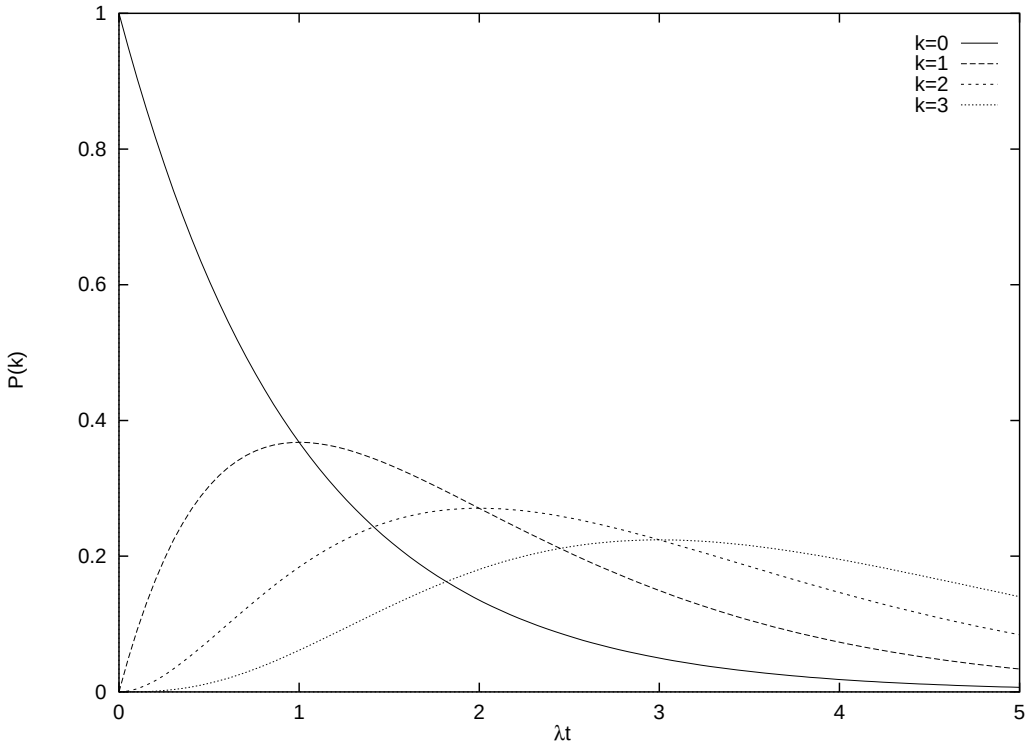
Teoria del traffico

Andamento della distribuzione di Poisson

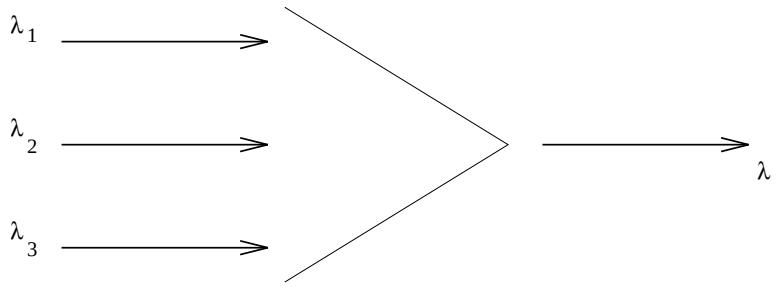
$$\mathrm{P}(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$



Proprietá

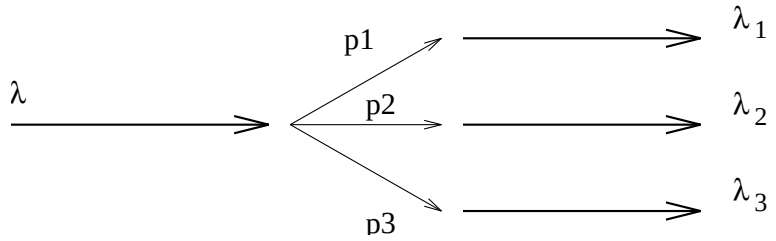
La sovrapposizione di h processi di Poisson aventi frequenze di arrivo λ_i $i = 1 \ldots h$ é ancora un processo di Poisson con

