Appendice B: Reti di code

B.1 INTRODUZIONE ALLE RETI DI CODE

B.1.1 Generalità

La trattazione della teoria delle code effettuata fino ad ora ha sempre considerato singoli sistemi a coda. Tuttavia, molto spesso i sistemi reali si possono presentare sotto forma di più sistemi a coda connessi fra di loro e si parla quindi di *reti di code.*

Le reti di code cono state utilizzate con successo nella modellizzazione di molti sistemi reali: reti di telecomunicazioni, sistemi di calcolo, sistemi manufatturieri. Una rete di code può essere descritta come un grafo orientato composto da un certo numero di nodi (m) ciascuno dei quali rappresenta un sistema a coda e gli archi rappresentano le rotte seguite dagli utenti da un sistema all'altro. In generale, si può pensare che ad una rete di code arrivino utenti dall'esterno in ciascun nodo della rete, così come da ciascun nodo un utente può lasciare la rete. Ovvero, un utente entra nella rete in un certo nodo, attraversa alcuni nodi della rete e poi presso un nodo lascia la rete. Può anche accadere che gli utenti possano ritornare presso nodi già visitati.

È facile immaginare che lo studio di una rete è in generale è molto più complicato dello studio di un singolo sistema a coda perché ovviamente si deve tenere conto di ogni sistema a coda componente la rete per analizzare il flusso totale degli utenti.

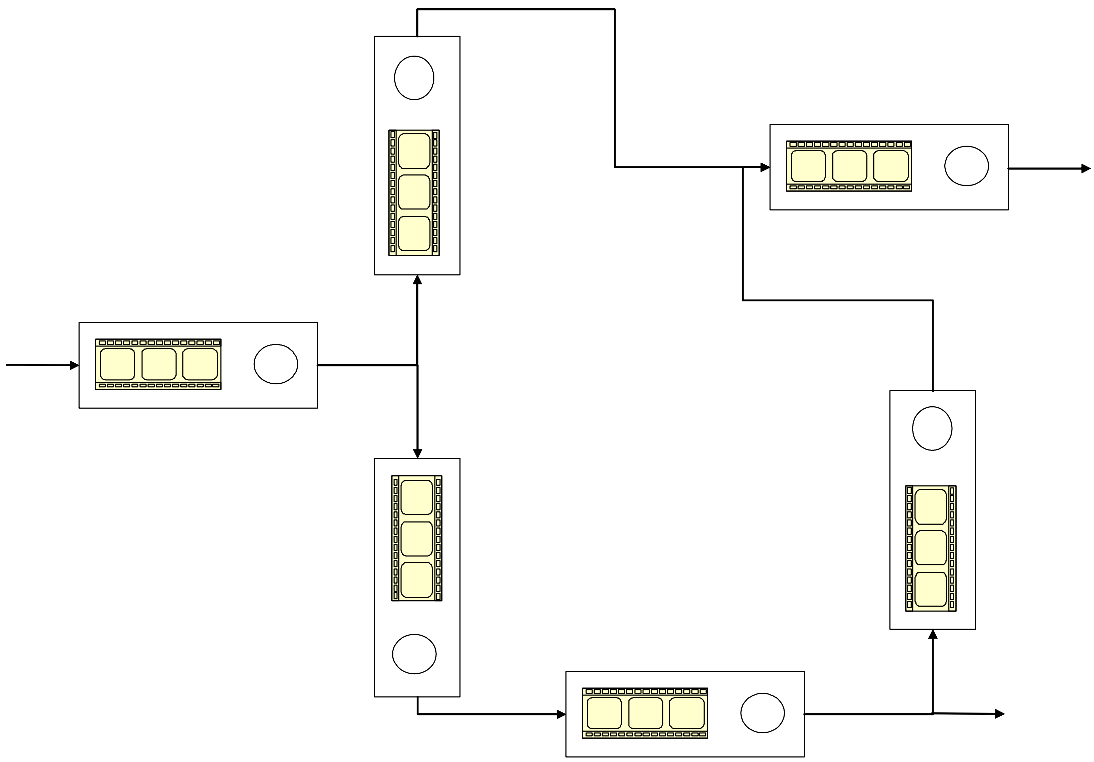

Fig. B.1.1 Esempio di rete aperta

Le reti di code si dividono

- *reti aperte* in cui sono possibili ingressi di utenti alla rete dall'esterno e uscite degli utenti dalla rete verso l'esterno;

- *reti chiuse* in cui il numero degli utenti all'interno della rete è fissato e gli utenti circolano all'interno della rete senza che ci sia possibilità di ingressi dall'esterno o uscite verso l'esterno.

Per caratterizzare completamente una rete di code devono essere assegnati:

- la topologia della rete

- le distribuzioni di probabilità dei tempi di interarrivo degli utenti presso i nodi che prevedono ingresso di utenti

- le distribuzioni di probabilità dei tempi di servizio presso ciascun nodo costituente la rete

- le regole di istradamento.

