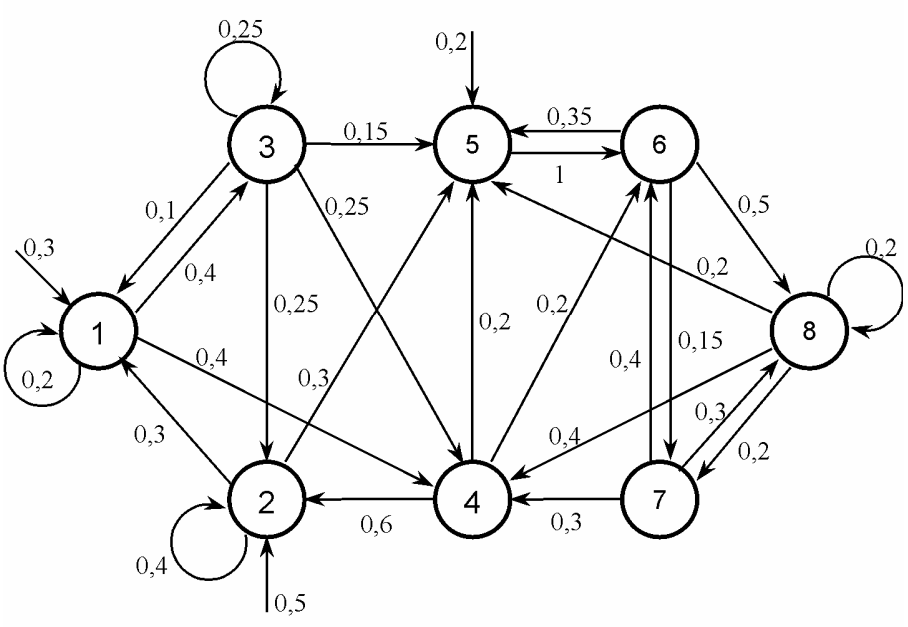
**UNIVERSITATEA TEHNICĂ A MOLDOVEI**

**Emilian GUȚULEAC**

# Lanțuri şi sisteme de aşteptare markoviene: Elemente teoretice şi aplicații




**Chişinău**
**2010**

**UNIVERSITATEA TEHNICĂ A MOLDOVEI**

**Catedra Calculatoare**

# Lanțuri şi sisteme de aşteptare markoviene: Elemente teoretice şi aplicații

## Ciclu de prelegeri

**Chişinău**
**U.T.M.**
**2010**

Prezentul ciclu de prelegeri la disciplina **"Procese stochastice"** este destinat studenţilor din anul I cu specializările 526.1 "*Calculatoare*" şi 526.2 "*Tehnologii informaţionale*", **Facultatea Calculatoare, Informatică şi Microelectronică**.
Tematica prelegerilor a fost stabilită în conformitate cu programa de învăţământ. Sunt prezentate unele consideraţii teoretice ale lanţurilor şi sistemelor de aşteptare markoviene şi semi-Markov, metode de analiză numerică a proprietăţii lor şi aspecte aplicative ale teoriei fenomenelor de aşteptare.
Scopul lucrării constă în familiarizarea studenţilor cu metodele de analiză a lanţurilor Markov şi a sistemelor de aşteptare care pot fi aplicate la modelarea şi evaluarea performanţelor calculatoarelor, sistemelor şi reţelelor de calculatoare. Pentru atingerea obiectivului respectiv poate fi utilizat pachetul de programe *QM* şi mediul de modelare *VPNP*.

**Elaborare**: conf. univ., dr. hab. Emilian GUŢULEAC

**Recenzent**: conf. univ., dr. . Sergiu ZAPOROJAN

**Aprobat la şedinţa Consiliului Ştiinţific al Facultăţii Calculatoare, Informatică şi Microelectronică din 20 octombrie 2009**

**Redactor responsabil**: conf. univ., dr. Victor ABABII

**Procesare computerizată:** conf. univ., dr. hab. Emilian GUŢULEAC

# Prefață

Sistemele de calcul, rețelele de calculatoare sau de telecomunicații sunt sisteme complexe formate dintr-o multitudine de sisteme elementare de tipuri diferite, interconectate după o structură convenabilă, având caracteristici proprii ce decurg atât din arhitectura lor fizică, precum şi din natura proceselor la care sunt supuse.

Teoria proceselor stochastice markoviene şi semi-Markov reprezintă un domeniu relevant în asamblul matematicilor aplicate, care necesită rezolvarea problemelor practice de modelare şi evaluare a performanțelor sistemelor de calcul cu stări şi evenimente discrete. Actualmente teoria proceselor stochastice ocupă o arie atât de mare încât este puțin probabil de a o percepe integral, ținând contul în mod deosebit de faptul că această teorie este în continuă dezvoltare.

Metodele *fenomenelor* de *aşteptare* descriu *sisteme* şi *procese* de *servire* cu caracter de masă care intervin în diferite domenii ale activității practice.

Teoria lanțurilor Markov şi a sistemelor de aşteptare este acea ramură a matematicii ce studiază fenomenele de aşteptare, principalele elemente ale căreia sunt: *sursa* - mulțimea unităților (cererilor, clienților) ce solicită un serviciu la un moment dat, care poate fi finită sau infinită; *sosirea* unităților în sistemul de aşteptare determină o variabilă aleatoare, care reprezintă numărul de unități ce intră în sistem în unitatea de timp. Este necesar să se cunoască repartiția acestei variabile aleatoare.

La originea *teoriei aşteptării* se găseşte *determinarea "încărcării" optime* a unei server. Pentru a rezolva această problemă, este necesar să se determine *cererile de servicii* (apelurile) care sosesc în mod întâmplător şi să se înregistreze timpul necesar pentru prelucrarea acestora. Un astfel de model în care se urmăreşte satisfacerea cât mai promptă a cererilor de servicii în condiții economice cât mai avantajoase se numeşte *model* (*sistem*) *de aşteptare* (*servire*).

Încercările de a prezenta esența şi materia respectivă a acestor teorii într-un volum relativ mic sunt supuse întru totul gusturilor, preferințelor autorilor şi programei de

învățământ a disciplinii predate. Astfel şi noi am fost forțați să selectăm anumite elemente de considerații teoretice din această teorie pentru a le aduce la cunoştință studenților înainte de a efectua anumite aplicații practice prevăzute în formă de lucrari de laborator.

Fiind, în general, subordonate unor anumite programe analitice, noțiunile şi conceptele prezentate în acest volum apar, în mod firesc, într-o succesiune logică şi sunt supuse unor restricții temporale şi de spațiu inevitabile care conduc adeseori la dezvoltări teoretice limitate.

Vom mulțumi anticipat acelora, care vor dori să facă observații constructive asupra prezentei lucrări şi vor manifesta înțelegere pentru eventualele abateri remarcate în text, formule sau figuri.

# Cuprins

# 1. Procese stochastice şi lanțuri Markov

## 1.1. Noțiuni şi definiții generale

Procesele stochastice permit modelarea matematică a numeroaselor componente ale sistemelor tehnice, informatice, economice, sociale etc.

În cele ce urmează vom reda succint principalele definiții şi proprietăți ale proceselor stochastice şi ale lanțurilor Markov ( *LM* ). Pentru o prezentare mai detaliată a noțiunilor redate succint se pot consulta lucrările [1,5,9,17].

**Definiția 1.1.** *Un proces stochastic X* este o familie de variabile aleatoare $(X_\tau)_{\tau\in\tau}$ definite pe acelaşi spațiu de probabilitate cu valori reale în acelaşi spațiu de valori $\Omega$ şi indexate după un parametru $\tau\in\tau\subseteq R$ . □

Un proces stochastic se reprezintă prin:

$$\{X_\tau\in\Omega,\ \tau\in\tau\ \} \tag{1.1}$$

De obicei precizarea mulțimii $\tau$ coincide cu intervalul de timp al evoluției diverselor clase de procese stochastice. Astfel, dacă $\tau=\{\tau_1,\tau_2,...,\tau_n\}$ este o mulțime finită, atunci procesul stochastic este echivalent cu un vector aleator , care determină vectorul de stare al sistemului studiat.

În termeni probabilistici, a descrie evoluția unui proces stochastic înseamnă cunoaşterea probabilităților tuturor evenimentelor de forma : " *la momentul $\tau$ procesul stochastic se găseşte în starea ($X_\tau=x$)* ", precum şi a probabilităților de realizare simultană a unui număr de astfel de evenimente pentru diverse momente $\tau_i\in\tau$ şi diverse mulțimi $e_i\subseteq R$, $1\le i\le n$ . Cu alte cuvinte, este necesar să fie cunoscute probabilitățile de forma :

$$Pr(X_{\tau_1}\in e_1,\ldots,X_{\tau_n}\in e_n) \tag{1.2}$$

pentru orice $n \in N$, orice $\tau_i \in \tau$ şi orice $e_i \subseteq R$, $1 \leq i \leq n$ . Acest fapt se manifestă prin cunoaşterea funcţiilor de repartiţie *n* - dimensionale

$$X_{\tau_1 \ldots \tau_n}(x_1, x_2, \ldots, x_n) = Pr(X_{\tau_1} \leq x_1, \ldots, X_{\tau_n} \leq x_n) \tag{1.3}$$

În acest context se mai spune că legea probabilistică a unui proces stochastic este dată de legea de repartiţie a tuturor vectorilor aleatori cu probabilităţile (1.2).

În ipoteza că parametrul $\tau \in \tau$ este timpul, se poate face şi presupunerea particulară că momentele $\tau_0, \tau_1, \ldots, \tau_n$ sunt ordonate şi anume că $\tau_0 < \tau_1 < \tau_2 < \ldots < \tau_n < \tau$, fapt care apare natural. Într-o astfel de situaţie, dacă observăm procesul stochastic la momentul $\tau_n$, pe care îl considerăm ca "prezent", putem presupune "trecutul" procesului pentru $\tau_i < \tau$, $0 \leq i \leq n$ şi în mod firesc ne interesează "viitorul" acestui proces pentru $\tau$. Un astfel de interes ne conduce în mod natural şi direct la evaluarea probabilităţilor condiţionate de forma

$$Pr(X_\tau \leq x \mid X_{\tau_n} = x_n, \ldots, X_{\tau_0} \leq x_0), \tag{1.4}$$

care înseamnă probabilitatea, ca procesul stochastic să se afle la momentul viitor $\tau$ în starea $X_\tau = x$, condiţionat de faptul că la momentele trecute $\tau_0 < \tau_1 < \ldots < \tau_n < \tau$ s-a aflat succesiv în stările $X_{\tau 0} = x_0, \ldots, X_{\tau_n} = x_n$ starea $x_0$ fiind starea iniţială a acestui proces.

Probabilităţile de forma $\pi_x(\tau) = Pr(X_\tau = x)$, $\tau_0 < \tau_1 < \ldots < \tau_n < \tau$ se numesc, în contextul de mai sus, *probabilităţi absolute de stare* şi se referă la evenimente de forma: " procesul se găseşte la momentul $\tau$ în starea $X_\tau = x$ ", fără a se face vreo referire la trecutul procesului stochastic.

Ca şi alte concepte matematice, nici procesele stochastice nu pot fi studiate global, fiind necesară o clasificare a acestora după anumite criterii.

O primă clasificare poate fi făcută pe baza mulţimii parametrilor procesului stochastic:

a) dacă $\tau$ este un interval mărginit al dreptei reale , atunci avem procese stochastice de tip continuu în raport cu parametrul timp;

b) dacă $\tau$ este o mulțime discretă, spre exemplu $\tau=Z$ (numere întregi) sau $\tau=R^+$ (mărimi reale), atunci avem procese stochastice de tip discret în raport cu parametrul $\tau$ şi care se mai numesc lanțuri.

O altă clasificare poate fi făcută în funcție de mulțimea valorilor procesului şi anume:

a) dacă mulțimea valorilor procesului este nenumărabilă (spre exemplu, un interval real), atunci avem procese cu valori continue;

b) dacă mulțimea valorilor procesului este cel mult numărabilă, atunci avem procese cu valori discrete.

Cel mai important criteriu de clasificare se referă la modul în care sunt legate între ele variabilele aleatoare $X_\tau$, ceea ce permite şi un studiu adecvat al diferitelor tipuri de procese stochastice.

**Definiția 1.2.** Un proces stochastic $X$ este un proces Markov ( sau markovian ) dacă şi numai dacă are loc relația numită proprietatea lui Markov:

$$\forall n \in N^+ \ .:\ \forall \tau_0<\tau_1<...<\tau_n<\tau, \ .:\ \forall (x_0, x_1,..., x_n, x\ ),$$

$$Pr(X_\tau \le x \mid X_{\tau_n} = x_n,\ldots,X_{\tau_0} \le x_0) = Pr(X_\tau \le x \mid X_{\tau_n} = x_n)$$

Un lanț Markov este un proces Markov cu un spațiu discret de stări.

$N^+$ este mulțimea numerelor naturale □

La baza conceptului de proces Markov se află imaginea pe care o avem despre un *sistem dinamic fără postacțiune*, adică un sistem al cărui evoluție viitoare (la momentul $\tau$) nu depinde decât de starea prezentă a procesului (cea de la momentul $\tau_n$ ) dar şi de ceea ce s-a petrecut în trecutul său (la momentele premergătoare $\tau_0<\tau_1<...<\tau_{n-1}$). Altfel spus, pentru astfel de procese stochastice, ultima stare cunoscută determină complet, din punct de vedere probabilistic, comportarea viitoare

a sistemului. Pentru exemplificare putem considera cazul transmisiei informației când, în anumite condiții, noul semnal care este emis are în vedere semnalul precedent emis şi nu toată emisiunea realizată până în acel moment. Cu alte cuvinte, *procesele Markov sunt procese fără memorie*.

Se spune că un proces ( sau lanț ) Markov X este *omogen* ( adică cu probabilități de trecere staționare) dacă, $\Pr(X_\tau \le x \mid X_{\tau n} = x_n) = \Pr(X_{\tau-\tau_n} \le x \mid X_0 = x_n)$ ceea ce implică faptul că aceste probabilități de trecere staționare nu depind explicit de timpul considerat $\tau$, ci numai de ecartul $\tau$ - $\tau_n$. În continuare vom folosi pentru studiul comportării sistemelor de calcul numai lanțuri Markov omogene (*LM*).

Probabilitatea ( condiționată ) de trecere spre starea *j* la momentul τ, ştiind că lanțul Markov se află în starea *i* la momentul *s*, este $r_{ij}(s,\tau) = Pr(X_t = j / X_s = i)$ sau $r_{ij}(\tau) = r_{ij}(s, s+\tau)$ dacă lanțul este omogen. Prin convenție se presupune că $r_{ij}(0,0)=1$ dacă $i=j$. Vom nota $R(s,\tau)$ şi $R(\tau)$ matricele stochastice respective care verifică proprietățile următoare:

$$\forall s, \tau \in \tau, \ \ \forall i, j \in \Omega, \ \ r_{ij}(s,\tau) \ge 0, \ \ \sum_{j \in \Omega} r_{ij}(s,\tau) = 1 ,$$

precum şi relația *Chapman-Kolmogorov* :

$$\forall s \le u \le \tau \in \tau, \ \forall i, j \in \Omega,$$

$$r_{ij}(s,\tau) = \sum_{k \in \Omega} r_{ik}(s,u) \cdot r_{kj}(u,\tau)$$

Probabilitatea (necondiționată) ca lanțul *LM* să se afle în starea *i* la momentul τ este $\pi_i(\tau)$. Vom nota vectorul-linie respectiv al probabilităților de stare, iar - distribuția inițială a probabilităților de stare a lanțului.

Timpul petrecut de un lanț *LM* într-o stare dată *i*, numită *durata de aflare* în starea *i* are o lege de distribuție exponențial-negativă pentru lanțuri Markov în timp continuu (LMTC) şi o lege de distribuție geometrică pentru lanțuri Markov în timp discret (LMTD). Durata de aflare în orice stare a lui LMTD este strict egală cu o unitate de

timp, numită epocă, perioadă sau tact. Numai aceste legi de distribuție posedă proprietatea de a fi "fără memorie":

$$\forall x>0,\ \forall y>0,\ Pr(X\le x+y \,/\, x\ge y)=Pr(X\le x).$$

De aceia lanțurile LMTD şi LMTC sunt folosite foarte des ca modele matematice simple de analizat ce descriu funcționarea diferitor sistema cu evenimente discrete [?]

Studiul comportării lanțurilor Markov în funcție de timp poate fi efectuat în două direcții mari:

- studierea în regim tranzitoriu, adică determinarea probabilităților de aflare în stare starea sau o submulțime de stări pentru orice $\tau>0$. Pentru aceasta se vor folosi şi matricele stocastice $R(s,\tau)$ pentru orice $(s,\tau)$, în particular $R(0,\tau)$ şi relațiile :

$$\pi_i(\tau)=\sum_{k\in\Omega}\pi_k(0)\cdot r_{ki}(0,\tau)\ ;$$

- studierea în regim de echilibru, adică se va căuta o distribuție a probabilităților staționare de stare $\overline{\pi}=(\pi_i)_{ie\Omega}$, astfel încât pentru orice $i$:

$$\lim_{\tau\to\infty}\pi_i(\tau)=\pi_i$$

În cele ce urmează ne vom interesa numai de studiul regimului de echilibru al lanțurilor Markov ce descriu comportarea sistemelor de calcul respective. Un lanț Markov care posedă o astfel de distribuție - limită a probabilităților staționare de stare independent de distribuția lor inițială este numit lanț Markov *ergodic*.

### 1.2. Clasificarea stărilor unui lanț Markov

Stările unui lanț Markov sunt clasificate în conformitate cu modul cum ele sunt "vizitate" în cursul timpului funcționării lanțului.

Prima clasificare este fondată pe momentele de reîntoarcere în starea dată. Notăm $\xi_i$ momentul de a $i$-mă schimbare de stare, iar $\xi_{ij}=min\{\tau/\tau>\xi_i \wedge X\tau=j\ /X_0=i\}$ momentul primei treceri în starea $j$ din starea $i$.

**Definiția 1.3.** O stare $i$ este: a) tranzitorie dacă $Pr(\xi_{ii}<\infty)<1$; b) recurent - nulă dacă $Pr(\xi_{ii}<\infty)=1$ şi $E[\xi_{ii}]=\infty$; c) recurent - nenulă dacă $Pr(\xi_{ii}<\infty)=1$ şi $E[\xi_{ii}]\neq\infty$, unde $E[\xi_{ii}]$ este numită durata medie de recurență a stării $i$.

Pentru o stare recurentă $i$ a unui lanț Markov în timp discret, dacă $\delta$ este cel mai mare divizor comun (c.m.d.c.) al numerelor întregi $n$, astfel încât $Pr(X_n=i/X_0=i)>0$, atunci starea $i$ este numită *periodică* de perioada $\delta$, dacă $\delta>1$, şi *aperiodică* dacă $\delta=1$.

Fără a menționa contrariul, în continuare vom folosi lanțuri Markov în timp discret aperiodice, adică stările căruia sunt toate stări aperiodice.

A doua clasificare este fondată pe mulțimea trecerilor dintr-o stare în alta. Astfel, subansamblul de stări este numit *închis* dacă:

$$\forall i \in \Omega',\ \ \exists \tau > 0,\ \ r_{ij}(\tau) \neq 0 \Rightarrow j \in \Omega'$$

adică este imposibil de a părăsi. Fie $\mathsf{E}$ o relație în $\Omega$: $i\mathsf{E}j$ dacă şi numai dacă lanțul poate trece din $i$ la $j$ şi invers, adică dacă există cel puțin un $\tau\geq 0$ astfel că $r_{ij}(\tau)>0$ şi un astfel că , ceea ce determină o relație de echivalență $\mathsf{E}$, adică fiecare clasă constituie un ansamblu de stări astfel încât din fiecare se poate, pe parcursul timpului, să se atingă toate alte stări ale acestei clase.

**Definiția 1.4.** O stare $i$ este *absorbantă* dacă ea este singurul element din clasa sa de echivalență $\mathsf{E}$ şi că această clasă este închisă.

Un lanț Markov este *ireductibil* dacă ansamblul de stări $\Omega$ formează o singură clasă de echivalență $\mathsf{E}$.

Asfel, îndată ce un lanț atinge o stare absorbantă, acolo pentru totdeauna el şi va rămâne.

Legătura între aceste două clasificări este dată de faptul că toate stările unei clase de echivalență pentru $\mathsf{E}$ sunt de acelaşi tip, adică tranzitorii, recurent - nule sau recurent - nenule. Pentru un lanț Markov în timp discret ele, de asemenea, toate sunt aperiodice sau periodice de aceeaşi perioadă.

Restricția unui lanț la o clasă de echivalență E închisă duce la un lanț ireductibil. Din aceste considerente, în afara unei mențiuni explicite, în continuare ne vom restrânge la lanțuri ireductibile.

### 1.3. Lanțuri Markov în timp discret

Un lanț Markov în timp discret este totalmente determinat de distribuția inițială şi matricea sa stochastică R(1), notată simplu R şi numită matrice de probabilități de trecere ale lanțului: $r_{ij}=Pr(X_{n+1}=j / X_n=i)$. Astfel, pentru orice $n>0$ : $R(n)=R^n$.

Dacă ne vom interesa de un regim tranzitoriu, atunci ne vom folosi de relația [1,8,9]:

$$\overline{\pi}(n+1)=\overline{\pi}(n)\cdot R \text{ sau } \overline{\pi}(n)=\overline{\pi}(0)\cdot R^n$$

ceea ce redau ecuațiile lui Kolmogorov, care descriu comportarea unui lanț *DLM*.

Teorema următoare dă posibilitate de a determina un criteriu de ergodicitate pentru lanțurile *DLM* [17].

**Teorema 1.1.** Orice lanț *DLM* omogen, aperiodic cu un spațiu finit de stări discrete şi ireductibil este un *lanț ergodic*. Distribuția limită $\pi$ a probabilităților de stare este independentă de distribuția inițială: $\pi_i=1/E[\xi_{ii}]$ este mărimea inversă a duratei medii de recurență în starea *i*. Ea este unica soluție a sistemului de ecuații Kolmogorov:

$$\overline{\pi}=\overline{\pi}*R, \sum_{i\in\Omega}\pi_i=1. \qquad (1.5)$$

Distribuția probabilităților de stare care verifică $\vec{\pi}=\vec{\pi}*R$ este numită *staționară*, deoarece pentru orice *n* este verificată relația $\vec{\pi}*R^n=\vec{\pi}.R.R^{n-1}=\vec{\pi}$. Deci dacă o distribuție $\vec{\pi}$ este staționară, atunci pentru orice moment de timp *n* are loc relația : $Pr(X_n=i)=\pi_i$, $i=\overline{1,|\Omega|}$, care dă posibilitatea de a facilita studiul comportării a unui astfel de lanț.

