Reti di Telecomunicazioni 1

Corso “on-line” - AA2005/06
“Sistemi a coda” Blocco E2 *v2*

Ing. Stefano Salsano
e-mail: stefano.salsano@uniroma2.it

- **Definizione di traffico e utilizzazione di un sistema**

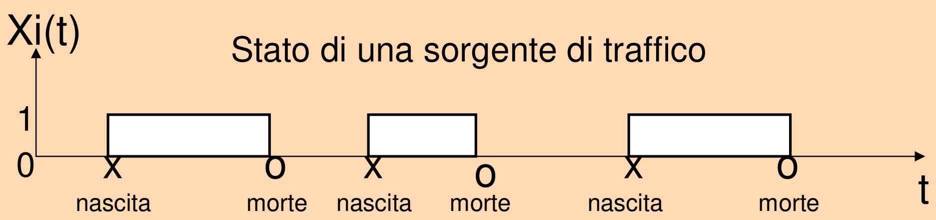
In questa sezione discuteremo la definizione del concetto di traffico.

Nella slide successiva si mostra (in alto) un diagramma con l'evoluzione temporale dello stato di una sorgente. La sorgente alterna periodi di attività ("chiamate") a periodi di inattività. L'attività è rappresentata con 1 e l'inattività con 0 e lo stato della sorgente corrisponde ad un "processo aleatorio" che vale 1 o 0. Un periodo di attività inizia con un evento di "nascita" e termina con un evento di "morte".

Se consideriamo più sorgenti (nella slide successiva, in basso, si dà il caso con 4), possiamo sommare i processi aleatori che rappresentano lo stato di una sorgente ed ottenere un diagramma che rappresenta il "traffico offerto" complessivo delle quattro sorgenti. In un dato istante il numero di sorgenti attive varia da zero al numero di sorgenti.

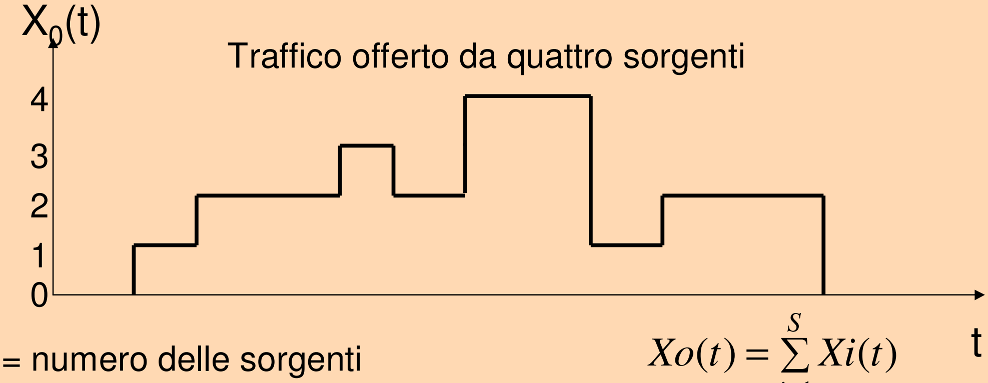
Processo di offerta del traffico su base chiamata

Stato di una sorgente di traffico



$\{Xi(t), t \geq 0\}$ Processo aleatorio che caratterizza l'i-esima sorgente

Traffico offerto da quattro sorgenti



S = numero delle sorgenti

$$Xo(t) = \sum_{i=1}^{S} Xi(t)$$

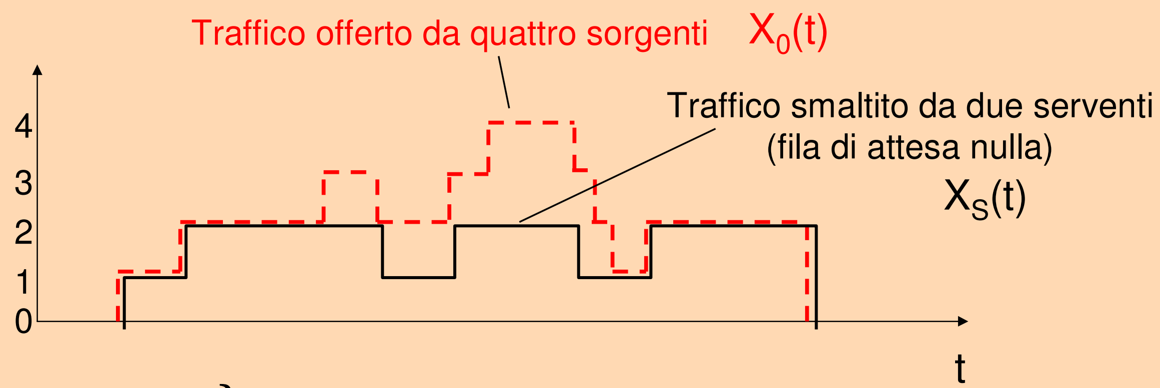
La slide precedente mostra una rappresentazione del “traffico offerto” originato da 4 terminali. Il numero di sorgenti “potenzialmente attive” evolve nel tempo nell’intervallo tra 0 e 4.