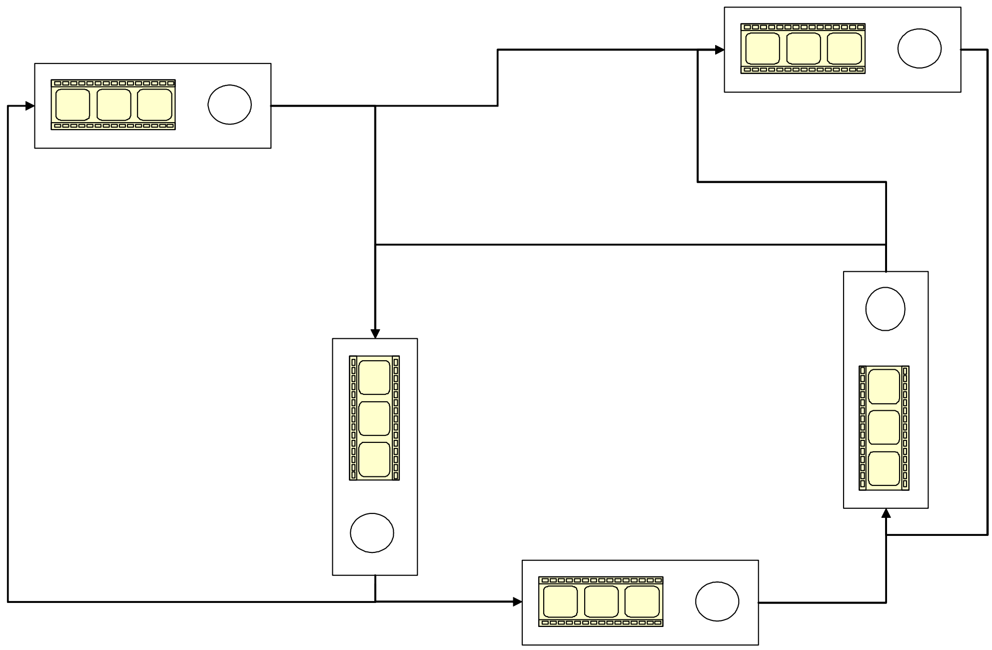

Fig. B.1.2 Esempio di rete chiusa

Lo stato n della rete è definito dal vettore $n = (n_1, \ldots, n_m)$, dove le n_i, $i = 1, \ldots, m$ rappresentano il numero di utenti del sistema $i-$esimo.

B.1.2 Il processo delle partenze per i sistemi M/M/s

Preliminare allo studio delle reti di code c'è l'analisi del processo delle partenze (uscite) dal sistema degli utenti che hanno usufruito del servizio. È molto importante conoscere le caratteristiche di questo processo per studiare le reti di code in quanto il processo delle partenze da un sistema coincide con il processo degli arrivi nel sistema successivo. Riportiamo un risultato che vale per sistemi M/M/s noto come *Teorema di Burke.*

Teorema di Burke

Teorema B.1.1 *Si consideri un sistema M/M/s con $s \geq 1$ (incluso $s = \infty$) con frequenza media di arrivo pari a λ in condizioni di stazionarietà. Allora*

i) il processo della partenze dal sistema è un processo di Poisson di parametro λ;

ii) ad ogni istante di tempo t, il numero di utenti presenti nel sistema è indipendente dalla sequenza dei tempi di partenza prima di t.

Quindi in un sistema a coda con arrivi poissoniani e tempi di servizio distribuiti esponenzialmente, anche il processo delle partenze dal sistema è un processo di Poisson, ovvero il processo delle partenze ha le stesse caratteristiche del processo degli arrivi, nonostante il servizio che è avenuto e il relativo tempo. Si osservi inoltre che questo risultato è indipendente anche dalla disciplina della coda.
L'utilità di questo risultato all'interno della teoria delle code è evidente: se gli utenti che escono dal un sistema M/M/s entrano in un altro sistema a coda, gli arrivi a questo secondo sistema saranno ancora poissoniani. Quindi, assumendo che i tempi di servizio del secondo sistema siano distribuiti esponenzialmente, il sistema a coda del secondo nodo si comporta come un sistema M/M/s e può essere studiato indipendentemente dal primo nodo. Quindi in virtù del teorema di Burke, possiamo collegare in una rete sistemi a coda M/M/s e, purchè non ci siano cicli, ovvero purché gli utenti non possano rivisitare nodi già precedentemente visitati (*reti feedforward*), si può analizzare la rete decomponendola nodo a nodo. L'assenza di cicli è necessaria altrimenti si potrebbe perdere la natura poissoniana dei flussi in ingresso ai nodi. La dimostrazione di questo teorema si basa su una proprietà detta *reversibilità* dei processi di nascita e morte che permette di vedere i processi delle partenze corrispondenti agli arrivi del processo invertito.

reti feedforward

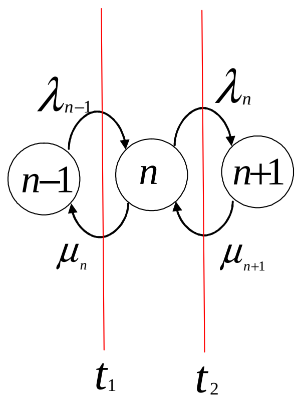

Fig. B.1.3 Diagramma degli stati

Senza entrare nel dettaglio, notiamo solamente che oltre le equazioni di bilancio *globale* date dalle (1.4.6) e (1.4.7), nei processi di nascita e morte, si possono anche considerare le seguenti equazioni di bilancio del flusso tra due stati adiacenti (dette equazioni di bilancio *dettagliato*). Si considerino i due tagli t_1 e t_2 in Figura B.1.3. Relativamente al flusso che attraversa il taglio t_1 si ha

$$\lambda_{n-1}p_{n-1} = \mu_n p_n, \tag{B.1.1}$$

mentre relativamente al taglio t_2 si deve avere

$$\lambda_n p_n = \mu_{n+1}p_{n+1}.$$

Ovviamente, sommando membro a membro queste due equazioni si ottengono le equazioni di bilancio globale (1.4.6). Queste due equazioni di bilancio del flusso ora ottenute tra stati adiacenti mettono in evidenza la proprietà dei processi di nascita e morte che va sotto in nome di *reversibilità* che è alla base del Teorema di Burke.