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



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

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



Teorema 1 *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 2 *La varianza del numero di arrivi in un tempo t é:*

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

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 3

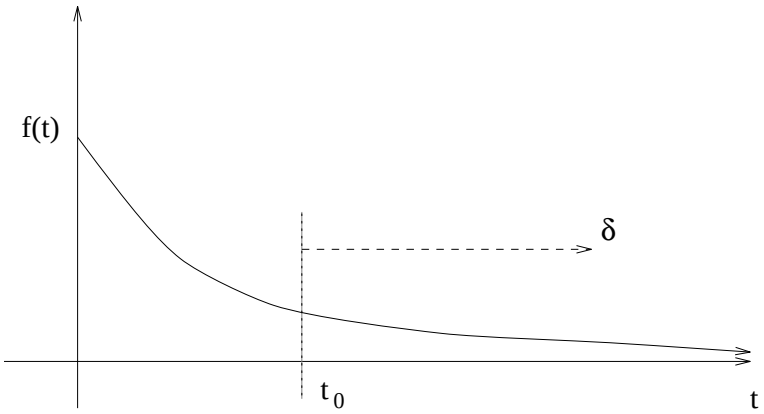
$$E[t] = \frac{1}{\lambda}$$

Teorema 4

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

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.



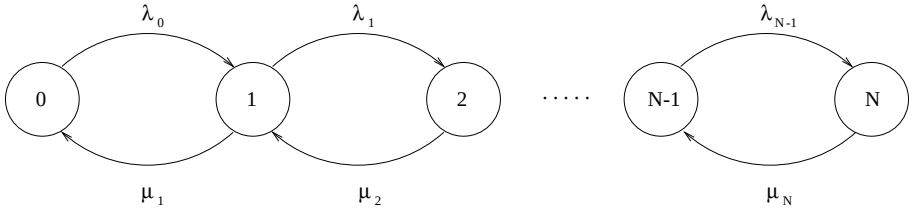
$$
\begin{aligned}
F(\delta) &= \mathrm{P}(t \le t_0 + \delta \mid t > t_0) \\
&= \frac{\mathrm{P}(t_0 < t \le t_0 + \delta)}{\mathrm{P}(t > t_0)} \\
&= \frac{\mathrm{P}(t \le t_0 + \delta) - \mathrm{P}(t \le t_0)}{\mathrm{P}(t > t_0)} \\
&= \frac{1 - e^{-\lambda(\delta + t_0)} - (1 - e^{-\lambda t_0})}{1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

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

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

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:

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

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

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:

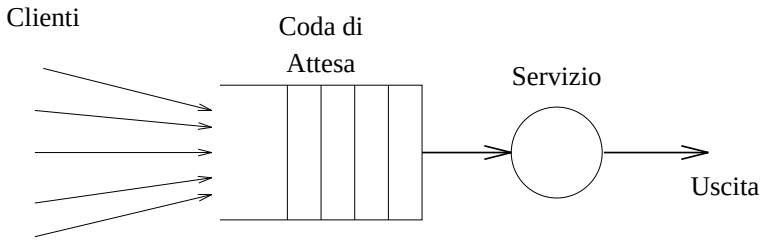
$$P_k = P_0 \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}$$

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

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.

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.

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.

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.

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.

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

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

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).

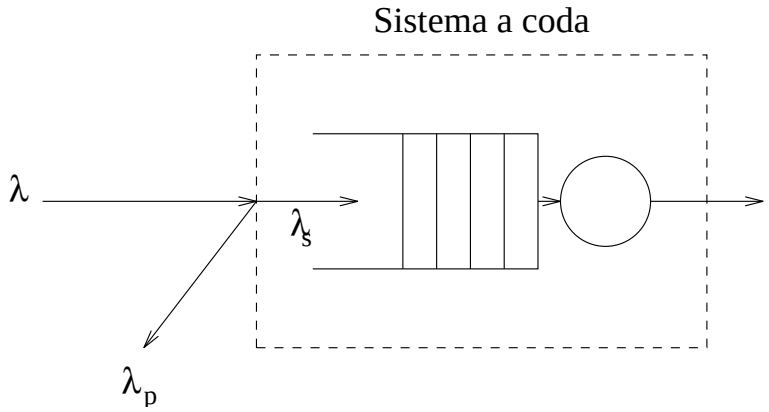
Sistemi con perdita

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.