## 1.4. Lanțuri Markov în timp continuu

Echivalentul de matrice $\mathsf{R}$ al lanțurilor *LMTD* pentru lanțuri Markov în timp continuu *LMTC* este noțiunea de generator sau matrice dinamică.

**Teorema 1.2.** Fie *X* un lanț *LMTC* omogen cu un spațiu finit de stări discrete, redat de o matrice $\mathsf{R}(\tau)$, ce este derivabilă la dreapta în 0. Matricea $D = \frac{dR(0)}{d\tau}$ este numită *matrice generatoare* sau *matrice dinamică* a lui *X*, astfel încât :

$$\frac{d\mathsf{R}(\tau)}{d\tau} = D \cdot \mathsf{R}(\tau) = \mathsf{R}(\tau) \cdot D$$

de unde : $R(\tau)=exp(D.\tau)$. □

Astfel, un lanț *LMTC* este totalmente definit de matricea sa dinamică *D* şi distribuția probabilităților de stare inițială. Matricea dinamică *D* verifică relațiile :

$$\forall i, j \in \Omega,\ \ i \neq j,\ \ 0 \leq d_{ij} < +\infty,$$
$$\forall i \in \Omega,\ \ d_{ij} = -\sum_{\substack{j \in \Omega \\ i \neq j}} d_{ij},\ \ d_{ii} < 0,$$
$$\forall i, j \in \Omega,\ \ 0 = \sum_{j \in \Omega} d_{ij},$$

unde elementul $d_{ij}$ este interpretat ca rata de trecere din starea *i* spre starea *j*, iar durata de aflare în starea *i* este o variabilă aleatorie distribuită conform legii exponențial - negative de parametrul $\lambda_i=-1/d_{ii}$, numită rata de ieşire din starea *i*. Notăm că suma elementelor situate pe fiecare linie a matriciei *D* este egală cu 0.

O matrice care verifică aceste proprietăți este numită *D - matrice dinamică*.

Dacă se va studia un regim tranzitoriu al unui lanț *LMTC* se vor utiliza relațiile :

$$\frac{d\vec{\pi}(\tau)}{d\tau} = \vec{\pi}(\tau) \cdot D$$

şi deci $\overline{\pi}(\tau) = \overline{\pi}(0) \cdot e^{D\tau}$.

Condițiile suficiente de ergodicitate al lanțurilor *LMTC* sunt rezumate de următoarea teoremă [3,12].

**Teorema 1.3.** Orice lanț *LMTC* omogen cu un spațiu finit de stări discrete şi ireductibil este ergodic. Distribuția limită a probabilităților de stare este independentă de distribuția inițială. Ea este unica soluție a sistemului de ecuații Kolmogorov:

$$\vec{\pi} \cdot D = 0, \ \sum_{i \in \Omega} \pi_i = 1, \ \ 0 < \pi_i < 1 \quad .$$

Cu orice lanț *CLM* se pot asocia două tipuri de lanțuri *DLM* : un lanț inclus şi/sau un lanț uniformizat.

**Definiția 1.5.** Fie $\tau_n$ momentul *n* al schimbării stării procesului *X*. Procesul $X^{(E)}$ definit de $X_n^{(E)} = X_{\tau_n}$ este un lanț *DLM* numit *lanț inclus* (Embedded Markov Chain) de *X* redat de o matrice dinamică *U*.

Se poate demonstra [12,14] că $U = I + diag\left\{\frac{1}{\lambda_i}\right\} \cdot D$, unde $diag\{a_i\}$ este o matrice diagonală, elementele căreia sunt $a_i$, iar *I* este o matrice unitară. Lanțul $X^{(E)}$ corespunde schimbărilor de stare ale lui *X*, ignorând timpul decurs de *X* în fiecare stare (care este distribuit conform legii $exp(-\lambda_i \cdot \tau)$). Dacă se va nota $\pi^{(E)}$ distribuția de echilibru a lanțului $X^{(E)}$, obținem:

$$\pi_i = \frac{\pi_i^{(E)} \cdot \frac{1}{\lambda_i}}{\sum_{i \in \Omega} \pi_i^{(E)} \cdot \frac{1}{\lambda_i}}$$

Fie pe de altă parte, $\lambda \geq \max_{\forall i} \{\lambda_i\}$ şi $X^{(Z)}$ este procesul stochastic obținut pentru lanțul Markov $Z_\lambda$ cu o matrice dinamică $K_\lambda = I + \frac{1}{\lambda_i} \cdot D$ subordonată unui proces tip Poisson [12] de parametrul λ, matricea stochastică de trecere $\mathsf{R}^{(Z)}(\tau)$ a căruia este definită de:

$$\mathsf{R}^{(Z)}(\tau) = \sum_{\forall n \geq 0} \frac{(\lambda\tau)^n \cdot exp(-\lambda\tau)}{n!} \cdot K_\lambda^n$$

**Definiția 1.6.** Procesul $X^{(Z)}$ are aceleaşi probabilități de trecere în regim de echilibru ca şi procesul $X$. Lanțul *DLM* definit de $Z_\lambda$ este numit un lanț uniformizat de $X$. □

Lanțul $Z_\lambda$ corespunde unei discretizări sau unei uniformizări ale lanțului $X$, în instantanee distincte de intervale de timp $\Delta\tau$ suficient de mici pentru ca probabilitatea a mai mult de o schimbare de stare a lui $X$ în acest interval de timp $\Delta\tau$ să fie mică, de ordinul $O(\Delta\tau)$, astfel încât $\lim_{\forall\tau\to 0}\frac{O(\Delta\tau)}{\Delta\tau}=0$. Durata de aflare în starea $i$ a procesului $X^{(Z)}$ este o variabilă aleatoare distribuită după legea $exp(-\lambda.\tau)$, iar distribuția de echilibru $\vec{\pi}$ de $X$ este, de asemenea, ca şi a celui de $Z_\lambda$ astfel încât: $\overline{\pi}=\overline{\pi}\cdot(I+\frac{1}{\lambda_i}\cdot D)$, ceea ce deseori dă posibilitate de a uşura analiza acestor tipuri de procese.

## 1.5. Agregare markoviană

Principiul de *agregare exactă*, constă în a *partiționa* spațiul de stări $\Omega$ ale unui lanț în subansambluri $(\Omega^k)_{k=1,\dots,K}$ în aşa mod încât comportamentele de stări a unor şi aceleaşi $\Omega^k$ să fie stochastic echivalente.

La evaluarea performanțelor unui sistem se va folosi o agregare, subansambluri de stări ale căreia au o interpretare în termeni de comportare a acestui sistem, considerată ca o macrostare. Metoda de agregare, în particular, este folosită în analiza markoviană a unor modele ale proceselor de calcul [9,14].

**Definiția 1.7.** Un lanț Markov $X$ poate fi agregat, urmând partiția $(\Omega^k)_{k=1,\dots,K}$, dacă procesul $X^{(A)}$ cu un spațiu de stări definit astfel încât:

$$\forall\tau\geq 0,\ X_\tau^{(A)}=\Omega^k,\ \ X_\tau\in\Omega^k$$

este , de asemenea, un lanț Markov. □

*I. G. Kemeny* şi *J. L. Snell* [9] au demonstrat rezultatul următor, care determină condiția de agregare a unui lanț Markov.

**Teorema 1.4.** Un lanț Markov poate fi agregat, oricare ar fi distribuția inițială, dacă

$$\forall h,k \in \{1,\ldots,K\},\ \ \forall w,w' \in \Omega^{(k)}:$$
$$\sum_{w_h \in \Omega^{(h)}} r_{w,w_h} = \sum_{w_h \in \Omega^{(h)}} r_{w',w_h} = \widetilde{r}_{k,h} \quad \text{pentru un } DLM$$
$$\sum_{w_h \in \Omega^{(h)}} d_{w,w_h} = \sum_{w_h \in \Omega^{(h)}} d_{w',w_h} = \widetilde{d}_{k,h} \quad \text{pentru un } CLM$$

Matricea stochastică de probabilități de trecere a unui lanț agregat *DLM* este $\widetilde{R} = [\widetilde{r}_{k.k}]$, iar matricea dinamică a unui lanț *CLM* este $\widetilde{D} = \left[\widetilde{d}_{k,k}\right]$. □

În general, are loc relația $K << \sum_{\forall k} |\Omega^{(k)}|$, astfel încât calculul probabilităților de stare în regim de echilibru să fie mai uşor de efectuat pentru $\widetilde{R}$ ( respectiv $\widetilde{D}$ ) decât pentru $R$ ( respectiv $D$ ). Invers, aceste probabilități nu permit, decât în cazuri particulare, determinarea probabilităților pentru fiecare stare a lanțului original în regim de echilibru.

## 1.6. Procese semi-Markov

În practică deseori se întâlnesc sisteme dinamice cu stări discrete, pentru care durata de aflare într-o oarecare stare *i*, fiind o variabilă aleatoare, depinde de această stare şi *de starea următoare de trecere şi ea nu este necesar distribuită conform legii exponențial-negative.* Evoluția sistemului este astfel definită de starea curentă şi de *starea ce urmează.* Acest tip de procese stochastice sunt numite procese semi-Markov.

**Definiția 1.8.** Un proces stochastic *X*, cu un spațiu finit Ω de stări discrete, este numit *proces semi-Markov* (Semi-Markov Process, *SPM*) dacă există o suită crescătoare de instantanee $(\tau_n)_{n \in N}$ astfel încât este o variabilă aleatoare definită ca :

$$\forall i,j \in \Omega,\ \ \forall n \in N,\ \ \forall \tau_0 < \tau_1 < \ldots < \tau_n,\ \ \forall \tau \geq 0,\ \ \forall i_0,\ldots,i_{n-1} \in \Omega,$$
$$Pr(X_{\tau_{n+1}} = j, \tau_{n+1} - \tau_n \leq \tau / X_{\tau_n} = i, X_{\tau_{n-1}} = i_{n-1},\ldots,X_{\tau_0} = i_0)$$
$$= Pr(X_{\tau_{n+1}} = j, \tau_{n+1} - \tau_n \leq \tau / X_{\tau_n} = i),$$
$$\lim_{n \to +\infty} \tau_n = +\infty$$

Această probabilitate de trecere este notată $h_{ij}(\tau)$, iar matricea stochastică $H(\tau)=(h_{ij}(\tau))$ este numită nucleu semimarkovian. □

Notăm că $\tau_n$ este instantaneul *n* de tranziție dintr-o stare oarecare, însă procesul *X* totuşi poate să rămână în aceeaşi stare pe parcursul acestei tranziții.

Studiul în regim de echilibru al proceselor semi-Markov *SPM* este facilitat de existența, ca şi pentru lanțuri *CLM*, al unui lanț *DLM* derivat de *SPM*, numit, de asemenea, lanț inclus, care corespunde observării procesului *X* în instantaneele de tranziție.

**Propoziția 1.1.** Procesul stochastic în timp discret $X^{(E)}$ definit de $X_n^{(E)} = X_{\tau_n}$ este un lanț *DLM* cu o matrice stochastică de treceri $R = \lim_{\tau \to +\infty} H(\tau)$, numit lanț inclus ( *LMI* ) al procesului *X*. □

Proprietățile stărilor : tranzitorii, recurent - nule sau recurent - nenule, astfel ca şi caracterul ireductibil sau nu, sunt aceleaşi pentru procesul semi-Markov *X* şi lanțul său inclus. Din contra, periodicitatea unei stări este o proprietate distinctă pentru *X* şi $X^{(E)}$ : starea *i* poate fi periodică pentru *X* de perioada :

$$\delta = max\{d > 0 / (\nexists n \in N, n \neq 0, \tau = nd) \Rightarrow Pr(X_\tau = i / X_0 = i) = 0\},$$

însă nu şi pentru lanțul său inclus şi invers.

Următoarea teoremă, demonstrată de *E. CINLAR* [5] (teorema 5.2.2, p. 342) oferă un criteriu de ergodicitate a proceselor semi-Markov.

Teorema 1.5. Orice proces semi-Markov aperiodic ce are un lanț Markov inclus $X^{(E)}$ omogen cu un spațiu finit de stări discrete, ireductibil şi aperiodic este un *proces SPM ergodic*. Distribuția - limită $\pi$ a probabilităților de stare este independentă de distribuția inițială. Dacă $\pi^{(E)}$ este distribiția de echilibru de $X^{(E)}$, iar $E(\Theta_i)$ este durata medie de aflare în starea *i*, atunci avem :

$$\pi_i = \frac{\pi_i^{(E)} \cdot E(\Theta_i)}{\sum_{j \in \Omega} \pi_i^{(E)} \cdot E(\Theta_i)} \tag{1.6}$$

Rezultatele acestei teoreme, care vor fi folosite la analiza semimarcoviană a proceselor de calcul, permit calcularea distribuției probabilităților de stare $\pi$ în regim de echilibru al procesului *SPM*, referindu-ne la lanțul său inclus.

Cum deja s-a menționat procesele semi-Markov descriu mai adecvat funcționarea sistemelor reale. Un proces markovian cu probabilitățile de trecere $r_{ij}$ dintr-o stare $i$ în altă stare $j$, ($i,j \in J$) devine un proces semi-Markov dacă distribuția duratei de aflare în fiecare stare este $F_i(t)$. Cu scopul de a determina probabilitățile staționare $q_i$ de aflare a unui lanț semi-Markov în starea $i \in J$ cu $n=| J |$ stări finite vom considera un lanț Markov *DLM* cu aceleaşi probabilități de treceri.

Pentru acest lanț *DLM* sunt verificate ecuațiile (1.6) cu condiția de normalitate: $\sum_{i=1}^{n} \pi_i = 1$ .

Vom considera în lanțul semi-Markov cu un număr $N$ de treceri suficient de mare. În timpul efectuării a $N$ treceri lanțul *DLM* în mediu $N_i=\pi_i.N$ ori se va afla în starea $i \in J$. Dacă este cunoscută durata medie $\Theta_i$ de aflare a lanțului semi-Markov în starea $i$:

$$0 < \Theta_i = \int_0^{\infty} (1 - F_i(\tau)) d\tau \qquad < +\infty \qquad ,$$

atunci se poate de calculat durata medie de aflare $\bar{\tau}_i$ a acestui lanț în starea $i$ pentru acelaşi număr $N$ de treceri:

$$\bar{\tau}_i = \pi_i \cdot N \cdot \Theta_i \quad . \tag{1.7}$$

Durata medie necesară lanțului semi-Markov pentru a efectua $N$ treceri este:

$$\bar{\tau} = \sum_{j=1}^{n} \bar{\tau}_j = N \cdot \sum_{j=1}^{n} \pi_j \cdot \Theta_j \quad . \tag{1.8}$$

Deoarece $q_i$ este probabilitatea staționară că lanțul semi-Markov se va afla în starea $i \in J$, ceea ce înseamnă că pe această durată , durata medie de aflare în starea $i$ a acestui lanț este:

$$\bar{\tau}_i = q_i \cdot \bar{\tau} = q_i \cdot N \cdot \sum_{j=1}^{n} \pi_j \cdot \Theta_j \quad . \tag{1.9}$$

Din relațiile (1.7) şi (1.9), obținem:

$$\pi_i = \frac{q_i}{\Theta_i} \cdot \sum_{j=1}^{n} \pi_j \cdot \Theta_j \quad , \tag{1.10}$$

Substituind expresia lui $\pi_i$ din (1.10) în ecuația Kolmogorov (1.6) şi divizând la (cu $\Theta_i \neq 0$, ) ambele părți ale ecuațiilor astfel obținute, determinăm sistemul de ecuații algebrice pentru calcularea probabilităților staționare $q_i$ ale lanțului semi-Markov:

$$\begin{aligned} &\frac{q_i}{\Theta_i} = \sum_{k=1}^{n} r_{ki} \cdot \frac{q_k}{\Theta_k} \quad , \quad i = 1,2,\ldots,n; \\ &\sum_{i=1}^{n} q_i = 1 \quad . \end{aligned} \tag{1.11}$$

Este evident că condiția de existență a soluției unice a ecuațiilor (1.11) este aceeaşi ca şi condiția de ergodicitate a lanțului Markov cu matricea stochastică a probabilităților de trecere $R=(r_{ij})_{n\times n}$.

Uneori este mai uşor de a rezolva sistemul de ecuații (1.6) decât sistemul (1.11). Atunci în acest caz din (1.10) putem determina probabilitatea staționară $q_i$ a lanțului semi-Markov:

$$q_i = \frac{\pi_i \cdot \Theta_i}{\sum_{j=1}^{n} \pi_j \cdot \Theta_j} \quad , \qquad i = 1,2,\ldots,n. \tag{1.12}$$

Pentru un studiu mai detaliat al proceselor semi-Markov în aplicarea lor la modelarea unor procese de calcul cititorul poate consulta lucrările [13,17].

Pentru graful unui lanț semi-Markov, nodurile căruia sunt ponderate cu durata medie $\Theta_i$ de aflare în starea respectivă $i \in J$, putem determina condițiile de conservare a fluxului de probabilitate prin metoda tăieturilor.

*Metoda tăieturilor grafului de treceri al unui lanț semi-Markov se reduce la următoarea procedură.* Notăm mulțimea stărilor $J$, $|\ J\ | = n$ ale grafului $G=(J,A)$

lanțului semi-Markov, iar mulțimea arcelor acestui graf este $A$. Astfel, dacă $i,j \in J$ şi într-un pas poate avea loc o trecere din starea $i$ în starea $j$, adică într-o unitate de timp, atunci arcul $(i,j) \in A$. Fie $r(i,j)$ - probabilitatea de trecere într-un pas din starea $i$ în starea $j$, iar $\Theta_i$ este durata medie de aflare a lanțului semi-Markov în starea $i$. Dacă arcul $(i,j) \in A$, atunci $r(i,j)>0$ şi deci $\sum_{i \in j} r(i,j) = 1$. Notăm $q_i$ - probabilitatea staționară că procesul semi-Markov în orice moment de timp se va afla în starea $i$ şi este verificată relația de normalitate: $\sum_{i \in j} q_i = 1$

În aceste condiții este necesar de a determina probabilitățile staționare $q_i$ - mărimi necunoscute de aflare a lanțului semi-Markov, în starea $i$ exprimate prin mărimile cunoscute ale probabilităților de treceri $r(i,j)$ ale acestui lanț şi a duratei medie $\Theta_i$ de aflare în starea $i \in J$.

Pentru aceasta vom introduce următoarea definiție.

**Definiția 1.9.** Vom numi $U$ - tăietură a mulțimii arcelor grafului $G=(J,A)$, ce are următoarele proprietăți:

1) $U \subset A$; 2) Înlăturarea tuturor arcelor acestei $U$ - tăieturi din graful $G$ va transforma acest graf în două subgrafuri $G_1=(J_1,A_1)$ şi $G_2=(J_2,A_2)$, care nu sunt conectate între ele, adică $I=I_1 \cup I_2$, $J_1 \cap J_2 = \varnothing$ şi $A=A_1 \cup A_2$, $A_1 \cap A_2 = \varnothing$; 3) dacă $(i,j) \in U$, atunci nu este o tăietură a grafului $G$. □

Din această *definiție* reiese că o $U$ - tăietură nu va conține nici o buclă, adică $(i,i) \notin U$. Deci o $U$ - tăietură poate fi redată în modul următor:

- $U=U_1 \cup U_2$, $U_1 \cap U_2 = \varnothing$;
- dacă $(i,j) \in U_1$, atunci $j \in J_2$, $i \in J_1$;
- dacă $(i,j) \in U_2$, atunci $i \in J_2$, $j \in J_1$.

În baza definiției 1.9 putem demonstra următoarea teoremă.

**Teorema 1.6** Pentru orice $U$ - tăietură $U=U_1 \cup U_2$ în graful lanțului semi-Markov este verificată relația:

(1.13)

$$\sum_{(i,j)\in U_1} r(i,j)\cdot \frac{q_i}{\Theta_i} = \sum_{(k,l)\in U_2} r(k,l)\cdot \frac{q_k}{\Theta_k}$$

□

*Demonstrație.* Vom considera un lanț Markov cu aceeaşi mulțime de stări $J$ şi cu aceleaşi probabilități de treceri $r(i,j)$, $(i,j\in J)$, $(i,j)\in A$ ca şi în lanțul semi-Markov dat. Conform [3] pentru orice $S$ - tăietură în graful de treceri al lanțului Markov este verificată următoarea relație:

$$\sum_{(i,j)\in S_1} r(i,j)\cdot \pi_i = \sum_{(k,l)\in S_2} r(k,l)\cdot \pi_k \quad , \tag{1.14}$$

unde $\pi_i$ este probabilitatea staționară că în orice moment de timp, lanțul Markov se va afla în starea $i\in J$.

Legătura dintre probabilitățile staționare $\pi_i$ respective ale lanțului Markov şi a probabilităților staționare $q_i$ ale lanțului semi-Markov este determinată de relația [1]:

$$\pi_i = (q_i / \Theta_i)\cdot \sum_{j\in J} \pi_j \cdot m_j \quad . \tag{1.15}$$

Substituind expresia (1.15) în relația (1.14) obținem relația (1.13).

Formula (1.13), folosită pentru diferite tăieturi posibile, dă posibilitatea de a obține ecuații recursive de un rang mai mic în raport cu expresiile similare obținute din ecuațiile de stare tradiționale ce descriu comportarea lanțului semi-Markov [1,17]:

$$\frac{q_j}{\Theta_j} = \sum_{i\in J} r(i,j)\cdot \frac{q_i}{\Theta_i} \quad , \qquad j\in J \quad . \tag{1.16}$$

Este uşor de verificat că expresia (1.16) este un caz particular al ecuației (1.13) în cazul când mulțimea $J$ de stări în tăietura considerată degenerează într-o singură stare a grafului de treceri al lanțului semi-Markov.

## 1.7. Rezolvarea numerică a lanțurilor Markov

Problema constă, fiind dat un lanț *CLM* cu *matricea sa dinamică D* de dimensiunea ($n \times n$), în a calcula vectorul probabilităților staționare de $R^n$, astfel încât :

$$\vec{\pi} \cdot D = 0,$$
$$\vec{\pi} \geq 0, \quad \sum_{i=1}^{n} \pi_i = 1 \tag{1.17}$$

Mărimea $\pi_i$ este probabilitatea staționară că sistemul se va afla în starea *i*.