B.2 SERIE DI CODE

La più semplice rete di code che si può costruire consiste nell'avere un numero fissato (m) di sistemi a coda *in serie*, in cui non ci siano limiti sulla capacità della coda di ogni singolo sistema componente la rete.
Questa situazione si verifica nella pratica, ad esempio, nei sistemi manufatturieri (linee di assemblaggio in cui i prodotti devono passare attraverso una serie di stazioni di lavoro); nei sistemi ospedalieri dove un paziente può essere sottoposto, in sequenza, ad un certo numero di accertamenti; nelle procedure di tipo amministrativo che prevedono più fasi successive (registrazione, pagamento, ritiro documenti, ...). Assumiamo che al primo sistema arrivino utenti secondo un

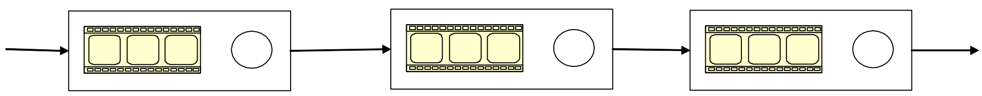

Fig. B.2.1 Esempio di serie di code

processo di Poisson di parametro λ e che in ciascun sistema componente ci siano s_i serventi, $i = 1, \dots, m$, ciascuno operante con tempi di servizio distribuiti esponenzialmente con parametro μ_i con $\lambda < s_i\mu_i$ per ogni $i = 1, \dots, m$. Si assume inoltre che tali tempi di servizio siano indipendenti. In condizioni stazionarie, per il Teorema di Burke, ciascuno dei sistemi ha arrivi poissoniani con parametro λ e quindi i sistemi possono essere analizzati come tanti sistemi M/M/s_i isolati. Il caso più semplice corrisponde ad avere due sistemi in serie (sistema tandem) con singolo servente. In questo caso è molto semplice dimostrare che la probabilità congiunta che n_1 utenti sono nel primo sistema e n_2 utenti nel secondo sistema è data dal prodotto delle singole probabilità, ovvero

$$P(n_1, n_2) = p_{n_1} p_{n_2} = \rho_1^{n_1}(1-\rho_1) \quad \rho_2^{n_2}(1-\rho_2),$$

con $\rho_1 = \lambda/\mu_1 < 1$ e $\rho_2 = \lambda/\mu_2 < 1$.
Infatti, per la prima parte del Teorema di Burke il processo delle partenze dal secondo sistema è di Poisson e per l'indipendenza dei tempi di servizio il secondo

sistema può essere visto come un singolo sistema M/M/1, e quindi si ha

$$p_{n_1} = \rho_1^{n_1}(1-\rho_1), \qquad\qquad p_{n_2} = \rho_2^{n_2}(1-\rho_2).$$

Per la seconda parte del Teorema di Burke, il numero degli utenti presenti nel primo sistema è indipendente dalla sequenza degli arrivi al secondo sistema precedenti e quindi anche dal numero degli utenti nel secondo sistema. Quindi si ha

$$P(n_1, n_2) = \rho_1^{n_1}(1-\rho_1)\ \ \rho_2^{n_2}(1-\rho_2).$$

Questo si estende ad una serie di un numero finito m di sistemi a coda in serie. Quindi la probabilità congiunta che la rete sia allo stato $n = (n_1, \ldots, n_m)$ si può scrivere nella forma

$$P(n) = P(n_1, n_2, \ldots, n_m) = p_{n_1}p_{n_2}\cdots p_{n_m} = \rho_1^{n_1}(1-\rho_1)\rho_2^{n_2}(1-\rho_2)\cdots\rho_m^{n_m}(1-\rho_m),$$

dove $\rho_j = \lambda/\mu_j < 1$, $j = 1, \ldots, m$. Questa rappresenta la cosiddetta *soluzione in forma prodotto.*
Il tempo medio totale di permanenza nell'intero sistema, ovvero nella rete, e il numero medio di utenti presenti nella rete si possono calcolare semplicemente sommando le corrispondenti quantità calcolate in riferimento ai singoli sistemi. Si osservi che le considerazioni fino ad ora riportate non valgono se la capacità dei singoli sistemi a coda componenti la rete fosse finita.

Esempio B.2.1 Supponiamo di avere un sistema formato da due stazioni di lavoro monoserventi in serie: la prima è una stazione di lavorazione, la seconda una stazione di collaudo. I pezzi arrivano alla prima stazione secondo un processo di Poisson di parametro $\lambda = 10$. I tempi di servizio dei serventi sono distribuiti esponenzialmente con $\mu_1 = 12$ e $\mu_2 = 15$.
Per quanto esposto, gli arrivi alla seconda stazione sono poissoniani di parametro $\lambda = 10$ e l'analisi può essere condotta studiando singolarmente i due sistemi M/M/1. Poiché risulta

$$\rho_1 = \frac{5}{6} < 1 \qquad\qquad \rho_2 = \frac{2}{3} < 1$$

esiste una distribuzione stazionaria della rete. Analizzando i due sistemi singolarmente si ha

$$T_1 = \frac{1}{\mu_1 - \lambda} = \frac{1}{2} \ \text{ ora}, \qquad\qquad T_2 = \frac{1}{\mu_2 - \lambda} = \frac{1}{5} \ \text{ ora}$$

e

$$N_1 = \lambda T_1 = 5 \text{ pezzi}, \qquad\qquad N_2 = \lambda T_2 = 2 \text{ pezzi}.$$

Per quanto riguarda la rete si ha $N = N_1 + N_2 = 7$ pezzi e $T = T_1 + T_2 = 7/10$ ora, ovvero 42 minuti.

