



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Protocolli: Prestazioni

Franco CALLEGATI

Dipartimento di Informatica: Scienza e Ingegneria



Affidabilità

- Controllo dell'errore
 - Rivelazione e correzioni
 - CRC e Internet checksum
- Eventuale recupero dell'errore
 - ARQ e ritrasmissione
- Controllo di flusso e sequenza
 - Ack e ARQ
- Questi temi sono stati trattati lo scorso anno

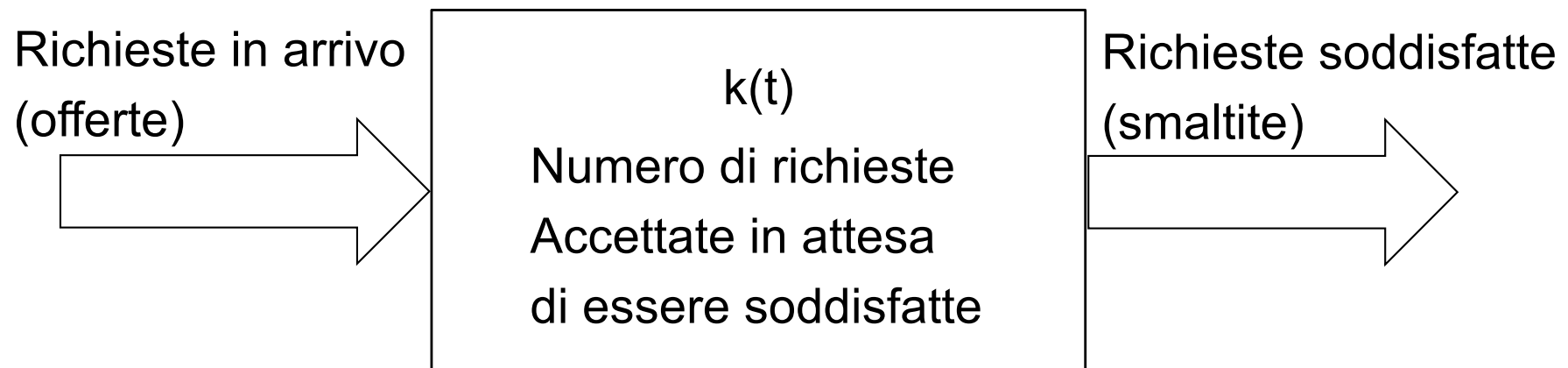


Funzionalità e prestazioni

- I protocolli sono progettati per garantire
 - Funzionalità
 - La trasmissione dati deve poter avvenire risolvendo i problemi che si riscontrano nell'accesso e nell'uso del canale
 - Prestazioni
 - La trasmissione deve avvenire con successo utilizzando per quanto possibile la capacità messa a disposizione dallo strato fisico

Le prestazioni in generale

- Un *sistema* deve *smaltire* del *lavoro* che gli viene *offerto* dall'esterno
- Esempio nel caso specifico delle reti di tlc
 - Livello N+1 invia PDU al livello N tramite la relativa interfaccia (e un opportuno SAP)
 - Livello N impiega un certo tempo per soddisfare la richiesta

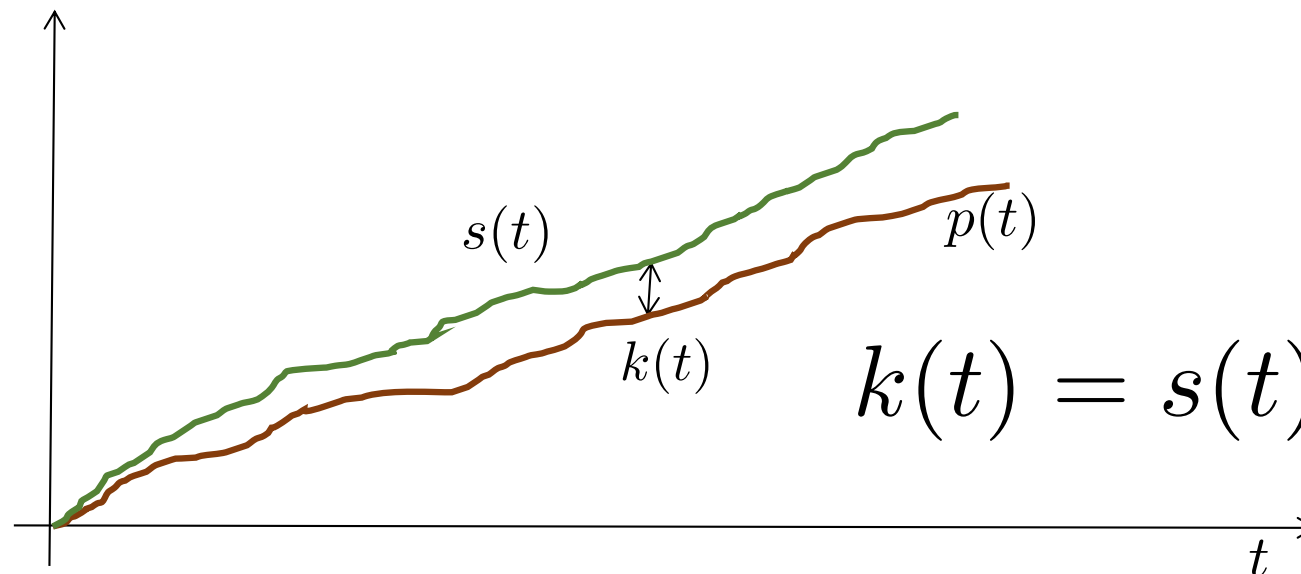


Arrivi e partenze

$a(t)$ • Numero di richieste di servizio giunte al tempo t

$s(t)$ • Numero di richieste accettate al tempo t

$p(t)$ • Numero di partenze dal sistema al tempo t



Richieste offerte e smaltite

- Frequenza media delle richieste offerte

$$\lambda = \lim_{t \rightarrow \infty} \frac{a(t)}{t}$$

- Frequenza media delle richieste smaltite

$$\lambda_s = \lim_{t \rightarrow \infty} \frac{p(t)}{t}$$

- Se il sistema in oggetto non produce lavoro ma lo riceve solamente dall'esterno

$$\lambda_s \leq \lambda$$



Richieste perdute

$$\lambda_s = \lambda \quad \text{Implica} \quad s(t) = a(t)$$

- Tutte le richieste vengono accettate dal sistema e prima o poi soddisfatte

$$\lambda_s < \lambda \quad \text{implica} \quad r(t) = a(t) - s(t)$$

- Dove $r(t)$ rappresenta le richieste che non vengono accettate e sono *rifiutate o perdute* dal sistema

Analogamente

- Posso definire

Frequenza media
delle richieste perse ← $\lambda_p = \lim_{t \rightarrow \infty} \frac{r(t)}{t}$

- Da cui consegue

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

Utente e servizio

- In una rete a pacchetto considerare il semplice «bit rate» del canale non è del tutto corretto

- I bit sono raggruppati in pacchetti
- L'unità di servizio è il pacchetto non il bit

→ L'unità di riferimento per il servizio è il pacchetto, perché solo quando l'intero pacchetto è trasmesso e ricevuto può essere utile.