Pentru un lanț *DLM*, cu matricea sa stochastică a probabilităților de trecere $R$ , se poate formula căutarea soluției de echilibru a sistemului de ecuații (1.17) sub aceeaşi formă, dacă vom pune $D=R - I_n$, unde $I_n$ este matricea unitară de dimensiunea ($n \times n$).

Reamintim că, în majoritatea cazurilor *n* este foarte "mare" (de exemplu, $n \geq 10^2$) şi deci matricea *D* are o talie "mare". Însă în general, ea posedă multe elemente nule şi astfel deseori se vor implementa scheme de stocare a numai elementelor nenule ale matricei dinamice *D*. Există numeroase metode de rezolvare. Recenta carte a lui *W. J. STEWART* [15,16] conține unele din lucrările de referință în acest domeniu. Aici vom descrie numai în linii mari metodele numerice ce se impun, luând seama de situațiile care se vor întâlni în lucrarea dată.

Astfel, analiza lucrărilor în acest domeniu dă posibilitate de a distinge trei mari clase de metode ce pot fi folosite pentru matricele dinamice *D,* fără a menționa anumite proprietăți particulare: metode *directe*, metode *iterative*, metode *de proiecție* şi metode *specifice*.

La folosirea metodelor directe: metoda eliminării *GAUSS*, factorizării *LU*, de iterare inversă etc, se vor calcula dintr-o dată valorile lui $\pi_i$.

Metodele iterative sunt mai simple şi mai uşor accesibile decât metodele directe. Ele se folosesc, în general, pentru sisteme de ecuații mari ( $n>50$ ) şi în mod special pentru sisteme cu matrice rare, respectiv şi cu mulți coeficienți nuli. Ideea de bază a acestor metode constă în folosirea unei expresii de tipul $\vec{\pi}^{(k+1)} = f(\vec{\pi}^{(k+1)})$, unde $\vec{\pi}^{(k)}$

este aproximația mărimii $\vec{\pi}$ obținută la pasul *k*. Funcția *f* determină metodele folosite: 1) metoda puterii iterate, 2) metoda *GAUSS-SEIDEL*, 3) metoda *JACOBI*, 4) metoda suprarelaxării (*SOR*) etc.

Metoda puterii iterate constă în a regăsi valoarea proprie cu modulul cel mai mare a unei matrice *R* şi un vector propriu asociat ei.

Deoarece *R* este o matrice stochastică, se ştie că această valoare proprie este 1 şi dacă matricea este ireductibilă, atunci există un singur vector propriu la stânga π>0, ei asociat, care verifică relația $\sum_{i=1}^{n} \pi_i = 1$. Se obține astfel $\vec{\pi}$ , utilizând iterațiile

$$\vec{\pi}^{(k+1)} = \vec{\pi}^{(k)} \cdot \mathsf{R} .$$

În cazul lanțurilor *DLM* incluse matricea R este definită de relația:

$$R = I + \frac{1}{\alpha + \varepsilon} \cdot D ,$$

unde marimea α este aleasă astfel ca *R* să fie o matrice stochastică, adică $\alpha \geq \max_{\forall i} \sum_{j=1, j \neq i}^{n} d_{ij}$ , iar ε este o mărime mică reală pozitivă.

Astfel metoda puterii iterate constă în a folosi iterațiile :

$$\vec{\pi}^{(k+1)} = \vec{\pi}^{(k)} + \frac{1}{\alpha + \varepsilon} \vec{\pi}^{(k)} \cdot D .$$

Pe plan teoretic dacă R este ireductibilă atunci convergența este asigurată, însă ea poate fi lentă şi depinde de modulul cel mai mare al valorilor proprii ce sunt subunitare (inferioare lui 1): cu cât mai aproape de 1 este valoarea acestui modul, cu atât convergența va fi mai lentă.

Notăm că calculul direct de $\vec{\pi}^{(k)} = \vec{\pi}^{(0)} \cdot \mathsf{R}^k$ , în general, nu este practicabil pentru acest tip de matrice, deoarece ele sunt rarifiate şi în acest caz matricea $R^k$, din contra, va conține din ce în ce mai puține elemente nule.

La rezolvarea sistemului de ecuații (1.17) sunt preferabile metodele iterative sau de proiecție, deoarece ele sunt folosite din următoarele considerente:

- în aceste metode se folosesc numai produse vector cu matricea $D$ sau matrice precondiționate de $D$. Acest tip de operații propagă caracterul rarifiat al matricei $D$ la matricele introduse, care permit astfel de a efectua implementări reale, ținând cont de acest caracter. Aceasta este mult mai dificil pentru metodele directe ;

- eventual se poate folosi o valoare inițială $\vec{\pi}^{(0)}$, luând în considerație proprietățile sistemului de ecuații, care duce la o convergență rapidă a acestor metode ;

- este posibilă oprirea calculului printr-o metodă iterativă îndată ce precizia dorită este atinsă, pe când trebuie de terminat calculul cu o metodă directă ;

- matricea $D$ nu este niciodată modificată, ceea ce împiedică producerea şi creşterea erorilor de rotunjire.

Însă inconvenientul major al metodelor iterative este convergența posibil lentă şi faptul că nu se ştie de a enunța condiția suficientă de convergență, în afară de cea a metodei puterei iterate sau pentru unele cazuri particulare de matrice $D$.

Metodele specifice sunt fondate pe proprietățile matricei dinamice $D$ (sau matricei stochastice $R$ ) deduse, fie direct din analiza lui $D$, fie din proprietățile modelului, care generează $D$. Ele pot fi particulare unui tip dat de matrice sau constituie adaptări la unele metode generale.

În [14] autorii menționează trei mari clase de proprietăți: matrice aproape complet decompozabile (descompuse), matrice $\delta$- ciclice şi matrice ce posedă o *expresie tensorială*.

## 1.8. Algebra Kronecker şi lanțuri Markov

Metoda de compunere *Kronecker* a matricelor dinamice generatoare ale proceselor stocastice pentru rezolvarea lor în regim de echilibru a fost folosită în [13, 14]. În continuare vom vedea că ele permit de a scrie aceste matrice cu ajutorul unei colecții de matrice de dimensiuni cu mult mai mici şi de a folosi această scriere în formă de *produs* sau *sumă Kronecker* pentru rezolvarea numerică a lanțurilor *LMTC,* fără a manipula matricea dinamică globală.

Pentru a aplica rezultatele sublanțurilor *LMTC*, folosind algebra *Kronecker*, este indispensabil de a elabora o metodă de nivel înalt care va genera automat astfel de compuneri ale submodelelor.

În cele ce urmează vom considera $K$ procese stochastice $(X_{1,t}),(X_{2,t}),...,(X_{K,t})$, fiecare din ele primeşte valori într-un spațiu finit $S_K$, de cardinalul $l_K$ ( pentru a facilita scrierea în continuare, $K$ va însemna, în dependență de context, ca şi aici, un întreg $K$ sau un ansamblu de $K$ numere întregi $\{1,...,K\}$ ), dacă nu intervine nici o confuzie. Pentru a reda interacțiunile proceselor, este necesar de a considera procesul rezultant $X_\tau = (X_{1,\tau}, X_{2,\tau},..., X_{K,\tau})$ care primeşte valorile sale în spațiul produsului cartezian ale acestora, adică $S = \prod_{k=1}^{K} S_K$. Vom ordona, în mod lexicografic, componentele lui $S$. O stare $e = (e_1,..., e_K)$ a lui $S$, componentele căreia sunt $e_k$ indexate $i_k$, are ca indice un număr întreg:

$$i = \sum_{k=1}^{K-1} (i_k - 1)(\prod_{j=k+1}^{K} \alpha_j) + i_K .$$

Pentru aceasta vom identifica stările cu numere întregi, astfel încât starea $i$ va fi scrisă ca $(i_1,...,i_K)$ într-o "multibază" $(a_1,...,a_K)$. În acelaşi mod proiecția lui $i$ pe $S_K$ va fi notată $i_k$.

De asemenea, vom preciza două tipuri de tranziții (treceri) în evoluția procesului $X$.

**Definiția 1.10.** O tranziție $t$ este *locală* în $X$, dacă $i \xrightarrow{t} j \Rightarrow \forall k' \neq k$, $j_{k'} = i_{k'}$. O tranziție $t$ este de *sincronizare* dacă: $i \xrightarrow{t} j \Rightarrow \exists k', k''$, $k' \neq k''$, astfel încât $j_{k'} \neq i_{k'}$, $j_{k''} \neq i_{k''}$. □

Toate matricele dinamice considerate în continuare au valori reale. Vom nota mulțimea acestor matrice de dimensiunea $n \times m$ prin $D^*_{[n,m]}$.

*Produs Kronecker şi lanțuri Markov.* Produsul *Kronecker* permite de a obține o expresie a matricei stocastice prin probabilitățile de schmbare a stărilor proceselor

markoviene în timp discret cu $K$ componente. Însă, în acelaşi timp, el traduce, la nivelul generatoarelor lanțurilor *LMTC*, efectele tranzițiilor de sincronizare. Anume această proprietate va fi folosită, deoarece vom studia modele ale sistemelor de calcul cu un spațiu de stări discrete ce evoluează în timp continuu.

**Definiția 1.11.** (*Produsul Kronecker*). Fie date două matrice $A \in D^{*}_{[n_1,m_1]}$ şi $B \in D^{*}_{[n_2,m_2]}$. *Produsul Kronecker* $(\otimes)$ al matricei $A$ cu matricea $B$ este matricea $C \in D_{[n_1 \cdot n_2,\, m_1 \cdot m_2]} : C = A \otimes B$ cu $C_{i,j} = a_{i_1,j_1} \cdot b_{i_2,j_2}$, unde $i = (i_1, i_2)$ este în multibaza $(n_1, n_2)$ şi $j = (j_1, j_2)$ în multibaza $(m_1, m_2)$. □

*Exemplul 1.1*. Dacă $A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix}$ şi $B = \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} & b_{1,4} \\ b_{2,1} & b_{2,2} & b_{2,3} & b_{2,4} \end{bmatrix}$,

atunci: $A \otimes B = [a_{i,j} \cdot B]$, $\quad i, j = 1, 2$.

Astfel, produsul Kronecker reprezintă procesul de înmulțire a fiecărui element al unei matrice cu fiecare element al altei matrice [18]. Fie două matrice $A \in IR^{m \times n}$ şi $B \in IR^{p \times q}$:

$$A = \begin{bmatrix} a_{0,0} & \cdots & a_{0,n-1} \\ \vdots & \ddots & \vdots \\ a_{m-1,0} & \vdots & a_{m-1,n-1} \end{bmatrix} \text{ şi } B = \begin{bmatrix} b_{0,0} & \cdots & b_{0,q-1} \\ \vdots & \ddots & \vdots \\ b_{p-1,0} & \vdots & b_{p-1,q-1} \end{bmatrix}.$$

Produsul Kronecker $A \otimes B \in IR^{mp \times nq}$ este:

$$A \otimes B = \begin{bmatrix} a_{0,0}B & \cdots & a_{0,n-1}B \\ \vdots & \ddots & \vdots \\ a_{m-1,0}B & \cdots & a_{m-1,n-1}B \end{bmatrix} =$$

$$\begin{bmatrix} a_{0,0}b_{0,0} & \cdots & a_{0,0}b_{0,q-1} & \cdots & a_{0,n-1}b_{0,0} & \cdots & a_{0,n-1}b_{0,q-1} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{0,0}b_{p-1,0} & \cdots & a_{0,0}b_{p-1,q-1} & \cdots & a_{0,n-1}b_{p-1,0} & \cdots & a_{0,n-1}b_{p-1,q-1} \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ a_{m-1,0}b_{0,0} & \cdots & a_{m-1,0}b_{0,q-1} & \cdots & a_{m-1,n-1}b_{0,0} & \cdots & a_{m-1,n-1}b_{0,q-1} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m-1,0}b_{p-1,0} & \cdots & a_{m-1,0}b_{p-1,q-1} & \cdots & a_{m-1,n-1}b_{p-1,0} & \cdots & a_{m-1,n-1}b_{p-1,q-1} \end{bmatrix}.$$

În loc să memorăm $A \otimes B$ explicit, ceea ce necesită o memorie de $O(m \cdot n \cdot p \cdot q)$, noi am putea stoca $A$ şi $B$ şi apoi calcula elementele $A \otimes B$ după necesitate în conformitate cu formula:

$$(A \otimes B)[i, j] = A\left[\left[\frac{i}{p}\right], \left[\frac{j}{q}\right]\right] B[i \bmod p, j \bmod q],$$

ce necesită doar o memorie de $O(m \cdot n + p \cdot q)$ cu preţul căreia se va complica procesul de calcul.

Acest proces poate fi generalizat prin produsul Kronecker a $K$ matrice:

$$A = A_K \otimes \cdots \otimes A_1 = \bigotimes_{k=1}^{k} A_k$$

Expresia pentru un element al lui $A$ este în relaţie cu noţiunea de sistem de reprezentare a numerelor în baza mixtă. Dacă matricea $A_k$ are $r_k$ rânduri şi $c_k$ coloane, atunci un element din $A$ poate fi calculat prin:

$$A[i, j] = \prod_{k=1}^{K} A_k[i_k, j_k], \qquad (1.18)$$

unde $i = [i_K, \ldots, i_1]$ este reprezentarea în baza mixtă a lui $i$ cu respectarea bazei $r = [r_K, \ldots, r_1]$, iar $j = [j_K, \ldots, j_1]$ este reprezentarea în baza mixtă a lui $j$ cu respectarea bazei $c = [c_K, \ldots, c_1]$.

În figura 1.1 este prezentat un exemplu de produs Kronecker

$$A_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 4 & 0 & 5 & 6 \\ 7 & 8 & 0 & 9 \end{bmatrix} \quad A = A_3 \otimes A_2 \otimes A_1$$

$$A = \begin{bmatrix}
0 & 1 & 2 & 3 & 0 & \frac{1}{2} & 1 & \frac{3}{2} & 0 & \frac{1}{3} & \frac{2}{3} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 5 & 6 & 2 & 0 & \frac{5}{2} & 3 & \frac{4}{3} & 0 & \frac{5}{3} & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
7 & 8 & 0 & 9 & \frac{7}{2} & 4 & 0 & \frac{9}{2} & \frac{7}{3} & \frac{8}{3} & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} & 0 & \frac{1}{5} & \frac{2}{5} & \frac{3}{5} & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & \frac{5}{4} & \frac{3}{2} & \frac{4}{5} & 0 & 1 & \frac{6}{5} & \frac{2}{3} & 0 & \frac{5}{6} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{7}{4} & 2 & 0 & \frac{9}{4} & \frac{7}{5} & \frac{8}{5} & 0 & \frac{9}{5} & \frac{7}{6} & \frac{4}{3} & 0 & \frac{3}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 0 & \frac{1}{2} & 1 & \frac{3}{2} & 0 & \frac{1}{3} & \frac{2}{3} & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 5 & 6 & 2 & 0 & \frac{5}{2} & 3 & \frac{4}{3} & 0 & \frac{5}{3} & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 7 & 8 & 0 & 9 & \frac{7}{2} & 4 & 0 & \frac{9}{2} & \frac{7}{3} & \frac{8}{3} & 0 & 3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} & 0 & \frac{1}{5} & \frac{2}{5} & \frac{3}{5} & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \frac{5}{4} & \frac{3}{2} & \frac{4}{5} & 0 & 1 & \frac{6}{5} & \frac{2}{3} & 0 & \frac{5}{6} & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{7}{4} & 2 & 0 & \frac{9}{4} & \frac{7}{5} & \frac{8}{5} & 0 & \frac{9}{5} & \frac{7}{6} & \frac{4}{3} & 0 & \frac{3}{2}
\end{bmatrix}$$

Figura 1.1. Exemplu de produs Kronecker

Un exemplu de produs Kronecker a trei matrice este prezentat în Fig. A3.1. Menționăm că stocarea explicită a lui $A$ necesită un spațiu de 288 valori în virgulă mobilă, pe când stocarea matricelor $A_3$, $A_2$ şi $A_1$ necesită doar 22 valori în virgulă mobilă. Calcularea unui element din $A$ necesită 2 multiplicări în virgulă mobilă şi câteva calcule aritmetice cu numere întregi pentru a determina reprezentarea în baza mixtă. Este evident, că odată cu mărirea numărului de matrice implicate în produsul Kronecker, creşte atât volumul de memorie necesar pentru stocarea lor, cât şi cheltuielile de prelucrare a acestora.

În continuare, vom considera doar produsul Kronecker a matricelor pătrate. În particular, considerăm produsul Kronecker a două matrice unitare:

$$I_n \otimes I_m = \begin{bmatrix} I_m & 0 & \cdots & 0 \\ 0 & I_m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_m \end{bmatrix}.$$

Aşadar, produsul Kronecker al matricelor unitare este o matrice unitară, a cărei dimensiune este produsul dimensiunilor matricelor inițiale:

$$I_{n_K} \otimes \cdots \otimes I_{n_1} = I_{\prod_{k=1}^{K} n_k}.$$

Menționăm aici doar că această operație nu este comutativă.

*Compunerea lanțurilor Markov în timp discret* (*LMTD*) *independente*. Fie $X_1, X_2, ..., X_K$ sunt $K$ sublanțuri *LMTD* independente cu matricele stocastice respective ale probabilităților de tranziție $R_k(n)$. Se poate demonstra [18] că *LMTD* rezultant $X = (X_1, X_2, ..., X_K)$ de asemenea, este un lanț *LMTD*, matricea stocastică a căruia este $R = \overset{K}{\underset{k=1}{\otimes}} R_k(n)$.

*Suma Kronecker şi lanțuri Markov în timp continuu.*

**Definiția 1.12.** (*Suma Kronecker*). Fie $A \in D^*_{[n]}$ şi $B \in D^*_{[m]}$ două matrice pătratice. Suma Kronecker $(\oplus)$ a matricei $A$ cu matricea $B$ este o matrice $C \in D_{[n,m]} : C = A \oplus B = A \otimes I_m + I_n \otimes B$, astfel încât:

$$c_{ij} = \begin{cases} a_{i_1,j_1} + b_{i_2,j_2}, & (i_1 = j_1) \,\&\, (i_2 = j_2) \\ b_{i_2,j_2} & (i_1 = j_1) \,\&\, (i_2 \neq j_2) \\ a_{i_1,j_1} & (i_1 \neq j_1) \,\&\, (i_2 = j_2) \\ 0 & (i_1 \neq j_1) \,\&\, (i_2 \neq j_2) \end{cases},$$

unde $I_m$ şi $I_n$ sunt matrice unitare de dimensiunea respectivă. □

Suma *Kronecker* permite de a determina o expresie compactă a generatorului lanțurilor *LMTC*, constituite din *K* componente. Ea descrie funcționarea autonomă a componentelor *LMTC* care rezultă din tranzițiile locale.

În [18] sunt prezentate mai detaliat proprietățile acestei operații şi unele exemple de aplicare. De asemenea, suma Kronecker se extinde şi la *K* matrice $D_K$:

$$\bigoplus_{k=1}^{K} D_k = \sum_{k=1}^{K} \bigotimes_{1\le k'<k} I_{\alpha_{k'}} \otimes D_k \bigotimes_{k<k'<K} I_{\alpha_{k'}} = \sum_{k=1}^{K} I_{\alpha_{k'}} \otimes D_k \otimes I_{U_k},$$

unde $\alpha_k = \prod_{k'<k} \alpha_{k'}$ şi $U_k = \prod_{k'>k} \alpha_{k'}$. *Compunera lanţurilor LMTC independente*. Fie $X_1, X_2, ..., X_K$ sunt *K* lanţuri *LMTC* independente cu generatorul respectiv $D_K(\tau)$. Se poate demonstra [13] că procesul markovian rezultant $X = (X_1, X_2, ..., X_K)$ areun lanţ *LMTC* rezultant, generatorul căruia este $D(\tau) = \bigoplus_{k=1}^{K} D_k(\tau)$.

Într-un lanţ *LMTC* cu un generator *D,* componentele căruia nu mai sunt independente, tranziţiile *locale* generează în expresia lui *D* o *sumă Kronecker*, care traduce independenţa acestor tranziţii, iar tranziţiile de *sincronizare* generează un produs *Kronecker* care traduce modificarea simultană a mai multor componente. Următoarele rezultate stau la baza descrierilor structurale ale generatoarelor lanţurilor *LMTC,* folosite pentru compoziţia modelelor de reţele Petri stocastice.

*Generatorul unui lanţ LMTC cu K* componente *dependente*. Fie $X = (X_1, ..., X_K)$ un lanţ *LMTC* şi $T_s$ mulţimea tranziţiilor sale de sincronizare. Generatorul lui *X* este:

$$D = \bigoplus_{k=1}^{K} D'_k + \sum_{t \in T_s} \lambda(t) \cdot \left[ \bigotimes_{k=1}^{K} C_k - \bigotimes_{k=1}^{K} A_k \right], \qquad (1.18)$$

unde $D'$ este restricţia matricei dinamice D la tranziţii locale în $S_k$ şi, dacă durata de tranziţie a $t \in T_s$ este distribuită conform legii exponenţial-negative $\exp(-\lambda(t) \cdot \tau)$ şi $i \xrightarrow{t} j$, atunci:

$$C_k = A_k = I_{\alpha_k}, \text{ dacă } t(i_k) = i_k, \quad C_k = 1_{\alpha_k}(i_k, j_k) \text{ şi } A_k = 1_{\alpha_k}(i_k, i_k), \text{ dacă}$$

$t(i_k) = j_k \neq i_k$, unde $1_n(j, j')$ este o matrice de $D^*_{[n]}$, toţi termenii căreia sunt nuli, în afara celui cu indice $1(j, j')$, care este egal cu 1. Matricile $I_{\alpha_k}$ "propagă" în *D* salturile de ieşire din fiecare $S_k$, unde tranziţia *t* produce o schimbare de stare.

În general, *compunerea Kronecker* reprezintă o structură constituită din mai multe submodele *RMG* ale subsistemelor ce interacționează între ele, fiecare posedând totodată o autonomie, mai mică sau mai mare, de funcționare. Modelul este atunci construit, ținând cont de această structură şi de condițiile care trebuie impuse pentru a obține o expresie tensorială factorizată.