B.3 RETI DI JACKSON APERTE

Una tipologia di reti di code molto studiate e che continua a prevedere l'utilizzo del modello M/M/s sono le cosiddette *reti di Jackson aperte.* A differenza dei sistemi a coda in serie, gli utenti visitano i nodi in un ordine qualsiasi e in ogni nodo ci possono essere utenti che arrivano sia dall'esterno sia da altri nodi. Formalmente si ha la seguente definizione.

Rete di Jackson

Definizione B.3.1 *Una rete di code aperta si dice* rete di Jackson aperta *se*

i) *gli arrivi dall'esterno della rete ad un nodo i della rete ($i = 1, \dots, m$) sono poissoniani di parametro γ_i*

ii) *i tempi di servizio di ciascun servente degli s_i serventi presenti ad ogni nodo i sono indipendenti e distribuiti esponenzialmente di parametro μ_i*

iii) *le probabilità che un utente che ha completato il servizio al nodo i andrà presso il successivo nodo j ($i, j = 1, \dots, n$) (probabilità di routing) è pari a p_{ij} ed è indipendente dallo stato del sistema*

Assumeremo, inoltre, che i sistemi a coda componenti la rete siano a capacità illimitata (ovvero coda infinita). Naturalmente può accadere che $\gamma_i = 0$ per qualche i, ovvero che non ci siano arrivi dall'esterno al nodo $i-$esimo, ma si richiede che $\gamma_i > 0$ per almeno un i.
Dopo che un utente viene servito presso il nodo i, esso può procedere verso un nodo j con probabilità p_{ij} o può uscire dalla rete con probabilità $1 - \sum_{j=1}^{m} p_{ij}$. Le probabilità p_{ij} possono essere schematizzate in una matrice quadrata di ordine m

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}$$

che viene chiamata *matrice di routing.*

È chiaro che la frequenza media effettiva degli arrivi degli utenti presso un nodo si ottiene sommando gli arrivi dall'esterno del sistema (che sono poissoniani di parametro γ_i) e gli arrivi dai nodi interni alle rete (che non sono necessariamente poissoniani). Ovvero, se indichiamo con λ_j la frequenza media effettiva degli arrivi al nodo j si ha che essa è data da

$$\lambda_j = \gamma_j + \sum_{i=1}^{m} \lambda_i p_{ij} \qquad j = 1, \ldots, m. \qquad \text{(B.3.1)}$$

Questo perché $\lambda_i p_{ij}$ rappresenta il contributo agli arrivi nel nodo j da parte del nodo i. Poiché le γ_i sono assegnate, come anche le p_{ij}, la (B.3.1) rappresenta un sistema lineare di m equazioni in m incognite (le λ_j) che quindi ammette soluzione (unica) se la matrice dei coefficienti è non singolare. Definendo i vettori $\Lambda = (\lambda_1, \ldots, \lambda_m)^T$ e $\Gamma = (\gamma_1, \ldots, \gamma_m)^T$, la (B.3.1) può essere scritta in forma vettoriale

$$\Lambda = \Gamma + P^T \Lambda,$$

ovvero, nell'ipotesi che la matrice $(I - P^T)$ sia invertibile, $\Lambda = (I - P^T)^{-1}\Gamma$, dove P è la matrice di routing e I è l'identità $m \times m$.

Si osservi che l'espressione (B.3.1) per la determinazione della frequenza effettiva vale per una generica rete di code aperta, non solamente per reti di Jackson.

Anche per le reti di Jackson aperte è stato dimostrato che, analogamente al caso delle reti in serie, la probabilità congiunta dello stato n è data dalla produttoria delle probabilità di stato dei singoli sistemi (probabilità marginali). Questo significa che si continua ad avere la soluzione in forma prodotto anche per reti che non sono feedforward come invece era richiesto per l'applicazione del Teorema di Burke. Riportiamo ora in dettaglio questo risultato che va sotto il nome di *Teorema di Jackson.*

B.3.1 Caso monoserevente

Supporremo inizialmente che i singoli sistemi a coda componenti la rete siano monoserventi ($s_i = 1$, $i = 1, \ldots, m$) e che il fattore di utilizzazione del servente sia dato da $\rho_j = \lambda_j / \mu_j$ dove λ_j è la soluzione del sistema (B.3.1), supponendo che $\rho_j < 1$, $j = 1, \ldots, m$.

Teorema di Jackson

Teorema B.3.1 – Teorema di Jackson. *Sia data una rete aperta di Jackson composta da m nodi ciascuno dei quali è a singolo servente. Allora la distribuzione di probabilità congiunta che la rete di trovi allo stato* $n = (n_1, \ldots, n_m)$ *si fattorizza nel prodotto delle distribuzioni di probabulità marginali, ovvero*

$$P(n) = P(n_1, \ldots, n_m) = p_{n_1} p_{n_2} \cdots p_{n_m} \qquad \text{(B.3.2)}$$

dove

$$p_{n_j} = \rho_j^{n_j}(1 - \rho_j).$$

Dimostrazione: Consideriamo inizialmente il solo nodo $j-$esimo della rete, supponendo che negli altri nodi non avvenga alcuna transizione di stato. Relativamente a questo nodo l'equazione di bilancio (B.1.1) nel passaggio dallo stato $n_j - 1$ allo stato n_j oppure dallo stato n_j allo stato n_{j-1}, si può riscrivere

$$\lambda_j p_{n_j - 1} = \mu_j p_{n_j}. \tag{B.3.3}$$

Avendo assunto che negli altri nodi diversi dal $j-$esimo non ci sono transizioni di stato, la (B.3.3) si può riscrivere

$$\lambda_j P(n_1, \ldots, n_j - 1, \ldots, n_m) = \mu_j P(n_1, \ldots, n_j, \ldots, n_m).$$

ovvero

$$P(n_1, \ldots, n_j, \ldots, n_m) = \rho_j P(n_1, \ldots, n_j - 1, \ldots, n_m).$$