- In altre parole trasmettere una porzione di bit appartenenti ad un pacchetto non ha senso di per sé

- **Il riferimento è il tempo di servizio dell'intero pacchetto che solo se completato produce un risultato «utile» per l'utente**

→ Il tempo necessario per fornire il servizio al pacchetto

v

Tempo richiesto dal servizio di un generico cliente (pacchetto dati, PDU)

- Servizio aleatorio

- Si fa riferimento in prima battuta al tempo medio

- Servizio deterministico

- Tempo di servizio costante ed uguale al suo valore medio



La lunghezza del pacchetto

- L lunghezza del pacchetto in bit
- C capacità del canale in bit per secondo (bitrate)
- Ovviamente

$$\bar{v} = \frac{L}{C}$$

↳ Tempo di trasmissione medio, ovvero il tempo medio necessario per trasmettere l'intero pacchetto sul canale.

Frequenza di servizio

- L'inverso del tempo medio di servizio viene detto frequenza media di servizio

$$\mu = \frac{1}{\bar{\vartheta}}$$

- La frequenza media di servizio è ovviamente legata alla presenza di utenti del sistema
 - Se non vi sono richieste di servizio ovviamente la frequenza di servizio è nulla
 - Se vi sono richieste di servizio il parametro da indicazione di quanto velocemente esse vengono soddisfatte



Il ruolo di μ

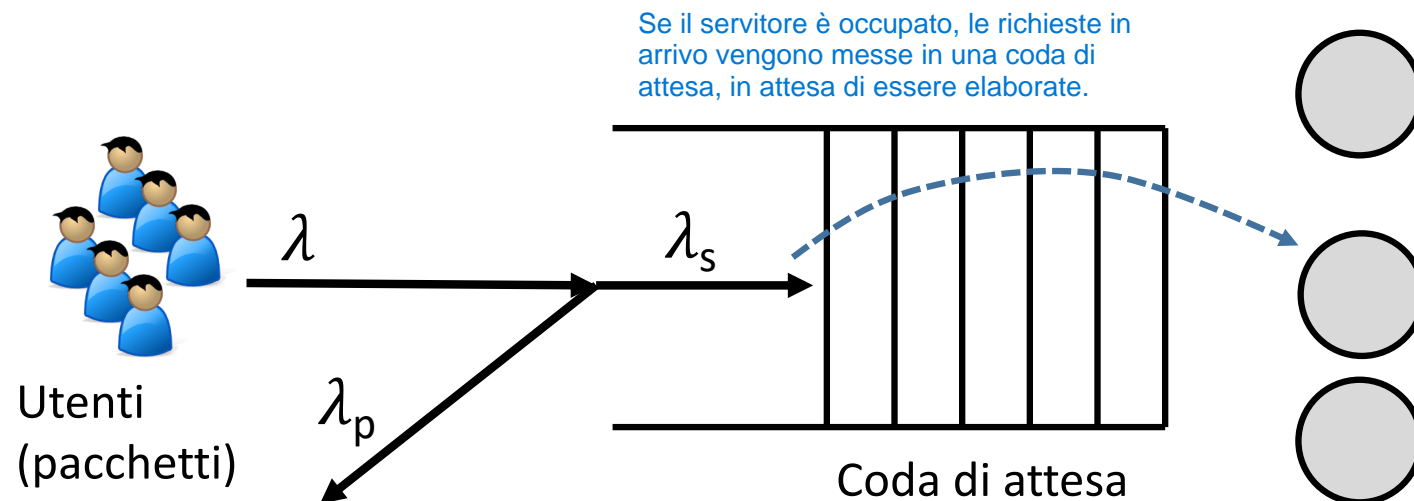
- Di fatto m può essere interpretato come una sorta di capacità massima del servitore
- Se $\bar{\vartheta} = 0.5$ s ne consegue che il servitore al più smaltirà $\mu = 2$ pacchetti/s (numero massimo di pacchetti che il servitore può elaborare in un'unità di tempo)
- Possiamo dire che per il singolo servitore

$$\lambda_s^{\max} = \mu$$

Frequenza massima media delle richieste smaltite in un'unità di tempo. È il tasso massimo di arrivo delle richieste che un servitore può accettare senza che si verifichi congestione. Questo indica che il servitore può elaborare richieste a una frequenza massima pari alla sua capacità.

Di quale sistema parliamo

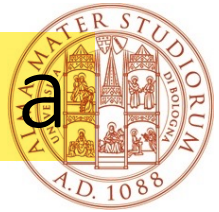
- Per le reti hanno particolare importanza i sistemi a coda



Uno o più servitori
(sistemi di trasmissione)

Nella reti a pacchetto 1 solo servitore
è il caso di riferimento

Cosa accade all'utente in un sistema a coda?



- L'utente permane nel sistema per un tempo che tiene conto dell'attesa in coda e del tempo di esecuzione del servizio
- $\bar{\delta}$ tempo medio totale speso dal singolo utente nel sistema a coda che è composto da
 - $\bar{\vartheta}$ tempo ^{medio} effettivo di servizio (l'utente fa le cose per le quali è entrato nel sistema)
 - T_A tempo speso in coda, ossia tempo di attesa prima di essere effettivamente servito

$$\bar{\delta} = \bar{\vartheta} + \bar{T}_A$$

Tempo medio di Attesa

Il traffico

- Le prestazioni del sistema che fornisce il servizio dipendono:
 - Dalla numerosità degli arrivi, tipicamente frequenza media di arrivo (utenti per secondo)
 - Dalla durata del servizio, tipicamente la frequenza media di servizio (utenti al secondo) o il tempo medio di servizio (secondi)
- Si definisce **traffico** il numero medio di utenti presenti nel sistema in un dato momento. ^(o richieste)
- Si dimostra che il prodotto fra frequenza di arrivo e tempo medio di permanenza nel sistema (teorema di Little) dà il traffico

$$A = \lambda \bar{\delta}$$

Rappresenta il tempo medio che una richiesta trascorre nel sistema. Include sia il tempo medio di attesa in coda sia il tempo medio di elaborazione.