Pentru orice matrice $A$, $B$, $C$ şi $D$ produsul lor Kronecker, dacă aceste expresii au un sens practic, are următoarele proprietăți :

$$(A \otimes B) \otimes C = A \otimes (B \otimes C);; (A + B) \otimes C = (A \otimes C) + (B \otimes C)$$

$$(A \cdot B) \otimes (C \cdot D) = (A \otimes B) \cdot (C \otimes D);\ (A \otimes B)^n = A^n \otimes B^n;$$

$$(A \cdot B)^{\otimes n} = A^{\otimes n} \cdot B^{\otimes n};\ (A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

*Suma Kronecker* este definită [29] în termeni ai produsului Kronecker ce implică matrice, ca:

$$\bigoplus_{k=1}^{K} A_k = \sum_{k=1}^{K} I_{n_K} \otimes \cdots \otimes I_{n_{k+1}} \otimes A_k \otimes I_{n_{k-1}} \otimes \cdots \otimes I_{n_1} = \sum_{k=1}^{K} I_{\prod_{i=k+1}^{K} n_i} \otimes A_k \otimes I_{\prod_{i=1}^{k-1} n_i}, \qquad (1.19)$$

unde matricea $A_\mathrm{k}$ este o matrice pătratică de dimensiunea $n_k$. Un exemplu de sumă Kronecker a trei matrice este prezentat în figura 1.2, în care se indică fiecare termen al sumei din ecuația (1.19) şi suma lor totală.

*Reprezentarea Kronecker rarefiată.* **U**tilizarea operațiilor Kronecker permite reprezentarea totală a matricelor voluminoase prin stocarea unor matrice mult mai mici. Vom examina efectul alegerii metodei stocării matricelor parțiale (mici) asupra complexității prelucrării lor.

Examinând exemplul din Fig. 1.2, putem observa că matricea $A$ este destul de rarefiată – mai mult de jumătate din elemente sunt nule. Un element nul apare atunci, când unul din elementele matricelor parțiale este nul. De fapt, blocuri largi de elemente nule apar ca rezultat a două elemente nule din matricea $A_3$. Din punct de vedere a necesităților de memorie pentru acest exemplu, alegerea unei structuri de date pentru matricele parțiale (mici) (de exemplu completă, rarefiată pe rânduri sau rarefiată pe coloane) nu este critică, deoarece matricele parțiale vor necesita un volum

de memorie foarte mic, indiferent de structura de date utilizată. Cu mărirea dimensuiunii matricei $A_k$ această alegere, desigur, va avea un impact apreciabil.

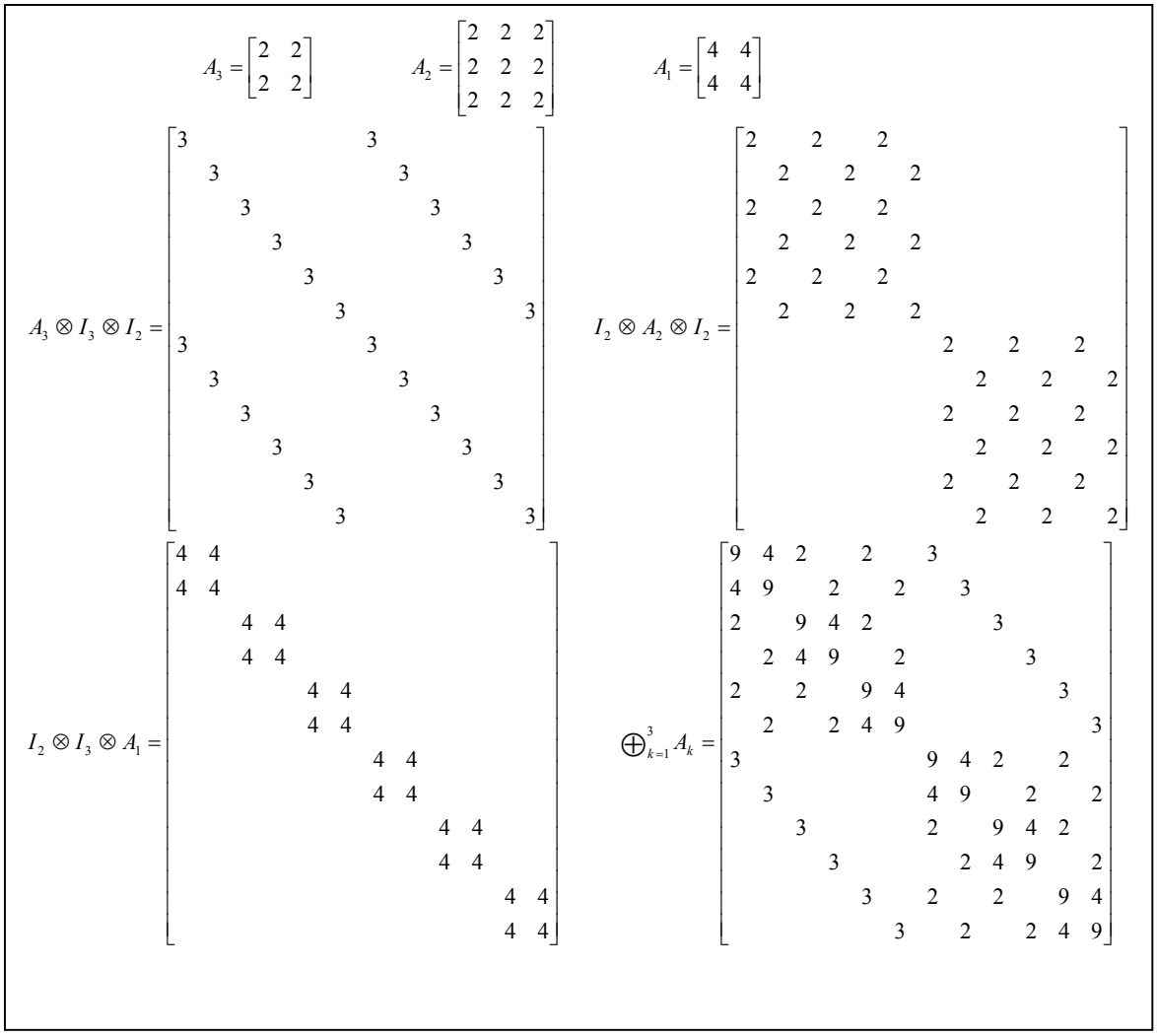



Figura 1.2. Exemplu de produs Kronecker

Utilizarea stocării matricelor rarefiate poate aduce la o diminuare considerabilă a complexității multiplicărilor vector-matrice. Aceasta este în special adevărat pentru o reprezentare produs Kronecker. De exemplu, presupunem că trebuie să înmulțim un vector cu o matrice $A = A_K \otimes \cdots \otimes A_1$, unde $A \in IR^{N \times N}$. Pentru simplitate, vom utiliza metoda direct-înainte, în care fiecare element al coloanelor matricei $A$ este calculat explicit şi înmulțit cu elementul-vector cel mai apropiat. Dacă fiecare matrice $A_k$ este complet stocată, atunci costul multiplicării vector-matrice este exact $KN^2$ înmulțiri în

virgulă mobilă. În cazul în care vom memora matricele $A_k$ utilizând metoda coloanelor rarefiate, costul multiplicării vector-matrice este exact $K\eta(A)$ înmulțiri în virgulă mobilă. Aşadar, pentru orice element nul din $A$, vom efectua cu $K$ înmulțiri mai puține, în cazul în care alegem stocarea rarefiată a matricelor parțiale.

Desigur, necesitatea accesării elementelor matricei $A$ impune metoda de stocare a matricelor parțiale $A_k,\ldots, A_1$. Dacă dorim accesul la $A$ pe rânduri (coloane), vom stoca fiecare matrice $A_k$ pe rânduri (coloane). În cazul în care avem nevoie de a accesa matricea $A$ atât pe rânduri, cât şi pe coloane, va trebui să utilizăm un format rarefiat, ce permite accesul pe rânduri şi coloane pentru orice matrice $A_k$. Stocarea completă va fi necesară, doar dacă avem nevoie de acces la anumite elemente ale matricei $A$. La implementarea lui poate fi utilizat atât accesul pe rânduri, cât şi pe coloane. Aşadar, vom considera că fiecare matrice $A_k$ este stocată sau prin metoda rândurilor rarefiate sau prin metoda coloanelor rarefiate.

# 2.Elemente de teoria aşteptării

## 2.1. Generalități

Rețeaua de telecomunicații sau de calculatoare este un sistem complex format dintr-o multitudine de sistemele elementare interconectate după o structură convenabilă.

Sisteme elementare sunt de tipuri diferite, având caracteristici proprii ce decurg atât din constituția lor fizică, precum şi din natura proceselor la care sunt supuse.

Metodele *fenomenelor* de *aşteptare* descriu *sisteme* şi *procese* de *servire* cu caracter de masă care intervin în diferite domenii ale activității practice.

Teoria aşteptării (teoria firelor de aşteptare sau teoria cozilor) este acea ramură a matematicii, ce studiază fenomenele de aşteptare. Enumerăm acum principalele elemente ale problemei fenomenului de aşteptare.

*Sursa* este mulțimea unităților (cererilor, clienților) ce solicită un serviciu la un moment dat. Ea poate fi finită sau infinită. Sosirea unităților în sistemul de aşteptare

determină o variabilă aleatoare *X* care reprezintă numărul de unități ce intră în sistem în unitatea de timp. Este necesar să se cunoască repartiția variabilei *X*.

La originea *teoriei aşteptării* se găseşte *determinarea "încărcării" optime* a unei centrale telefonice. Pentru a rezolva această problemă, este necesar să se determine *cererile de servicii* (apelurile) care sosesc în mod întâmplător şi să se înregistreze timpul necesar pentru obținerea legăturilor telefonice. Un astfel de model în care se urmăreşte satisfacerea cât mai promptă a cererilor de servicii în condiții economice cât mai avantajoase se numeşte *model* (*sistem*) *de aşteptare* (*servire*).

În sistemul de aşteptare există un *flux de unități* (*cereri*) pentru servire numit *flux de intrare* caracterizat prin numărul de cereri care intră în sistem în unitatea de timp. Într-un *sistem de aşteptare* există elemente care efectuează *serviciile*, numite *servere* sau *canale de servire*. Pentru servirea fiecărei *unități* (*cerere, client*), este necesar un timp oarecare în cursul căruia serverul este ocupat şi nu poate servi alte unități. *Durata servirii* este *întâmplătoare* (*aleatoare*). Un *sistem de aşteptare* este descris complet de următoarele elemente: *flux de intrare, şirul* (*firul*) *de aşteptare, serverul* (*serverii*) *de servire* şi *fluxul de ieşire*. Cu ajutorul fluxului de intrare putem determina modul în care sosesc unitățile în sistemul de aşteptare. Presupunem că *intrările* (*sosirile*) în sistem sunt *întâmplătoare* şi *independente*. Deci probabilitatea ca o unitate (cerere) să sosească în sistem este independentă atât de momentul în care se produce sosirea cât şi de numărul de unități existente deja în sistem sau de numărul de unități ce vor veni. Probabilitatea ca în intervalul de timp $(t, t+\Delta t)$, $t>0$, să se producă o intrare în sistem reprezintă *numărul mediu de intrări* (*sosiri*) în unitatea de timp $\Delta t$ şi este egală cu $1/\lambda_i$, în ipoteza că sosirile urmează un proces Poisson de parametru $\lambda_i$, $(0 < \lambda_i < \infty)$. Să presupunem că $t \geq 0$ şi să notăm cu $t_0, t_1, \ldots, t_n, \ldots$ momentele succesive în care sosesc unitățile în sistem. Vom admite că *intervalele de timp* dintre două unități consecutive $\tau_n = t_{n+1} - t_n$, $t_0 = 0$, n = 1,2,... sunt *variabile aleatoare pozitive independente cu funcția de repartiție*

$$F(x) = P(\tau_n \le x),\ 0 < x < \infty,\ n = 1,2,\ldots \quad ,$$

iar $\tau_n$, $n = 1,2,\ldots$ , fiind variabile aleatorii identic repartizate. Timpul necesar pentru servirea unei unități se numeşte *timp de servire*. Presupunem că *duratele de servire* sunt variabile aleatoare pozitive, identic repartizate, independente şi, de asemenea, independente de $\tau_1,\tau_2,\ldots$ Notând cu $u_n$ timpul de servire al cererii de-a *n*-a unități, *funcția de repartiție a timpului de servire* este

$$H(x) = P(u_n \le x),\ 0 \le x < \infty.$$

*Şirul de aşteptare* este determinat de numărul locurilor de aşteptare şi numărul unităților care aşteaptă şi poate fi finit sau infinit (limitat, respectiv nelimitat).

*Serverul* (*unitatea de serviciu*) poate fi un lucrător (vânzător din magazin, casierul de la autoservire, mecanicul de întreținere şi reparații), o maşină, un calculator etc., care efectuează serviciul solicitat. Timpul de servire a unei unități de către un server este o variabilă aleatoare *Y*.

Practica pune la dispoziție valori empirice pentru variabilele aleatoare *X* şi *Y*, cu ajutorul cărora determinăm legile de repartiție şi parametrii pentru variabilele aleatoare *X* şi *Y*, de exemplu, prin ajustarea unei distribuții empirice la o distribuție teoretică după metodele pe care teoria probabilităților şi statistica matematică le oferă.

Studierea fenomenelor de aşteptare are drept scop stabilirea structurii optime a sistemelor tehnice astfel încât cheltuielile ocazionale de aşteptări să fie minime.

Un fenomen de aşteptare se caracterizează, deci, prin următoarele aspecte:

- Existența unităților (clienților) care intră în sistem într-un număr limitat sau nelimitat şi formează şiruri sau cozi de aşteptare.

- Prezența pe traseele de aşteptare a unuia sau mai multor serveri (stații de serviciu) care îndeplinesc anumite prestații.

- Corelația strânsă dintre sosirile unităților, formarea şirurilor (cozilor) de aşteptare şi servirile clienților din şirul de aşteptare. Pot exista unul sau mai multe şiruri paralele de aşteptare şi unul sau mai mulți serveri (canale) de servire. Servirea

clienților se poate face în *paralel,* în *cascadă* sau *în serie-paralel*. Disciplina de servire a unităților poate fi "primul venit, primul servit", disciplina prin excepție şi disciplina după gradul de urgență etc.

- Schemele de aşteptare pot fi cu *circuit închis* sau cu *circuit deschis*. Dacă unitățile care părăsesc sistemul nu se reântorc la punctul de intrare, atunci schema sistemului de aşteptare este cu circuit deschis. În caz contrar schemele sistemelor de aşteptare sunt cu circuit închis.

- Grupurile funcționale ale teoriei aşteptării sunt: sursele care generează şi trimit după anumite legi în sistem unitățile care participă la fenomenul de aşteptare; şirurile de aşteptare care se formează din cauza neregularităților dintre sosiri şi serviri şi servirile care pot fi individuale în grup şi în masă.

Modelele de aşteptare denumite şi *sisteme de aşteptare* (*SA*) se regăsesc în diverse sisteme tehnice, sub forme caracteristice cum ar fi: aşteptările mesajelor pentru a fi prelucrate de diverse calculatoare; depanarea programelor etc. Funcționarea optimă a sistemelor în care apar fenomene de aşteptare se realizează pe baza minimului cheltuielilor de funcționare în condiții restrictive impuse. Aceasta reclamă reglamentarea fluxului de intrare, micşorarea cozilor de aşteptare, verificarea serverilor în vederea funcționării continue la un ritm impus şi coordonarea cantității şi calității serverilor cu cerințele formulate de beneficiari.

Modelele *SA* de servire a unităților care aşteaptă pot fi cu unul sau mai mulți serveri. Când există doi sau mai mulți serveri, aceştea pot funcționa în paralel sau în serie (cascadă). Servirile se execută fie în ordinea intrărilor unităților în sistem fie în ordinea de gravitate a defecțiunilor.

Fluxul de intrare descrie modul în care sosesc în sistem unitățile ce solicită servicii. Rata sosirilor şi servirilor unităților în sistem într-o unitate de timp se notează ($\lambda$) şi respectiv ($\mu$). Numărul unităților din fluxul de intrare poate fi finit sau infinit. Sistemele de aşteptare a unităților solicitante pot fi cu sau fără pierderi. Sistemele de aşteptare cu pierderi pot fi cu refuzuri directe şi refuzuri întârziate. În primul caz

unitatea solicitantă care soseşte în sistem observă că serverul este ocupat şi neavând timp să aştepte părăseşte sistemul fără să revină vreodată la punctul de intrare în sistem. În cazul al doilea, când se aplică schema mixtă de aşteptare, unitatea solicitantă dispune de un anumit timp pentru aşteptare şi dacă nu este servită în acest interval ea părăseşte sistemul. Aceste situații se studiază cu ajutorul sistemelor de aşteptare cu restricții de timp şi (sau) de loc (când în sistem nu există loc pentru noile unități care vin şi au timp să aştepte).

În cazul mai multor şiruri de aşteptare, unitățile se servesc după următoarele reguli: "*primul venit, primul servit*", "*ultimul venit, primul servit*" şi pe baza regulilor de prioritate relativă sau absolută.Organizarea servirilor poate fi realizată individual, în grup şi servire în masă. Dacă fluxul ieşirilor din stațiile de servire nu este organizat corespunzător atunci la sistemele de aşteptare şi servire în cascadă apare fenomenul de blocaj.

*Aplicarea teoriei aşteptării* la diverse situații de gestionare a resurselor sistemelor informatice implică pe de o parte construirea modelelor matematice, iar pe de altă parte stabilirea fluxului unităților din sistem, capacitatea de funcționare, numărul de serveri cu scopul determinării soluției optime la care cheltuielile au valoare minimă. Mărimile care intră în structura modelelor matematice ale SA se stabilesc pentru procese nestaționare şi staționare. Probabilitățile după care se desfăşoară evenimentele în cadrul proceselor cu aşteptare se determină fie cu ajutorul relațiilor analitice fie prin aplicarea metodelor de simulare statistică tip Monte-Carlo sau metodele de simulare tip "joc". Probabilitățile se determină atât pentru variabile cu structură discretă cât şi pentru variabile cu structură continuă. Legile de repartiție ale probabilităților pentru structuri discrete sunt de tip *Bernoulli, Poisson, Pascal* etc., iar pentru structuri continue sunt de tip exponențial-negative, *Erlang-k* ($E_k$), *Hiperexponențial-k* ($H_k$), *Cox-k, Weibull*, etc. Dacă în situațiile practice intervin distribuții cu mai multe variabile, atunci se pot calcula probabilitățile cu relații de tip *Dirichlet, Wishard, Pareto* etc. [1,2,4,7].

Aplicațiile practice ale teoriei aşteptării se regăsesc în situații diverse şi anume: ciclul lung între emiterea ideilor noi şi aplicarea lor în vederea asimilării de noi produse, aşteptarea mesajelor prelucrate etc.

Procesele de aşteptare pot fi staționare şi nestaționare. Pentru cunoaşterea legii după care au loc sosirile şi servirile probabiliste ale unităților ce sosesc în sistem trebuie studiate fenomenele din punct de vedere statistic, precizându-se legea de probabilitate care le guvernează.

Fenomenele de aşteptare din cadrul sistemelor informatice generează cheltuieli care grevează prețul de cost al prelucrării informației. Stabilirea structurii acestor cheltuieli pentru diverse situații din cadrul sistemelor de aşteptare şi precizarea mărimilor care au ponderi semnificative în modelul cheltuielilor se face analizând relațiile de cost specificate de beneficiar.

Restricțiile pentru aplicarea modelelor matematice se referă la respectarea cantității şi calității serviciilor şi la evitarea fenomenului de blocaj. Dacă o unitate vine în sistem şi constată că timpul de aşteptare este mai mare decât timpul de care dispune, atunci ea părăseşte sistemul şi nu mai revine niciodată. În acest caz modelul de aşteptare este un *SA* cu pierderi. Când există diferențe mari între timpul de serviciu şi intervalul dintere două sosiri se formează şiruri mari de aşteptare, ceea ce creează condiții favorabile blocajului.

Cele mai frecvente situații de aşteptare întâlnite în practică sunt: un sistem *SA* cu număr limitat sau nelimitat de clienți, mai mulți serveri cu un număr limitat şi nelimitat de clienți. Cele patru situații se regăsesc atât în procese staționare cât şi în procese nestaționare.

Relațiile de calcul ale probabilităților $\pi(n)$ se deduc din studierea proceselor generalizate de naştere şi moarte ale cererilor şi pot fi de tip *Poisson, Erlang etc.*, caracteristicile numerice ale cărora pot fi determinate prin soluționarea ecuațiilor *Chapman-Kolmogorov* ce descrie comportarea sistemului *SA* analizat.

## 2.2. Procese de reînnoire

Timpii care se scurg între producerea evenimentelor succesive sunt variabile aleatoare independente şi identic repartizate, supunându-se unei legi exponențiale de parametru $\lambda$. O generalizare naturală a unui proces Poisson este cea în care, renunțând la cazul particular al repartiției exponențiale, se presupune că timpii dintre evenimentele succesive sunt variabile aleatoare independente şi identic repartizate, având o lege de repartiție arbitrară. Un asemenea proces de numărare este cunoscut sub numele de proces de reînnoire, denumirea de "reînnoire" însemnând apariția (producerea) unui nou eveniment.

**Definiția 2.1.** Un proces de numărare $\{N(t),\ t < 0\}$, pentru care timpii inter-sosire sunt variabile aleatoare independente şi identic repartizate cu funcția de repartiție arbitrară $F(t)$, se numeşte proces de reînnoire. □

Aşa cum am arătat mai sus, procesul Poisson a cărui repartiție a timpilor inter-sosire este repartiția exponențială:

$$F(t) = 1 - e^{-\lambda t},\ t \geq 0$$

reprezintă cazul tipic de proces de reînnoire.

*Exemplul 2.1.*Să considerăm cazul unui nod de comutare a pachetelor de date într-o rețea de calculatoare, la care timpii de prelucrare sunt independenți şi identici repartizați. Să presupunem că atunci când un pachet părăseşte nodul, cel ce aşteaptă la rând intră imediat. În aceste condiții, procesul $\{N(t),\ t \geq 0\}$, unde $N(t)$ reprezintă numărul pachetelor până la momentul $t$, reprezintă un proces de reînnoire.

