$$

e quindi

$$P_n = (1-\rho)\rho^n$$

Slide 121



Slide 122

Numero medio di utenti nel sistema

Il numero di utenti nel sistema $\bar{N}$ risulta:

$$\begin{aligned}
\bar{N} &= \mathrm{E}[n] = \textstyle\sum_{n=0}^{\infty} nP_n \\
&= \textstyle\sum_{n=0}^{\infty} n(1-\rho)\rho^n \\
&= \rho(1-\rho)\textstyle\sum_{n=0}^{\infty} n\rho^{n-1} \\
&= \rho(1-\rho)\tfrac{d}{d\rho}\left(\textstyle\sum_{n=0}^{\infty} \rho^n\right) \\
&= \rho(1-\rho)\tfrac{d}{d\rho}\left(\tfrac{1}{1-\rho}\right) \\
&= \rho(1-\rho)\tfrac{1}{(1-\rho)^2} \\
&= \tfrac{\rho}{1-\rho}
\end{aligned}$$

Il tempo medio $\bar{\delta}$ speso nel sistema da ogni utente si ricava dal risultato di Little come segue:

$$\bar{\delta} = \frac{\bar{N}}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{\frac{1}{\mu}}{1-\rho}$$

Infine la probabilitá di avere piú di k utenti nel sistema é:

$$\mathrm{P}(n \geq k) = \sum_{i=k}^{\infty} P_i = \sum_{i=k}^{\infty}(1-\rho)\rho^i = \rho^k$$

Slide 123

Numero medio di utenti in coda

Il numero medio di utenti in coda si ottiene come segue:

$$\begin{aligned}
\bar{N}_c &= \textstyle\sum_{n=1}^{\infty} nP_{n+1} \\
&= \textstyle\sum_{n=1}^{\infty} n(1-\rho)\rho^{n+1} \\
&= \rho^2(1-\rho)\tfrac{d}{d\rho}\textstyle\sum_{n=0}^{\infty} \rho^n \\
&= \tfrac{\rho^2(1-\rho)}{(1-\rho)^2} \\
&= \tfrac{\rho^2}{1-\rho}
\end{aligned}$$

Il tempo medio $\bar{\eta}$ di attesa in coda per ogni utente si ricava dal risultato di Little come segue:

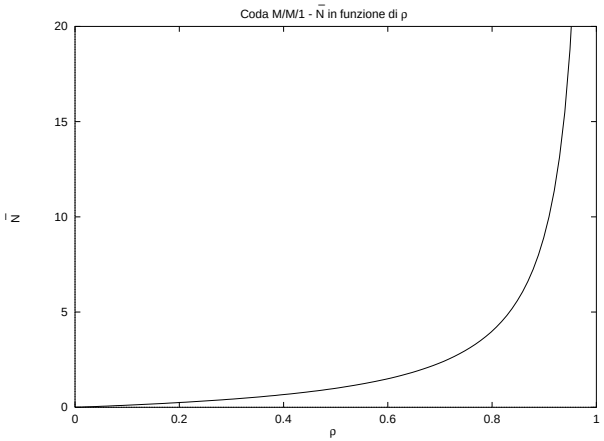
$$\bar{\eta} = \frac{\bar{N}_c}{\lambda} = \frac{\rho^2}{\lambda(1-\rho)}$$

Si noti anche che:

$$\begin{aligned}
\bar{\delta} - \bar{\eta} &= \tfrac{\rho}{\lambda(1-\rho)} - \tfrac{\rho^2}{\lambda(1-\rho)} \\
&= \tfrac{\rho}{\lambda(1-\rho)}(1-\rho) = \tfrac{\rho}{\lambda} = \tfrac{1}{\mu} = \bar{\theta}
\end{aligned}$$

per cui come era lecito attendersi il tempo medio di attesa totale é dato dalla somma fra tempo medio di attesa in coda e durata media del servizio.

Slide 124



Slide 125

Distribuzione del tempo di permanenza nel sistema δ

Un utente entra nel sistema e vi trova n altri utenti di cui uno in servizio ed $n-1$ in attesa. Prima di lasciare il sistema deve attendere che:

1. lasci il sistema il cliente che é in corso di servizio al momento dell'arrivo,
2. vengano serviti gli $n-1$ utenti in attesa,
3. venga servito lui stesso.