Alcune definizioni conseguenti

- Per analogia si definiscono

$$A_0 = \lambda \bar{\vartheta} \rightarrow \text{Tempo di trasmissione medio}$$

- Traffico offerto (occupazione media di un sistema ideale che serve subito tutti gli utenti senza attesa)

$$A_s = \lambda_s \bar{\vartheta}$$

- Traffico smaltito (occupazione media dei servitori del sistema)
- Ci ricorda che se λ_s utenti mediamente entrano per unità di tempo entrano allora λ_s devono uscire

$$A_p = \lambda_p \bar{\vartheta}$$

- Traffico perduto (occupazione media di un sistema che serve gli utenti che invece sono stati rifiutati)

Il traffico smaltito

- A_s da una valutazione della capacità di servizio del sistema considerato
- Se i servitori sono impegnati al 100% il sistema esprime tutta la sua capacità di servizio
- Se i servitori sono m allora

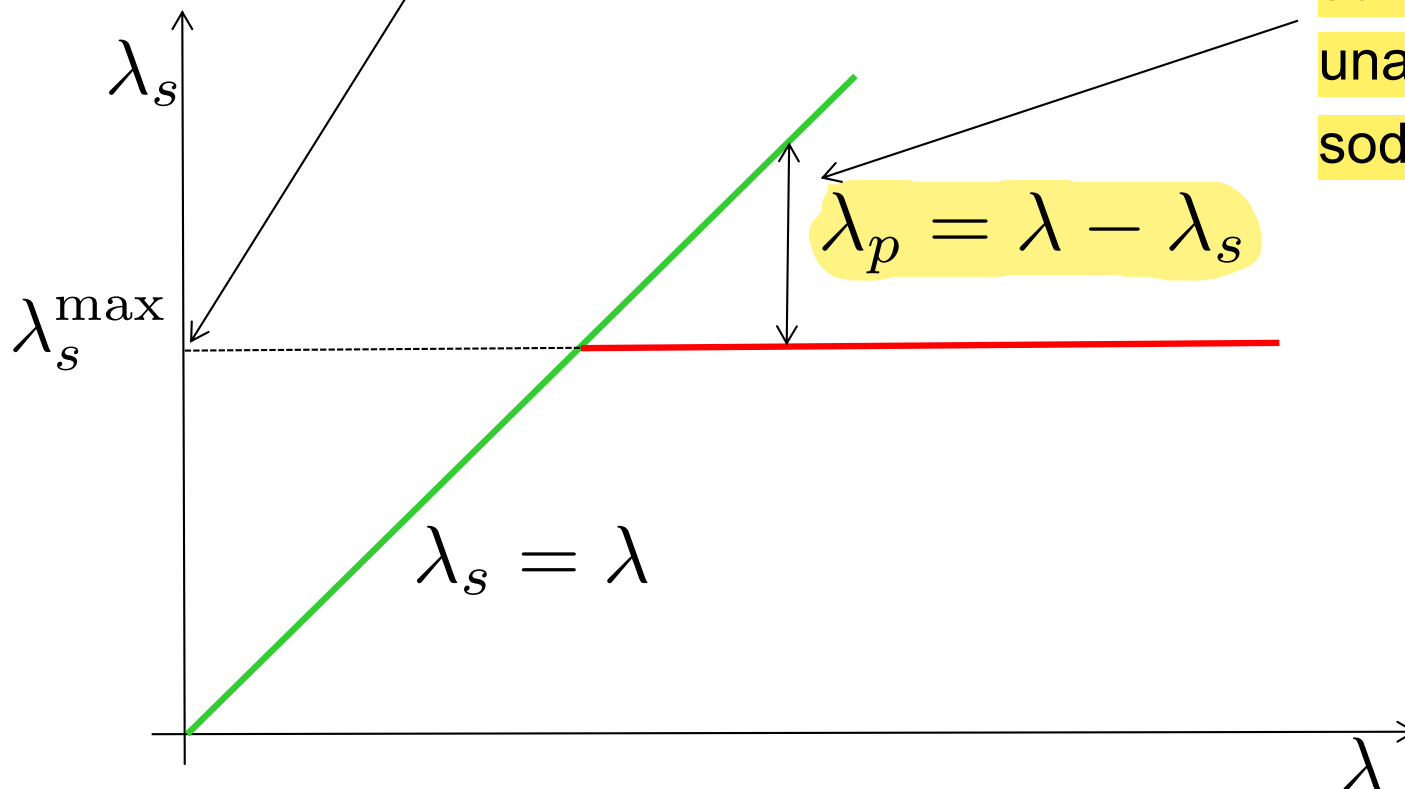
$$0 \leq A_s \leq m$$

- A_s è spesso indicato con il nome di *throughput*

Throughput: A_s è spesso indicato come throughput, che misura la quantità effettiva di richieste soddisfatte in un dato intervallo di tempo.

In un sistema ideale

Il sistema ha una capacità massima finita
di smaltire richieste
(dipende dalle condizioni in cui opera)



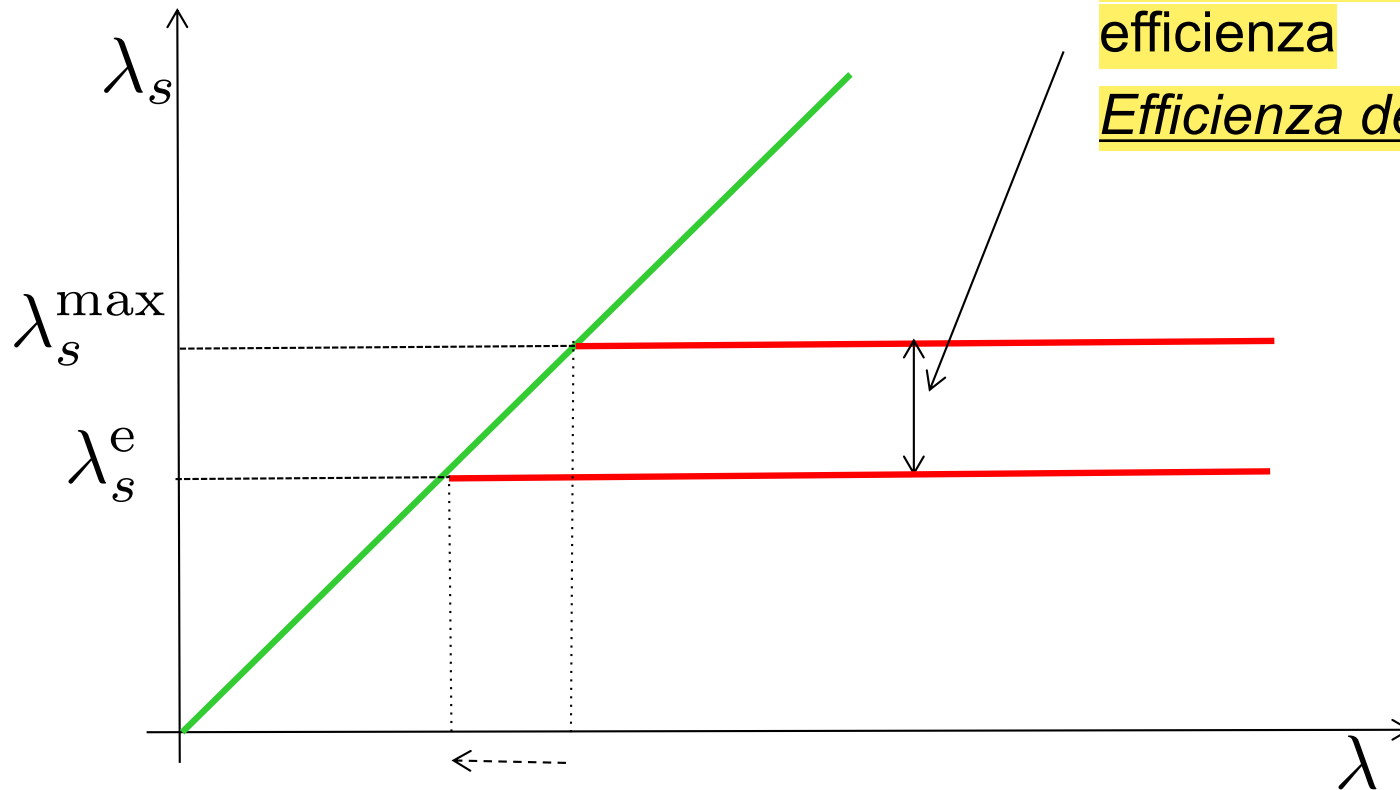
Se le richieste offerte
sono eccessive
una parte non può essere
soddisfatta