La fel ca şi în cazul particular al procesului Poisson, vom vorbi atât de timpii inter-sosire $T_n,\ n = 1,2,\ldots,$ cât şi de timpii de aşteptare $S_n = T_1+T_2+\ldots+T_n,\ S_0 = 0,$ interesul nostru focalizându-se pe repartițiile acestora. În figura 2.1 ilustrăm grafic reprezentarea timpilor specifici unui proces de reînnoire.

În cele ce urmează vom presupune că funcția de repartiție a timpilor inter-sosire este $F(t) = P\{T_n \leq t\},\ n = 1,2,\ldots,$ cu $F(0) = P\{T_n = 0\} < 1$ şi să notăm cu $\mu = E[T_n]$

media corespunzătoare acestora. Deoarece $S_n$ reprezintă suma a *n* variabile nenegative, independente şi identic repartizate, rezultă că funcţia de repartiţie corespunzătoare este dată de:

$$P\{S_n \le t\} = F^{(n)}(t) = F*F*...*F,$$

unde * reprezintă operatorul de convoluţie Stieltjes.

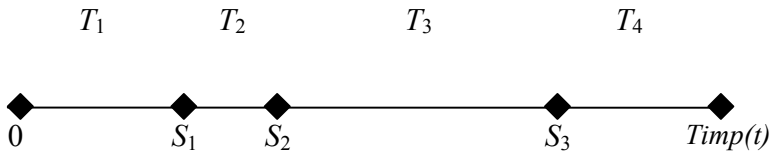



Figura 2.1. Timpii inter-sosire şi timpii de aşteptare ai unui proces de reînnoire.

Un proces de reînnoire se referă, în primul rând, la procesul de numărare *{N(t), t ≥ 0}*. Pentru a găsi repartiţia acestuia, să reamintim echivalenţa menţionată şi în cazul particular al procesului Poisson, relaţia ce leagă variabilele *N(t)* şi $S_n$.

$$S_n \le t \Leftrightarrow N(t) \ge n, \text{ care implică:}$$

$$P\{N(t) = n\} = P\{N(t) \ge n\} - P\{N(t) \ge n+1\} = F^{(n)}(t) - F^{(n+1)}(t)$$

Tot în acest context vom remarca şi faptul că:

$$N(t) = max\{n / S_n \le t\}$$

Media variabilei *N(t)*, notată *m(t) = E[N(t)],* se numeşte *funcţie de reînnoire* şi reprezintă numărul mediu de "reînnoiri" înregistrate până la momentul *t*. Printr-un simplu calcul, observăm că:

$$m(t) = \sum_{n=1}^{\infty} P\{N(t) \ge n\} = \sum_{n=1}^{\infty} P\{S_n \le t\} = \sum_{n=1}^{\infty} F^{(n)}(t)$$

Funcţia de reînnoire este poate mai importantă din alt punct de vedere decât acela că reprezintă media unei variabile aleatoare, ea determinând în mod unic procesul de reînnoire, în sensul că există o corespondenţă biunivocă între funcţia de repartiţie *F* şi

funcția de reînnoire *m(t)*. Fără a intra în amănunte privind demonstrația acestui rezultat, vom menționa doar ecuația integrală:

$$m(t) = F(t) + \int_0^t m(t-x)dF(x)$$

numită şi *ecuația de reînnoire*, în care funcția de reînnoire m(t) este necunoscută, precum şi relația dintre transformatele Laplace-Stieltjes ale celor două funcții:

$$m^*(s) = \frac{F^*(s)}{1-F^*(s)}$$

relația ce ilustrează afirmația făcută mai sus privind faptul că fiecare dintre ele o determină unic pe cealaltă.

*Exemplul 2.2.* a) Să considerăm că repartiția timpilor inter-sosire este gamma de parametri *(λ, 2)*, deci $F(t) = 1-(1+\lambda t)e^{-\lambda t}$. Rezultă că funcția de reînnoire este:

$$m(t) = \frac{\lambda t}{2} - \frac{1}{4} + \frac{1}{4}e^{-2\lambda t},$$

Să ne reamintim că transformata Laplace-Stieltjes a repartiției de mai sus este dată

de relația: $F^*(s) = (\frac{\lambda}{\lambda+s})^2$, ceea ce implică: $m^*(s) = \frac{\lambda}{2s} - \frac{1}{4}\cdot\frac{2\lambda}{s+2\lambda}$,

b) Să considerăm că de unde $m(t) = \frac{\lambda t}{2} - \frac{1}{4} + \frac{1}{4}e^{-2\lambda t}$. funcția de reînnoire este dată de: $m(t) = 2t,\ t \geq 0$. Deoarece funcția de reînnoire corespunde unui proces Poisson, având rata $\lambda = 2$, rezultă că funcția de repartiție a timpilor inter-sosire este o repartiție exponențială de medie *1/2*.

În general, deoarece transformata Laplace-Stieltjes a repartiției exponențiale:

$F(t) = 1-e^{-\lambda t}$, $t \geq 0$ este dată de: $F^*(s) = \frac{\lambda}{s+\lambda}$, obținem: $m^*(s) = \frac{\lambda}{s}$ $m^*(s) = \frac{\lambda}{s}$

care implică $m(t) = \lambda t$.

Dacă considerăm $N(t)/t$ ca fiind rata de reînnoiri până la momentul $t$, rezultă că aceasta converge, cu probabilitatea egală cu $1$, către numărul $1/\mu$ ,cunoscut sub denumirea de *rata* procesului de reînnoire, adică rata numărului mediu de reînnoiri converge, de asemenea, la rata $1/\mu$ a procesului.

O generalizare a noțiunii de proces de reînnoire este următoarea. Să presupunem existența unui proces de reînnoire $\{N(t), t \geq 0\}$, cu următoarea proprietate: de fiecare dată când apare o "reînnoire" se oferă o recompensă $R_n$. Să presupunem că recompensele $R_n$ reprezintă variabile aleatoare independente şi identic repartizate. De asemenea, nu vom ignora cazul în care $R_n$ poate depinde de $T_n$. Un asemenea proces se numeşte *proces de reînnoire cu recompensă*.

Se poate arăta că, dacă notăm: $R(t) = \sum_{n=1}^{N(t)} R_n$ recompensa totală obținută până la

momentul $t$, atunci cu probabilitatea $1$ avem:

$$\lim_{t\to\infty} \frac{R(t)}{t} = \frac{E[R_n]}{E[T_n]} \text{ şi } \lim_{t\to\infty} \frac{E[R(t)]}{t} = \frac{E[R_n]}{E[T_n]}.$$

În cazul proceselor de reînnoire cu recompensă este introdusă noțiunea de *ciclu*, privit ca momentul în care apare un nou eveniment (vom spune că "un ciclu s-a încheiat" atunci când în cadru procesului un nou eveniment şi-a făcut apariția). Rezultatul de mai sus se poate traduce prin aceea că, pentru un interval de timp suficient de mare, astfel încât procesul să fie stabilizat, recompensa medie pe unitatea de timp este egală cu raportul dintre media recompensei obținute într-un ciclu şi media duratei ciclului, ceea ce, intuitiv, era de aşteptat.

*Exemplu 2.3:* Să presupunem că într-un nod de comutație a pachetelor a unei rețele de calculatoare sosesc pachete conform unui proces de reînnoire, având media timpilor inter-sosire egală cu μ. De câte ori în nod se găsesc $N$ pachete, valoarea lui $N$

reprezentând un "prag" de limitare a numărului de pachete în aşteptarea prelucrării lor. Să presupunem că pentru un număr de *n* pachete sosite, costul pe unitate de timp este de *nc* lei. Se cere costul mediu cerut de sistem, în regim stabilizat de funcționare.

Vom presupune în acest caz că un ciclu este încheiat atunci când pachet este deservit, un loc se eliberează. Rezultă că media duratei unui ciclu este egală cu produsul dintre numărul necesar de locuri ce implică o deservire şi media timpului dintre două sosiri a pachetelor.

$$E[\text{durata unui ciclu}] = N\mu$$

Dacă notăm cu $U_n$ timpul trecut între cea de-a *n*-a şi cea de-a *(n+1)* sosire dintr-un ciclu, atunci media costului pentru un ciclu este dată de:

$$E[\text{costul per ciclu}] = E[cU_1+2cU_2+\ldots+(N-1)cU_{N-1}],$$

de unde, ținând cont că $E[U_n] = \mu$, rezultă că:

$$E[\text{cos}\,tul\ \ per\ ciclu] = c\mu\frac{N}{2}(N-1).$$

Conform rezultatului de mai sus, rezultă că raportul dintre costul mediu al unui ciclu şi media duratei ciclului reprezintă costul mediu corespunzător nodului, deci:

$$\text{cos}\,tul\ mediu\ al\ nodului = \frac{c(N-1)}{2}.$$

Ca o aplicație numerică pentru situația de mai sus, să considerăm cazul în care rata de sosire a pachetelor pentru internare este de $\mu = 2.4$ pachete / *s*, costul normalizat pe unitate de timp este *c = 10 u.m* (unități monetare), iar de fiecare dată când un loc se eliberează, se "acordă o recompensă" de *48 u.m*. Se cere, în acest context, să se găsească numărul *N* de p pachete care să minimizeze costul mediu al nodului, în regim stabilizat de funcționare.

Pentru a răspunde la această întrebare, să observăm că funcția ce dă costul mediu pe unitate de timp este, în ipoteza de mai sus:

$$C = \frac{5N^2-5+20}{N}$$

Aplicând acum aparatul clasic al Analizei matematice, obținem valoarea lui *N* care

minimizează costul mediu, $N = 2$, costul mediu pe unitate de timp astfel obținut fiind egal cu 15 u.m.

## 2.3. Sistem de aşteptare elementar

În forma sa cea mai generală un sistem de aşteptare elementar compus din surse, şiruri şi serveri, plasați în zona de servire, cu circuit închis sau deschis se poate urmări ca în fig. 2.2.

Un număr de *unități* intră în *sistem* pentru a obține un serviciu. Ele sunt produse de *surse* exterioare sistemului, surse ce pot fi într-un număr finit sau infinit. Sursele funcționează deci ca un generator, ce poate furniza un număr *limitat* sau *nelimitat* de unități. Unitățile intră în sistem şi găsesc aici, în zona de servire, un număr de elemente ce trebuie să le asigure serviciul, numite *serveri* (*resurse*). Dacă numărul acestora este limitat, atunci unitățile pot fi îndrumate spre o aşteptare, formând unul sau mai multe *şiruri de aşteptare*. Şirurile sunt organizate după anumite reguli de *priorități*, iar maniera de prelucrare a unităților din şir în vederea prelucrărilor de către resurse este asemenea variabilă. Unitățile pot fi considerate *clienți* ce solicită serverul şi uneori chiar aşa sunt denumite.

Îndată ce unitatea a fost prelucrată, ea părăseşte sistemul prin *ieşire*, putând în unele situații de exploatare să fie repoziționată la *intrarea* sistemului, o dată sau chiar de mai multe ori, înainte de a regăsi sursa-generatoare.

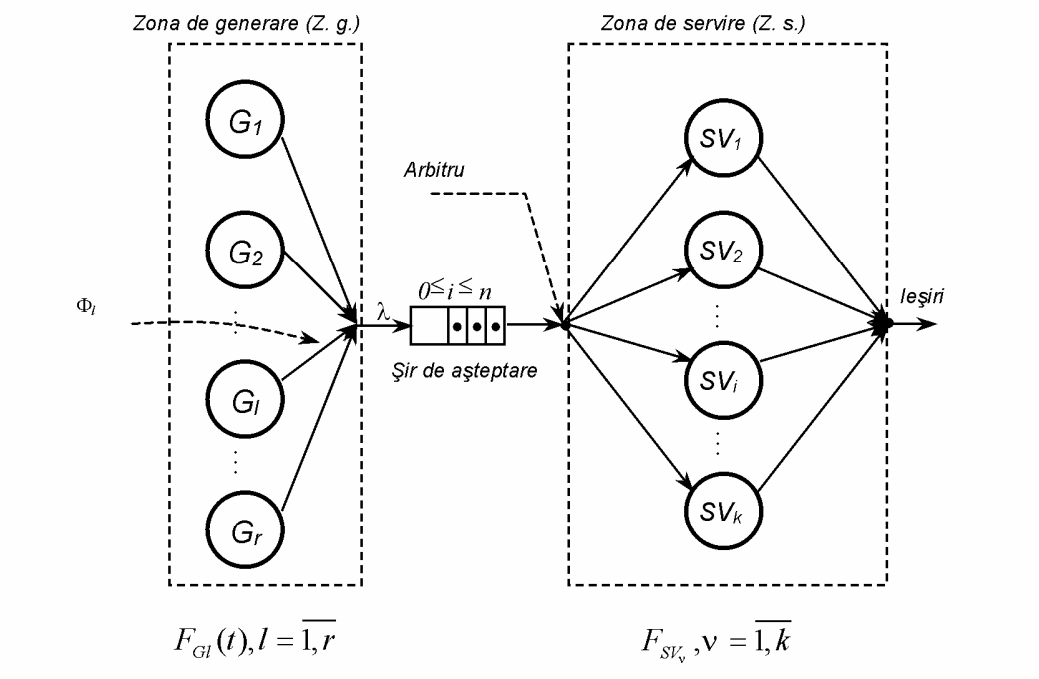
Din exterior, sistemul SA elementar este perceptibil doar prin intrările şi ieşirile sale, eventual prin identitatea unităților care intră sau care ies din el. Funcționarea sistemului se manifestă deci prin transferarea unităților de la intrare la ieşirea sa.

Analiza matematică a unui sistem SA elementar (denumit simplu în cele ce urmează *sistem SA*) trebuie să considere următoarele elemente:

- secvența momentelor $0 \le t_1 \le t_2 \le ... \le t_j$ ... de sosire a unităților în sistem;

- secvența $u_1, u_2, ..., u_j, ...$ a duratelor de servire a unităților $1,2,...,j,...$;

- disciplina de servire, care precizează ordinea în care sunt serviți clienții ce ajung în şir (*FIFO*, *LIFO* etc.)




$\Phi_l$ – *fluxul parțial de cereri generate de generatorul* $G_l$;
$F_{Gl}(t)$ – *funcția de repartiție a intervalelor de timp ale intersosirilor cererilor în SA generate de* $G_l$;
$F_{SV}(t)$ – *funcția de repartiție a duratei de servire a cererii de catre serverul* $SV_v$.

Fig. 2.2. Schema generală a unui sistem de aşteptare elementar

Putem face o clasificare a fenomenelor de aşteptare fie după numărul unităților din sursă şi numărul serverilor, fie după natura sosirilor sau a servirilor. Astfel avem situațiile: un şir, un server; un şir, mai mulți serveri; mai multe şiruri, un server; mai multe şiruri, mai mulți serveri sau sosiri constante, servicii constante; sosiri constante, servicii aleatoare; sosiri aleatoare, servicii constante, servicii aleatoare etc.

Datorită diversității lor, sistemele au trebuit să fie categorisite şi în acest scop se foloseşte clasificarea şi notația Kendall, corespunzător căreia fiecărui sistem i se ataşează o formulă cu 6 caractere alfanumerice *SA* : *A / B / k / m / N /* (*Disciplină*), care au următoarele semnificații:

- $A$: natura procesului de sosire a cererilor,
- $B$: natura procesului de servire a cererilor,
- $k$: numărul serverilor în zona de servire a sistemului,
- $N$: numărul maxim al unităților în sursa generatoare,
- $m$: numărul total al locurilor disponibile în şirul de aşteptare,
- *Disciplină*: disciplina de servire a cererilor în sistem.

Pentru a defini principalele procese de intrare sau de servire se folosesc simbolurile:

- $M$: lege exponențială, $C_v = 1$.
- $D$: lege constantă, $C_v = 0$.
- $E_k$: lege Erlang de ordin $k$, $0 < C_v < 1$.
- $H_r$: lege Hiperexponențială de ordin $r$, $0 < C_v < +\infty$.
- $G$: lege generală, $0 \le C_v < +\infty$,

unde $C_v = \sqrt{D[X]} / M[X]$ este coeficientul de variație respectiv, $D[X]$ este dispersia, iar $M[X]$ este speranța matematică a legii de repartiție a variabilei aleatoare $X$.

Principalele discipline de servire a clienților sunt:

*FiFo* (“first in first out”) sau (*paps*) - primul ajuns, primul servit (“premier arrive premier servi”),

*LiFo* (“last in first out”) sau (*daps*) - ultimul sosit, primul servit (“dernier arrive premier servi”),

*FiRo* (“first in random out”) sau *a* - aleatoriu (“aleatoire”).

Numărul combinațiilor tuturor acestor categorii de proces şi disciplinile de servire este considerabil de mare şi de aceea a fost necesară această clasificare riguroasă tip Kendall, unanim acceptată.

În cele ce urmează ne propunem să analizăm câteva modele de sisteme de aşteptare, întâlnite în practică, determinând în acelaşi timp diferiți indicatori numerici de performanță legați de astfel de modele.

Într-o problemă de aşteptare se întâlnesc următorii indicatori numerici de performanţă principali:

1) $m$ - numărul locurilor de aşteptare ale unităţilor populaţiei din sursă care sosesc în sistem, $m$ poate fi finit sau infinit;

2) $k$ - numărul serverilor (unităţilor de serviciu);

3) $\pi_n(t)$ - probabilitatea ca în sistemul de aşteptare să se găsească $n$ unităţi la momentul $t$ oarecare. Pentru simplificarea scrierii în cele ce urmează vom nota doar prin $\pi_n$.

4) $n(t)$ - numărul de unităţi ce se găsesc în sistemul de aşteptare (şir + serviciu) la momentul $t$ şi care este o variabilă aleatoare cu distribuţia discretă

$$n(t):\begin{pmatrix} 0 & 1 & 2 & \dots & n & \dots & m \\ \pi_0 & \pi_1 & \pi_2 & \dots & \pi_n & \dots & \pi_m \end{pmatrix}$$

5) $\overline{n}_{SA}(t)$ - numărul mediu de unităţi în sistem la un moment $t$, care este valoarea medie a variabilei $n(t)$, adică:

Evident când $m=\infty, \overline{n}_{SA}(t)=\sum_{n=0}^{m} n\pi_{n,}$ ceea ce impune ca această serie să fie convergentă.

5) $n_ş(t)$ - numărul de unităţi din şirul de aşteptare, la un moment $t$; $n_ş(t)$ este o variabilă aleatoare cu distribuţia:

$$n_ş(t):\begin{pmatrix} 0 & 1\dots n-k\dots m-k \\ \pi_0+\pi_1+\dots+\pi_k & \pi_{k+1}\dots\pi_n\dots\pi_m \end{pmatrix},$$

deoarece există unităţi în şirul de aşteptare atunci când $n$ depăşeşte numărul de serveri $k$.

5) $\overline{n}_л(\mathrm{t})$ - numărul mediu de unităţi ce se află în şir şi are forma:

$$\overline{n}_л(t)=\sum_{n=s+1}^{m}(n-k)\pi_n.$$

Pentru $m = \infty,\ \bar{n}_{ş}(t) = \sum_{n=s+1}^{m}(n-k)\pi_n$ şi seria va trebui să fie convergentă.

6) $n_s(t)$ - numărul de unităţi ce sunt servite la un moment *t*. Evident $n_s(t) = n_{SA}(t) - n_{ş}(t)$, deci $n_s(t)$ este o variabilă aleatorie.

6) $\bar{n}_s(t)$ - numărul mediu de unităţi ce sunt servite la momentul *t* şi $\bar{n}_s(t) = \bar{n}_{SA}(t) - \bar{n}_{ş}(t)$ , din proprietatea mediei a unei variabile aleatoare.

7) $\pi(n(t) > l)$ - probabilitatea ca numărul unităţilor în sistem la momentul *t* să fie mai mare decât *l*. Dar

$$\pi(n(t) > l) = 1 - \pi(n(t) \le l) = 1 - (\pi_0+\pi_1+...+\pi_l).$$

8) $\bar{t}_{ş}$ - timpul mediu de aşteptare a unei unităţi în şir.

9) $\bar{t}_{SA}$ - timpul mediu de aşteptare a unei unităţi în sistemul de aşteptare.

Ultimii doi indicatori depind de legile serviciului şi a sosirilor şi vor fi determinaţi pentru fiecare model în parte în conformitate cu legea lui Little a SA:

$\bar{t}_{SA} = \bar{n}_{SA} / \bar{\lambda}$, unde $\bar{\lambda} = \sum_{i=1}^{m}\lambda_i \cdot \pi_i$

este rata medie de sosire a unităţilor (cererilor, clienţilor, tasck-urilor) în *SA*.

## 2.3. Legi probabilistice ale sosirilor şi servirilor

Fie *X* variabilă aleatoare discretă ce reprezintă numărul de unităţi sosite în unitatea de timp, într-un sistem de aşteptare. Ne propunem să cercetăm repartiţia variabilei *X* în următoarele condiţii:

a) probabilitatea sosirii unei unităţi la un moment dat este constantă şi nu depinde de ceea ce s-a întâmplat anterior.

b) probabilitatea unei sosiri într-un interval de timp foarte mic ($t$, $t+\Delta t$) este proporţională cu lungimea intervalului $\Delta t$, oricare ar fi *t*, adică este

$\lambda\Delta t + 0(\Delta t)$, unde $\lim_{\Delta t\to 0} 0(\Delta t) = 0$ şi $\lim_{\Delta t\to 0}\frac{0(\Delta t)}{\Delta t} = 0.$

c) probabilitatea ca în intervalul de timp ($t$, $t+\Delta t$) să avem mai mult decât o sosire este aproximativ egală cu zero, când $\Delta t$ îndeplineşte condiția b).

**Propoziția 2.1.** *Variabila aleatoare X ce reprezintă numărul de unități sosite pe unitatea de timp într-un sistem de aşteptare are repartiția Poisson.* □

*Demonstrație.* Vom determina probabilitatea evenimentului $A_n(t)$, ca într-un interval de timp (0, $t$) să avem $n$ sosiri, adică $P_n(t) = P(X = A_n(t))$, procedând din aproape în aproape pentru $n$= 0,1,...