Nella realtà può accadere che una sorgente richieda servizio ma non ci siano risorse per esaudire la richiesta di servizio. La richiesta di servizio viene scartata. In queste situazioni si parlerà di traffico “smaltito” per indicare la parte di traffico che il sistema riesce a smaltire e di traffico “rifiutato” per indicare la parte di traffico che il sistema non riesce a smaltire e quindi scarta.

Ad esempio nella slide successiva si mostra una situazione in cui vi sono solo 2 serventi disponibili e si evidenzia in rosso una parte del traffico che non riesce ad essere smaltita dal sistema e deve quindi essere scartata.

Traffico offerto e traffico smaltito

- Se il sistema di servizio è composto da un numero limitato di risorse il numero di risorse contemporaneamente occupate è diverso dal numero di sorgenti che sono (o che vorrebbero essere) attive.
- Si parla quindi di traffico offerto e di traffico smaltito.



$$\{Xs(t), t \geq 0\}$$

Processo aleatorio relativo alle risorse impegnate = traffico smaltito

In questo caso ci si rende conto di come parlare di traffico offerto sia una astrazione concettuale: nel momento in cui una richiesta di servizio viene rifiutata, in molti casi non si può misurare la durata del tempo di servizio relativa a quella richiesta. Ad esempio se le richieste sono chiamate telefoniche non si può sapere quanto sarebbe durata una chiamata che viene rifiutata dalla rete.

Nonostante questo, possiamo utilmente far riferimento al concetto di traffico offerto utilizzando tecniche simulative e/o analitiche.

Passiamo quindi nella prossima slide a definire ed analizzare alcuni concetti utili, considerando indifferentemente un processo di arrivo o di servizio.

Definiamo anzitutto la proprietà di "stazionarietà" per i processi di "arrivo" e di "servizio". Un processo è "stazionario" se la distribuzione di probabilità delle sue componenti non varia nel tempo. Le componenti del processo sono le variabili aleatorie X(t), per ogni t.

Il valore medio A=E{X(t)}di una qualunque delle componenti del processo aleatorio (dato che sono tutte "uguali" per l'ipotesi di stazionarietà…) viene detto intensità media di traffico.

Traffico offerto e traffico smaltito

$\{Xo(t), t \geq 0\}$ Processo aleatorio relativo alle sorgenti attive = traffico offerto

$\{Xs(t), t \geq 0\}$ Processo aleatorio relativo alle risorse impegnate = traffico smaltito

- **Supponiamo tali processi stazionari: la distribuzione del numero X di occupazioni non dipende dal tempo**

$$p_X(k) = P\{X(t) = k\}$$ **Per *k*= 0, 1, 2, per ogni t**

- **Il valor medio della distribuzione è indicato come intensità media di traffico**

$$A = E\{X(t)\} = \sum_k k p_X(k)$$

Unità di misura del traffico: l'Erlang

$$A = E\{X(t)\} = \sum_k k p_X(k)$$

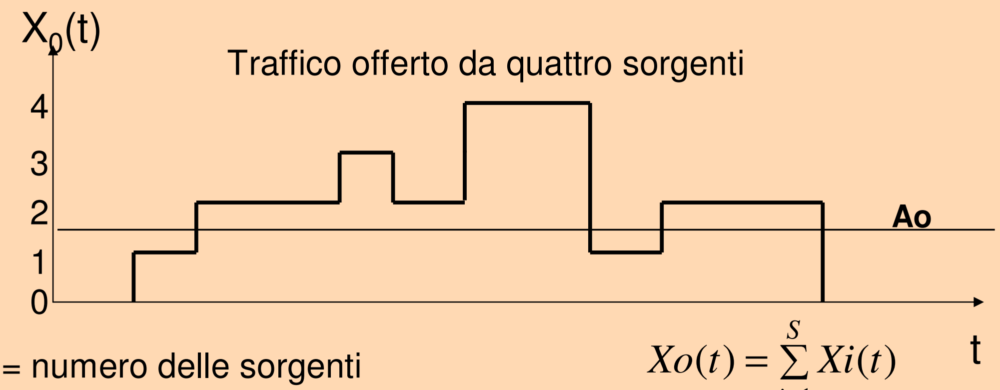
- **L'intensità media di traffico A si esprime in Erlang. Dimensionalmente è un numero puro.**
- **Si applica sia al traffico offerto che al traffico smaltito**