In un sistema reale

La riduzione di capacità si interpreta come perdita di efficienza

Efficienza del protocollo

$$\eta = \frac{\lambda_s^e}{\lambda_s^{\max}} \leq 1$$



La riduzione di capacità obbliga il livello superiore a limitare la proprie richieste pena la perdita dei dati

Capacità massima ed efficienza

- Quali sono le prestazioni ideali per un protocollo data link?
 - Poiché il protocollo invia i bit dello strato 3 sul canale la sua capacità massima teorica è la velocità del canale C

- Il tempo medio di servizio minimo possibile sarebbe quindi

$$\bar{\vartheta} = \frac{L}{C} = \frac{1}{\mu}$$

- Se il protocollo richiede maggiore tempo per la completa trasmissione della trama allora

$$\bar{\vartheta}_e = \frac{L}{C_e} > \frac{1}{\mu}$$

C_e

- La capacità effettiva dipende dal protocollo
- Se le funzionalità richieste o una situazione non ideale richiedono più tempo per ogni PDU allora parte della capacità risulta inutilizzabile per i dati degli utenti
 - PCI necessarie per la segnalazione
 - Errori di trasmissione
 - Ritrasmissioni
 - Tempi morti legati alle dinamiche del protocollo
 - Tempi morti in attesa di accedere al canale



Unità di misura

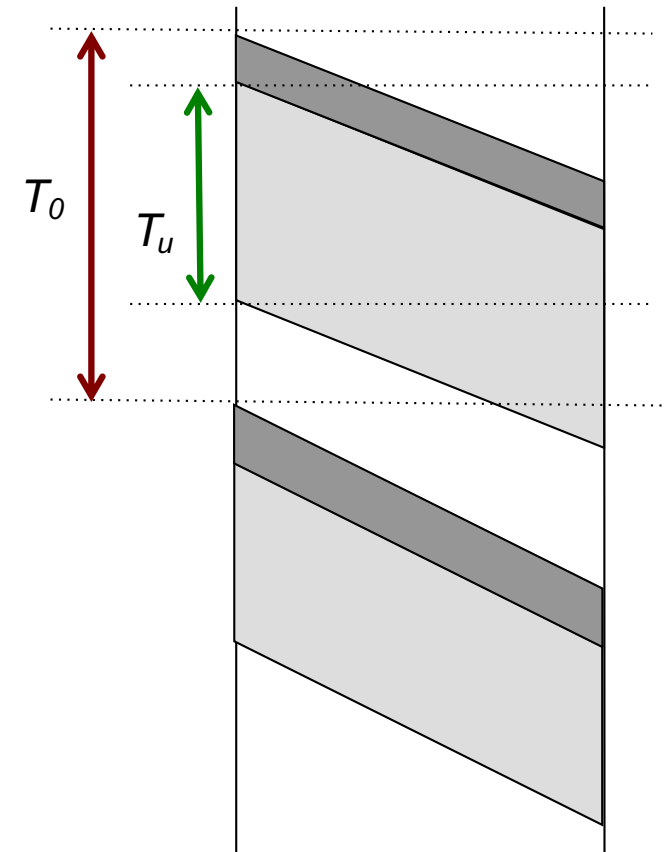
- Il traffico è una grandezza adimensionale
 - Moltiplico richieste di servizio per unità di tempo (secondo) per la durata del servizio in unità di tempo (secondi)

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

- Formalmente si misura con una unità fittizia detta Erlang

Valutazione efficienza

- Per valutare l'efficienza di solito si fa riferimento alla PDU
- Si confronta:
 - La quantità di tempo strettamente utilizzato per inviare i soli dati d'utente (SDU)
 - La quantità di tempo utilizzato complessivamente per completare correttamente l'invio della PDU
 - In funzione delle regole del protocollo
- L'efficienza è data dal rapporto fra queste due quantità
- Questo problema l'abbiamo studiato lo scorso anno studiando i protocolli ARQ