Să notăm cu $A_0(t+\Delta t)$ evenimentul ca în intervalul de timp de lungime $t+\Delta t$ să avem zero sosiri. Acest eveniment are loc atunci, când nu avem nici o sosire în intervalul de timp (0 + $t$) şi ($t$, $t+\Delta t$) şi scriem:

$$A_0(t+\Delta t) = A_0(t) \cap A_0(\Delta t).$$

Din condiția a) rezultă:

$$P(A_0(t+\Delta t)) = P(A_0(t)) \cap P(A_0(\Delta t)) \qquad (2.1)$$

Din c) rezultă:

$$P(A_0(\Delta t)) = 1 - P(A_1(\Delta t)),$$

iar din b) deducem că:

$$P(A_0(\Delta t)) = 1 - \lambda \Delta t.$$

Atunci relația (2.1) se va scrie

$$P_0(t+\Delta t) = P_0(t)(1-\lambda \Delta t),$$

de unde

$$P_0(t+\Delta t) - P_0(t) = -\lambda \cdot P_0(t)\Delta t,$$

deci

$$\frac{P_0(t+\Delta t) - P_0(t)}{\Delta t} = -\lambda \cdot P_0(t).$$

Pentru $\Delta t \to 0$ obținem ecuația diferențială lineară de ordinul întâi în $P_0(t)$: $P_o'(t) = -\lambda \cdot P_0(t)$. (2.2)

a cărei soluție este:

$$P_0(t) = Ce^{-\lambda t}. \tag{2.3}$$

Pentru determinarea constantei $C$ din ecuația (2.3), ținem seama că la timpul $t = 0$, evenimentul de a nu avea nici o sosire este eveniment sigur, deci $P_0(0) = 1$. Înlocuim în (2.3) pe $t = 0$ şi avem

$$P_0(0) = C, \text{ deci } C = 1.$$

Aşadar, probabilitatea de a avea zero sosiri în intervalul $(0, t)$ este

$$P_0(t) = Ce^{-\lambda t}.$$

Să determinăm acum $P_n(t+\Delta t)$, adică probabilitatea ca în intervalul de timp $(0, t+\Delta t)$ să avem $n$ sosiri. În intervalul $(0, t+\Delta t)$ conform condiției c) poate sosi una sau nici o unitate, de aceea vom scrie:

$$A_n(t + \Delta t) = (A_n(t) \cap A_0(\Delta t) \cup A_{n-1}(t) \cap A_1(\Delta t)),$$

de unde

$$P_n(t + \Delta t) = P_n(t) \cdot P_0(\Delta t) + P_{n-1}(t) \cdot A_1(\Delta t)$$

sau

$$P_n(t + \Delta t) = P_n(t)(1 - \lambda\Delta t) + P_{n-1}(t)(\lambda\Delta t + 0(\Delta t)).$$

Prelucrând această egalitate, împărțind prin $\Delta t$ şi trecând la limită avem:

$$P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}(t). \tag{2.4}$$

Pentru $n = 1$ obținem

$$P_1'(t) + \lambda P_1(t) - \lambda e^{-\lambda t} = 0,$$

ecuație diferențială de ordinul întâi în $P_1(t)$ a cărei soluție este:

$$P_1(t) = e^{-\int \lambda dt}(C + \int \lambda e^{-\lambda t} e^{\int \lambda dt} dt) = e^{-\lambda t}(C + \lambda t).$$

Pentru determinarea constantei $C$, folosim egalitatea $P_1(0) = 0$, adică evenimentul că la momentul inițial $t = 0$ să avem o unitate sosită este imposibil. Atunci, făcând $t = 0$ în soluția generală, avem:

$P_1(0) = e^{\lambda 0}(C + \lambda \cdot 0)$, deci $0 = 1.C$, de unde $C = 0$.

Atunci

$$P_1(t) = \lambda t e^{-\lambda t}.$$

Analog pentru $n = 2$ obținem

$$P_2'(t) + \lambda P_2(t) - \lambda P_1(t) = 0$$

sau

$$P_2'(t) + \lambda P_2(t) - \lambda^2 t e^{-\lambda t} = 0,$$

a cărei soluție generală este

$$P_2(t) = e^{-\lambda t}(c + \frac{1}{2}\lambda^2 t^2)$$

şi cum

$P_2(0)=0$, avem $P_2(t) = \frac{1}{2}\lambda^2 t^2 e^{-\lambda t}.$

Prin generalizare putem scrie:

$$P_n(t) = \frac{1}{n!}\lambda^n t^n e^{-\lambda t}. \qquad (2.5)$$

Pentru a accepta această formă, vom recurge la inducția matematică. Şi cum etapa de verificare a fost efectuată, presupunem adevărata relație (2.5) pentru $n$ natural oarecare.

Să demonstrăm că este adevărată şi pentru $n + 1$, pentru care relația (2.4) se va scrie astfel:

$$P_{n+1}^{'}(t) = -\lambda P_{n+1}(t) + \lambda P_n(t)$$

sau înlocuind $P_n(t)$ din (2.5) obținem:

$$P'_{n+1}(t) + \lambda P_{n+1}(t) - \frac{1}{n!}\lambda^{n+1} t^n e^{-\lambda t} = 0,$$

ecuație diferențială liniară de ordinul întâi, cu soluția

$$P'_{n+1}(t) = e^{-\lambda t}(C + \frac{1}{(n+1)!}\lambda^{n+1}\lambda^{n+1}).$$

Şi cum $P_{n+1}(0) = 0$, rezultă că $C = 0$, şi deci

$$P_{n+1}(t) = \frac{1}{(n+1)!}\lambda^{n+1} t^{n+1} e^{-\lambda t},$$

formulă ce confirmă valabilitatea relației (2.5) pentru orice *n* natural.

Cu expresia obținută pentru $P_n(t)$, putem afirma că într-un fenomen de aşteptare în care sunt îndeplinite condițiile a), b), c) numărul de sosiri în intervalul de timp (0, *t*) este o variabilă aleatoare poissoniană, cu parametrul $\lambda t$.

Evident pentru $t = 1$ (unitatea de timp) avem:

$$P_n(1) = \frac{1}{n!}\lambda^n e^{-\lambda t}$$

şi rezultă că λ coeficientul de proporționalitate introdus prin condiția b) reprezintă numărul mediu de unități sosite în unitatea de timp.

*Observație*. Acceptând condițiile a), b), c) şi pentru numărul unităților servite de către un server ce lucrează fără întrerupere, obținem printr-un raționament analog, că numărul de servicii ce pot fi făcute de un server într-un timp *t*, este o variabilă poissoniană. Şi dacă μ este coeficientul de proporționalitate introdus porin b), el reprezintă numărul mediu de unități servire în unitatea de timp.

În continuare să cercetăm relația variabilei *Y*, care reprezintă timpul dintre două sosiri consecutive. Evident *Y* este o variabilă aleatoare continuă.

*Propoziția 2.2***.** *Variabila aleatoare Y are repartiția exponențială.* □

*Demonstrație.* Considerând că momentul inițial, momentul în care a sosit prima din cele două unități, să calculăm probabilitatea ca timpul *Y* dintre cele două sosiri să fie mai mare decât un timp oarecare $t > 0$. Această probabilitate va fi evident $P_0(t)$ adică probabilitatea ca în timpul *t* să nu avem nici o sosire. Deci

$$P(Y > t) = P_0(t) = e^{-\lambda t},$$

de unde

$$P(Y \le t) = 1 - e^{-\lambda t}$$

egalitate ce ne spune că funcția de repartiție a variabilei $Y$ este

$$F(t) = 1 - e^{-\lambda t}$$

specifică variabilei $Y$ de repartiție exponențială, cu densitatea de repartiție

$$f(t) = F'(t) = \lambda e^{-\lambda t}$$

şi deci

$$M(Y) = \int_0^\infty t f(t) dt = \int_0^\infty t \lambda e^{-\lambda t} dt = \frac{1}{\lambda}$$

Aşadar, intervalul mediu dintre două sosiri consecutive este inversul numărului mediu de sosiri în unitatea de timp.

*Observație.* Printr-un raționament analog putem deduce că timpul dintre două servicii consecutive are o repartiție exponențială şi că intervalul mediu dintre două servicii consecutive este inversul numărului mediu de servicii în unitatea de timp (în cazul nostru $1/\mu$).

*Exemplu*. Cu ajutorul metodelor statistice s-a stabilit că sosirile mesajelor la un calculator sunt poissoniene cu numărul mediu de 120 mesaje pe oră, iar timpul mediu de prelucrare a unui mesaj este de 20 secunde. Să se determine intervalul mediu dintre două sosiri consecutive, numărul mediu de mesaje prelucrate într-un minut şi probabilitatea că în 75 de secunde să nu sosească nici un mesaj.

*Rezolvare.* Considerăm drept unitate de timp minutul. Numărul mediu de sosiri pe minut va fi de 2 mesaje, deci $\lambda = 2$. Intervalul mediu dintre două sosiri va fi $1/\lambda = 1/2$ minute, adică 30 secunde.

Timpul mediu de prelucrare a unui mesaj fiind de 20 secunde = 1/3 minute = $1/\mu$, rezultă că numărul mediu de mesaje ce pot fi prelucrate într-un minut este $\mu = 3$ mesaje.

Deoarece 75 de secunde fac 5/4 minute, rezultă, folosind relația (2.5), că

$$P_0(\frac{5}{4}) = \frac{1}{0}2^0(\frac{5}{4})^0 e^{-2*5/4} = e^{-2.5} = 0.083$$

*Observație.* Nu trebuie de înțeles că în orice fenomen de aşteptare sosirile şi serviciile se supun numai repartițiilor Poisson şi exponențială. Există cazuri că ele se supun şi altor legi cum ar fi: uniformă, normală, $\chi^2$, Erlang, Hiperexponențială, COX, etc. Dar deaorece mai frecvente sunt primele vom prezenta câteva modele de aşteptare pentru care sosirile sunt poissoniene şi timpul de sosire este exponențial.

## 2.4. Deducerea ecuațiilor de stare pentru un fenomen de aşteptare în regim staționar

În cele ce urmează vom construi sistemul ecuațiilor de stare în cazul general, respectiv în cazul procesului Poisson de naştere şi de moarte, presupunând cunoscute următoarele:

1) probabilitatea de trecere din starea $s_n$ (în sistemul de aşteptare există $n$ unități) în starea $s_{n+1}$ (în sistemul de aşteptare există $n$+1 unități) în intervalul de timp ($t$, $t$+$\Delta t$) este: $\lambda_n \Delta t + 0(\Delta t)$ şi corespunde unei sosiri în sistem;

2) probabilitatea de trecere din starea $s_n$ în starea $s_{n-1}$ (în sistemul de aşteptare există $n$-1 unități) în intervalul de timp ($t$, $t$+$\Delta t$) este:

$$\mu_n \Delta t + 0(\Delta t)$$

şi corespunde unei plecări (serviri);

3) probabilitatea de trecere din starea $s_n$ în starea $s_{n+k}$ sau într-o stare $s_{n-k}$, cu $k \in N$, $k$>1 este $0(\Delta t)$;

4) probabilitatea ca în intervalul ($t$, $t$+$\Delta t$) să nu aibă loc nici o modificare de stare este:

$$1 - (\lambda_n + \mu_n)\Delta t + 0(\Delta t).$$

Observăm că probabilitatea evenimentului contrar celui din cazul 1 (în sistemul de aşteptare există *n* unităţi şi nu soseşte nici o unitate în intervalul de timp $(t, t+\Delta t)$) va fi evident

$$1 - \lambda_n \Delta t + 0(\Delta t),$$

iar probabilitatea evenimentului contrar celui din cazul 2 (în sistemul de aşteptare există *n* unităţi şi nu pleacă nici o unitate în intervalul de timp $(t, t+\Delta t)$) va fi

$$1 - \mu_n \Delta t + 0(\Delta t).$$

Din prezentarea indicatorilor unui fenomen de aşteptare am văzut că majoritatea dintre ei se exprimă în funcţie de $\pi_n(t)$ - probabilitatea că în sistemul de aşteptare, la momentul *t* să existe *n* unităţi. De aceea ne propunem determinarea acestei probabilităţi pentru diferite valori ale lui $n = 0,1,2,...$

*Cazul* $n = 0$. Să calculăm probabilitatea ca în sistemul de aşteptare să nu existe nici o unitate la momentul $t + \Delta t$, ţinând cont de situaţia posibilă la momentul *t*.

În sistem nu există nici o unitate la momentul $t + \Delta t$ când avem una din situaţiile: a) (nu există nici o unitate la momentul *t*) şi (nu soseşte nici o unitate în intervalul $(t, t + \Delta t)$) sau b) (există o singură unitate la momentul *t*) şi (pleacă o unitate în intervalul $(t, t + \Delta t)$) şi (nu soseşte nici o unitate în intervalul $(t, t + \Delta t)$).

Situaţiile a) şi b) de mai sus reprezintă evenimente incompatibile care sunt formate din evenimente independente şi neglijând funcţiile $0(\Delta t)$ care tind către zero, când $\Delta t$ este foarte mic, se obţine probabilitatea de stare căutată, adică

$$\pi_0(t + \Delta t) = \pi_0(t)(1 - \lambda_0 \Delta t) + \pi_1(t) * \mu_1 \Delta t(1 - \lambda_1 \Delta t).$$

Se trece în stânga $\pi_0(t)$ şi se obţine

$$\pi_0(t + \Delta t) - \pi_0(t) = \lambda_0 \pi_0(t)\Delta t + \mu_1 \pi_1(t)\Delta t - \lambda_1 \mu_1 \pi_1(t)(\Delta t)^2.$$

Se împarte apoi cu $\Delta t$ şi se trece la limită

$$\lim_{\Delta t \to 0} \frac{\pi_0(t + \Delta t) - \pi_0(t)}{\Delta t} = -\lambda_0 \pi_0(t) + \mu_1 \pi_1(t)$$

sau

$$\pi_0^{'}(t) = -\lambda_0 \pi_0(t) + \mu_1 \pi_1(t). \qquad (2.6)$$

*Cazul* $n > 0$. Să calculăm probabilitatea ca în sistemul de aşteptare să nu existe $n$ unităţi la momentul $t + \Delta t$; ţinând cont de situaţia posibilă la momentul $t$.

În sistem există $n$ unităţi la momentul $t + \Delta t$ când avem una din situaţiile: a) (la momentul $t$ există în sistem $n$ unităţi) şi (nu soseşte nici o unitate în intervalul $(t, t + \Delta t)$) şi (nu pleacă nici o unitate în intervalul $(t, t + \Delta t)$) sau b) (există în sistem $n + 1$ unităţi la momentul $t$) şi (pleacă o unitate în intervalul $(t, t + \Delta t)$) şi (nu soseşte nici o unitate în intervalul $(t, t + \Delta t)$) sau c) (există $n$ - 1 unităţi în sistem la momentul $t$) şi (soseşte o unitate în intervalul $(t, t + \Delta t)$) şi (nu pleacă nici o unitate în intervalul $(t, t + \Delta t)$) sau d) (există $n$ unităţi în sistem la momentul $t$) şi (soseşte o unitate în intervalul $(t, t + \Delta t)$) şi (pleacă o unitate în intervalul $(t, t + \Delta t)$).

Deci evenimentul de a exista $n$ unităţi la momentul $t + \Delta t$ în sistemul de aşteptare este reuniunea a patru evenimente incompatibile şi fiecare dintre acestea este intersecţie a câte trei evenimente independente. Neglijând funcţiile $0(\Delta t)$, probabilitatea căutată va fi

$$\pi_n(t + \Delta t) = \pi_n(t)(1 - \lambda_n \Delta t)(1 - \mu_n \Delta t) + \pi_{n+1}(t)(1 - \lambda_{n+1} \Delta t)\mu_{n+1} \Delta t +)^2.$$
$$+ \pi_{n-1}(t)\lambda_{n-1} \Delta t(1 - \mu_{n-1} \Delta t) + \pi_n(t)\lambda_n \mu_n (\Delta t)^2.$$

Se trece în membrul stâng $\pi_n(t)$, se împarte la $\Delta t$ şi obţinem:

$$\lim_{\Delta t \to 0} \frac{\pi_n(t + \Delta t)}{\Delta t} = -(\lambda_n + \mu_n)\pi_n(t) + \mu_{n+1}\pi_{n+1}(t),$$

adică

$$\pi_n^{'}(t) = \lambda_{n-1}\pi_{n-1}(t) - (\lambda_n + \mu_n)\pi_n(t) + \mu_{n+1}\pi_{n+1}(t). \qquad (2.7)$$

Relaţiile (2.6) şi (2.7) formează sistemul ecuaţiilor de stare care în cazul unui regim staţionar în care probabilităţile $\pi_n(t)$ nu depind de momentul $t$, deci sunt constante la orice moment de timp, adică $\pi_k(t) = \pi_k$, şi deci , $k = 0,1,2,...$ devine

$$\begin{cases} -\lambda_0\pi_0 + \mu_1\pi_1 = 0 \\ \lambda_{n-1}\pi_{n-1} - (\lambda_n + \mu_n)\pi_n + \mu_{n+1}\pi_{n+1} = 0 \end{cases} \tag{2.8}$$

cunoscut sub numele de *sistemul ecuațiilor de stare în regim staționar*. Convenim ca în acest caz nici ceilalți indicatori să nu se mai scrie funcții de *t*.

Din prima ecuație a sistemului (2.8) avem

$$\pi_1 = \frac{\lambda_0}{\mu_1}\pi_0,$$

iar dacă în a doua ecuație facem $n = 1$, avem

$$\pi_2 = \frac{\lambda_0\lambda_1}{\mu_1\mu_2}\pi_0.$$

Presupunând că pentru $n = k$ -1 şi $n = k$, pentru $k$ număr natural oarecare, avem

$$\pi_k = \frac{\lambda_0 \ldots \lambda_{k-1}}{\mu_1 \ldots \mu_{k+2}}\pi_0. \quad \text{şi} \quad \pi_{k+1} = \frac{\lambda_0 \ldots \lambda_k}{\mu_1 \ldots \mu_{k+1}}\pi_0.$$

atunci pentru $n = k + 1$ din ecuația a doua a sistemului (2.8) obținem

$$\pi_{k+2} = \frac{\lambda_0 \ldots \lambda_{k+1}}{\mu_1 \ldots \mu_{k+2}}\pi_0.$$

deci este verificată relația

$$\pi_n = \frac{\lambda_0 \ldots \lambda_{n-1}}{\mu_1 \ldots \mu_n}\pi_0, n = 1,2,... \tag{2.9}$$

ceea ce dă posibilitate de a evalua indicatorii numerici de performanță.

## 2.5. Modele de aşteptare

*Procese de naştere* (*sosiri*) *şi moarte* (*serviri*). Procesul de naştere şi moarte se caracterizează prin faptul că *sosirile* şi *serviciile* sunt poissoniene. Ecuațiile diferențiale corespunzătoare sunt

$$\pi_n'(t) = -(\lambda_n + \mu_n)\pi_n(t) + \lambda_{n-1}\pi_{n+1}(t) + \mu_{n+1}\pi_{n+1}(t), n > 0$$

$$\pi_0'(t) = -\lambda_0\pi_0(t) + \mu_1\pi_1(t),$$

unde $\lambda_n$ şi $\mu_n$ sunt funcţii de *n*. Din aceste ecuaţii rezultă diverse cazuri particulare ale fenomenelor de aşteptare.

*Modelul unui sistem cu un şir de aşteptare, un singur server şi populaţie infinită.* În acest caz $\lambda_n = \lambda$, $\mu_n = \mu$. Vom presupune că *probabilităţile* $\pi_n$ sunt *independente* de *t*, adică *procesul* corespunzător este *staţionar* şi *permanent*. Astfel obţinem sistemul liniar de ecuaţii

$$\lambda\pi_{n-1} + \mu\pi_{n+1} - (\lambda + \mu)\pi_n = 0, n > 0,$$
$$\lambda\pi_0 + \mu\pi_1 = 0$$

cu condiţia $\lambda < \mu$. Ţinând cont că $\pi_0+\pi_1+...=1$ deducem

$$\pi_0 = 1 - \rho, \rho = \frac{\lambda}{\mu}, \rho < 1,$$
$$\pi_n = \rho^n(1 - \rho), \rho < 1,$$

unde ρ se numeşte *factorul traficului* ($0 < \rho < 1$). Ţinând cont de expresia *densităţii de probabilitate*

$$\pi_n = \rho^n(1 - \rho), 0 < \rho < 1,$$

rezultă că

$$\max \pi_n(\rho) = (\frac{n}{n+1})^n \frac{1}{n+1}.$$

Probabilitatea $P(k \le n)$ este de forma

$$P(k \le n) = 1 - \rho^{n+1},$$

deci

$$P(k > n) = 1 - (1 - \rho^{n+1}) = \rho^{n+1},$$

*Numărul mediu de unităţi din sistemul de aşteptare*, adică din şirul de aşteptare şi în curs de servire, este ş

$$\overline{n}_{SA} = \sum_{n-1}^{\infty} n\pi_n = \frac{\lambda}{\mu - \lambda} = \frac{\rho}{1 - \rho}.$$

*Numărul mediu de unităţi din şirul de aşteptare* este

$$\overline{n}_{\text{â}} = \sum_{n=2}^{\infty}(n-1)\pi_n = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{\rho^2}{1-\rho}$$

*Numărul mediu de serveri activi din sistemul de aşteptare* este

$$\overline{n}_s = 1 - \pi_0 = \rho$$

*Timpul mediu de aşteptare în şir* este

$$\bar{t}_n = \frac{\overline{n}_{\hat{a}}}{\lambda} = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{\rho}{\mu(1-\rho)},$$

iar *timpul mediu de aşteptare în sistem* este

$$\bar{t}_s = \frac{\overline{n}_{SA}}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu(1-\rho)},$$

de unde *timpul mediu de servire* este

$$\bar{t}_s = \bar{t}_{SA} - \bar{t}_s = \frac{1}{\mu}.$$

Probabilitatea ca o unitate să aştepte în şirul de aşteptare un timp mai mare decât un timp *t* dat este

$$P(\bar{t}_s > t) = \frac{\lambda}{\mu} e^{-(\mu-\lambda)t},$$

Să considerăm un alt caz particular al procesului de naştere şi moarte corespunzător unei rate medii de sosire constantă şi unei rate medii de servire proporțional cu numărul de unități din sistemul de aşteptare, adică $\lambda_n = \lambda$, $\mu_n = n\mu$. În regim *permanent* avem

$$\pi_0 = e^{-\rho}, \pi_n = \frac{\rho^n e^{-\rho}}{n!}, \rho = \frac{\lambda}{\mu}.$$

*Modelul unui sistem cu un şir de aşteptare cu mai mulți serveri şi populație infinită*. Fie *n* numărul de unități din sistem şi *s* numărul de serveri care asigură servirea. Presupunem că *sosirile* sunt *poissoniene*, iar *serviciul* este *exponențial* cu $\rho < s$. În acest caz

$$\lambda_n = \lambda,$$

$$\mu_n = n\mu,\ 0 \le n < s,$$

$$\mu_n = s\mu,\ n \geq s.$$

În regim *permanent* $\pi_n(t) = \pi_n = const.$ pentru orice $n$. Din sistemul liniar de ecuații:

$$\lambda\pi_0 = \mu\pi_1,$$

$$(\lambda + n\mu)\pi_n = \lambda\pi_{n-1} + (n+1)\mu\pi_{n+1}, 1 \leq n < s,$$

$$(\lambda + s\mu)\pi_n = \lambda\pi_{n-1} + s\mu\pi_{n+1}, n \geq s,$$

rezultă

$$\pi_n = \frac{\rho}{n!}\pi_0, \rho = \frac{\lambda}{\mu}, 1 \leq n < s,$$

$$\pi_n = \frac{\rho^n}{s!s^{n-s}}\pi_0, \rho = \frac{\lambda}{\mu}, n \geq s.$$

Ținând cont că $\pi_0+\pi_1+...=1$, deducem

$$\pi_0 = \frac{1}{\sum_{n=0}^{s-1}\frac{\rho^n}{n!} + \frac{\rho^s}{s!(1-\rho/s)}},$$

de unde $\lim_{s\to\infty}\pi_0 = e^{-\rho}$.

