Analisi operazionale

M. Arrigoni Neri & *P. Borghese*

indice

- principi e metodi della *analisi operazionale*
  - modelli a rete di code
  - modello aperto/chiuso
  - leggi dell'analisi operazionale
  - limiti asintotici
  - limiti bilanciati
  - MVA per sistemi chiusi (cenni)
    - estensione ai sistemi multi-classe
    - MVA approssimato

  - E. D. Lazowska, J. Zahorian, G. S. Graham, K. C. Sevcik: Quantitative System Performance - *Prentice-Hall*
    si può scaricare da:
    http://www.cs.washington.edu/homes/lazowska/qsp/

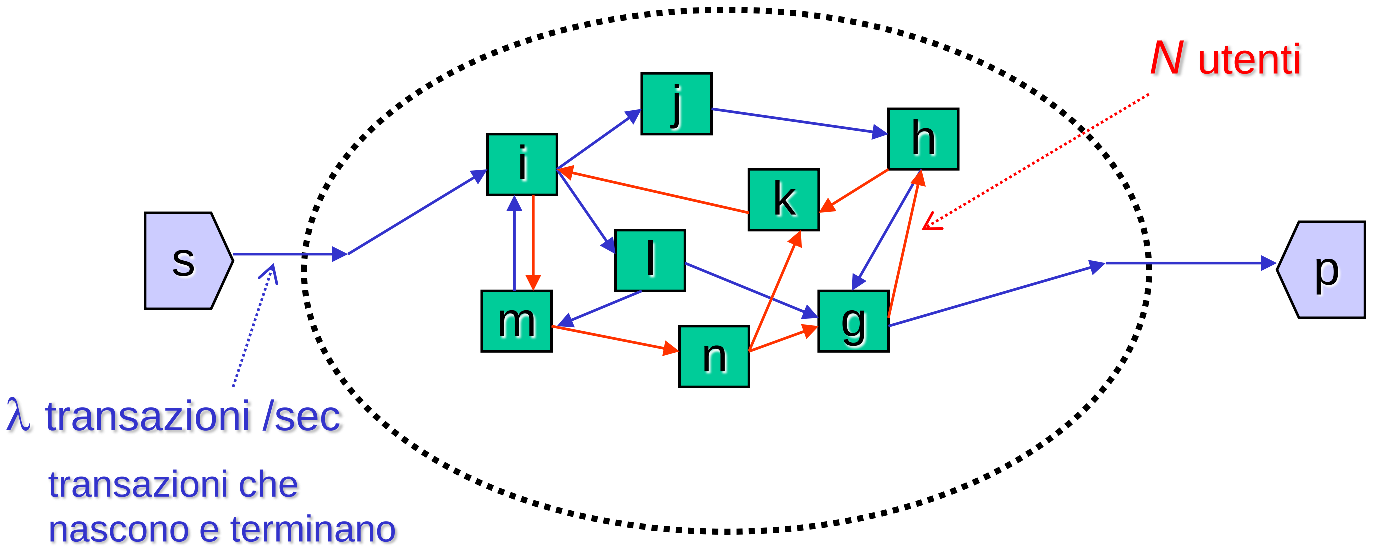
principi e metodi della analisi operazionale

Per dare un significato effettivo a una nozione e non a un'apparenza di questa in senso puramente verbale o metafisico, è richiesta una definizione operativa cioè basata su un criterio che ci permetta di misurarla

modelli a rete di code

- rappresentano in *modo naturale* un sistema di componenti interconnessi visitati da *transazioni*
- transazione (unità di carico)
  - entità che risiede nel sistema per un certo periodo di tempo (response time) e passa istantaneamente da un nodo all'altro
- stazione di servizio (nodo - componente attivo del sistema)
  - riceve e accoda le *transazioni* provenienti dagli altri *nodi* della rete ed esegue uno specifico servizio con una data disciplina;
  - in una stazione di servizio operano uno o più *serventi*;
  - al completamento del servizio corrente la transazione viene instradata verso un'altra stazione (*scelta secondo una probabilità fissata*) dove riceverà il prossimo servizio, oppure abbandona la rete.

modelli a rete di code (cont.)



- ipotesi semplificatrice: una *transazione* in ogni *istante* si trova in uno *stato* ben definito in uno specifico *nodo* della rete *(il passaggio da un nodo al successivo è istantaneo)*
- problema fondamentale:
  - *dato un sistema reale come si costruisce la configurazione della rete che lo rappresenta?*
  - *non c'è una risposta univoca - (esistono anche costruttori automatici di modelli)*

modelli a rete di code (cont.)

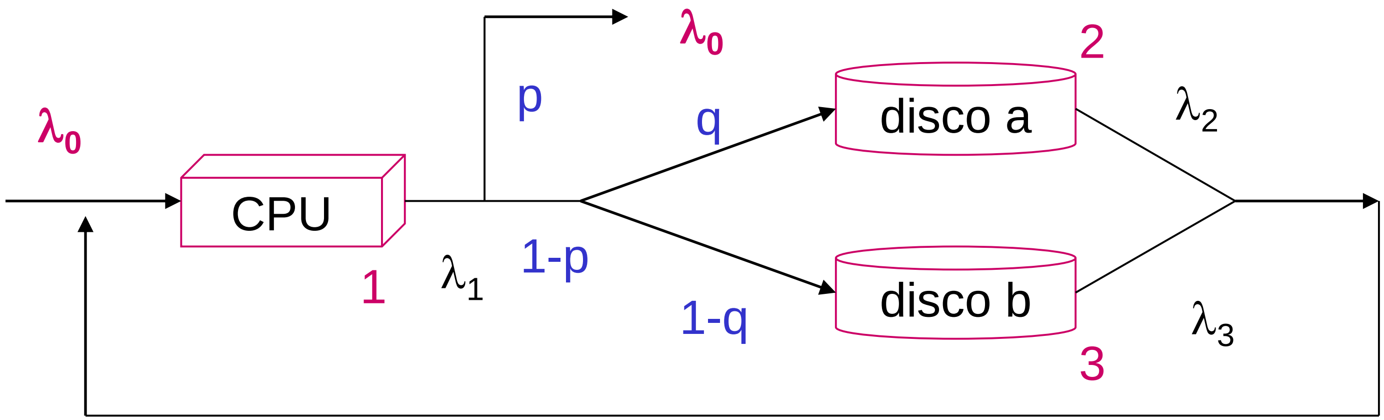
- lo *stato* del sistema è rappresentato dal vettore del numero di elementi presenti in ogni componente: $(n_1, n_2, \dots n_M)$
- se $n_1 + n_2 + \dots n_M = N$ (popolazione) = costante, il modello è *chiuso* cioè isolato dal modo esterno
- in caso contrario è *aperto*, le transazioni provengono da sorgenti esterne e lasciano il sistema al loro completamento

- in condizioni di equilibrio i *flussi* di transazioni sono *bilanciati* localmente e globalmente ($\lambda_1^- = \lambda_1^+ \dots$)
- una *soluzione* del modello permette la previsione delle grandezze di prestazione relative a una certa situazione stazionaria di *carico*

- se le classi di carico sono più di una allora: N, $\mathbf{n_i}$, $\boldsymbol{\lambda_i}$ sono vettori di componenti: N_c , $n_{i,c}$, $\lambda_{i,c}$ (c = classe; i = nodo; $N_c = \Sigma_i \, n_{i,c}$)

modello aperto

- *stazioni*: risorse (CPU, dischi, linee) - *clienti*: transazioni - job
- λ_0 (*variabile indipendente*): flusso da e verso l'esterno

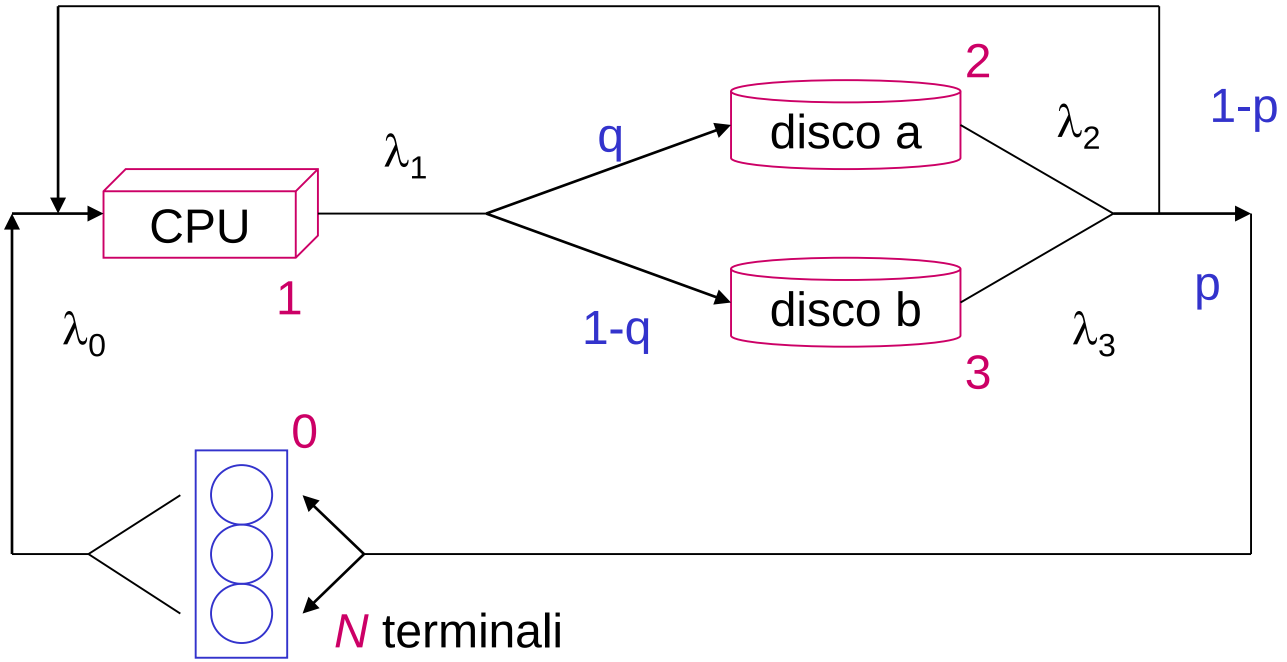


- in condizioni di *equilibrio*:
  - λ_1 p = λ_0
  - numero di I/O (per transazione):

    λ_1 (1-p) / λ_0 = (1-p) / p = 1/p -1

modello chiuso

- *N* (*variabile indipendente*): popolazione di utenti



- in condizioni di *equilibrio*:
  - λ_1 (1-p) + λ_0 = λ_1 λ_1 p = λ_0
  - numero di I/O (per transazione): 1/p (visite ai dischi)

modelli a rete di code (cont.)