Applicando n_j volte questa relazione si ottiene

$$P(n_1, \ldots, n_j, \ldots, n_m) = \rho_j^{n_j} P(n_1, \ldots, n_{j-1}, 0, n_{j+1}, \ldots, n_m).$$

Ripetendo il procedimento per ogni $j = 1, \ldots, m$ si ha

$$P(n_1, \ldots, n_m) = \prod_{j=1}^{m} \rho_j^{n_j} P(0, \ldots, 0). \tag{B.3.4}$$

Rimane da calcolare il valore di $P(0, \ldots, 0)$ che è la probabilità che tutti i nodi siano allo stato zero, ovvero senza alcun utente presente. Per ottenere tale valore è sufficiente sommare tutte le probablità di stato $P(n_1, \ldots, n_m)$ date dalla (B.3.4) e imporre che si ottenga 1, ovvero

$$\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_m=0}^{\infty} \prod_{j=1}^{m} \rho_j^{n_j} P(0, \ldots, 0) = 1$$

da cui, nell'ipotesi che le serie presenti convergano ($\rho_j < 1$, $j = 1, \ldots m$), si ha

$$\begin{aligned}
P(0, \ldots, 0) &= \frac{1}{\displaystyle\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_m=0}^{\infty} \prod_{j=1}^{m} \rho_j^{n_j}} \\
&= \frac{1}{\displaystyle\sum_{n_1=0}^{\infty} \rho_1^{n_1} \sum_{n_2=0}^{\infty} \rho_2^{n_2} \cdots \sum_{n_m=0}^{\infty} \rho_m^{n_m}} \\
&= \frac{1}{\dfrac{1}{1-\rho_1} \dfrac{1}{1-\rho_2} \cdots \dfrac{1}{1-\rho_m}} \\
&= (1-\rho_1)(1-\rho_2) \cdots (1-\rho_m).
\end{aligned}$$

□

Il Teorema di Jackson fornisce la soluzione in forma prodotto in quanto la probabilità che la rete si trovi allo stato $n = (n_1, \dots, n_m)$ risulta pari al prodotto delle proabilità che i singoli nodi si trovino rispettivamente negli stati n_j indipendentemente dagli altri nodi.
In virtù di questo teorema la rete può essere analizzata scomponendola in m sistemi a coda M/M/1 indipendenti, ovvero la rete si comporta come se ciascun nodo fosse un nodo isolato indipendente M/M/1 con frequenza media di arrivo λ_j e tempi medi di servizio pari a μ_j, in quanto le p_{n_j} nella (B.3.2) sono proprio le probabilità che n_j utenti sono presenti nel nodo $j-$esimo considerato come sistema M/M/1. Si ribadisce che questo risultato è vero nonstante il fatto che il flusso entrante ad un generico nodo $j-$esimo non è necessariamente poissoninano di paramero λ_j.
La condizione sotto la quale una rete di code ammette distribuzione stazionaria è che la "capacità del servizio" di ogni singolo nodo sia strettamente maggiore della frequenza media effettiva degli arrivi, ovvero $\rho_j = \lambda_j/\mu_j < 1$, $j = 1, \dots, m$.

Quindi, in virtù del Teorema di Jackson, per studiare una rete di code aperta di Jackson nel caso monoservente è sufficiente determinare le frequenze medie effettive degli arrivi a ciascun nodo j date da λ_j e analizzare indipendente ogni singolo nodo. Quindi si verifica che risulti $\rho_j = \lambda_j/\mu_j < 1$ e si determinano le misure di prestazione ad ogni singolo nodo. Queste verranno poi aggregate opportunamente per avere indicazioni sulla rete. Quindi per ottenere il numero medio di utenti nella rete N, si sommano tutti i valori N_j di ogni singolo nodo, ovvero

$$N = \sum_{j=1}^{m} N_j,$$

dove $N_j = \lambda_j/(\mu_j - \lambda_j)$ è il numero medio di utenti nel nodo $j-$esimo. Per determinare il tempo medio di permanenza nella rete si usa il Teorema di Little e quindi

$$T = \frac{N}{\lambda}, \qquad \text{dove} \qquad \lambda = \sum_{i=1}^{m} \gamma_i.$$

Se inoltre volessimo determinare il tempo medio di permanenza in un nodo j, si deve porre attenzione al fatto che esso, in generale, non è pari a $T_j = 1/(\mu_j - \lambda_j)$, Questo perché le T_j rappresentano il tempo medio di permanenza nel nodo j ogni volta che l'utente viene processato nel nodo $j-$esimo e quindi coincide con il valore che stiamo cercando solo se l'utente visita il nodo j una sola volta,

Visit count

altrimenti andrà moltiplicato per il valore atteso del numero delle visite di un utente al nodo j (visit count) che indichiamo con ν_j. Il calcolo di tale numero è molto semplice: si consideri, infatti, il rapporto $\gamma_j / \sum_{i=1}^{m} \gamma_i = \gamma_j/\lambda$ che è la frazione di utenti che, proveniendo dall'esterno, visitano come primo nodo il nodo