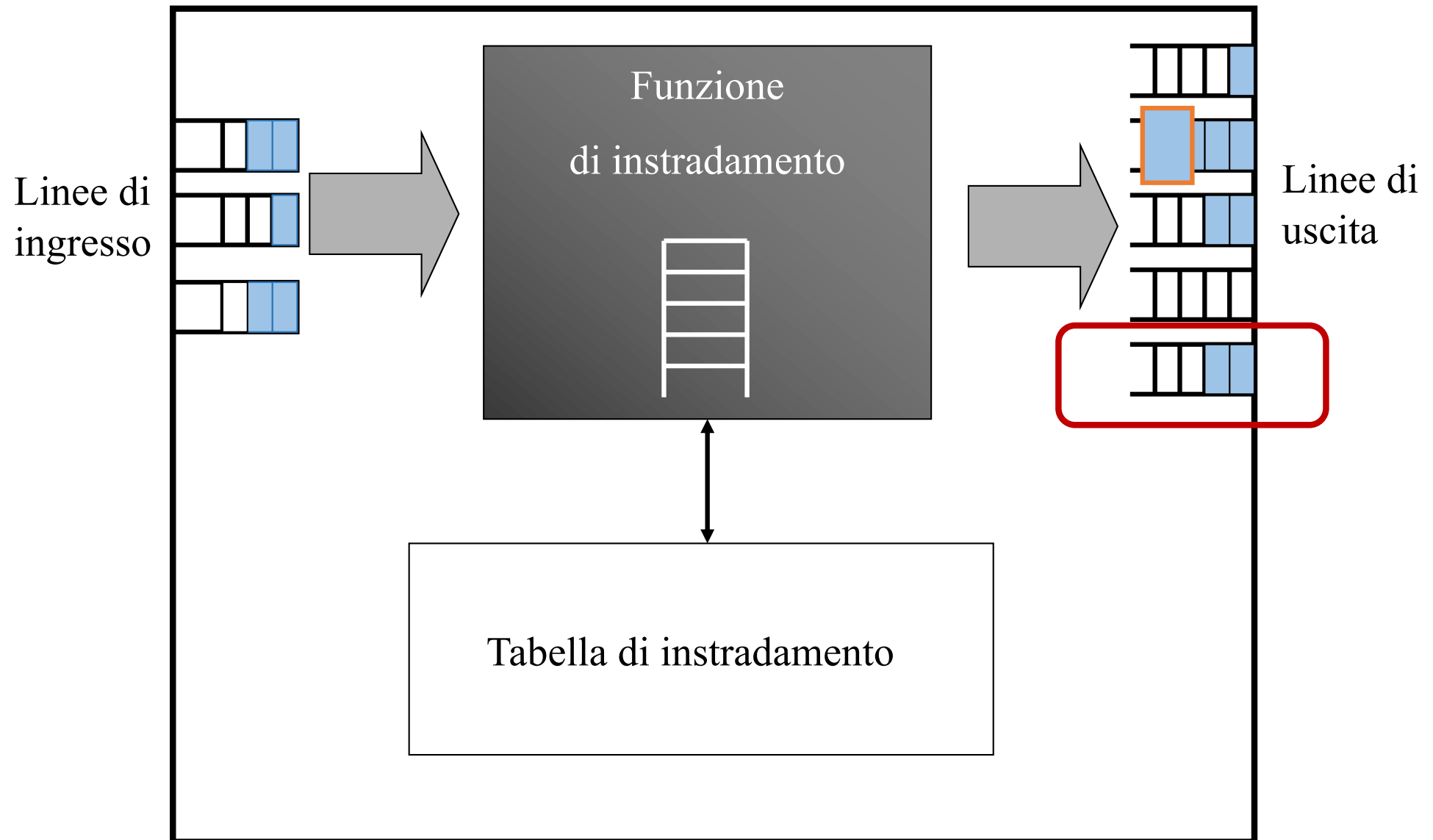
$$\eta = \frac{T_u}{T_0} = \frac{\bar{\vartheta}}{\bar{\vartheta}_e}$$



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

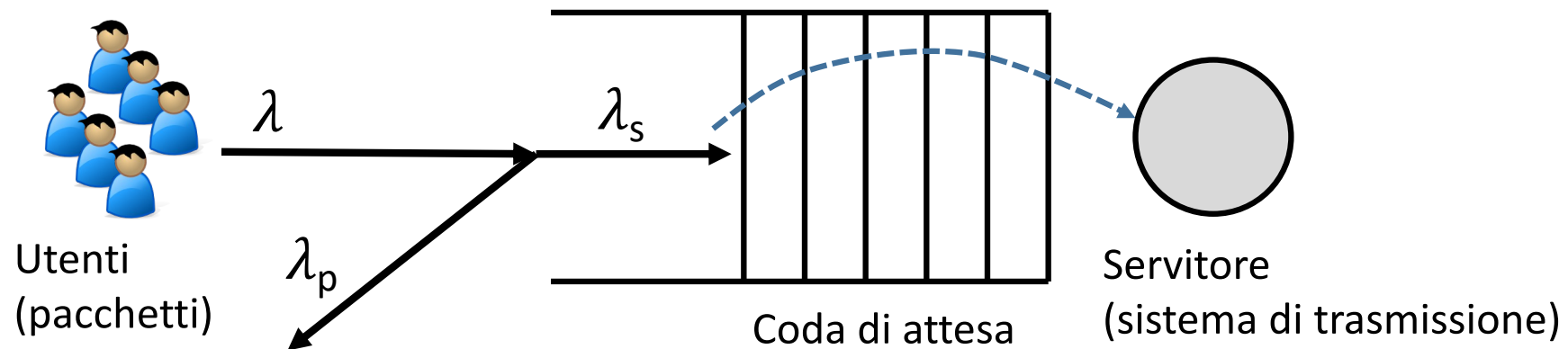
Reti commutate: il sistema a coda con singolo servitore

Il nodo di commutazione a pacchetto



Coda singolo servitore

- Ogni collegamento in uscita viene schematizzato come sistema coda singolo servitore





Alcune ipotesi semplificative

- Le perdite di pacchetti in prima approssimazione sono trascurabili (coda infinita)

$$\lambda_p = 0$$

$$\lambda_s = \lambda$$

- I pacchetti arrivano causalmente con distribuzione di Poisson

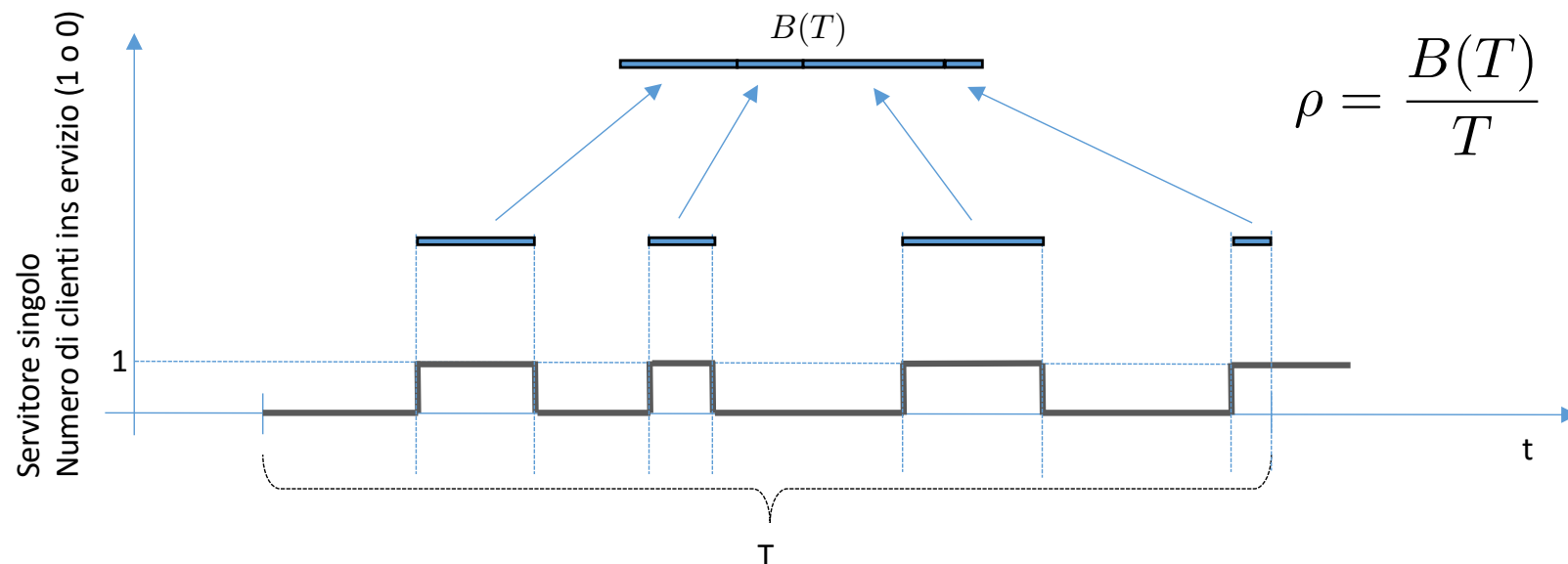
$$\Pr \{k \text{ arrivals during } t\} = P(k, t) = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$$