- parametri di prestazione (misure - indici)
  - **s_i** = $1/\mu_i$ = tempo medio di servizio al nodo (i)
    [$s_i(n)$; n numero di utenti nel nodo]; μ_i = tasso di servizio
  - **λ_0** = tasso di arrivi dall'esterno (o dai terminali)
  - **λ_i** = throughput - numero medio di completamenti di servizio nell'unità di tempo [λ_i (N); N: numero di utenti nella rete]
  - **r_i** = tempo medio di residenza (speso in (i))
  - **w_i** = tempo medio di attesa servizio in (i)
  - **L_i** = numero medio di utenti nella stazione (i) esclusi quelli in servizio
  - **Q_i** = numero medio di utenti nella stazione (i) compresi quelli in servizio
  - **U_i** = utilizzo, frazione di tempo in cui la stazione (i) è occupata (busy)
  - **$P_i(n)$** = probabilità che ci siano n utenti in (i)

ipotesi dell'analisi operazionale

- i *comportamenti osservati* durante il *periodo* T di osservazione possiedono le seguenti proprietà:
  - *omogeneità*

    arrivi: λ tasso di arrivo ad una stazione dall'esterno è indipendente dal numero di utenti presenti in essa;

    servizi: μ tasso di servizio è indipendente dal numero di utenti presenti;

    nota: μ deve essere 0 quando la stazione è vuota (analogamente il tasso di arrivo λ alla stazione i in un modello chiuso, deve essere 0 se tutti gli N clienti sono concentrati nella stazione i);

    cammini: le probabilità di *routing* (instradamento) sono indipendenti dallo stato
  - *conservazione del flusso (bilanciamento)*

    il numero totale di arrivi nell'intervallo T è uguale al numero delle partenze (localmente e globalmente) – le transazioni sono unità "atomiche"
  - *one step behavior*

    i cambiamenti di stato avvengono fra stati contigui - in ogni istante lo stato deve poter essere definito senza ambiguità;

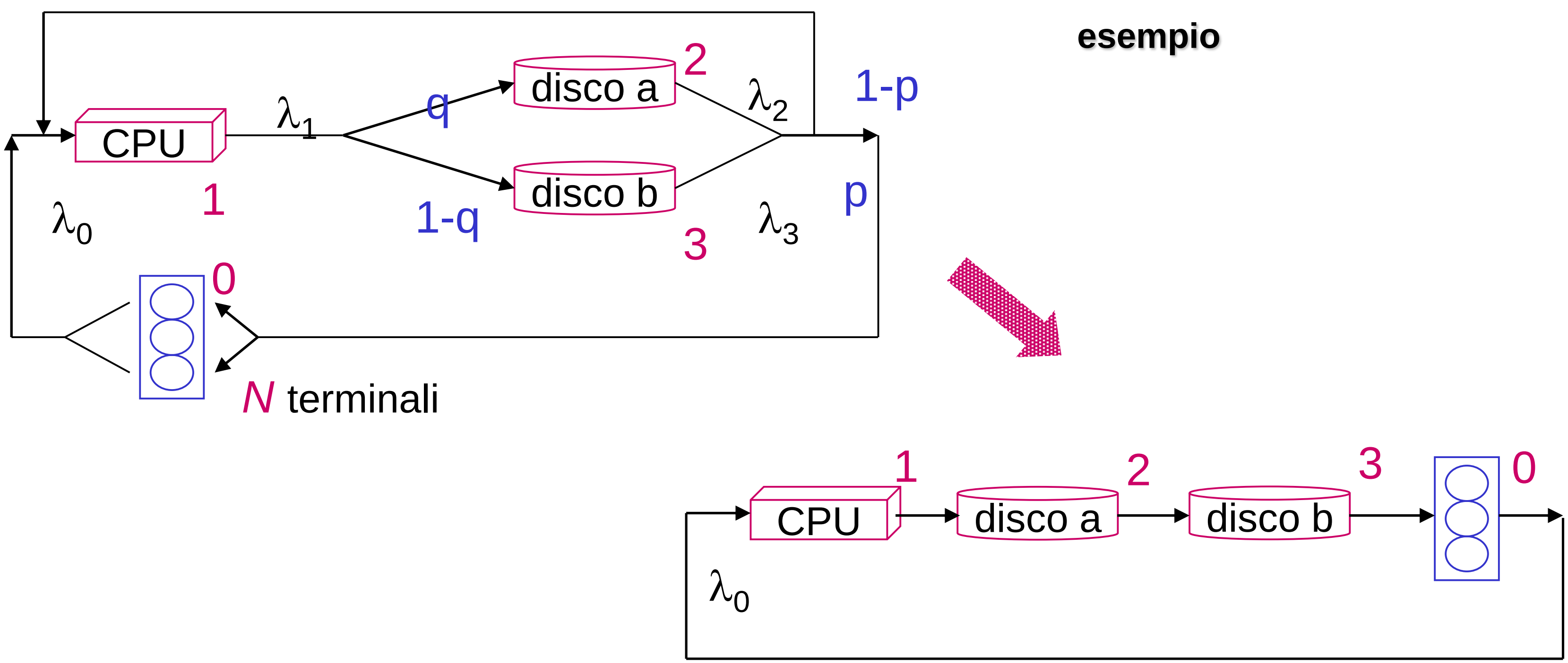
    (gli arrivi non coincidono con le partenze - in ogni istante può verificarsi solo un arrivo o una partenza)

analisi operazionale (cont.)

l'analisi operazionale fornisce per un sistema in equilibrio:

- i *valori medi* delle grandezze di prestazione
- mediante la soluzione di un modello semplificato del sistema ottenuto da quello di partenza

esempio



relazioni operazionali (misure relative all'intervallo T)

$$\begin{cases} A_i(n) \\ C_i(n) \\ T_i(n) \end{cases}$$

- numero degli arrivi a (i) che trovano *n* clienti presenti
- numero delle partenze che lasciano *n* -1 clienti in (i)
- durata del tempo in cui ci sono *n* clienti nel nodo (i)

$$T - T_i(0) = \sum_{n=1}^{N} T_i(n) = B_i$$

- tempo di *busy* di (i)

$$\sum_{n=0}^{N-1} A_i(n) = A_i \qquad \sum_{n=1}^{N} C_i(n) = C_i$$

- numero totale di arrivi e partenze

$$X_i = A_i / T \qquad \lambda_i = C_i / T$$

- flussi in entrata e uscita

relazioni operazionali (misure nell'intervallo T) (cont.)

$$P_i(n) = T_i(n)/T$$

- probabilità che vi siano n utenti in (i)

$$U_i = \frac{T - T_i(0)}{T} = 1 - P_i(0)$$

- utilizzo della stazione (i) (probabilità di trovare la stazione occupata)

$$U_i(N) = \frac{B_i}{T} = \frac{B_i}{C_i} \cdot \frac{C_i}{T} = s_i \cdot \lambda_i(N)$$

- *legge dell'Utilizzo*
  (in questa forma vale quando $s_i(n)$ non dipende da n)

$$U_i = \frac{\lambda_i}{\mu_i}$$

- $\mu_i = 1/s_i$
  rappresenta il flusso in uscita da (i) quando (i) è occupato (n >0)

relazioni operazionali (misure nell'intervallo T) (cont.)

$$A_i(n) = C_i(n+1)$$

- condizioni di equilibrio

$$PA_i(n) = A_i(n)/A_i$$

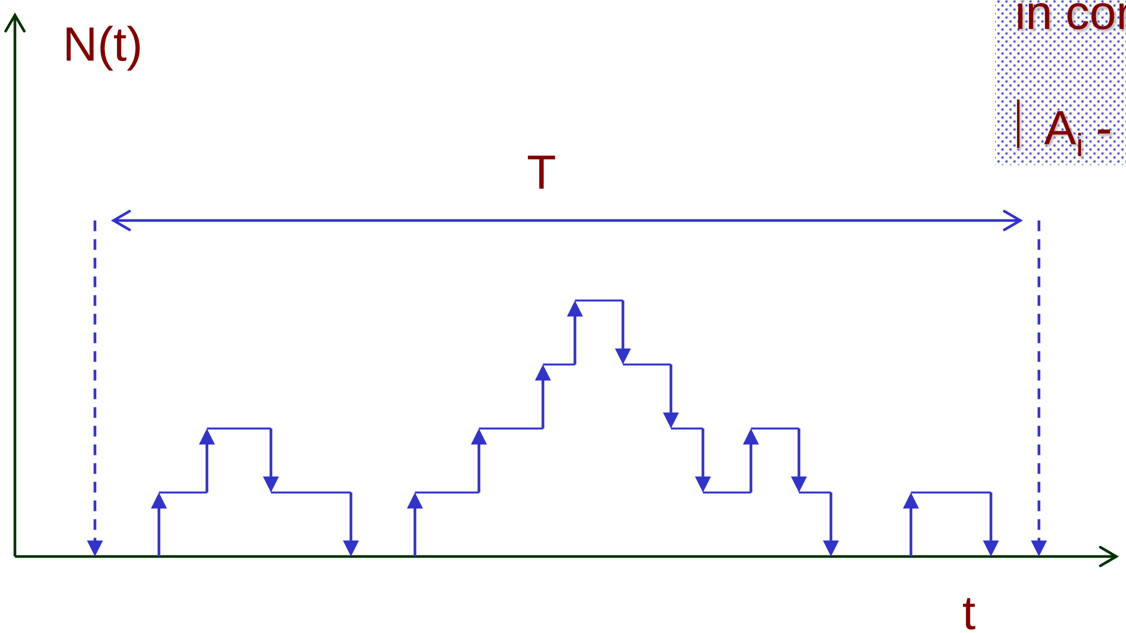
$$PC_i(n) = C_i(n+1)/C_i$$

- probabilità di trovare n utenti in (i) all'arrivo
- probabilità di lasciare n utenti in (i) alla partenza

(a rigore si tratta di frequenze che poi saranno usate come probabilità nelle soluzioni)

analisi operazionale (cont.)