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:

$$A_s = A_0 \;\; A_p = 0$$

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.

Coda M/M/m/0

- m servitori
- Coda nulla

Probabilitá di stato

Le frequenze di nascita e di morte risultano:

$$\begin{cases} \lambda_k = \lambda \ \ , 0 \leq k < m \\ \\ \lambda_k = 0 \ \ , k = m \end{cases}$$

$$\mu_k = k\mu \ \ , 0 < k \leq m$$

da cui

$$P_k = P_0 \frac{\left(\frac{\lambda}{\mu}\right)^k}{k!} = P_0 \, \frac{A_0^k}{k!}, \ \ 0 \leq k \leq m$$

con fattore di normalizzazione:

$$P_0 = \frac{1}{\sum_{k=0}^{m} \frac{A_0^k}{k!}}$$

per qualsiasi valore del traffico offerto il sistema risulta stazionario perchè può riempirsi solo fino ad m.

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 $\Pi_t \neq \Pi$.

Legame fra Call Congestion e Time Congestion

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

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

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

per cui dal Teorema di Bayes:

$$Pr(A|B)Pr(B) = Pr(B|A)Pr(A)$$

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 arrivi dallo stato del sistema é sempre vera se gli arrivi sono poissoniani. (Proprietà PASTA: Poisson Arrivals See Time Averages)

Principale conseguenza : per l'analisi della congestione è sufficiente studiare le probabilità di stato P_k.

Probabilità di congestione: la formula B di Erlang

Per il sistema $M/M/m/0$ la probabilità di congestione è data dalla probabilità che un utente arrivi e trovi il sistema pieno e quindi venga respinto: probabilità di perdita o di blocco (π_p).

Questa coincide per la proprietà PASTA con la probabilità di sistema pieno P_m:

$$\pi_p = P_m = \frac{\frac{A_0^m}{m!}}{\sum_{k=0}^{m} \frac{A_0^k}{k!}} = \mathbf{B}(m, A_0)$$

nota come *formula B di Erlang*.

Importante proprietà: vale per qualunque distribuzione del tempo di servizio.

Trova applicazione nel dimensionamento dei sistemi a commutazione di circuito in cui il numero dei circuiti diponibili è finito e viene usato in modo esclusivo da una particolare connessione.

Se una chiamata arriva e tutti i circuiti in uscita dalla centrale sono impegnati, questa viene rifiutata o bloccata (sistema a chiamate bloccate cancellate).

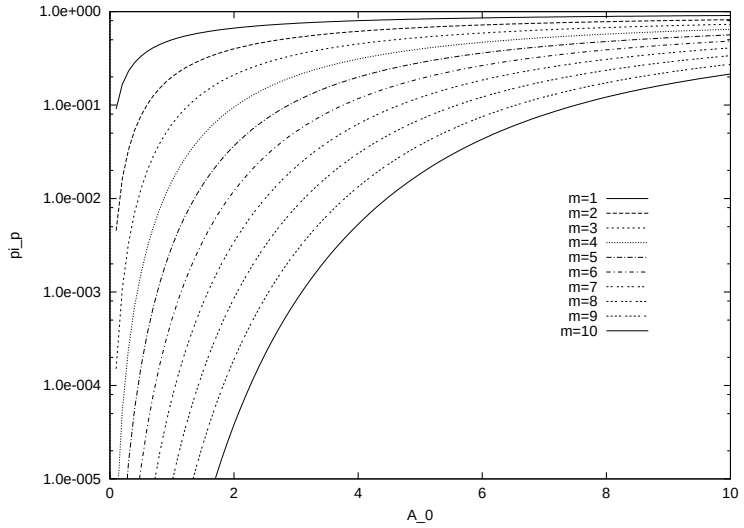


Figure 1: Andamento di π_p in funzione di A_0 per vari valori di m.