*Valoarea medie a numărului de unități în sistemul de aşteptare* este

$$\overline{n}_{SA} = \sum_{n=0}^{\infty} n\pi_n = \rho,$$

iar *valoarea medie a numărului de unități din şir* este

$$\overline{n}_{â} = \sum_{n=s*1}^{\infty}(n-s)\pi_n = \sum_{n=s+1}^{\infty}\frac{(n-s)\rho^n}{s!s^{n-s}}\pi_0 = \frac{\rho^{s+1}}{s*s!(1-\rho/s)}\pi_0.$$

*Numărul mediu de serveri neocupați* este

$$\overline{v} = \sum_{n=0}^{s}(n-s)\pi_n = s - \rho,$$

Are loc relația

$$\pi(>0) = P(n \geq s) = \sum_{n=s}^{\infty}\pi_n = \frac{\rho^s}{s!(1-\rho/s)}\pi_0.$$

*Probabilitatea ca o unitate să aştepte în sistem* este

$$\pi(>0) = P(n \geq s) = \sum_{n=s}^{\infty} \pi_n = \frac{\rho^s}{s!(1-\rho/s)} \pi_0.$$

*Timpul mediu* (*durata medie*) *de aşteptare în şir* se obține din relația $\bar{n}_{\hat{a}} = \lambda \bar{t}_{\hat{a}}$, de unde

$$\bar{t}_{\hat{s}} = \frac{\bar{n}_{\hat{s}}}{\lambda} = \frac{\rho^s}{s \cdot s!(1-\rho/s)^2} \pi_0,$$

*Timpul mediu de aşteptare în sistem* este

$$\bar{t}_{SA} = \bar{t}_{\hat{s}} + \frac{1}{\mu}.$$

*Modelul unui sistem cu un şir de aşteptare, server unic, populație finită* (*număr limitat de clienți*). Dacă *m* este numărul de clienți (solicitanți), atunci procesul de naştere şi moarte conține parametrii $\lambda_n$ şi $\mu_n$ pentru care

$$\lambda_n = m\lambda, \quad \mu_n = 0, \quad n = 0,$$

$$\lambda_n = (m-n)\lambda, \quad \mu_n = \mu, \quad 0 < n \leq m.$$

În *regim permanent*, adică $\pi_n(t) = \pi_n = const$, $n = 0,1,2,...,m$, ecuațiile corespunzătoare sunt

$$m\lambda\pi_0 = \mu\pi_1,$$

$$[(m-n)\lambda + \mu]\pi_n = (m-n+1)\lambda\pi_{n-1} + \mu\pi_{n+1}, \quad 0 < n < m,$$

$$\lambda\pi_{m-1} = \mu\pi_m.$$

Din relația de recurență

$$\pi_n = (m-n+1)\rho\pi_{n-1}, \quad 0 < n < m, \quad \rho = \lambda/\mu$$

rezultă

$$\pi_n = \frac{m!\rho^n}{(m-n)!}\pi_0, \quad 0 < n \leq m,$$

iar

$$\pi_0 = \frac{1}{1 + \sum_{n=1}^{m} \frac{m!\rho^n}{(m-n)!}}.$$

*Numărul mediu de unităţi în şir* este

$$\bar{n}_{\hat{s}} = \sum_{n=2}^{m} (n-1)\pi_n = m - \frac{1+\rho}{\rho}(1-\pi_0).$$

*Valoarea medie a inactivităţii serverului* este

$$\bar{\nu} = \sum_{n=0}^{1} (1-n)\pi_n = \pi_0.$$

Are loc relaţia $\bar{n}_{SA} = \bar{n}_{\hat{s}} + 1 - \bar{\nu}.$

*Timpul mediu de aşteptare în şir* se obţine din relaţia

$$\bar{n}_f = (m - \bar{n})\bar{t}_f,$$

de unde

$$\bar{t}_f = \frac{1}{\lambda(m - \bar{n}_{SA})} \sum_{n=2}^{m} (n-1)\pi_n = \frac{1}{\mu}\left( \frac{m}{1-\pi_0} - \frac{1+\rho}{\rho} \right),$$

iar *timpul de aşteptare în sistem* este

$$\bar{t}_s = \frac{n}{\lambda(m - \bar{n})} = \frac{1}{\mu}\left( \frac{m}{1-\pi_0} - \frac{1}{\rho} \right).$$

*Modelul unui sistem cu un şir de aşteptare cu mai mulţi serveri şi populaţie finită* (*număr limitat de clienţi*). Dacă notăm cu *m* numărul de clienţi (solicitanţi) şi *s* numărul de serveri (*m*>*s*), atunci procesul de naştere şi moarte cu parametrii $\lambda_n$ şi $\mu_n$ este dat de relaţiile

$$\lambda_n = m\lambda, \quad \mu_n = 0, \quad n > 0,$$

$$\lambda_n = (m-n)\lambda, \quad \mu_n = n\mu, \quad 1 \le n \le s,$$

$$\lambda_n = (m-n)\lambda, \quad \mu_n = s\mu, \quad s \le n \le m.$$

În *regim permanent*, adică $\pi_n(t) = \pi_n$, $n = 0,1,2,...,m$, obţinem sistemul liniar de ecuaţii

$$m\lambda\pi_0 = \mu\pi_1,$$

$$[(m-n)\lambda + n\mu]\pi_n = (m-n+1)\lambda\pi_{n-1} + (n+1)\mu\pi_{n+1}, \quad 1 \le n \le s,$$

$$[(m-n)\lambda + s\mu]\pi_n = (m-n+1)\lambda\pi_{n-1} + s\mu\pi_{n+1}, \quad s \le n \le m,$$

$$s\mu\pi_m = \lambda\pi_{m-1}.$$

Ţinând cont de formulele de recurenţă

$$\pi_n = \frac{m-n+1}{n}\rho\pi_{n-1}, \quad 0 \le n < s,$$

$$\pi_n = \frac{m-n+1}{s}\rho\pi_{n-1}, \quad s \le n \le m,$$

rezultă

$$\pi_n = C_n^m \rho^n \pi_0, \quad C_n^m = \frac{m!}{(m-n)!n!}, \quad n < s,$$

$$\pi_n = \frac{n!}{s!s^{n-s}} C_n^m \rho^n \pi_0, \quad s \le n \le m,$$

iar

$$\pi_0 = \frac{1}{\sum_{n=0}^{s-1} C_n^m \rho^n + \frac{s^s}{s!}\sum_{n=1}^{m} \frac{1}{(m-n)!}\left(\frac{\rho}{s}\right)^n}.$$

*Numărul mediu de unităţi din şirul de aşteptare* este

$$\bar{n}_s = \sum_{n=s+1}^{m}(n-s)\pi_n = \left(m - \frac{s}{\rho} - s\right)\left(1 - \pi_0\sum_{n=0}^{s-1} C_n^m \rho^n\right) + s\rho^{s-1}C_n^m\pi_0.$$

*Numărul mediu de unităţi în sistemul de aşteptare* este

$$\bar{n}_{SA} = \left[\frac{s\rho - s - m\rho}{\rho(1+\rho)}\sum_{n=0}^{s-1} C_n^m \rho^n + \frac{s\rho^{s-1}C_n^m}{1+\rho}\right]\pi_0 + \frac{m\rho - s}{\rho}.$$

## 2.6. Modele cu restricţii

Sistemele tehnice, informatice, sistemele de calcul şi reţelele de calculatoare în care atât timpul de aşteptare în şir (la coadă) sau timpul de aşteptare total (şir+server) cât şi numărul locurilor de aşteptarte sunt limitate se numesc *sisteme cu restricţii*.

Pentru sosiri Poisson şi serverii de forma $F(t)=(1-e^{-\mu x})$, când unitatea sosită în sistem poate aştepta un anumit timp constant şi mărginit, el se aşează la coadă şi

aşteaptă. Dacă timpul de aşteptare este mai mare decât timpul de care dispune unitatea sosită, atunci ea părăseşte sistemul. Uneori unitatea părăseşte sistemul fără să fie servită. *Sisteme SA* cu astfel de caracteristici sunt denumite *sisteme de aşteptare cu restricţii sau cu pierderi*.

Cele mai uzuale modele cu restricţii sunt următoarele:

- modele cu timp de aşteptare (constant sau variabil), în şir mărginit;
- modele cu timp de aşteptare cumulat (şir+server) dar mărginit;
- modele cu şir de aşteptare limitat.

Relaţiile de calcul al timpului pentru primele două cazuri se dau în literatură [1,2,7,8,10].

Modelele de aşteptare cu şir limitat în cazul unui server, când intrarile şi servirile respectă legea Poisson, operează cu următoarele mărimi:

$$\pi_n = \rho^n \cdot \pi_0 = \rho^n \cdot \frac{1-\rho}{1-\rho^{N+1}} .$$

În general valorile lui  (numărul mediu de unităţi în şir) şi  (numărul mediu de unităţi în *SA*) se calculează astfel:

$$\bar{n}_{ş} = \rho^2 \frac{(1-N)\rho^{N+1} + (N-1)\rho^N}{(1-\rho)(1-\rho^N)}$$

$$\bar{n}_{SA} = \rho \frac{1-(N+1)\rho^N + N\rho^{N+1}}{(1-\rho)(1-\rho^{N+1})}$$

în care: $N$ - numărul maxim admisibil de unităţi în sistem.

Modelele cu şir de aşteptare limitat şi cu mai mulţi serveri în paralel în cazul când intrările şi serverii sunt de tip Poisson operează cu relaţii de forma:

$$\pi_j = \frac{\pi_0 (\lambda/\mu)^j}{j!} \quad , \quad 1 \le j < S < N$$

$$\pi_0 = \left[ \sum_{k=0}^{S} \frac{\rho^k}{k!} + \frac{\rho^S}{S!} \sum_{i=0}^{N-S} \left( \frac{\rho}{S} \right)^i \right]^{-1}$$

în care: $j$ - numărul de unităţi din sistem la un moment dat.

Probabilitatea ca o unitate să părăsească sistemul se poate scrie astfel:

$$\pi_j = \frac{\rho^{s+j}}{S!S^j}\pi_0.$$

În cazul în care funcția de repartiție a duratei de servire nu este exponențial-negativă, adică coeficientul de variație $C_v \neq 1$, atunci sunt utilizate funcțiile de repartiție $E_k$ (*Erlang-k*), *Cox-k* şi $H_k$ (*Hiperexponențială* de ordinul *k*). Schema sistemului de aşteptare *SA* cu zonele de servire respective sunt prezentate în fig. 2.2.

Modelele de aşteptare ciclice cu (*S*) serveri în serie sau paralel, arată modul de stabilire a cheltuielilor cu aşteptarea când servirea este ciclică. Aceste tipuri de modele operează cu relații de forma:

$$\pi(n) = \frac{K+S-1}{(S-1)!K!}; \qquad \bar{t}_t^S = \frac{S-1}{K+S-1};$$

$$\bar{n}_{SA} = \frac{K(K-1)}{(S+K-1)};$$

$$\bar{t}_s = \frac{K-1}{\mu(S+K-1)}; \qquad t_c = \frac{1}{\mu}\left[S + \frac{K(K-1)}{S+K-1}\right]$$

în care: *K* - numărul total al unităților din şirul de aşteptare; - timpul cât un server este neocupat; $t_c$ - durata unui ciclu de prelucrare.

Modelele de aşteptare cu intrări şi serviri în grup apar în diverse situații practice cum ar fi: procesele de fisiune nucleară, procesele de numărătoare a particulelor elementare, procesele automate din cadrul centralelor de telecomunicații şi sitemelor informatice etc. Studierea acestor tipuri de procese se face, analizând în maniera teoriei aşteptării următoarele situații:

- modele cu intrări în grup şi serviri individuali;
- modele cu intrări individuale şi serviri în grup;
- modele cu intrări şi serviri în grup.

În modelele expuse şirul de aşteptare se formează în ordinea sosirilor unităților în sistem, iar servirea respectă principiul "primul venit, primul servit". Nu sunt rare

cazurile când disciplina şirului de aşteptare se stabileşte în funcţie de urgenţa şi importanţa lucrării, aplicându-se principiul “ultimul venit, primul servit”. Modelele în care disciplina de servire după alte criterii diferă de disciplina intrării unităţilor în sistem se numesc *modele cu prioritate*.

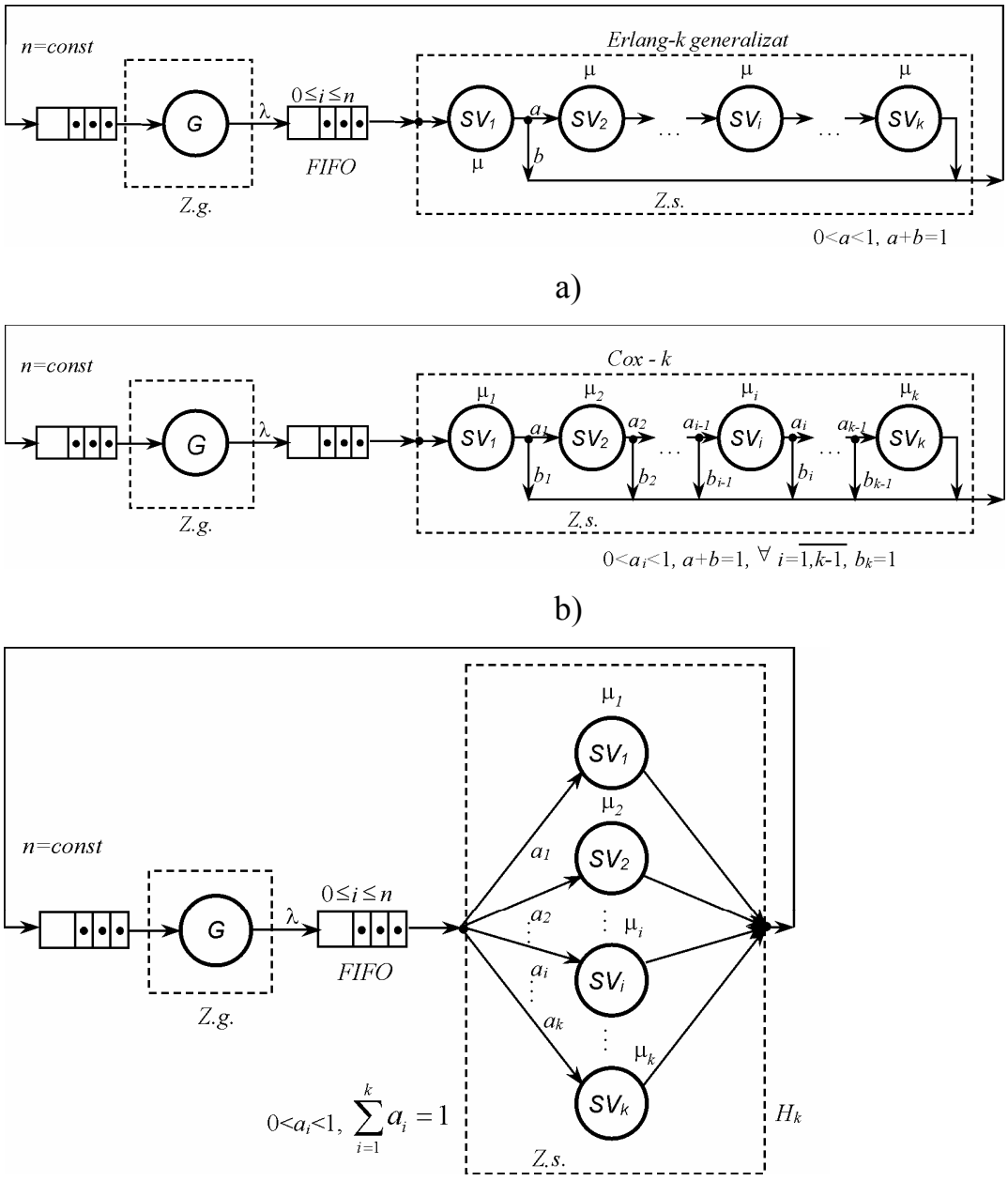



Figura 2.3 Schema sistemelor de aşteptare cu zone de servire

Modelele cu prioritate se împart în *modele cu prioritate relativă* şi *modele cu prioritate absolută*. În cazul modelelor cu prioritate parțială, unitatea sosită în sistem aşteaptă până ce stația se eliberează de unitatea în lucru şi apoi trece în fruntea şirului foramat anterior. Dacă modelele de aşteptare sunt cu prioritatea absolută, atunci în momentul sosirii unei unități în sistem se întrerupe servirea la o unitate în curs de servire şi se serveşte noua unitate servită în sistem.

Studiul sistemelor în care unitățile au prioritate absolută cuprinde diverse situații dintre care rețin atenția următoarelor: modele cu aşteptare care necesită un timp de orientare pentru a schimba tipul de unitate pe care trebuie s-o deservească, modele în care prioritatea se atribuie prin clasificarea unităților, modele în care unitățile sunt alese pentru servire în mod întâmplător şi modele în care servirea se face după principiul "ultimul sosit, primul servit".

Modelele cu mai mulți serveri în paralel pot fi cu şi fără informații asupra situațiilor serverilor. Dacă ne găsim în situația modelelor fără informație atunci unitățile care sosesc în sistem formează şiruri întâmplătoare în fața serverilor. Aceste situații se studiază ca procese Markov la care probabilitățile de trecere a sistemului dintr-o stare în alta se calculează cu relațiile obținute prin soluționarea ecuațiilor Chapman-Kolmogorov.

# 3. Aplicații

## 3.1. Lucrarea de laborator nr. 1

## Lanțuri Markov timp discret

### 3.1.1. Considerații teoretice

Cum s-a menționat mai înainte, practica oferă numeroase exemple, în care anumite valori caracteristice ale unui sistem, formând aşa numitele stări discrete ale sistemului, variază o dată cu timpul, astfel încât ele nu pot fi prevăzute cu exactitate. Un asemenea proces în care una sau mai multe valori caracteristice lui variază aleator în timp îl numim *"proces stochastic"*.

Lanțul aleator de tip Markov este un şir de variabile aleatoare care satisface condiția lui Markov şi anume: probabilitatea că sistemul discret la momentul ($k$+1) (deseori numită şi epocă sau perioadă), să se găsească în starea discretă ($i_{k+1}$), condiționată de faptul că sistemul s-a găsit respectiv la momentele 1,2,...,$k$-1,$k$ în stările $i_1,i_2,...,i_k$, nu depinde decât de ultima stare, adică

$$P_r(x_{k+1}=i_{k+1}/x_k=i_k,\ \ x_{k-1}=i_{k-1},\ldots,x_1=i_1)=P_r(x_{k+1}=i_{k+1}/x_k=i_k)\ .$$

Probabilitatea că sistemul să fie în starea $i$ la momentul $k$, o vom nota

$$\pi_i(k)=P_r(x_k=i) \text{ cu}$$

$$0\le\pi_i(k)\le 1,\ \ \sum_{i=1}^{n}\pi_i(k)=1.$$

Probabilitatea ca sistemul să treacă în starea $j$ la momentul ($k$+1), ştiind că în momentul precedent $k$ el s-a aflat în starea $i$, adică probabilitatea condiționată

$$P_r(x_{k+1}=j/x_k=i)=p_{ij},\ \ (i,j=1,2,...,n)$$

poartă numele de probabilitate de trecere.

Un lanț Markov este complet determinat dacă cunoaştem: mulțimea stărilor discrete $S=\{s_i,i=\overline{1,n}\}$, vectorul-linie al probabilităților de stare inițială şi matricea

stochastică a probabilităților de trecere $P=(p_{ij})$, ($i,j$ = 1,...,$n$),

$$0 \le p_{ij} \le 1, \sum_{j=1}^{n} p_{ij} = 1.$$

Relația prin care determinăm probabilitățile de stare la momentul ($k$+1), cu ajutorul probabilităților de trecere şi a vectorului de stare corespunzător momentului $k$, este descrisă de ecuația Kolmogorov [9]:

$$\pi_i(k+1) = \sum_{i=1}^{m} \pi_i(k) p_{ij}, \quad j = 1,\ldots,n; \quad k = 0,1,2,\ldots$$

Dacă la fiecare stare $j$ se va ataşa o funcție cost $c_j(k)$ de aflare a lanțului *DLM* în această stare, atunci costul mediu de funcționare a lanțului este:

$$\overline{C}(k) = \sum_{i=1}^{n} \left[ c_j(k) \cdot \pi_j(k) \right].$$

În fig. 3.1 este prezentat lanțul Markov *DLM*1 ergodic (ireductibil şi aperiodic), iar în fig. 3.2 un alt lanț *DLM*2 neergodic (reductibil şi aperiodic). Aceste lanțuri sunt redate fiecare de un graf orientat ponderat cu probabilitățile de trecere respective şi condițiile inițiale ale lui .