$j-$esimo e quindi rappresenta la probabilità che la prima visita di un utente sia al nodo $j-$esimo. Inoltre una frazione p_{ij} di utenti visitano il nodo $j-$esimo provenendo dal nodo $i-$esimo. Quindi si ha

$$\nu_j = \frac{\gamma_j}{\lambda} + \sum_{i=1}^{m} p_{ij}\nu_i$$

che ammette come soluzione (confrontare la (B.3.1))

$$\nu_j = \frac{\lambda_j}{\lambda}. \tag{B.3.5}$$

Note le μ_j, $j = 1, \ldots, m$, alternativamente si può calcolare T nella forma[1]

$$T = \sum_{j=1}^{m} \nu_j T_j.$$

Esempio B.3.2 Si consideri una semplice rete di code formata da due nodi: il primo nodo è una stazione di lavorazione, il secondo è una stazione di ispezione–collaudo. I pezzi arrivano dall'esterno alla stazione 1 secondo un processo di Poisson con media 10 pezzi l'ora e successivamente alla lavorazione nella stazione 1 procedono nella stazione 2 per il controllo. Dalla stazione 2 si ha che il 10% dei pezzi collaudati risultano difettosi e tornano alla stazione 1 per essere lavorati di nuovo. Le due stazioni sono monoserventi con tempi di servizio distributi esponenzialmente rispettivamente con media 12 pezzi l'ora e 15 pezzi l'ora. Determinare il numero medio di utenti presenti nella rete e il tempo medio di permanenza.

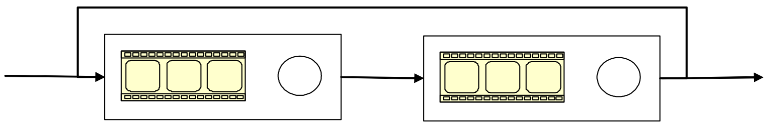

Fig. B.3.1 Rete dell'Esempio B.3.2

Calcoliamo innanzitutto le frequenze medie effettive (λ_1 e λ_2). Dalla (B.3.1) si ha il seguente sistema

$$\begin{cases} \lambda_1 = 10 + 0.1\lambda_1 \\ \lambda_2 = \lambda_1 \end{cases}$$

dal quale si ricava

$$\lambda_1 = \lambda_2 = 11.111.$$

Le condizioni per l'esistenza della distribuzione stazionaria sono ovviamente verificate in quanto risulta $\rho_1 = \lambda_1/\mu_1 < 1$ e $\rho_2 = \lambda_2/\mu_2 < 1$. Si ricavano

$$N_1 = \frac{\rho_1}{1-\rho_1} = 12.5, \qquad N_2 = \frac{\rho_2}{1-\rho_2} = 2.85, \qquad N = N_1 + N_2 = 15.35 \quad \text{pezzi},$$

[1] Nell'ambito dei sistemi manufatturieri i nodi di una rete di code vengono, di solito chiamati "*stazioni*" e T viene chiamato "*throughput time*".

da cui $T = N/\lambda = 1.535$ ore.
Per quanto riguarda il valore atteso del numero delle visite risulta

$$\nu_1 = \frac{\lambda_1}{\lambda} = 1.111, \qquad \nu_2 = \frac{\lambda_2}{\lambda} = 1.111.$$

Si ha inoltre

$$T_1 = \frac{1}{\mu_1 - \lambda_1} = 1.125 \quad \text{ore}, \qquad T_2 = \frac{1}{\mu_2 - \lambda_2} = 0.257 \quad \text{ore},$$

e quidi alternativamente si può calcolare

$$T = 1.111 \cdot 1.125 + 1.111 \cdot 0.257 = 1.535 \quad \text{ore}.$$

B.3.2 Caso multiservente

Il Teorema di Jackson enunciato per reti di code costituite da sistemi con singolo servente si estende anche al caso in cui ogni nodo della rete può avere più di un servente. Supponiamo, quindi, che $s_j \geq 1$ sia il numero dei serventi per ciascun nodo $j = 1, \ldots, m$, ed inoltre assumiamo che $\rho_j = \lambda_j/(s_j\mu_j) < 1$. Allora vale il seguente risultato.

Teorema B.3.2 – Tereoma di Jackson. *La distribuzione di probabilità congiunta che la rete si trovi allo stato $n = (n_1, \ldots, n_m)$ si fattorizza nel prodotto di tutte le distribuzioni marginali, ovvero*

$$P(n) = P(n_1, \ldots, n_m) = p_{n_1} p_{n_2} \cdots p_{n_m},$$

dove

$$p_{n_j} = \begin{cases} \dfrac{1}{n_j!}\left(\dfrac{\lambda_j}{\mu_j}\right)^{n_j}, & \text{per} \quad n_j < s_j \\ \dfrac{1}{s!s^{n_j - s_j}}\left(\dfrac{\lambda_j}{\mu_j}\right)^{n_j}, & \text{per} \quad n_j \geq s_j \end{cases}$$

Di questo teorema non forniamo per brevità la dimostrazione.

Questo teorema afferma che, in maniera del tutto analoga al caso monoservente, la probabilità congiunta si fattorizza nelle probabilità di avere n_j utenti nel nodo $j-$esimo considerato singolarmente come sistema M/M/s_j con frequenza media degli arrivi pari alla frequenza effettiva λ_j e tempo medio di servizio μ_j e questo indipendentemente dagli altri nodi della rete.
Come nel caso monoservente, le misure di prestazione di interesse possono essere calcolate aggregando i valori ottenuti nei singoli nodi.