- una rete è detta *ben formata* se ogni nodo è raggiungibile da tutti gli altri con probabilità > 0
- N(t) rappresenta la *popolazione* all'istante t (nell'intera rete o in una sua parte)



in condizioni di equilibrio:

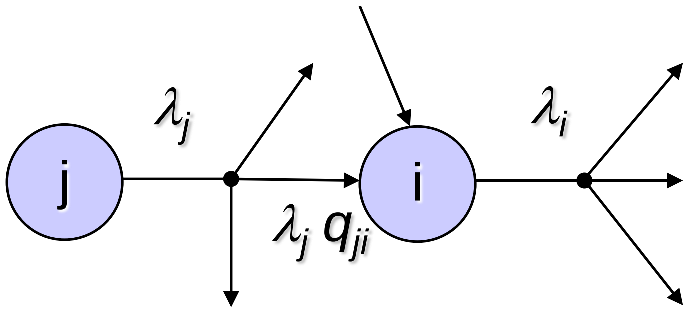
$|A_i - C_i| \ll A_i \Rightarrow \lambda_i \approx X_i$

$A_i = C_i$ se $N(0) = N(T)$

bilanciamento dei flussi

- imponendo che in ognuno degli M nodi della rete il flusso si conservi: $\lambda_i = X_i$

$$\lambda_i = \sum_{\substack{j=0 \\ j \neq i}}^{M} \lambda_j \cdot q_{ji}$$



- rete aperta:
  - λ_0 è il flusso entrante dall'esterno (e uscente) e q_{ji} la *probabilità di instradamento* da (j) a (i)
  - se la rete è *ben formata* per ogni valore di arrivi λ_0 esiste una soluzione λ_i (i =1,..M), *throughput* di ogni stazione (i)

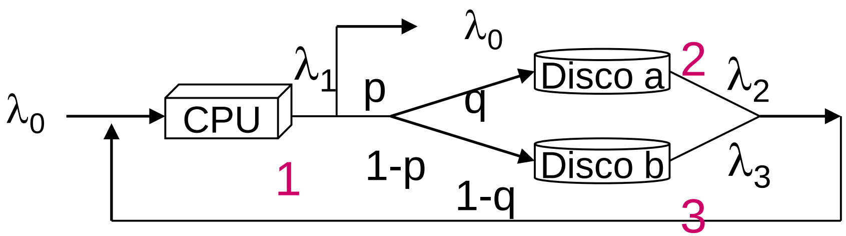
bilanciamento dei flussi (cont.)

- rete chiusa:
  - il rango del sistema omogeneo è M -1, abbiamo M -1 equazioni linearmente indipendenti e possiamo ricavare i throughput relativi alla stazione k di riferimento λ_i / λ_k

- facendo riferimento ai due modelli precedenti, possiamo scrivere i seguenti sistemi di equazioni nelle incognite λ:

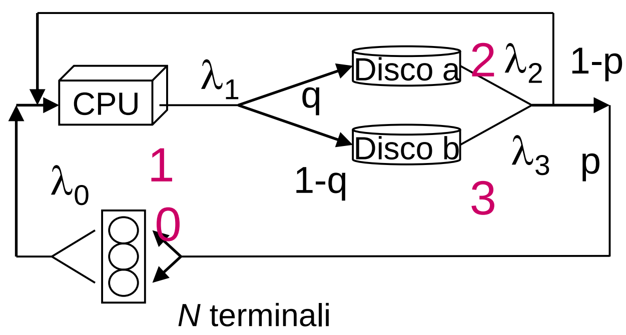
modello aperto

$$\lambda_0 = \lambda_1 \cdot p$$
$$\lambda_1 = \lambda_0 + \lambda_2 + \lambda_3$$
$$\lambda_2 = (1-p)\lambda_1 \cdot q$$
$$\lambda_3 = (1-p)\lambda_1 \cdot (1-q)$$



modello chiuso (k = 0: stazione di riferimento)

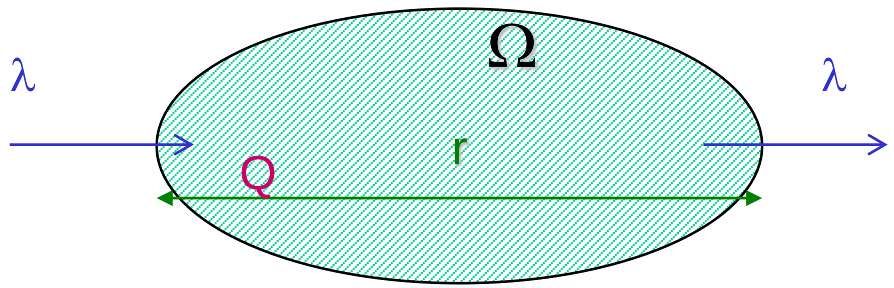
$$\lambda_1 = \lambda_0 + (\lambda_2 + \lambda_3)\cdot(1-p)$$
$$\lambda_2 = \lambda_1 \cdot q$$
$$\lambda_3 = \lambda_1 \cdot (1-q)$$



relazioni fondamentali: legge di Little

- sia Ω una regione chiusa che contiene una parte della rete, facendo riferimento a tale regione (che può anche essere l'intera rete, un solo nodo o una parte di questo) se:
  - λ: output rate (tasso di ingresso / uscita)
  - **Q**: numero di clienti che si trovano mediamente in Ω
  - **r**: tempo medio di residenza dei clienti in Ω fra un ingresso e una uscita

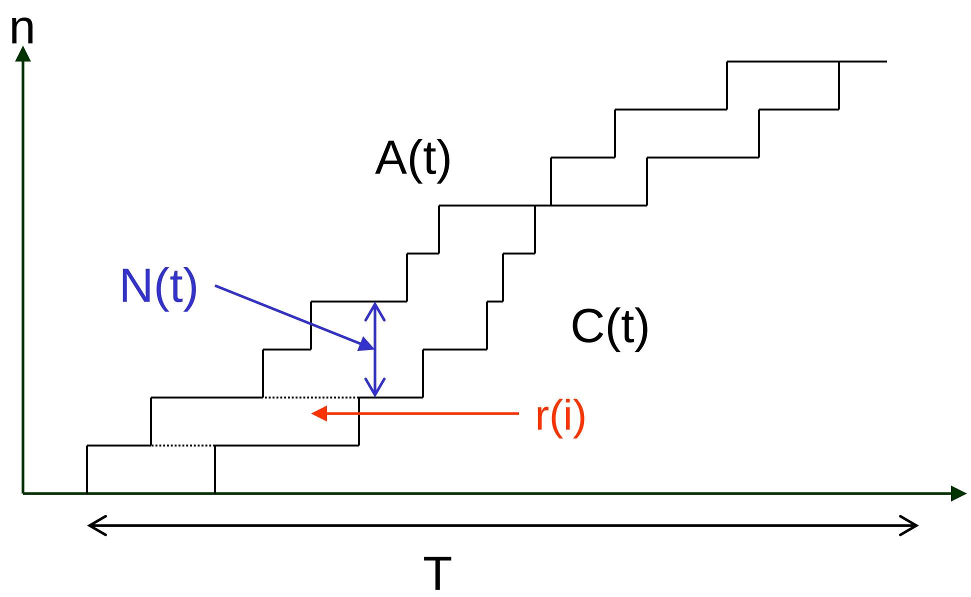
  allora:
  - **$Q = r\,\lambda$** Q = valore medio di N(t)



- **A(t)**: processo degli arrivi
- **C(t)**: processo delle partenze
- **N(t) = A(t) - C(t)** : numero di utenti al tempo t

(i nodi di ingresso / uscita possono essere multipli)

Legge di Little (cont.)



n arrivi nel tempo *T*
r(i) risposta della transazione i^{ma}

area : integrale dei clienti presenti nel sistema

area: Σ r(i)

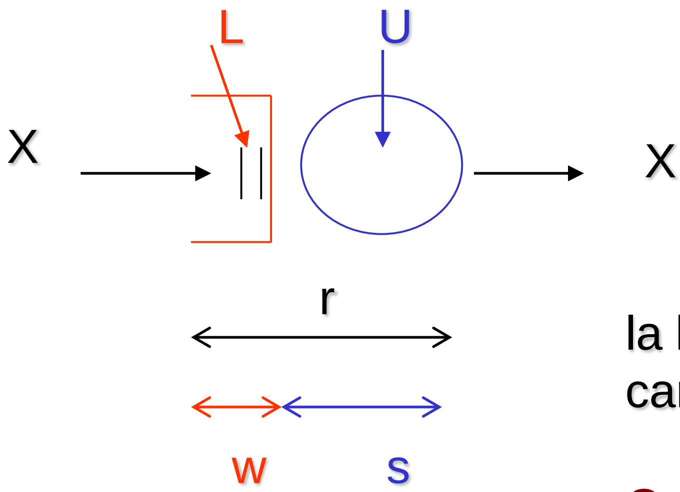
r: *larghezza media*
Q: *altezza media (dell'area)*

$$r = \frac{1}{n}\sum_{1}^{n} r(i) = \frac{area}{n}$$

$$Q = \frac{area}{T}$$

$$\Rightarrow \boxed{Q = \frac{area}{n} \cdot \frac{n}{T} = r \cdot \lambda}$$

Legge di Little (cont.)



$Q = L + U$
$r = w + s$
$Q = X\ r$
$L = X\ w$
$U = X\ s$

la legge vale globalmente e per ogni classe di carico - c = classe; i = nodo

$Q_{i,c} = X_{i,c}\ r_{i,c}$ $\quad Q_i = \Sigma\ Q_{i,c}$ $\quad X_i = \Sigma\ X_{i,c}$

media da Little

media pesata

$$r_i = \frac{Q_i}{X_i} = \frac{\sum_c r_{i,c} \cdot X_{i,c}}{\sum_c X_{i,c}}$$

alcune proprietà importanti

la legge di *Little* è:

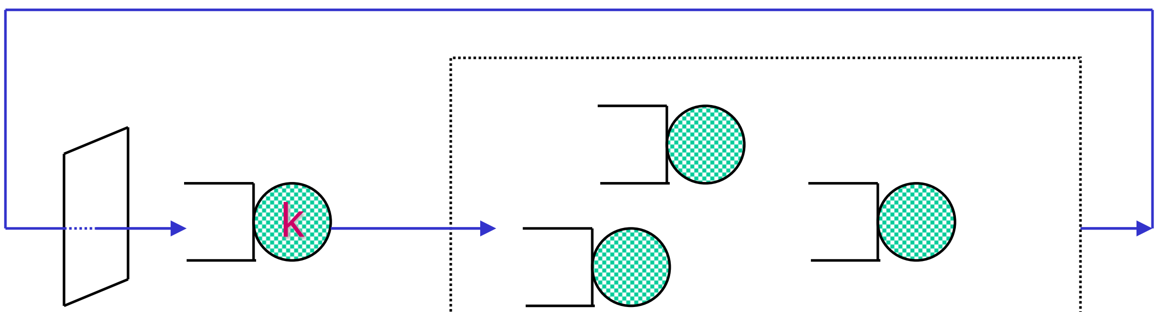
- indipendente dalla disciplina di servizio
  - discipline diverse a parità di λ possono dar luogo a diversi valori di r e Q ma:
  - Q/r = Q'/r' = Q"/ r"
- vale per il generico nodo (i)
- per la sola area di attesa
- per il solo servente (equivale alla *legge dell'utilizzo*)

$$\boxed{Q_i = \lambda_i \cdot r_i} \qquad \boxed{L_i = \lambda_i \cdot W_i} \qquad \boxed{U_i = \lambda_i \cdot s_i}$$

$\leftarrow$ numero degli utenti mediamente in servizio (questa interpretazione ha senso anche per numero di serventi > 1)

alcune proprietà importanti (cont.)