Traffico offerto: esprime il numero medio di sorgenti "potenzialmente" attive

Traffico smaltito: esprime numero medio di serventi occupati

Ricordiamo cosa significa il "potenzialmente" nella definizione del traffico offerto: significa che si considera un sistema "virtuale" che accolga tutte le richieste di traffico degli utenti senza perdite.

Intensità media del traffico offerto

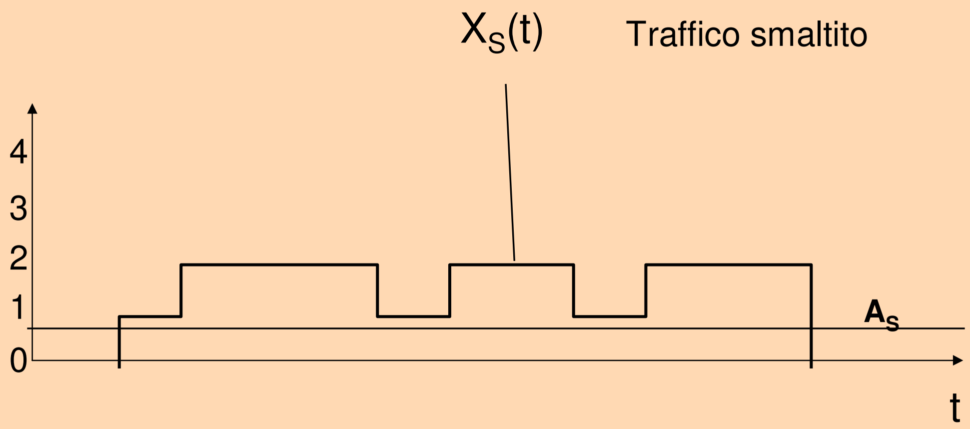


S = numero delle sorgenti

$$Xo(t) = \sum_{i=1}^{S} Xi(t)$$

Ao: intensità media del traffico offerto $$A_O = E\{X_O(t)\} = \sum_k k p_{X_O}(k)$$

Intensità media del traffico smaltito



$\{Xs(t), t \geq 0\}$ Processo aleatorio relativo alle risorse impegnate = traffico smaltito

A_S : intensità media del traffico smaltito $A_S = E\{X_S(t)\} = \sum_k k p_{X_S}(k)$

Esempio 1: Traffico offerto

- **Si osservi un sistema con 3 sorgenti che possano essere attive o non attive indipendentemente.**
- **Fissato un qualunque istante di tempo t si abbia la seguente distribuzione del numero di sorgenti attive:**

sorgenti	probabilità
0	P(0) = 0.4
1	P(1) = 0.3
2	P(2) = 0.2
3	P(3) = 0.1

Il numero medio di sorgenti attive è:

$$0 \cdot P(0) + 1 \cdot P(1) + 2 \cdot P(2) + 3 \cdot P(3) = 1$$

Il traffico offerto è pari a 1 Erlang… corrisponde ad una sorgente attiva per il 100% del tempo

Esempio 1: Traffico offerto (continua)

- **Se le sorgenti sono uguali tra loro, il traffico offerto da ciascuna sorgente sarà 1 / 3 = 0.333 Erlang:**
  ho una valutazione della attività della singola sorgente…

- **Quanto vale il traffico offerto da 10 sorgenti di questo tipo ?**

  A_o(10 sorgenti) = 10 * A_o(1 sorgente) = 3.33 Erlang

- **Corrisponde a 3.33 sorgenti <u>sempre</u> attive**

Commenti sulla definizione di Erlang

- **Quando diciamo ad esempio che 10 sorgenti con attività 0.333 "corrispondono" a 3.33 sorgenti sempre attive, ci riferiamo <u>esclusivamente</u> alla quantità media di traffico offerto.**

- **Le prestazioni (ritardo, perdita) esperite dalle sorgenti quando accedono ad un sistema di servizio NON sono equivalenti nei due casi…**

- **La <u>variabilità</u> del processo di arrivo è diversa nei due casi e intuitivamente ci aspettiamo che un processo a variabilità maggiore (10 sorgenti con attività 0.333) sia più difficile da servire e quindi riceva prestazioni peggiori di un processo a variabilità minore (3.33 sorgenti sempre attive)**