Esempio B.3.3 [Gross, Harris, 1998] Si consideri un call center al quale arrivano chiamate secondo la distribuzione di Poisson con media 35 l'ora. Alle chiamate che arrivano il gestore fornisce due opzioni: digitare il tasto 1 per il servizio reclami, oppure digitare 2 per il servizio informazioni. Si stima che il tempo di ascolto del messsaggio e della pressione del tasto sia esponenziale con media 30 secondi. Le chiamate che trovassero occupato vengono poste in attesa con l'assunzione che nessun utente si scoraggia per l'attesa e quindi aspetta comunque di usufruire del servizio. Il 55% delle chiamate chiedono di accedere al servizio reclami e le rimanenti chiedono di accedere al servizio informazioni. Il nodo del processo dei reclami ha 3 serventi (operanti in parallelo) che operano con tempi di servizio distribuiti esponenzialmente con media 6 minuti. Il nodo del processo delle informazioni ha 7 serventi (operanti in parallelo) che operano con tempi di servizio distribuiti esponenzialmente con media 20 minuti. I buffer di attesa si assumono illimitati. Inoltre circa il 2% dei clienti che hanno usufrutio del servizio reclami decidono di usufruire anche del servizio informazioni e l'1% dei clienti che hanno usufruito del servizio informazioni chiedono anche di usufruire del servizio reclami. Si vuole determinare la lunghezza media della coda in ciascun nodo e il tempo medio totale che un cliente passa nella in linea con il call center.

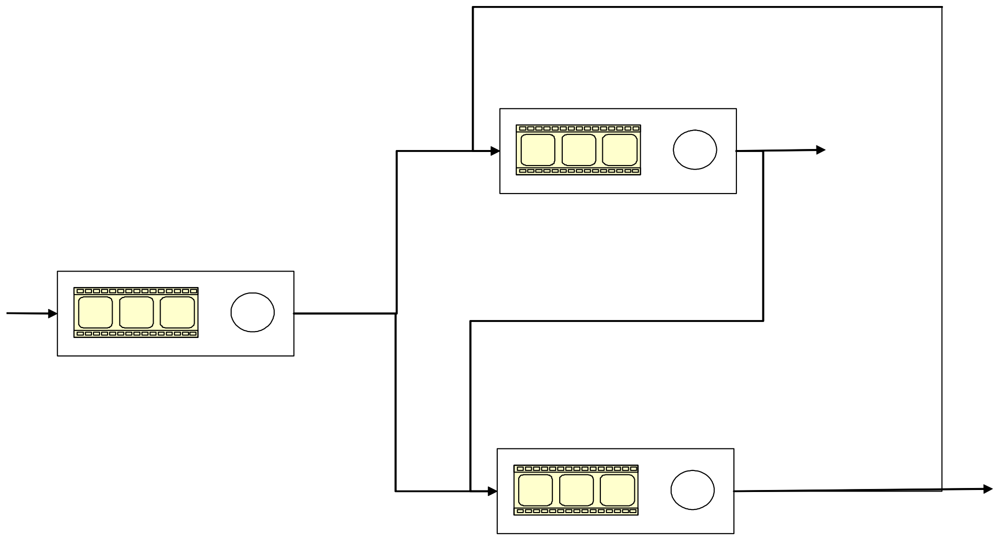

Fig. B.3.2 Rete dell'Esempio B.3.3

Risulta $s_1 = 1$, $s_2 = 3$, $s_3 = 7$ e $\mu_1 = 120$, $\mu_2 = 10$, $\mu_3 = 3$. Inoltre si ha $\gamma_1 = 35$ e $\gamma_2 = \gamma_3 = 0$. La matrice di routing è

$$\begin{pmatrix} 0 & 0.55 & 0.45 \\ 0 & 0 & 0.02 \\ 0 & 0.01 & 0 \end{pmatrix}$$

Calcoliamo la frequenza effettiva utilizzando il sistema (B.3.1):

$$\begin{cases} \lambda_1 = 35 \\ \lambda 2 = 35 \cdot 0.55 + \lambda_3 \cdot 0.01 \\ \lambda_3 = 35 \cdot 0.45 + \lambda_2 \cdot 0.02 \end{cases}$$

dal quale si ricavano

$$\lambda_1 = 35, \qquad \lambda_2 = 19,411, \qquad \lambda_3 = 16.138.$$

Si verifica immediatamente che risulta $\rho_j = \lambda_j/(s_j\mu_j) < 1$ per $j = 1, 2, 3$ e quindi esiste una distribuzione stazionaria per la rete in esame. A questo punto è sufficiente ricavare

- per il primo sistema, dal modello M/M/1 con $\lambda_1 = 35$ e $\mu_1 = 120$, $N_1^q = 0$ e $N_1 = 0.412$;

- per il secondo sistema, dal modello M/M/3 con $\lambda_2 = 19,411$ e $\mu_2 = 10$, $N_2^q = 0.765$ e $N_2 = 2.706$;
- per il terzo sistema, dal modello M/M/7 con $\lambda_3 = 16.138$ e $\mu_3 = 3$, $N_3^q = 1.402$ e $N_3 = 6.781$.

Si ottiene immediatamente $N = N_1 + N_2 + N_3 = 9.899$ e quindi

$$T = \frac{N}{\lambda} = \frac{N}{\gamma_1 + \gamma_2 + \gamma_3} = \frac{9.899}{35} = 0.283 \quad \text{ore},$$

ovvero $T = 17$ minuti circa.

B.4 RETI DI JACKSON CHIUSE

Si tratta di reti di code in cui non sono consentiti né arrivi, né partenze di utenti dalla rete e quindi il numero totale di utenti presenti nella rete risulta fissato. Quindi in una rete chiusa il valore di N non è più un valore medio da determinare, ma è un dato del problema.
Anche se il modello che si utilizza non prevede arrivi o partenze dalla rete, da un punto di vista pratico esso rappresenta bene quei casi in cui ogniqualvolta ci sia un utente che esce dalla rete, esso viene immediatamente rimpiazzato da un nuovo utente.
Questa tipologia di rete di code è stata introdotta da Gordon e Newell nel 1967 e ha applicazioni significative nei sistemi di calcolo time-sharing e multi-utente.
Ovviamente, in una rete chiusa il numero degli stati possibili è finito ed è uguale al numero dei modi che si hanno di disporre N utenti (indistinguibili tra di loro) negli m nodi, considerando che ogni nodo può avere al più $N + 1$ stati possibili (includendo lo stato 0). Tale numero è dato da