- una rete chiusa contiene N utenti ed M stazioni
  - **$CT_k(N)$**: tempo del primo passaggio; (tempo che un utente mediamente impiega per tornare la prima volta a k dopo la sua visita precedente - comprensivo di r_k-)



$$T = v_k \cdot CT_k$$

valore medio

$$CT_k(N) = \sum_{i=1}^{M} v_i \cdot \frac{r_i(N)}{v_k}$$

per uno specifico utente

$$T = \sum_{1}^{M} v_i \cdot r_i$$

v_i: numero di *visite* nell'intervallo T: di osservazione

alcune proprietà importanti (cont.)

- i rapporti $\lambda_i / \lambda_k = v_i / v_k$ sono costanti e indipendenti dal carico (ipotesi di omogeneità)

- considerando come riferimento, la stazione ($_0$) cioè posto:
  - $k = 0$, $v_0 = 1$
  - $r_0 = Z$ tempo di *thinktime* dei terminali

  (le stazioni sono ora M+1)

  v_i = numero di visite ad (i) nel tempo CT(N) cioè per ogni passaggio da (0)

$$CT(N) = \sum_{i=1}^{M} v_i \cdot r_i(N) + Z$$

$$CT(N) \cdot \lambda_0(N) = N$$

$$\lambda_j(N) = \lambda_0(N) \cdot v_j = \frac{N \cdot v_j}{\sum_{i=1}^{M} v_i \cdot r_i(N) + Z}$$

- applicando la legge di Little

$$\boxed{\lambda_0 = \lambda_j / v_j = N / (R + Z)}$$

$$\boxed{R = \Sigma R_i = \Sigma r_i v_i}$$

Legge del "tempo di risposta"

alcune proprietà importanti (cont.)

- *uguaglianza delle distribuzioni* viste dall'utente all'arrivo e alla partenza da un nodo

$$PA_i\left(n|N\right) = PC_i\left(n|N\right)$$

$A_i = C_i$ nelle relazioni a pag. 14

- *teorema degli arrivi* nei sistemi chiusi:

$$\boxed{PA_i\left(n|N\right) = P_i\left(n|N-1\right)}$$

l'utente che arriva a un nodo vede il sistema in equilibrio con se stesso rimosso

- lunghezza media della coda

$$Q_i\left(N\right) = \sum_{n=1}^{N} n \cdot P_i\left(n|N\right)$$

- *teorema del valore medio*

$$\boxed{\begin{aligned} r_i(N) &= s_i \cdot \left(1 + QA_i(N)\right) \\ &= s_i \cdot \left(1 + Q_i(N-1)\right) \end{aligned}}$$

alcune proprietà importanti (cont.)

- *teorema degli arrivi* nei sistemi aperti:

  l'utente che arriva vede il sistema in equilibrio

$$\boxed{PA_i(n \mid \lambda_i) = P_i(n \mid \lambda_i)}$$

$$Q_i(\lambda_i) = \sum_{n=1}^{\infty} n \cdot P_i(n \mid \lambda_i)$$

- perciò:

  (legge di Little)

$$r_i(\lambda_i) = s_i \cdot (1 + Q_i(\lambda_i))$$

$$r_i(\lambda_i) = s_i (1 + \lambda_i \cdot r_i(\lambda_i))$$

- tempo di risposta:

$$r_i(\lambda_i) \cdot (1 - s_i \cdot \lambda_i) = s_i \quad \Rightarrow \quad \boxed{r_i(\lambda_i) = \frac{s_i}{1 - U_i}}$$

analisi operazionale: note

$v_i = X_i / X$
$Q_i = X_i\, r_i = X\, v_i\, r_i = X\, R_i$

$\mathbf{R_k > R_h \Rightarrow Q_k > Q_h}$

$U_i = X_i\, s_i = X\, v_i\, s_i = X\, D_i$

$\mathbf{U_h / U_k = D_h / D_k}$

- (il sistema è in equilibrio perciò possiamo utilizzare X o λ indifferentemente)
- gli utenti tendono ad addensarsi dove permangono più a lungo
- *i rapporti fra gli utilizzi dei componenti sono indipendenti dal carico*

- è utile considerare due tipi di nodi:
  - stazioni di servizio *"normali"* in cui si formano linee di attesa
  - stazioni che si possono considerare *ritardi* (delay) come i terminali, o nodi equivalenti a gruppi di dispositivi in cui non si formano linee di attesa (sono detti serventi IS - *infinite server*)

analisi operazionale: note (cont.)

$v_i = X_i / X$ ←

- la costanza di questo rapporto (dovuta all'ipotesi di omogeneità) permette di calcolare le variabili prestazionali R_i e Q_i senza passare dai valori X_i ma dalla sola conoscenza di X e D_i

$D = \Sigma D_i$

$R = \Sigma R_i$

Domanda
e
tempo di risposta
totale

$R_i = R_i (D_i, X)$

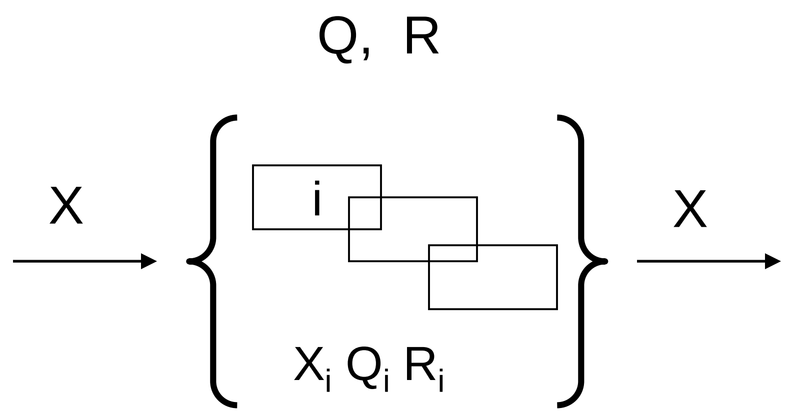
$R = R(D_1,\ldots D_M; X)$

- D_i domanda di risorse al nodo (i)
- R_i tempo di permanenza (residenza) della transazione nel nodo (i)
- X throughput globale delle transazioni
- se abbiamo diverse classi di transazioni, i calcoli sono più complessi ma in sostanza dello stesso tipo

ma attenzione!:

$R \neq R(D, X)$

riassunto: modello aperto



$Q = X\,R$

$Q_i = X_i\, r_i = X\, R_i$

$U_i = X_i\, s_i$

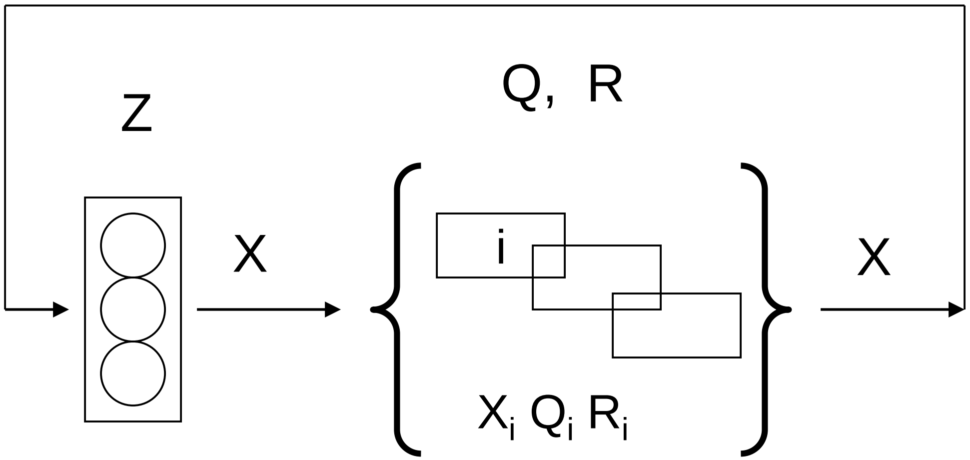
$X\,R = Q = \Sigma\, Q_i = \Sigma\, X_i\, r_i$

$R = \Sigma\,(X_i / X)\, r_i = \Sigma\, v_i\, r_i = \Sigma\, R_i$

Visite al nodo (i)

tempo di residenza in (i)
per ogni attraversamento
del sistema durante il quale
(i) viene visitato v_i volte

riassunto : modello chiuso



$Q = X\,R$
$Q_i = X_i\,r_i = X\,R_i$
$U_i = X_i\,s_i$
$X\,R = Q = \Sigma\,Q_i = \Sigma\,X_i\,r_i$

$X\,(R + Z) = N$
$R = N / X - Z$ ← Legge del tempo di risposta

$D_i = v_i\,s_i$
$D = \Sigma\,v_i\,s_i$ ← Domande di servizio

Nota: più che di modelli aperti o chiusi bisognerebbe parlare di classi di carico aperte o chiuse
i modelli generalmente sono “misti”

limiti di prestazione: analisi dei *colli di bottiglia*

b: indice del nodo strozzatura (collo di bottiglia, “*bottleneck*”)

$D_b = \max \{ D_i \}$

D: domanda totale

$D_a = D / M$: domanda media

$$D_b X = U_b \leq 1 \Rightarrow X \leq 1 / D_b = X_{MAX}$$

(attenzione ai serventi multipli: $X \leq m / D_b$)

limiti del tempo di risposta (modello chiuso)

$D + Z = R(1) + Z$: minimo tempo di ciclo

(eventuali delay sono compresi in Z)