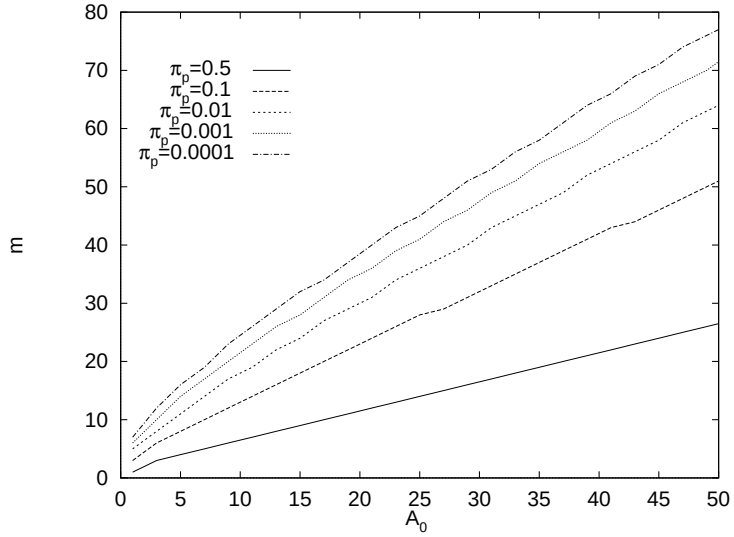


Figure 2: Andamento di m in funzione di A_0 per vari valori di π_p.

Utenti nel sistema e Utilizzazione dei servitori

Il numero medio di utenti nel sistema è uguale al numero medio di servitori occupati e quindi al traffico smaltito:

$$A_s = (1 - \pi_p)A_0 = [1 - \mathbf{B}(m, A_0)]A_0$$

Il tempo medio δ di permanenza nel sistema coincide con il tempo di servizio θ.

Utilizzazione dei servitori ora viene calcolata come:

$$\rho = \frac{A_s}{m} = \frac{A_0(1 - \pi_p)}{m}$$

La figura 3 mostra come all'aumentare di A_0, ρ tenda ad 1 in modo men che proporzionale ad A_0: a parità di π_p un sistema di grandi dimensioni fa un uso migliore dei suoi servitori, quindi si ha convenienza ad aggregare i flussi di traffico.

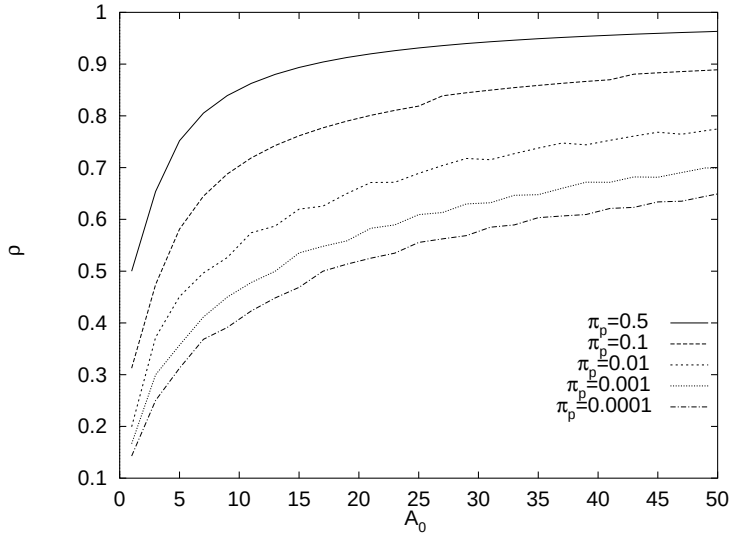


Figure 3: Andamento di ρ in funzione di A_0 per vari valori di π_p.