$$\binom{m + N - 1}{m - 1}.$$

Inoltre, poichè in una rete chiusa il numero degli utenti è limitato, una rete chiusa ammette sempre una distribuzione stazionaria.
Nell'analizzare questo tipo di reti di code, c'è un'importante differenza rispetto alle reti aperte, che sta nella determinazione della frequenza media effettiva di arrivo ad ogni singolo nodo. Infatti, poiché risulta $\gamma_j = 0$ per $j = 1, \dots, m$, il sistema di equazioni (B.3.1) diventa

$$\lambda_j = \sum_{i=1}^{m} \lambda_i p_{ij}, \qquad j = 1, \dots, m \tag{B.4.1}$$

che quindi ora è un *sistema omogeneo* che in forma matriciale si scrive

$$(I - P^T)\Lambda = 0,$$

dove $\Lambda = (\lambda_1, \ldots, \lambda_m)^T$ e P è la matrice di routing. Ora, poiché nessun utente può lasciare la rete deve risultare

$$\sum_{j=1}^{m} p_{ij} = 1,$$

e quindi la matrice $(I-P)$ non è a rango pieno e il sistema ha almeno un'equazione ridondante, ovvero le m equazioni del sistema non sono indipendenti. Supponendo che il rango della matrice $(I-P)$ sia pari a $m-1$, il sistema ammette infinite soluzioni. Quindi usando le $m-1$ equazioni indipendenti del sistema si possono determinare le λ_j, $j = 1, \ldots, m$ a meno di una costante moltiplicativa.
Siano $\{\bar{\lambda}_j \ , \ j = 1, \ldots, m\}$ una soluzione positiva del sistema (B.4.1) e μ_j la velocità media di servizio nel $j-$esimo nodo.
Esaminiamo inizialmente il caso monoservente, supponendo quindi che il sistema presente in ciascun nodo sia monoservente e sia $\rho_j = \bar{\lambda}_j/\mu_j$. Possiamo ora enunciare l'equivalente del Teorema di Jackson per reti chiuse.

Teorema B.4.1 *Sia data una rete di Jackson chiusa composta da m nodi a singolo sevente. Allora la distribuzione congiunta è data da*

$$P(n) = P(n_1, \ldots, n_m) = \frac{1}{G(N)} \rho_1^{n_1} \rho_2^{n_2} \cdots \rho_m^{n_m},$$

con $n_1 + n_2 + \cdots + n_m = N$ e dove

$$G(N) = \sum_{\substack{n_1, n_2, \ldots, n_m \\ n_1 + n_2 + \cdots n_m = N}} \rho_1^{n_1} \rho_2^{n_2} \cdots \rho_m^{n_m}.$$

Dimostrazione: La dimostrazione è analoga a quella del Teorema di Jackson per reti aperte e, in questo caso, si ottiene

$$P(n_1, \ldots, n_m) = \prod_{j=1}^{m} \rho_j^{n_j} C,$$

dove C è una costante che possiamo determinare imponendo che

$$\sum_{\substack{n_1, n_2, \ldots, n_m \\ n_1 + n_2 + \cdots n_m = N}} C \prod_{j=1}^{m} \rho_j^{n_j} = 1$$

dalla quale si ottiene

$$C = \frac{1}{\displaystyle\sum_{\substack{n_1, n_2, \ldots, n_m \\ n_1 + n_2 + \cdots n_m = N}} \rho_1^{n_1} \rho_2^{n_2} \cdots \rho_m^{n_m}}.$$

□

Il teorema si estende anche al caso multiservente. Supponiamo quindi che in ogni nodo siano presenti s_j serventi e si introduca la quantità

$$X_j = \frac{\nu_j}{\mu_j}$$

che rappresenta il tempo medio che un utente passa in servizio presso il nodo $j-$esimo e dove ν_j è il valore atteso del numero delle visite al nodo $j-$esimo che soddisfa l'equazione

$$\nu_j = \sum_{i=1}^{m} p_{ij}\nu_j.$$

Allora vale il seguente teorema.

Teorema B.4.2 – Teorema di Gordon–Newell. *La distribuzione congiunta che la rete si trova nello stato $n = (n_1, \ldots, n_m)$ è data da*

$$P(n) = P(n_1, \ldots, n_m) = \frac{1}{G(N)} \prod_{j=1}^{m} \hat{p}_{n_j},$$

dove

$$\hat{p}_{n_j} = \begin{cases} \dfrac{X_j^{n_j}}{n_j!}, & \text{per } \; n_j < s_j \\ \dfrac{X_j^{n_j}}{s_j! s_j^{n_j - s_j}}, & \text{per } \; n_j \geq s_j, \end{cases}$$

dove

$$G(N) = \sum_{\substack{n_1, n_2, \ldots, n_m \\ n_1 + n_2 + \cdots n_m = N}} \hat{p}_{n_1} \hat{p}_{n_2} \cdots \hat{p}_{n_m}.$$

È evidente l'analogia formale tra il Teorema di Jackson per le reti aperte e il Teorema di Gordon–Newell. Il coefficiente $G(N)$ è solamente un fattore di normalizzazione che garantisce che la somma di tutte le probabilità sia pari a 1. La determinazione di tale fattore di normalizzazione $G(N)$ può essere non banale per il fatto che la somma che lo definisce è estesa a tutti gli stati possibili della rete e tale numero può risultare molto elevato. Per ovviare a tale inconveniente sono stati definiti algoritmi ricorsivi la cui trattazione esula dallo scopo di queste note.