Esempio 2: Traffico smaltito

- **Si osservi un sistema con 2 serventi. Ad esempio un centralino di un'azienda con 2 operatori.**
- **Fissato un qualunque istante di tempo t si abbia la seguente distribuzione del numero di operatori attivo:**

operatori attivi	probabilità
0	P(0) = 0.7
1	P(1) = 0.2
2	P(2) = 0.1

Il numero medio di operatori attivi è:

$$0 \cdot P(0) + 1 \cdot P(1) + 2 \cdot P(2) = 0.4$$

Il traffico smaltito è pari a 0.4 Erlang… corrisponde ad un operatore attivo per il 40% del tempo

Riassumendo … Il traffico come processo aleatorio

- Il traffico è un processo aleatorio X(t), t≥0 che descrive il numero di occupazioni contemporanee in corso all'instate t
- Per un generico t fissato, X(t) è una variabile aleatoria caratterizzata da una distribuzione di probabilità il cui valor medio rappresenta l'intensità di traffico

- Si parla del <u>traffico offerto</u> se X(t) fa riferimento al numero di sorgenti contemporaneamente attive all'istante t
- Si parla di <u>traffico smaltito</u> se X(t) fa riferimento al numero di serventi contemporaneamente occupati all'istante t

- L'unità di misura del traffico è l'*Erlang* che corrisponde al traffico offerto da una sorgente sempre attiva o smaltito da un servente permanentemente occupato

Dopo aver definito i concetti di traffico offerto e smaltito e aver visto qual è il modo di misurarne l'intensità, consideriamo i parametri di un sistema di servizio come riportati nella slide seguente.

T(k) è il "tempo di coda" che abbiamo definito in precedenza per il k-esimo utente che entra nel sistema. T(k) include quindi il tempo di servizio e l'eventuale tempo speso nella fila di attesa.

λ è la frequenza di ingresso al sistema e si esprime in s^{-1}. Nella slide si tiene in conto delle eventuali perdite "a monte" quindi si definisce λ_o come la frequenza di arrivo al sistema (il pedice "o" sta per offerto) e λ_p come la frequenza di perdita. Ovviamente $\lambda_o = \lambda + \lambda_p$.

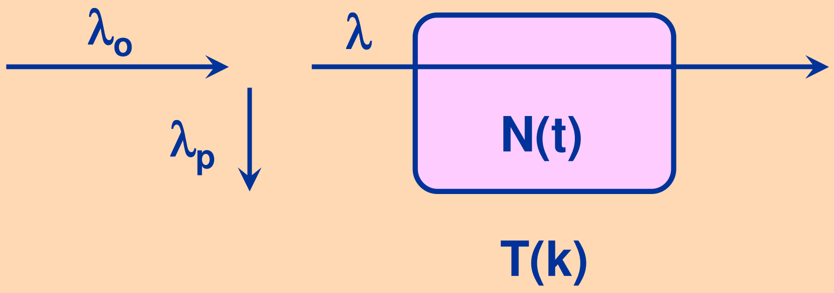
Dato che le perdite sono modellate "a monte", in questa convenzione la frequenza di ingresso equivale alla frequenza di attraversamento.

Si noti che talvolta si può utilizzare λ per indicare la frequenza di arrivo al sistema, se ci sono delle perdite si potrà denotare la frequenza di ingresso al sistema con λ_i e si avrà che $\lambda = \lambda_i + \lambda_p$

Se il sistema è senza perdite ovviamente la frequenza di arrivo e quella di ingresso coincidono.

Si definiscono infine i valori medi del numero di utenti presenti nel sistema E(N) e del tempo di attraversamento del sistema E(T), che costituiscono i due parametri principali per valutare un sistema a coda.

Parametri di un sistema di servizio

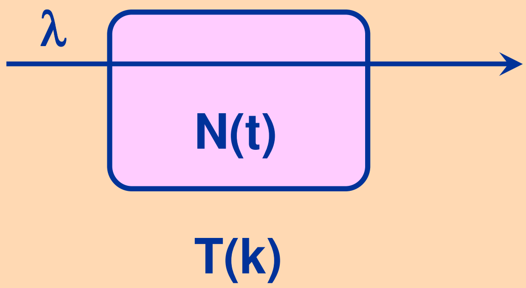


- N(t) è il numero di utenti presenti nel sistema all'istante t
- T(k) è il tempo che l'utente k-esimo ha impiegato all'interno del sistema
- λ è la frequenza media di arrivo al sistema (e di uscita dallo stesso, considerando che non si hanno perdite)
- E(N) è il numero medio di utenti nel sistema
- E(T) è il tempo medio di attraversamento del sistema

Nella slide successiva andiamo a considerare una relazione fondamentale tra la frequenza di ingresso ad un sistema senza perdita, il numero medio di utenti presenti nel sistema e il tempo di attraversamento del sistema stesso.