$R(1) = D$: minimo tempo di risposta

$N D + Z$: massimo tempo di ciclo

ND: massimo tempo di risposta

modello aperto (calcolo del tempo medio di risposta)

- il tempo di permanenza nel nodo (k) è dato da:
- $R_k = D_k / (1- U_k)$

  dove U_k è l'utilizzo del nodo (k)
- $R = \Sigma_k R_k$

caso multiclasse

- il tempo di permanenza nel nodo (k) per la classe (c) è dato da:
- $R_{c,k} = D_{c,k} / (1- U_k)$

  dove U_k è l'utilizzo del nodo (k) da parte di *tutte* le classi
- $R_c = \Sigma_k R_{c,k}$

Importante!

limiti asintotici (modello chiuso)

$$\frac{N}{ND+Z} \leq X(N) \leq \min\left(\frac{N}{D+Z}, \frac{1}{D_b}\right)$$

$$\max(D, ND_b - Z) \leq R(N) \leq ND$$

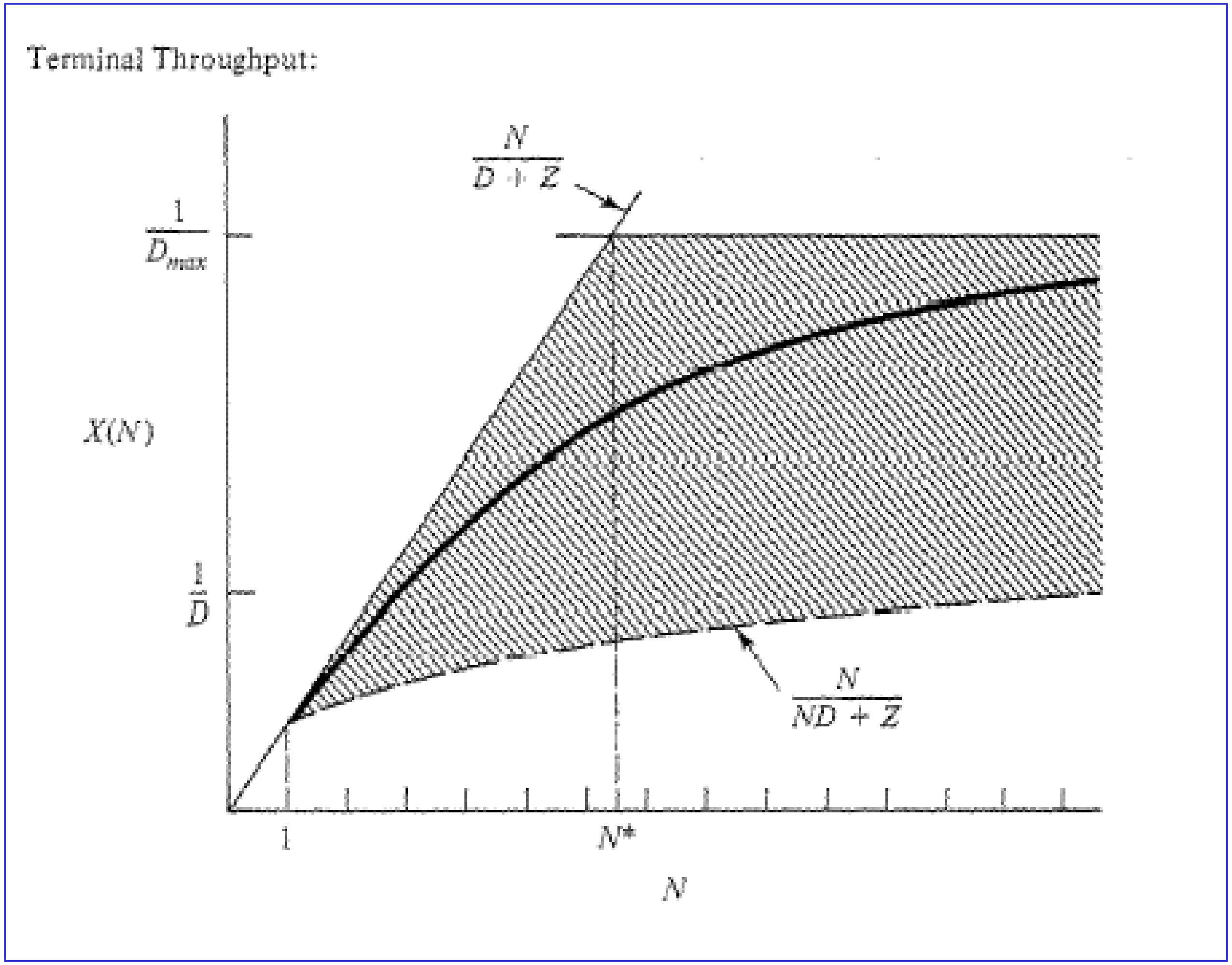
$$\boxed{N^* = (D+Z)/D_b}$$

intersezione delle rette:

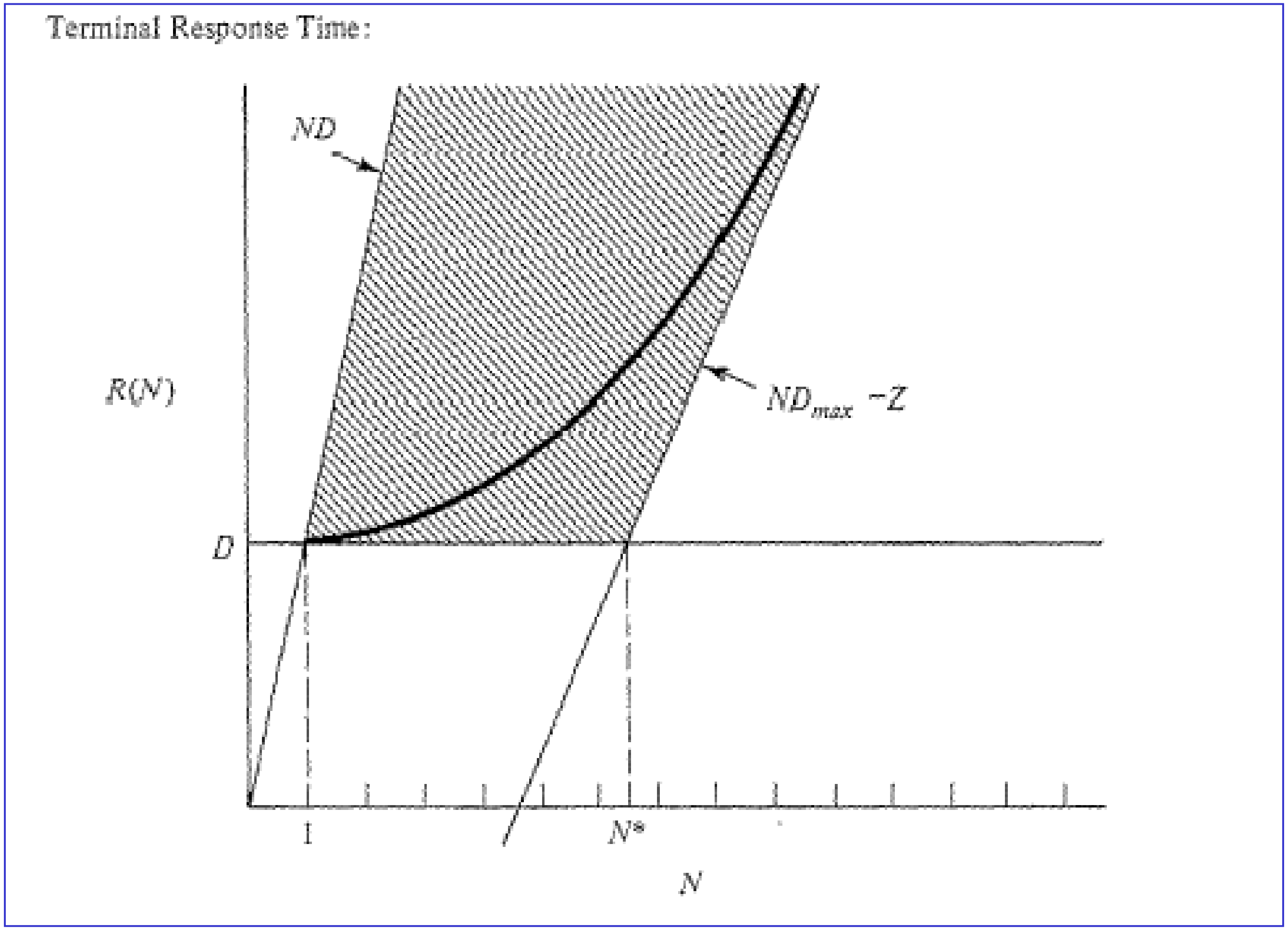
$$\begin{cases} X = N/(D+Z) \\ X = 1/D_b \end{cases}$$