- La dimensione dei pacchetti è casuale con distribuzione esponenziale uguale per tutti i pacchetti

$$\Pr\{\vartheta \leq t\} = F_{\vartheta}(t) = 1 - e^{-\frac{t}{\vartheta}}$$

Utilizzazione

- Il servitore alterna fasi di lavoro a fasi di pausa
 - Idealmente vorremmo il servitore sempre attivo
 - Il servitore in pausa è uno spreco di risorse
 - Il servitore può essere attivo solo se ci sono clienti da servire
 - I clienti arrivano in modo casuale e può capitare che ci sia un periodo di tempo senza arrivi





Servitore occupato

- In un sistema a servitore singolo l'utilizzazione è la percentuale di tempo per cui il servitore è impegnato
- In un sistema ergodico questa è anche la probabilità di trovare il servitore occupato in un istante qualunque
- Quindi

$$\rho = \Pr\{\text{Servitore occupato}\} = \Pr\{\text{Pacchetto accodato}\}$$

Traffico e utilizzazione

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

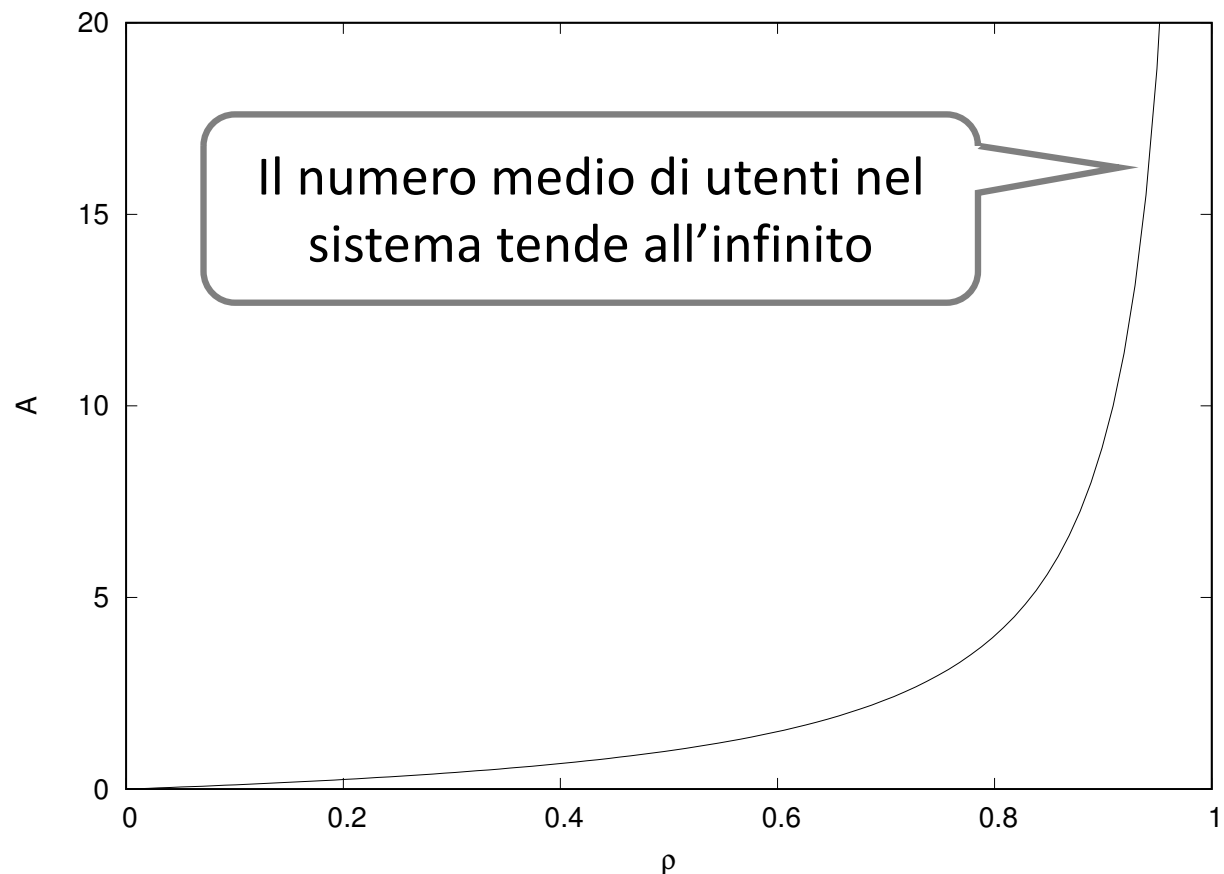
Questa grandezza è significativa perché confronta il ritmo di arrivo con quello di servizio