Complessivamente deve attendere $n+1$ tempi di servizio tutti distribuiti esponenzialmente con valor medio $\bar{\theta} = \frac{1}{\mu}$.

Sia

- θ_i il tempo di servizio del generico cliente i-esimo.
- $\eta/n = \sum_{i=1}^{n} \theta_i$ tempo di attesa condizionato al fatto di trovare all'arrivo n utenti nel sistema,

Slide 126

Si ricordi che: $f_{\theta_i}(t) = \mu e^{-\mu t}$.

La distribuzione di η_n é la convoluzione delle $f_{\theta_i}(t)$ e vale:

$$f_{\eta/n}(t) = \mu \frac{(\mu t)^{(n-1)} e^{-\mu t}}{(n-1)!}$$

La distribuzione del tempo speso nel sistema qualunque sia il numero di utenti incontrati all'arrivo si calcola mediante la formula:

$$\begin{aligned}
&f_\delta(t) = \textstyle\sum_{n=0}^{\infty} P_n f_{\theta_{n+1}}(t) \\
&= \textstyle\sum_{n=0}^{\infty} (1-\rho)\rho^n \mu (\mu t)^n \frac{e^{-\mu t}}{n!} \\
&= \mu(1-\rho) \textstyle\sum_{n=0}^{\infty} (\lambda t)^n \frac{e-\mu t}{n!} \\
&= (\mu - \lambda) e^{-\mu t} e^{\lambda t} \\
&= (\mu - \lambda) e^{-(\mu-\lambda)t}
\end{aligned}$$

e risulta essere un esponenziale con valor medio $\frac{1}{\mu-\lambda}$.

Slide 127

Distribuzione del tempo di attesa prima di essere serviti

Si devono distinguere due casi:

- si trova almeno un utente nel sistema e quindi si deve attendere, con probabilitá

$$1 - P_0 = \rho$$

- si trova il sistema vuoto quindi l'attesa é nulla, con probabilitá

$$P_0 = 1 - \rho$$

Con attesa la distribuzione del tempo di attesa é:

$$\begin{aligned}
&f_{\eta|attesa}(t) = \textstyle\sum_{n=1}^{\infty} P_n f_{\theta_n}(t) \\
&= \textstyle\sum_{n=0}^{\infty} (1-\rho)\rho^n \mu (\mu t)^{n-1} \frac{e^{-\mu t}}{(n-1)!} \\
&= \mu(1-\rho) \textstyle\sum_{n=0}^{\infty} (\mu t)^{n-1} \rho^n \frac{e^{-\mu t}}{(n-1)!} \\
&= (1-\rho)\rho e^{-\mu t} \mu e^{\lambda t} \\
&= \rho(\mu - \lambda) e^{-(\mu-\lambda)t} = \rho f_\delta(t)
\end{aligned}$$

Se non si deve attendere il tempo di attesa é deterministicamente nullo, per cui la relativa distribuzione é un impulso centrato in 0.

La distribuzione **complessiva** del tempo di attesa risulta:

$$f_\eta(t) = (1-\rho)\delta(t) + \rho(\mu - \lambda) e^{-(\mu-\lambda)t}$$

che ha una probabilitá non nulla concentrata al tempo 0.

Slide 128

Coda M/M/1/K

- Arrivi poissoniani con frequenza di arrivo λ
- Durata del servizio esponenziale con durata media $\bar{\theta} = \frac{1}{\mu}$
- Un solo servitore
- **Coda finita di lunghezza** K

Probabilitá di stato

Le frequenze di nascita e di morte risultano:

$$\begin{cases} \lambda_n = \lambda & 0 \leq n < K+1 \\ \\ \mu_n = \mu & 0 < n \leq K+1 \end{cases}$$

Ponendo $\frac{\lambda}{\mu} = A_0$

$$P_n = P_0 \left(\frac{\lambda}{\mu}\right)^n = P_0 \, A_0^n$$

Il valore di P_0 si ricava imponendo la condizione di normalizzazione:

$$P_0 \sum_{i=0}^{K+1} A_0^i = 1$$

Slide 129

e con:

$$\sum_{i=0}^{K+1} A_0^i = \begin{cases} \frac{1-A_0^{K+2}}{1-A_0} & A_0 \neq 1 \\ K+2 & A_0 = 1 \end{cases}$$

si ottengono le P_n:

$$P_n = \begin{cases} \frac{(1-A_0)A_0^n}{(1-A_0^{K+2})} & A_0 < 1 \\ \frac{1}{K+2} & A_0 = 1 \end{cases}$$

Nota: In questo caso, poiché il sistema ha coda finita esistono sempre le P_n per ogni valore di A_0, ossia il sistema é sempre stabile.