Questa relazione è nota come "legge di Little" o "risultato di Little" ed ha una validità molto generale.

Legge di Little



- In condizioni di equilibrio statistico, in un qualunque sistema senza perdite

$$E(N) = \lambda E(T)$$

Il numero medio di utenti nel sistema è uguale al prodotto della frequenza di interarrivo per il tempo di attraversamento

Legge di Little

- La legge di Little si applica anche ai casi “deterministici”

- Una linea di autobus si effettua una partenza ogni 10 minuti
  Il percorso dura 1 ora. Quanti autobus sono contemporaneamente in viaggio ?

$$N = \lambda T = 1/10\ (\text{min}^{-1}) \cdot 60\ (\text{min}) = 6 \text{ autobus}$$

- Una seggiovia ha 160 seggiolini (80 in salita e 80 in discesa). Passano 15 secondi tra un seggiolino e il successivo. Quanto dura il percorso ?

$$T = N / \lambda = 80 / 4\ (\text{min}^{-1}) = 20 \text{ minuti}$$

Carico / utilizzazione di un sistema

- Possiamo quindi esprimere il traffico offerto al sistema come il prodotto della frequenza media delle richieste per la loro durata media o “tempo medio di servizio”

$$A_o = \lambda_o \cdot E\{T_s\}$$

- Se confrontiamo A_o con il numero di serventi disponili m otteniamo una idea di quanto stiamo caricando il sistema

$A_o/m << 1 \Rightarrow$ sistema poco carico

$A_o/m < 1$ ma prossimo ad 1 $\Rightarrow$ sistema carico

$A_o/m \geq 1 \Rightarrow$ il sistema può operare solo con perdita

Avevamo visto che l’intensità traffico offerto al sistema, indicata con A_o, corrisponde al numero medio di utenti che sarebbero attivi in un sistema senza perdita. Grazie alla legge di Little, possiamo esprimere questo numero medio come il prodotto tra la frequenza di arrivo al sistema e il tempo medio di permanenza nel servizio $E\{T_s\}$.

In pratica possiamo ricavare l’intensità di traffico offerto note la frequenza media di arrivo e il tempo di servizio.

Nella seconda parte della slide precedente si mette in relazione l'intensità di traffico offerto con il numero dei serventi m del sistema.

Se l'intensità di traffico è molto minore del numero dei serventi, il sistema sarà "poco carico", cioè gli utenti generalmente troveranno dei serventi disponibili.

Quando l'intensità di traffico è minore ma prossima al numero dei serventi il sistema è "carico". Se il sistema opera con attesa, vuol dire che per un utente sarà probabile trovare i serventi occupati e dover aspettare nella fila di attesa. Più il sistema è carico e maggiore sarà il tempo medio speso nella fila di attesa. Se il sistema opera con perdita, più il sistema è carico e maggiore sarà la probabilità di perdita.

Infine, se l'intensità di traffico è maggiore del numero di serventi, il sistema dovrà necessariamente operare con perdita.

Carico / utilizzazione di un sistema

- Il traffico smaltito si esprime come prodotto della frequenza media di richieste accolte per il tempo medio di servizio

$$A_s = \lambda_s \cdot E\{T_s\}$$

- Se dividiamo il traffico smaltito per il numero di serventi m otteniamo la <u>utilizzazione</u> dei serventi stessi.

$$\rho = A_s / m \qquad \rho < 1$$

- L'utilizzazione è sempre minore di 1

Applichiamo infine la legge di Little al traffico smaltito. L'intensità media di traffico smaltito dal sistema, indicata con A_s, corrisponde al numero medio di utenti che stanno ricevendo servizio. Grazie alla legge di Little, possiamo esprimere questo numero medio come il prodotto tra la frequenza di ingresso al sistema e il tempo medio di servizio $E\{T_s\}$.

In pratica possiamo <u>ricavare l'intensità di traffico smaltito note la frequenza media di ingresso e il tempo medio di servizio</u>.

Se dividiamo il traffico smaltito per il numero m dei serventi, otteniamo la "utilizzazione" ρ dei serventi. L'utilizzazione corrisponde alla percentuale media di tempo in cui il servente è attivo.