$$\mu = \frac{1}{\vartheta}$$

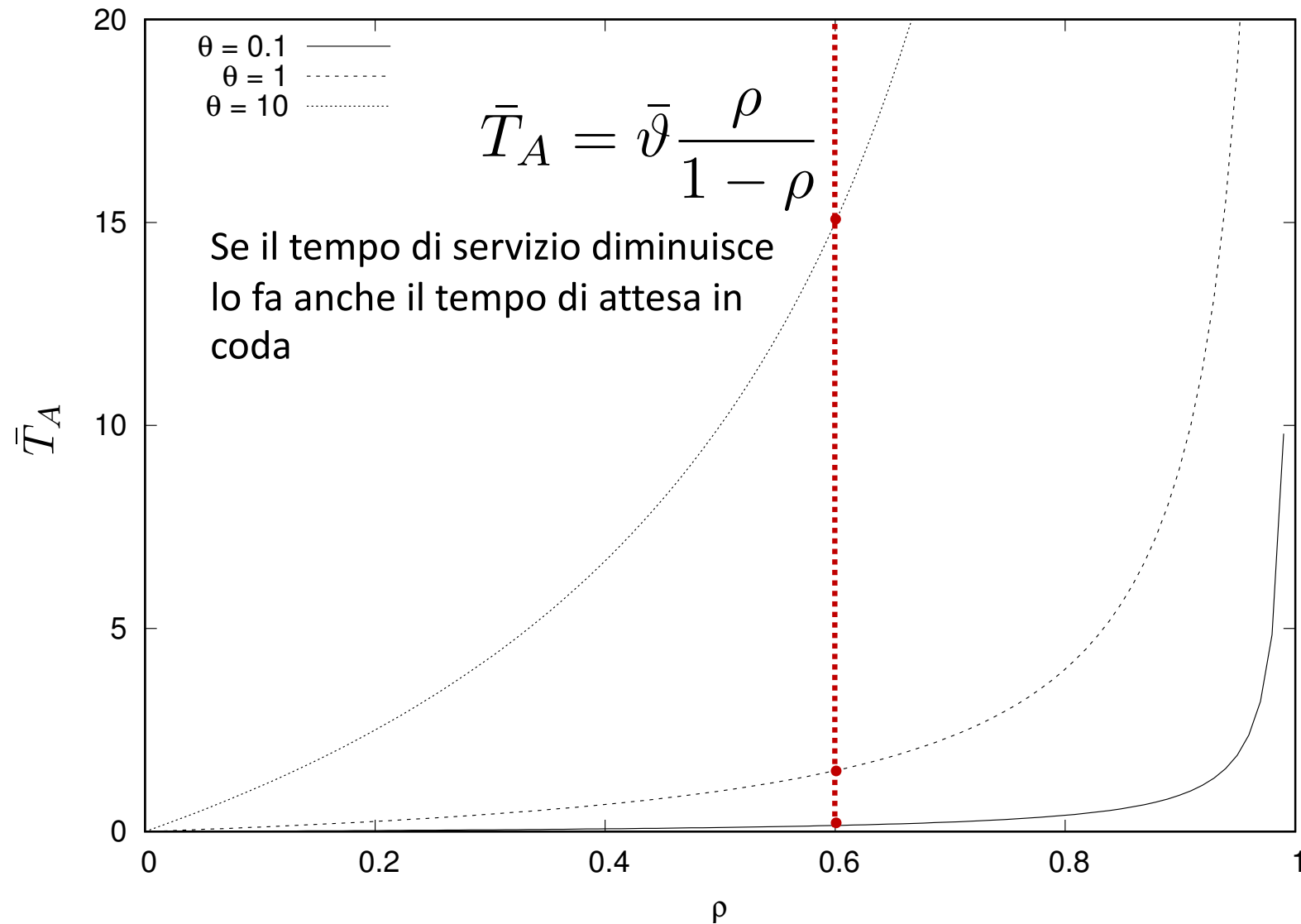
Questa grandezza è detta frequenza di servizio e ci dice mediamente quanti pacchetti possono essere serviti

$$\lambda < \mu$$

Non devo arrivare più pacchetti di quelli che possono essere serviti



Tempo medio di attesa in coda

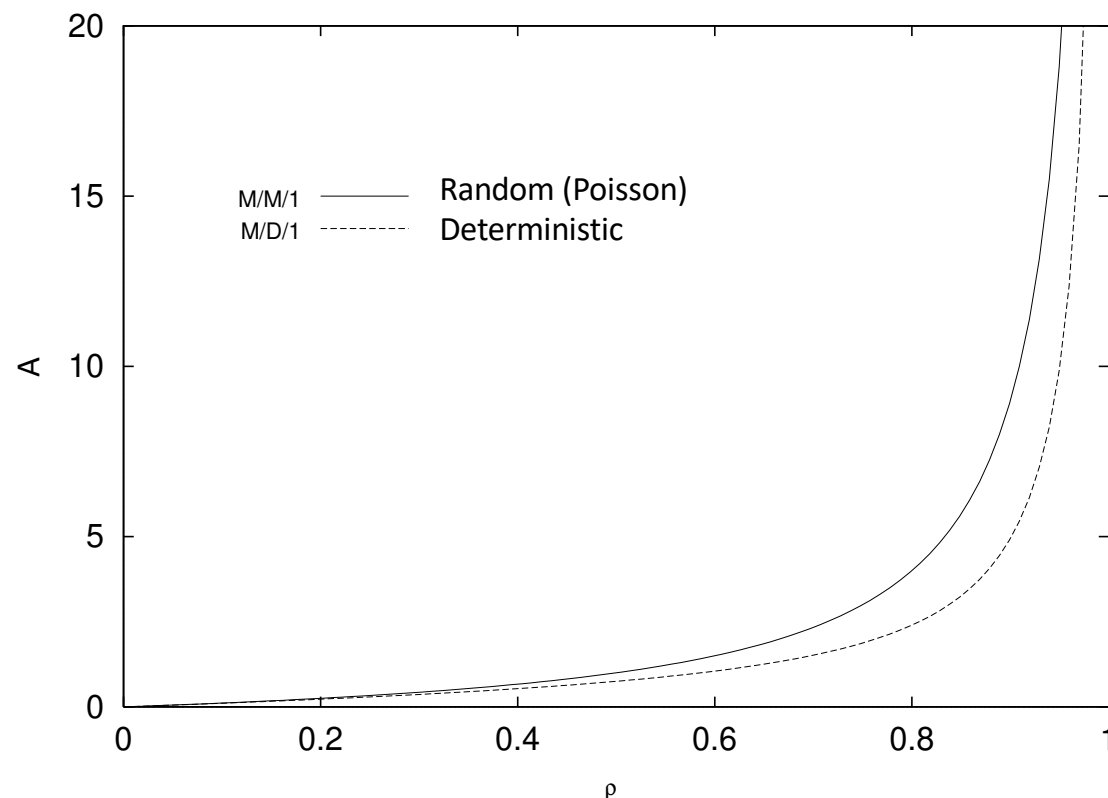


La statistica del tempo di servizio

- Il traffico nel sistema cambia se cambia la statistica del tempo di servizio
 - Servizio casuale con distribuzione esponenziale (M)
 - Servizio deterministico ossia sempre uguale (D)

$$\bar{T}_A = \bar{\vartheta} \frac{\rho}{1 - \rho}$$

$$\bar{T}_A = \bar{\vartheta} \frac{\rho}{2(1 - \rho)}$$





ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Local Area Network e Accesso multiplo