per N>N* si formano sicuramente accodamenti

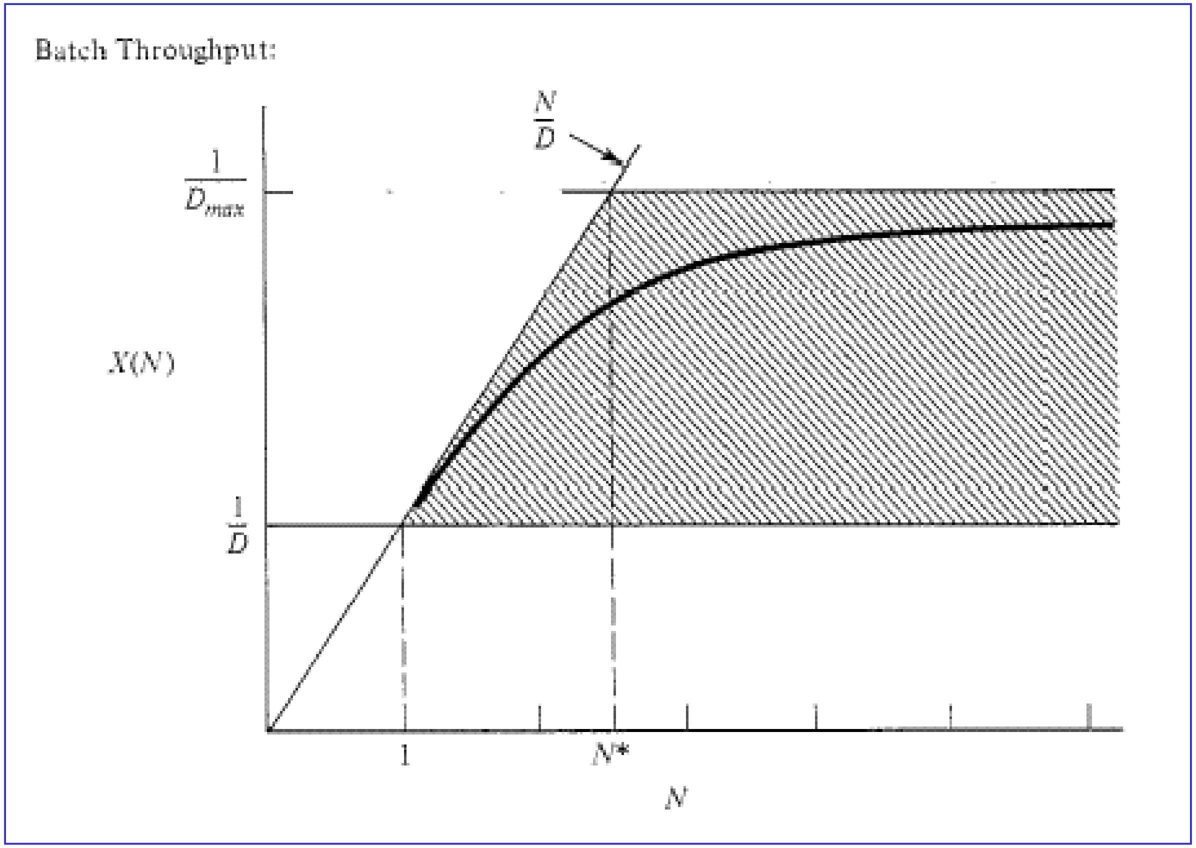
Bounds on Performance (Lazowska et al.)



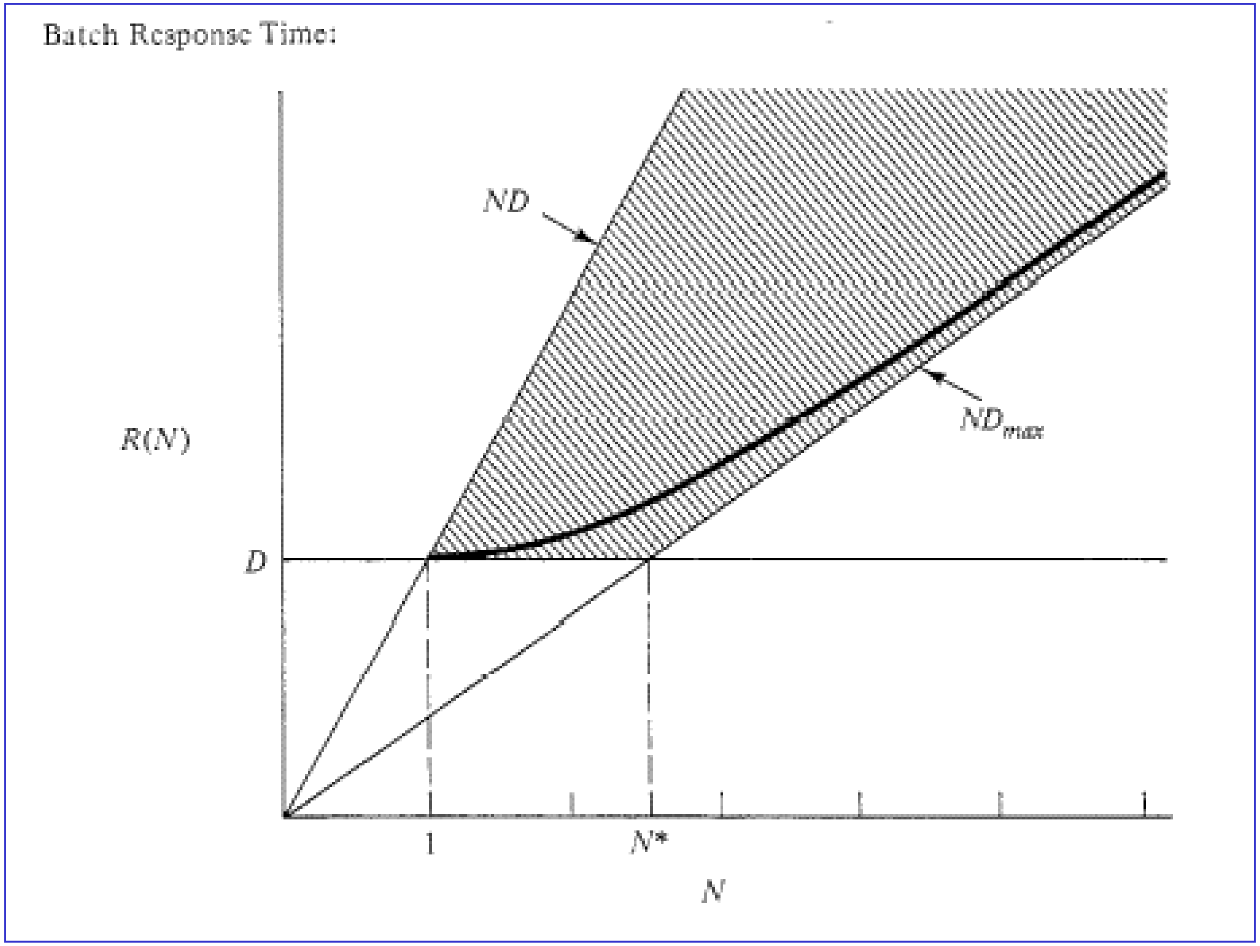
Bounds on Performance (Lazowska et al.) (cont.)



Bounds on Performance (Lazowska et al.) (cont.)



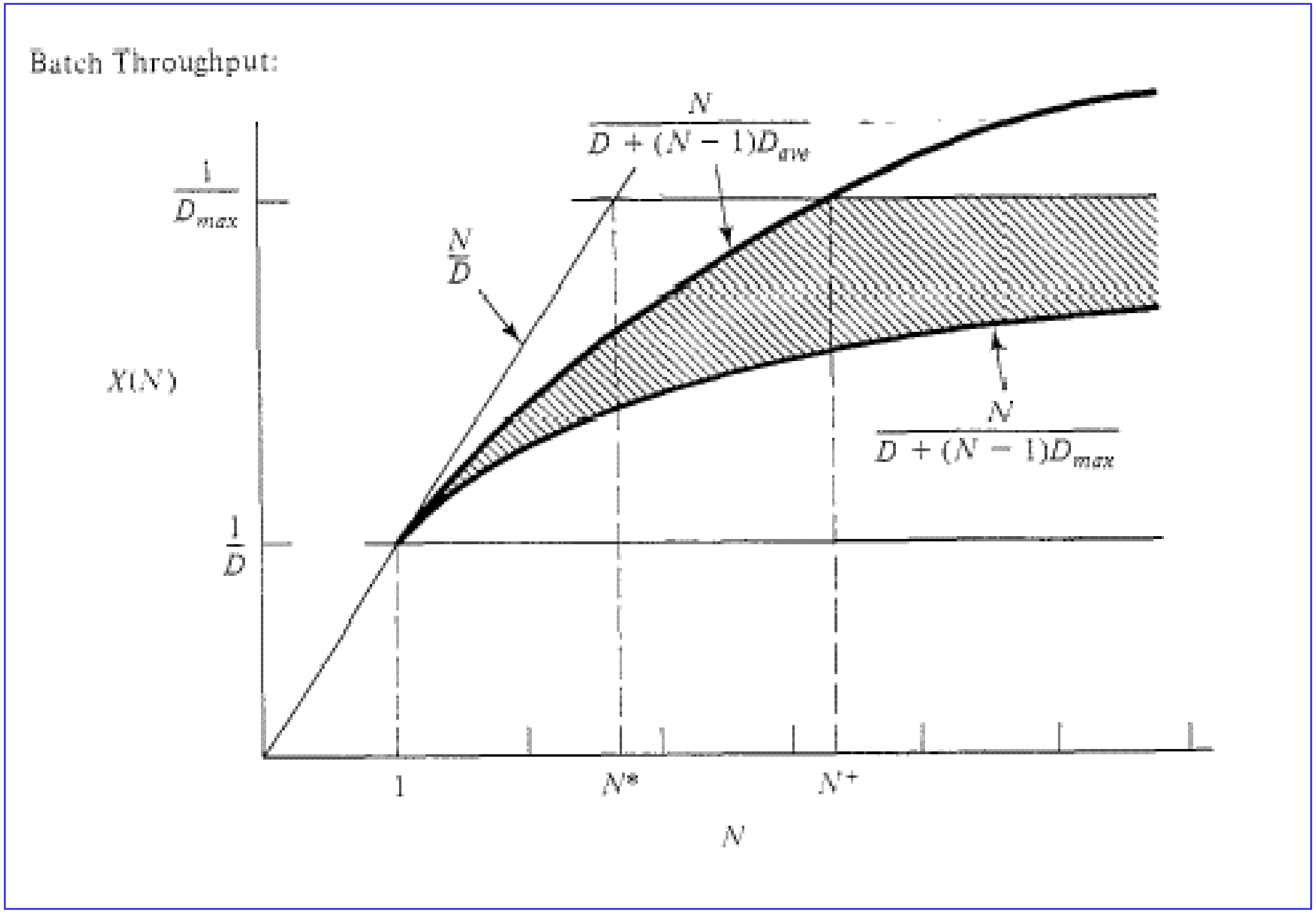
Bounds on Performance (Lazowska et al.) (cont.)