Se $\lambda \to \infty$ e quindi $A_0 \to \infty$, si vede che:

$$\lim_{A_0 \to \infty} P_n = \lim_{A_0 \to \infty} \frac{(1-A_0)A_0^n}{(1-A_0^{K+2})} = \begin{cases} 0 & \text{per } n < K+1 \\ 1 & \text{per } n = K+1 \end{cases}$$

quindi il sistema tende ad essere sempre pieno.

Slide 130

Time congestion e Call congestion

Definizione: Si definisce **probabilitá di congestione temporale (time congestion)** Π_T, la probabilitá che il sistema sia pieno e quindi non possa piú accettare clienti.

Rappresenta la percentuale di tempo durante il quale il sistema é completamente pieno.

Definizione: Si definisce **probabilitá di blocco di chiamata (call congestion)** Π, la probabilitá che un cliente arrivi e venga bloccato.

É la percentuale di utenti in arrivo al sistema che vengono rifiutati perché il sistema é pieno.

In generale $\Pi_t \neq \Pi$.

Slide 131

Legame fra Call Congestion e Time Congestion

Si possono scrivere le probabilitá di blocco di chiamata e di blocco temporale come segue:

$$\Pi = \mathrm{P}(\text{sistema bloccato}|\text{un arrivo al sistema})$$

$$\Pi_T = \mathrm{P}(\text{sistema bloccato})$$

per cui esiste fra le due il seguente legame:

$$\Pi = \Pi_T \frac{\mathrm{P}(\text{un arrivo al sistema}|\text{sistema bloccato})}{\mathrm{P}(\text{un arrivo al sistema})}$$

Se il processo degli arrivi é indipendente dallo stato del sistema allora:

$$\Pi = \Pi_T \frac{\mathrm{P}(\text{un arrivo al sistema})}{\mathrm{P}(\text{un arrivo al sistema})} = \Pi_T$$

per cui i due tipi di congestione coincidono.

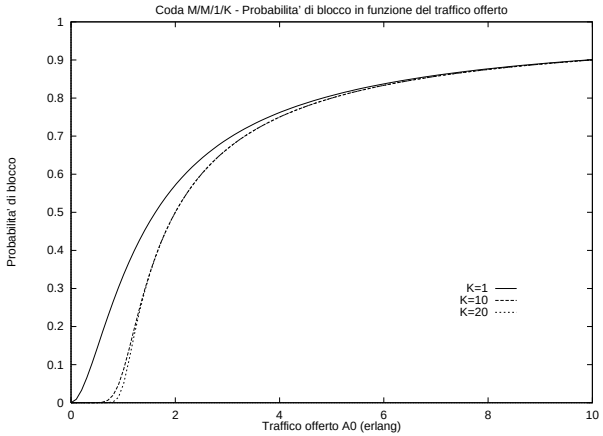
L'indipendenza degli arrrivi dallo stato del sistema é sempre vera se gli arrivi sono poissoniani.

Slide 132

Probabilitá di blocco

Per quanto detto risulta che:

$$\Pi = \Pi_T = P_{K+1} = \begin{cases} \frac{(1-A_0)A_0^{K+1}}{(1-A_0^{K+2})} & A_0 \neq 1 \\ \frac{1}{K+2} & A_0 = 1 \end{cases}$$



Slide 133

Numero medio di utenti nel sistema

$$\bar{N} = \frac{A_0}{1-A_0} - \frac{(K+2)A_0^{K+2}}{1-A_0^{K+2}}$$

Numero medio di utenti in coda

$$N_c = \bar{N} - \frac{A_0(1-A_0^{K+1})}{1-A_0^{K+2}}$$

Per valutare i tempi medi di ritardo e di attesa si applica il risultato di Little. In questo caso peró la frequenza di ingresso nel sistema non é λ, in quanto parte degli arrivi vengono rifiutati.

La frequenza di ingresso reale é:

$$\lambda_s = \lambda(1-\Pi)$$

Da questo segue anche che il **traffico smaltito** risulta:

$$A_s = \frac{\lambda_s}{\mu} = A_0(1-\Pi)$$

Slide 134