LAN: Local Area Networks

Infrastruttura di telecomunicazioni che consente ad apparati **indipendenti** (stazioni) di comunicare in un' area **limitata** attraverso **un canale fisico condiviso** ad **elevata bit rate** con **bassi tassi di errore**

- **INDIPENDENTI**: assenza di architetture master-slaves
- **LIMITATA**: un' area di dimensioni moderate è spesso privata, non soggetta a regulations
- **CANALE FISICO CONDIVISO**: potrebbe essere un unico mezzo fisico condiviso
- **ELEVATA BIT RATE**: uso esclusivo dell' intera banda anche se per brevi intervalli
- **BASSI TASSI DI ERRORE**: a causa delle piccole distanze si dispone di molta potenza



Scelte per le LAN

- Le LAN sono *reti di calcolatori* e devono essere implementate scegliendo *protocolli per tutti gli strati dell'OSI*
- Le **dimensioni limitate** rendono convenienti soluzioni particolari per gli **strati 1 e 2**
 - di questo si occupano gli standard per le LAN
- Occorre scegliere
 - Il mezzo trasmissivo
 - La topologia
 - Un eventuale protocollo di accesso

Mezzo trasmissivo

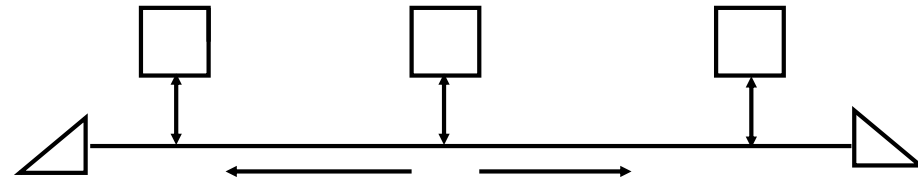
- In generale nelle reti moderne le fibre ottiche stanno progressivamente sostituendo il rame
 - Maggiore banda e distanza
 - Minore costo
 - Interconnessione più complessa e costosa
- Nelle LAN
 - Per le dimensioni limitate il costo del mezzo incide meno rispetto al costo dell'attacco per le stazioni
 - La penetrazione delle fibre ottiche è più lenta
 - Per gli ultimi metri fino all'attacco, potrebbero sopravvivere le coppie intrecciate (twisted pairs)
- Il mezzo radio, per motivi di affidabilità e di costi non è stato usato fino alla fine degli anni '90, ma sta acquistando un'importanza sempre crescente

Topologie

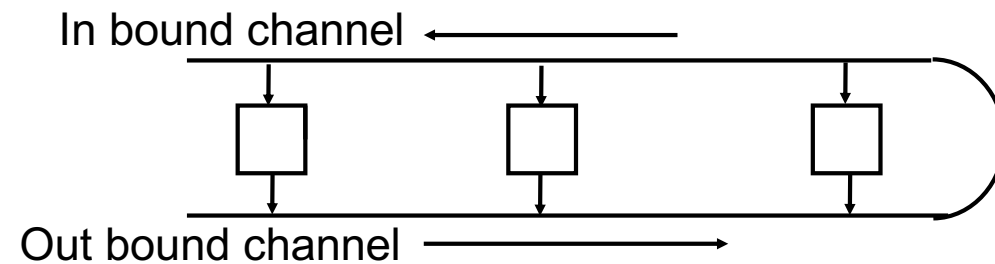
- Topologie tradizionalmente preferita nelle WAN
 - A stella
 - A maglia più o meno completa
 - Architettura gerarchica
- Topologie punto-multipunto
 - Non adatte per le WAN
 - Nelle WAN è importante avere i collegamenti più lunghi possibile
 - Le prese intermedie sottraggono potenza e aggiungono disturbi imponendo collegamenti più corti
 - Sono state preferite nelle prime proposte di LAN
 - Se i terminali sono pochi non servono nodi di commutazione

Topologie punto-multipunto

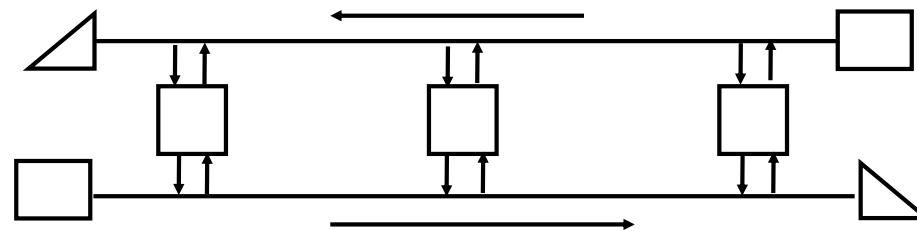
- Bus
bidirezionale



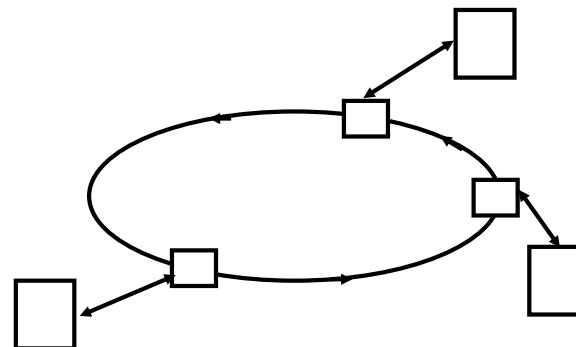
- Bus
unidirezionale



- Doppio bus
(dual bus)

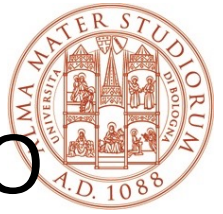


- Anello



Due caratteristiche

- Topologie punto-multipunto = mezzo di trasmissione condiviso = due caratteristiche peculiari
 - Broadcast
 - La LAN fornisce in modo nativo una comunicazione da uno a tutti (broadcast) sul mezzo condiviso
 - Per evitare che tutti i calcolatori leggano i dati di tutte le comunicazioni si deve introdurre un meccanismo di indirizzamento a livello di protocollo LAN
 - Si vuole evitare che lo strato 3 debba elaborare anche le conversazioni degli altri
 - Collisione
 - Su di un mezzo condiviso esiste la possibilità che più utenti inviino informazioni contemporaneamente
 - Interferenza – perturbazione o distruzione dell'informazione – perdita di trasparenza semantica



Accesso al canale di collegamento

- Collegamenti con canali punto-punto e commutati
 - Solo sorgente e destinazione hanno accesso al canale quindi la sorgente può liberamente impegnare tutta la capacità certa di raggiungere la destinazione
- Collegamenti con canali ad accesso multiplo
 - Più sorgenti possono accedere al canale contemporaneamente determinando quindi la «**collisione**»
- Collegamenti con accesso al canale controllato
 - Il canale è condiviso ma l'accesso viene controllato in modo centralizzato o distribuito per **evitare** fenomeni di collisione

Accesso multiplo: tassonomia