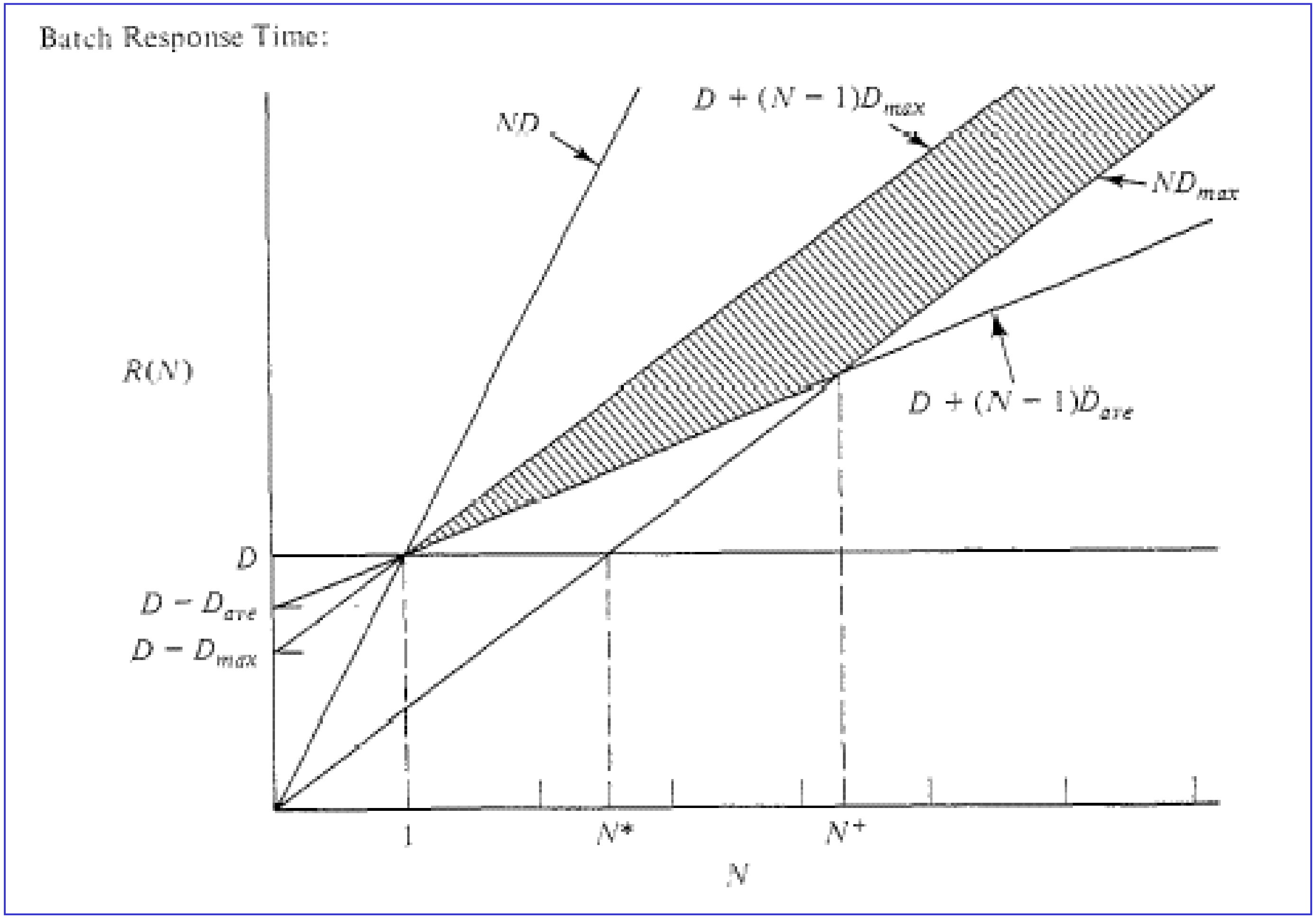
limiti di prestazione: bilanciamento

- l'analisi bilanciata (BJB) parte dalla relazione del valore medio
  - $\boxed{R_i(N) = D_i(1 + Q_i(N-1))}$
- supponiamo che non ci siano nodi *delay* e sommiamo i tempi su tutte le stazioni:
  - $R = CT(N) = D + \Sigma\, D_i\, Q_i(N-1)$
- dato che $D_i \leq D_b$ se tutti i nodi si comportano come quello più carico:
  - $\Rightarrow R(N) \leq D + D_b\, \Sigma\, Q_i(N-1) = D + D_b \times (N-1)$
- se il carico è bilanciato fra i nodi cioè:
  - $D_i = D_a = D/M$ allora:
  - $\Sigma\, D_i\, Q_i(N-1) \geq (1/M)\, \Sigma\, D_i\, \Sigma\, Q_i(N-1) = D_a \times (N-1)$
  - $\Rightarrow R(N) \geq D + D_a \times (N-1)$

Bounds on Performance (Lazowska et al.) (cont.)



Bounds on Performance (Lazowska et al.) (cont.)



limiti di prestazione: tecniche di calcolo BJB

- se ci sono nodi *delay* - la formula va modificata:

  CT(N) = Z + R

  $= Z + D + \Sigma D_i Q_i(N-1)$

non tutti gli altri N-1 clienti vengono incontrati nelle stazioni di servizio in cui si formano code, ma solo una frazione di questi contribuisce al ritardo

massima frazione $\longrightarrow$ $\boxed{\dfrac{ND}{ND+Z} = \dfrac{1}{1+Z/ND}}$

minima frazione $\longrightarrow$ $\boxed{\dfrac{D}{D+Z} = \dfrac{1}{1+Z/D}}$

limiti di prestazione: tecniche di calcolo BJB (cont.)

$$\max\left(ND_b - Z\ ,\ R^-(N)\right) \le R(N) \le R^+(N)$$

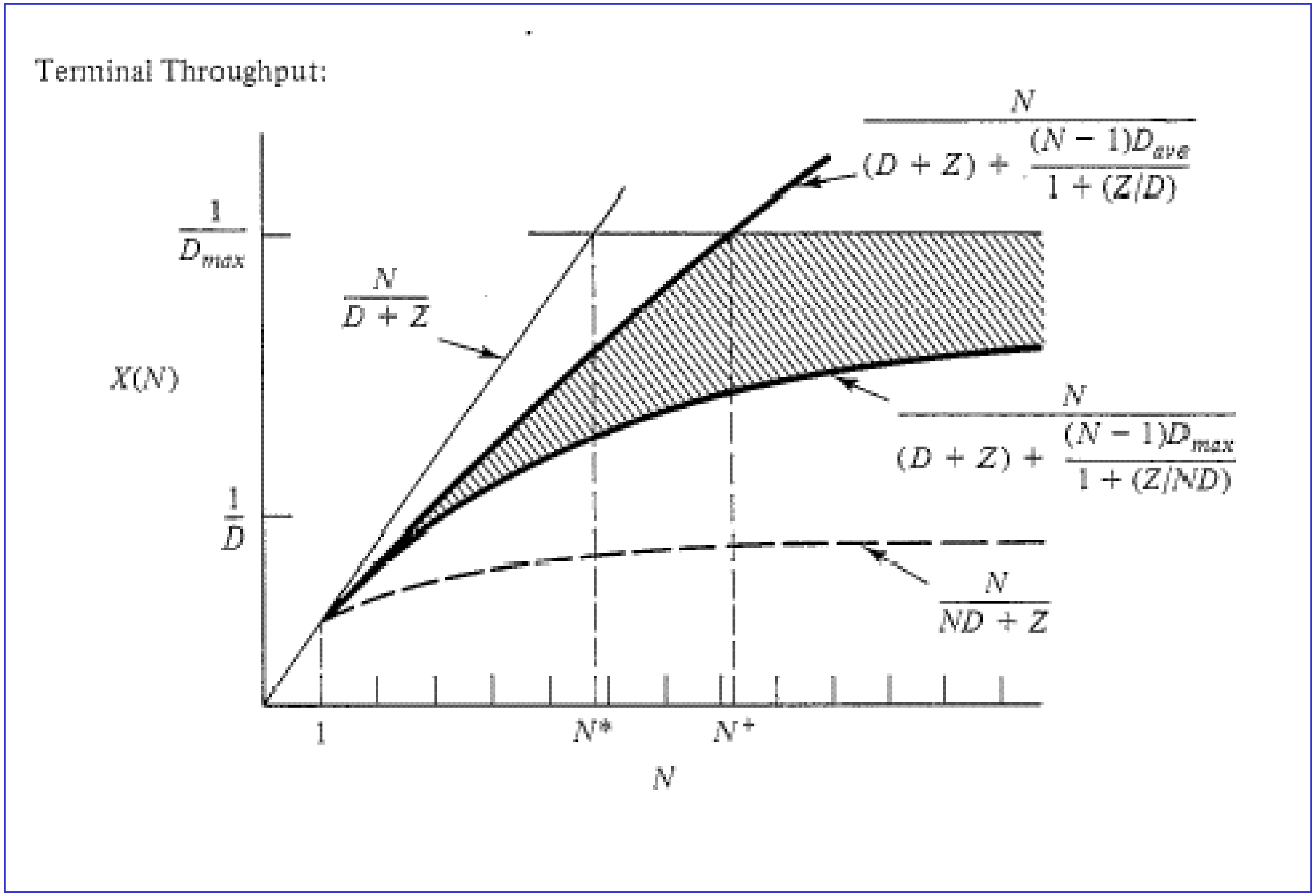
$$\frac{N}{Z + R^+(N)} \le X(N) \le \min\left(\frac{1}{D_b}\ ,\ \frac{N}{Z + R^-(N)}\right)$$

$$\boxed{\begin{aligned} R(N) &\le R^+(N) = D + \frac{D_b \cdot (N-1)}{1 + Z/ND} \\ R(N) &\ge R^-(N) = D + \frac{D_a \cdot (N-1)}{1 + Z/D} \end{aligned}}$$

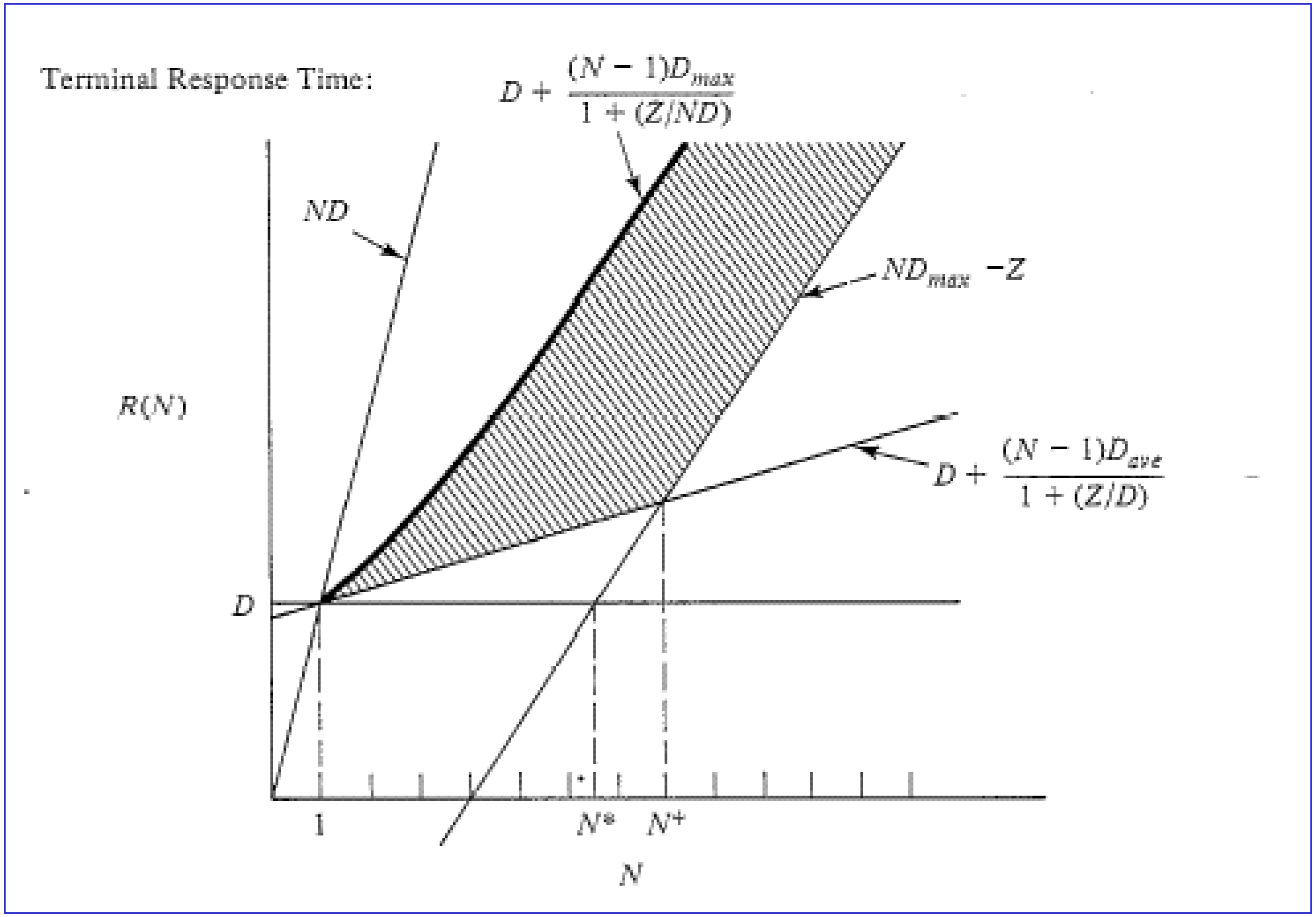
$$\boxed{N^+ = ((D+Z)^2 - DD_a)/((D+Z)D_b - DD_a)}$$

intersezione del limite ottimistico
bilanciato con quello asintotico
$N^+ > N^*$

Bounds on Performance (Lazowska et al.) (cont.)



Bounds on Performance (Lazowska et al.) (cont.)



metodo (esatto) della MVA

- dal teorema del valore medio si ottiene l'algoritmo della *mean value analysis*, vediamone una realizzazione nel caso in cui la rete contenga nodi di ritardo e code con un servente
  - nel primo caso il tempo di permanenza medio del cliente nel nodo k è: D_k
  - nel secondo è: $D_k\ (1 + Q_k^-)$ dove Q_k^- è la lunghezza della coda calcolata con un utente in meno nel sistema
  - alla soluzione con N utenti si perviene aumentando gli utenti di 1 alla volta da 1 a N

n = 0

$$n = 1, \quad Q_k^- = 0, \quad (k = 1, \cdots K) \Longrightarrow$$

valor medio

Little

$$\rightarrow R_k = \begin{cases} D_k \\ D_k \cdot (1 + Q_k^-) \end{cases} \rightarrow X = \frac{n}{\sum R_k} \rightarrow$$

$$n = n + 1 \leftarrow Q_k = X \cdot R_k \leftarrow$$

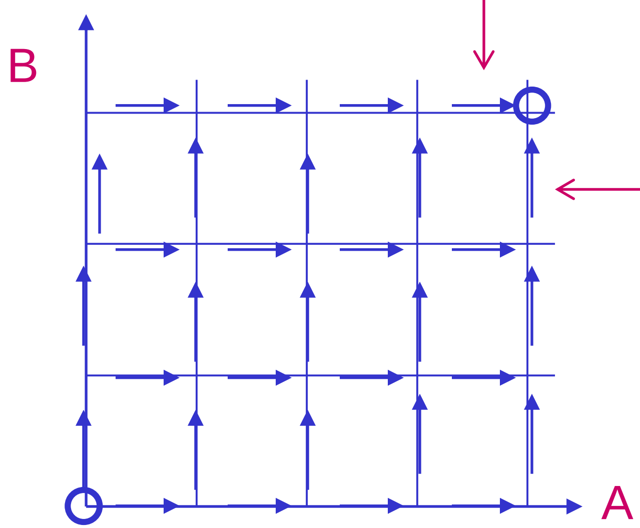
$R(n) = \Sigma\ R_k(n)$
$Q_k(n) = n \times R_k(n)/R(n)$

$\Longleftarrow$

n = N

MVA estensione al caso multiclasse (chiuse)

- il teorema del valore medio assume questa forma:
- $QA_{c,k}(N) = Q_k(N-1_c)$
  - dove il 1mo termine è la coda trovata all'arrivo da un utente appartenente alla classe c - il secondo è la lunghezza media della coda con un utente in meno (della sua classe c)
  - lo schema sottostante mostra il procedimento nel caso di due classi (A di 4 utenti e B di 3) - ogni punto richiede la valutazione di K stazioni



$R_A(4A,3B)$ dipende dallo stato (3A,3B)
$R_B(4A,3B)$ dipende dallo stato (4A,2B)

$N_A = 4$
$N_B = 3$
numero di punti $(N_A + 1)(N_B + 1) = 20$

Complessità: $O\left(K \cdot \prod_{c=1}^{C} N_c\right)$

metodo di valutazione dei sistemi misti

- se ci sono insieme classi aperte e chiuse, l'algoritmo di soluzione si basa sui seguenti punti che permettono di considerare il modello come se avesse alternativamente solo classi di carico chiuse e solo classi aperte:
- gli utenti delle classi chiuse vedono i nodi con una *velocità effettiva* ridotta del fattore: $1\text{-}U_O$ (dove U_O è l'utilizzo da parte delle classi aperte)
- la lunghezza della coda vista dagli utenti delle classi aperte comprende anche:

  Q_C (utenti che appartengono alle classi chiuse)

passi dell'algoritmo (sistema misto)

- per ogni nodo k, calcolo dell'utilizzo da parte delle classi aperte c $\in$ {O}

$$U_{\{O\},k} = \sum_{c \in \{O\}} \lambda_c \cdot D_{c,k} \qquad c \in \{O\}$$

- soluzione del modello con le sole classi chiuse c $\in$ {C} con domande di servizio: $D^*_{c,k}$ (ottenute da quelle originali $D_{c,k}$)

$$D^*_{c,k} = \frac{D_{c,k}}{1 - U_{\{O\},k}} \qquad c \in \{C\}$$

- calcolo dei tempi di residenza delle classi aperte con anche la coda delle classi chiuse ($Q_{\{C\},k}$)

$$R_{c,k} = \frac{D_{c,k}\left(1 + Q_{\{C\},k}\right)}{1 - U_{\{O\},k}} \qquad c \in \{O\}$$

sistema misto (colli di bottiglia)

- in un sistema misto, in cui il carico aperto da solo non saturi alcun nodo, valgono per il carico chiuso tutte le considerazioni fatte sui colli di bottiglia, comportamenti asintotici ecc. con la sostituzione di D(k) con D*(k)
- *gli utilizzi* ***devono*** *invece essere calcolati usando i valori D(k)*

attenzione!

$$U_{\{C\},k} \leq 1 - U_{\{O\},k}$$

in particolare se le classi chiuse si riducono ad una (la classe c)

$$U_{c,k} = X_c D_{c,k}$$

$$\frac{X_c D_{c,k}}{1 - U_{\{O\},k}} = X_c D^*_{c,k} \leq 1$$

$$\max X_c = \frac{1}{\max D^*_{c,k}}$$

tecnica di soluzione approssimata basata su MVA

- per N "grande" si assume valida la relazione seguente:

$$\boxed{\frac{Q_k(N-1)}{N-1} = \frac{Q_k(N)}{N}}$$

- si parte da una stima iniziale delle lunghezze delle code in ogni nodo,
- si calcolano i tempi di residenza e da questi una stima migliore della popolazione
- Il passo precedente viene ripetuto fino a che soluzioni consecutive risultano sufficientemente prossime

tecnica di soluzione approssimata basata su MVA (cont.)

$$Q_{c,k} = \frac{N_c \cdot D_{c,k}}{D_c}$$

- posizione iniziale
- il primo indice (c) si riferisce alla classe di carico
- il secondo (k) al nodo
- se manca è relativo alla somma su tutti i nodi
- se mancano entrambi è relativo alla somma di tutte le classi su tutti i nodi

$$QA_{c,k}\left(\overrightarrow{N}\right) = \frac{N_c - 1}{N_c} Q_{c,k}\left(\overrightarrow{N}\right) + \sum_{\substack{j=1 \\ j \neq c}}^{C} Q_{j,k}\left(\overrightarrow{N}\right)$$

$$R_{c,k}\left(\overrightarrow{N}\right) = \begin{cases} D_{c,k} \\ D_{c,k}\left[1 + QA_{c,k}\left(\overrightarrow{N}\right)\right] \end{cases}$$

$$X_c\left(\overrightarrow{N}\right) = \frac{N_c}{\sum_{1}^{K} R_{c.k}\left(\overrightarrow{N}\right)}$$

$$QA_k(N) = Q_k(N-1) \cong \frac{N-1}{N} Q_k(N)$$

se è presente una sola classe di carico

tecnica di soluzione approssimata basata su MVA (cont.)

$$Q_{c,k}\left(\overrightarrow{N}\right) = X_c\left(\overrightarrow{N}\right) \cdot R_{c,k}\left(\overrightarrow{N}\right)$$

$$Q_{,k}\left(\overrightarrow{N}\right) = \sum_c Q_{c,k}\left(\overrightarrow{N}\right)$$

$$\begin{tikzcd}
\{Q_{c,k}\} \arrow[r] & \{QA_{c,k}\} \arrow[r] & \{R_{c,k}\} \arrow[d] \\
 & & X_c \arrow[llu]
\end{tikzcd}$$

dalla "bontà" della ripartizione iniziale dei $Q_{c,k}$ dipende il numero di iterazioni necessarie per raggiungere l'approssimazione voluta