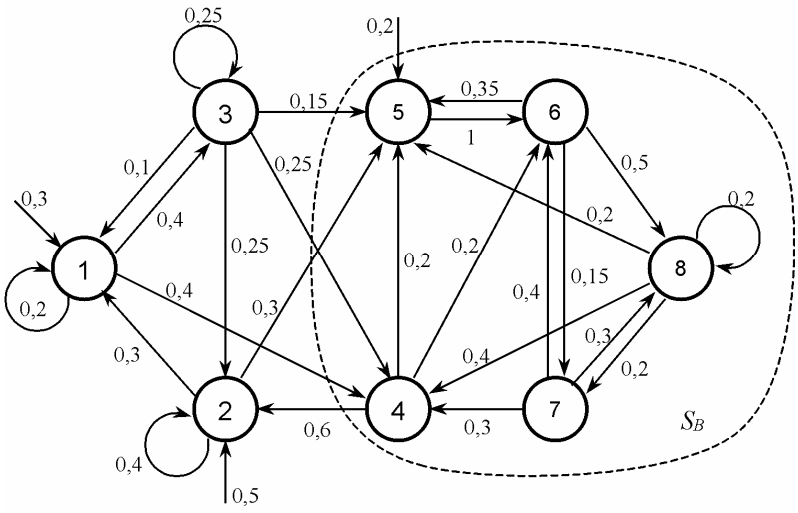



*Fig 3.1. Lanț Markov ergotic LMTD1.*

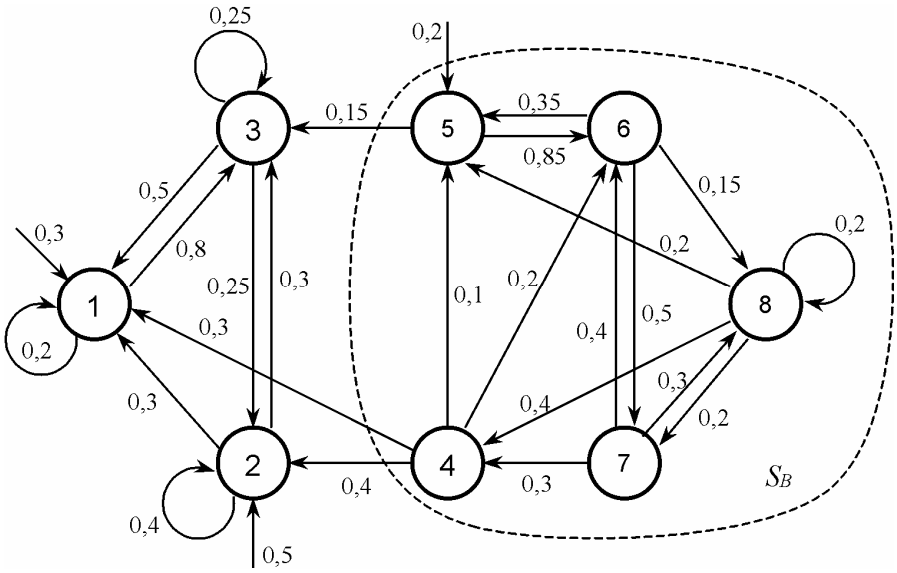



*Fig 3.2. Lanț Markov ergotic LMTD2.*

Deseori este necesar de a determina probabilitatea $\pi_{S_b}(k)$ şi costul mediu $\overline{C}_{S_B}(k)$ de aflare a lanțului *DLM* la momentul *k* într-o submulțime de stări $S_B \subset S$, astfel încât $S = S_B \cup S_R, S_B \cap S_R = \emptyset$. În acest caz $\pi_{S_B}(k) = \sum_{s_j \in S_B} \pi_j(k)$, iar $\overline{C}_{S_B}(k) = \sum_{s_j \in S_B} C_j(k) \cdot \pi_j(k)$.

La determinarea acestor caracteristici este necesar de a folosi sistemul instrumental *QM* pentru a calcula distribuția probabilităților de stare $\pi_j(k)$, $j$=1,2,...,$n$; $k$=0,1,...,$K$.

### 3.1.2. Scopul lucrării

Studierea metodelor de redare, descriere, analiză a proprietăților de comportare ale lanțurilor Markov timp discret (*LMTD*) şi evaluare a caracteristicilor numerice de performanță.

### 3.1.3. Ordinea îndeplinirii lucrării

Îndeplinirea lucrării prevede următoarele acțiuni:

- pentru varianta formulată de profesor de construit graful lanțului *DLM*;

- de determinat matricea stochastică a *DLM* şi de scris ecuaţiile Kolmogorov;

- de elaborat algoritmul şi programul de calcul numeric al repartiţiei probabilităţilor de stare la momentul *k*;

- de evaluat probabilitatea $\pi_{S_B}(k)$ de aflare în $S_B$ şi valoarea respectivă a costului mediu $\overline{C}_S(k)$ şi $\overline{C}_{S_B}(k)$ funcţie de durata funcţionării *DLM*.

### 3.1.4. Prezentarea şi susţinerea lucrării

Lucrarea se prezintă în formă de referat şi se susţine profesorului la calculator în mod practic

*Conţinutul referatului:*

- foaia de titlu cu denumirea lucrării;

- obiectivele lucrării de laborator, scurte date teoretice;

- calcularea parametrilor respectivi ai *DLM*;

- graficele varierii probabilităţilor de stare şi a costului mediu în funcţie de durata observării *DLM*;

- concluzii.

### 3.1.5. Întrebări şi teme

- matricea stochastică a *DLM*;

- clasificarea stărilor *DLM*;

- condiţii de ergodicitate ale *DLM*;

- metode de redare a unui *DLM*;

- ecuaţiile Kolmogorov ale unui *DLM*;

- metode numerice de soluţionare ale ecuaţiilor Kolmogorov.

## 3.2. Lucrarea de laborator nr. 2

## Analiza sistemelor de aşteptare multicanal

### 3.2.1. Consideraţii teoretice

Modelele fenomenelor de aşteptare descriu sisteme şi procese de aşteptare cu caracter de masă care intervin în diverse domenii ale activității practice. În sistemul *SA* există un flux de cereri (clienți) pentru servire, numit flux de intrare, caracterizat de numărul de cereri care intră în sistem într-o unitate de timp. Într-un *SA* există elemente care efectuează serviciile, numite stații de servire sau servere. Pentru servirea fiecărei unități (cereri), este necesar un timp oarecare, în cursul căruia stația este ocupată şi nu poate servi alte unități. Durata servirii este întâmplătoare (aleatoare). Un model *SA* este descris complet prin formula lui Kendall de următoarele elemente: fluxul de intrare, fluxul de aşteptare, stațiile de servire şi fluxul de ieşire. Cu ajutorul fluxului de intrare putem determina modul în care sosesc unitățile în *SA*. Presupunem că intrările (sosirile) în *SA* sunt întâmplătoare şi independente, deci probabilitatea ca o unitate (cerere) să vină în *SA* este independentă atât de momentul în care se produce sosirea, cât şi de numărul de unități existente deja în sistem sau numărul de unități ce vor veni. *Probabilitatea*, ca în intervalul de timp $(t, t+\Delta t)$, $t > 0$, să se producă o intrare în sistem, reprezintă *numărul mediu de intrări* în unitatea de timp $\Delta t$ şi este egală cu $1/\lambda$, în ipoteza că sosirile urmează un proces Poisson de parametru $\lambda$, $(0<\lambda<\infty)$. Să presupunem că $t\geq 0$ şi să notăm cu $t_1,t_2,...,t_n,...$ momentele succesive în care sosesc unitățile în sistem. Vom admite că intervalele de timp dintre intrările consecutive $\tau_n = t_{n+1}-t_n$, $t_0 = 0$, $n = 1,2,3,...$ sunt variabile aleatoare pozitive independente cu funcția de repartiție:

$$F(x) = P_r(\tau_n \leq x),\ \ 0 < x < \infty,\ \ n = 1,2,...,$$

$\tau_n$, $n = 1,2,...$ fiind variabile aleatoare identic repartizate. Durata necesară pentru servirea unei unități se numeşte durată de servire. Presupunem că duratele de servire sunt variabile aleatoare pozitive, identic repartizate, independente şi, de asemenea, independente de $\tau_1,\tau_2,...$ Notând cu $y_n$ durata de servire a celei de-a $n$-a unitate, funcția de repartiție a duratei de servire este $H(x) = P_r(y_n \leq x), 0 < x < \infty$. Într-o problemă de aşteptare se întâlnesc următorii indicatori de performanță principali:

$\pi_0$ - probabilitata staționară că în *SA* nu mai sunt unități pentru a fi deservite.

$\bar{n}_{SA}(t)$, (respectiv $\bar{n}_{ş}(t)$ ) - numărul mediu de unități în SA (în şirul de aşteptare) la un moment *t*, care este valoarea medie a variabilei $n_{SA}(t)$, (respectiv $n_{ş}(t)$).

$\bar{\tau}_{SA}$, (respectiv $\bar{\tau}_{ş}$) - durata medie de aşteptare a unei unități în *SA*.

Ultimii doi indicatori de performanță depind de legile serviciului şi ale sosirilor, şi aceştea vor fi determinați pentru fiecare model SA în parte.

### 3.2.2. Scopul lucrării

Studierea metodelor de descriere şi de evaluare a sistemelor de aşteptare multicanal elementare tip *GI/G/k/n*.

### 3.2.3. Ordinea îndeplinirii lucrării

Îndeplinirea lucrării prevede următoarele acțiuni:

- pentru sistemul SA, descris de formula Kendall dată de profesor, de construit graful ratelor de trecere ale lanțului *CLM*;

- de scris ecuațiile Chapman-Kolmogorov ale *CLM* pentru *SA* considerat.

- de elaborat algoritmul şi programul de calcul numeric pentru a determina repartizarea probabilităților staționare de stare;

- de evaluat indicatorii numerici ai *SA* pentru varianta definită de către profesor şi de efectuat modificările necesare, astfel încât $\bar{\tau}_{SA} < \bar{\tau}^*_{SA}$, unde este restricția specificată în prealabil.

### 3.2.4. Prezentarea şi susținerea lucrării de laborator

Lucrarea se prezintă în formă de referat şi se susține profesorului la calculator în mod practic

*Conținutul referatului:*

- foaia de titlu cu denumirea temei lucrării;

- obiectivele lucrării de laborator, scurte date teoretice;

- schema de servire şi graful ratelor de treceri al lanțului *CLM*;

- ecuațiile Chapman-Kolmogorov ale lanțului *CLM*;

- schema-bloc, programul şi rezultatele numerice ale indicatorilor de performanță ai *SA* în dependență de numărul de serveri;

- concluzie.

### 3.2.5. Întrebări şi teme

- interpretarea formulelor Kendall ale *SA*;

- echilibrul static local al fluxurilor de probabilitate în *SA*;

- funcții de repartiții ale duratelor de intersosire şi a duratei de servire a cererilor, criteriile de clasificare;

- legea lui Little ce descrie durata medie de aşteptare a unei unități în *SA*;

- ecuațiile Chapman-Kolmogorov ale lanțului *CLM* ce descrie funcționarea *SA* specificată.

## 3.3. Lucrarea de laborator nr. 3

## Analiza sistemelor de aşteptare prioritare

### 3.3.1. Considerații teoretice

Formarea şirurilor de aşteptare şi servirea cererilor în sistemul de aşteptare de regulă se face după disciplina de servire FIFO. În unele cazuri, este nevoie de a asigura o altă disciplină de formare a şirului de aşteptare în dependență de urgența cererii, având astfel o prioritate de a fi servită.

În continuare vom presupune că în sistemele de aşteptare prioritatea cererii creşte odată cu micşorarea indicelui clasei cărei aparține această cerere. Dacă $i<j$, cererile cu prioritatea $i$ vor avea o prioritate mai înaltă decât cele cu prioritatea $j$.

Modelele cu prioritate se împart în modele ale $SA: \overrightarrow{M}_r / \overrightarrow{M}_r / 1$ cu prioritate relativă şi cu prioritate absolută.

Servirea cererii conform priorităţii relative presupune manifestarea priorităţii numai în momentul eliberării serverului şi a selectării cererii, care va fi deservită din şirul de aşteptare, adică va fi deservită cererea care are cea mai mare prioritate.

Fie dat un *SA* cu *r* fluxuri de cereri $\Phi_k, k = \overline{1}, \overline{r}$ cu priorităţile respective de la 1 până la *r*. Fiecare flux $\Phi_k$ este de tip Poisson cu rata $\lambda_k$, ($k$ = 1,...,$r$). Fluxul sumar, de asemenea, este de tip Poisson cu rata $\lambda = \sum_{k=1}^{r} \lambda_k$. Vom presupune că durata de servire a cererii de prioritatea *k* are o distribuţie exponenţial-negativă cu parametrul $\mu_k$ şi deci durata medie de servire a unei cereri este egală cu $\tau_{ser} = 1/\mu_k$.

În cazul când la sosirea cererilor serverul este ocupat, în faţa ei se formează *r* şiruri de aşteptare şi în şirul *i* plasându-se cererile cu prioritatea *i*, ($i$ = 1,...,$r$).

Factorul sarcină exercitat de către fluxul $\Phi_k$ asupra sistemului *SA* este $\rho_k = \lambda_k / \mu_k$.

Vom nota

$$u_k = \sum_{i=1}^{k} \rho_i, i = \overline{1}, \overline{r}, \text{ iar } u_0 = 0.$$

Durata medie $\overline{\tau}_{qm}$ de aflare a cererilor în şirul de aşteptare cu prioritatea *m* este determinată de următoarea expresie [10]:

$$\overline{\tau}_{qm} = \frac{\sum_{k=1}^{m} \frac{\lambda_k}{\mu_k^2}}{(1-u_{m-1})(1-u_m)},$$

iar durata medie $\overline{\tau}_q$ de aşteptare a cererilor în SA şi numărul mediu de cereri $\overline{L}_q$ din SA este:

$$\overline{\tau}_q = \frac{\sum_{k=1}^{m} \lambda_k \cdot \overline{\tau}_{qk}}{\sum_{k=1}^{m} \lambda_k},$$

$$\overline{L}_q = \sum_{k=1}^{m} \lambda_k \cdot \overline{\tau}_q$$

Sistemul *SA* cu prioritate absolută presupune că servirea cererii curente va fi imediat întreruptă la sosirea unei cereri cu o prioritate mai înaltă şi astfel serverul va începe îndată deservirea cererii celei din urmă.

Durata medie $\overline{\tau}_j$ de aflare în sistem a cererilor cu prioritatea *j* este determinată de următoarea expresie:

$$\overline{\tau}_j = \frac{1}{1-u_{j-1}} \cdot \left[ \frac{1}{\mu_j} + \sum_{i=1}^{j} \left( \frac{\rho_i}{\mu_j} \right) \right],$$

$$u_j = \sum_{i=1}^{j} \rho_i, u_0 = 0 .$$

Ştiind durata medie $\overline{\tau}_{ser} = 1/\mu_i$ de servire a cererilor de prioritatea *i*, putem estima durata medie de aşteptare în şirul de aşteptare respectiv: $\overline{\tau}_{qi} = \overline{\tau}_t - \overline{\tau}_{ser}$.

### 3.3.2. Scopul lucrării

Studierea metodelor de descriere şi de evaluare a sistemelor de aşteptare prioritare.

### 3.3.3 Ordinea îndeplinirii lucrării

Îndeplinirea lucrării prevede următoarele acţiuni:

- pentru sistemul de aşteptare cu prioritate, redat de către formula Kendall SA: $\overrightarrow{M}_r / \overrightarrow{M}_r / 1$, de scris ecuaţiile Chapman-Kolmogorov. De construit graful lanţului *DLM* ce descrie comportarea sistemului SA: $\overrightarrow{M}_3 / \overrightarrow{M}_3 / 1$.

- de elaborat algoritmul şi programul de calcul numeric pentru determinarea repartizării probabilităţilor staţionare de stare ale acestui sistem *SA*.

- de evaluat indicatorii de performanță numerici ai *SA* cu proiritate relativă şi prioritate absolută, astfel încât $\overline{\tau}_{qm} < \overline{\tau}^{*}_{qm}$, unde $\overline{\tau}^{*}_{qm}$ este restricția specificată în prealabil.

### 3.3.4. Prezentarea şi susținerea lucrării de laborator

Lucrarea se prezintă în formă de referat şi se suține profesorului la calculator în mod practic

*Conținutul referatului:*

- foaia de titlu cu denumirea lucrării;
- obiectivele lucrării de laborator, scurte date teoretice;
- schema-bloc, programul şi rezultatele numerice ale indicatorilor de performanță;
- concluzii.

### 3.3.5. Întrebări şi teme

- disciplinele de servire ale cererilor în *SA* cu mai multe clase de cereri;
- *SA* prioritare multicanal;
- *SA* cu pierderi ale cererilor;
- legea Little pentru fluxuri de cereri prioritare;
- graful ratelor de treceri ale lanțului *CLM* pentru *SA* cu priorități.

## 3.4. Lucrarea de laborator nr. 4

## Analiza rețelelor stochastice model Jakson

### 3.4.1. Considerații teoretice

3.4.1.1. Rețele stochastice deschise model Jackson.

În studiul sistemelor de calcul, rețelelor de telecomunicații şi al rețelelor de calculatoare mărimile fundamentale de interes decurg în considerarea cererii de serviciu (oferta de apel) şi a duratei de serviciu.

Astfel, rețeaua de calculatoare este un ansamblu de sisteme elementare. Fiecare sistem este o entitate independentă şi poate avea o manieră specifică de comportament față de cererile de serviciu, fiind un sistem cu pierderi sau cu aşteptare, cu servire exponențială sau cu durată constantă de serviciu, aşa cum este expus în [2,10]. În cele ce urmează se va prezenta modelul general în formă de rețea stochastică discretă cu subsisteme de aşteptare (*RSA*) model Jackson, în care aceste sisteme pot coopera în cadrul unui sistem global.

Structura rețelei ce conține sistemele $SA_i$ elementare $i = 0,1,2,3,...,n$ se poate reprezenta printr-un graf orientat, ponderat, ca în fig.3.3. Nodurile grafului sunt sistemele $SA_i$ elementare componente. Ele se leagă prin joncțiuni indexate prin câte doi indici $(i,j)$ stabiliți în funcție de sistemul $SA_i$ *sursă* a intercomunicației şi de sistemul $SA_j$ de *destinație* şi reprezintă transferul de cereri (apeluri) în graf prin arcele corespunzătoare (fig. 3.4). Se recomandă folosirea indicelui 0 pentru nodul sursă de apel din orice rețea.

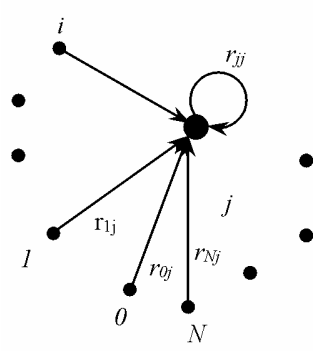



Figura 3.3. Tranziții între noduri i.

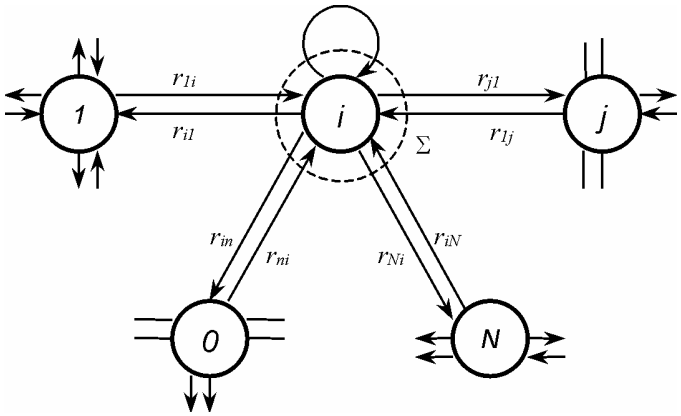



Figura 3.4. Graf general al RSA

Un apel ce parcurge rețeaua, de la nodul sursă până la cel căruia îi este adresat, este dirijat fie pe un traseu dinainte stabilit, corespunzător unor restricții ferme, fie la întâmplare pe un traseu ales în momentul solicitării, în funcție de condițiile prezentate. Înseamnă că apelul trebuie să traverseze mai multe noduri, în fiecare din ele executând o alegere dirijată sau întâmplătoare a direcției următoare de parcurs.

Există în consecință un transfer de apeluri de la un nod la altul, $r_{ij}$ reprezentând posibilitatea transferului unei unități de prelucrare (apel) de la nodul *i* la nodul *j*.

Dar, dacă se consideră întreaga rețea, atunci posibilitatea ca în intervalul de timp ($\tau$, $\tau+\Delta\tau$) să sosească un apel în nodul *j* este o combinație liniară de coeficienți constanți $r_{ij}$ ai tuturor probabilităților de ieşire în ($\tau$, $\tau+\Delta\tau$) a unui apel din toate nodurile *i* ale rețelei (*i* = 0,1,2,...,*n*). Înseamnă că, pe ansamblul rețelei, se poate forma o *matrice R de transfer*:

$R = \{r_{ij}\}$ pentru $i,j = 0,1,2,...,n$, astfel încât $0 \le r_{ij} \le 1 \sum_{j=0}^{n} r_{ij} = 1.$

Deoarece dacă un apel se află în nodul *i* la momentul $\tau$, atunci este sigur că la ($\tau$, $\tau+\Delta\tau$) el va fi găsit într-un nod *j* oarecare al rețelei.

Este evident că dacă prin rețea nu există legătura (*i,j*) atunci $r_{ij} = 0$, iar dacă apelul din *i* nu poate evalua decât spre *j* atunci $r_{ij} = 1$.

Dacă pentru un sistem elementar $SA_i$ se consideră că apelul *i* ce-l traversează au o rată medie $\lambda_i$, atunci raportat la o suprafață $\Sigma$ de flux nul de probabilități (fig. 3.4), se obține:

$$\sum_{j=0}^{n} \lambda_i \cdot r_{ij} = \sum_{j=0}^{n} \lambda_j \cdot r_{ji}.$$

Ceea ce înseamnă că:

- fluxul total de apeluri recepționate de un nod oarecare *i* din rețea (*traficul total recepționat*) este compus din *n* fracțiuni $\lambda_{ij}$ diferite:

$$\lambda_i = \sum_{j=0}^{n} \lambda_{ji} = \sum_{j=0}^{n} \lambda_j \cdot r_{ji};$$

- fluxul total de apeluri transmise rețelei de un nod oarecare *i* (*traficul total transmis*) este compus din *n* fracțiuni $\lambda_{ij}$ diferite:

$$\lambda_i = \sum_{j=0}^{n} \lambda_{ji} = \sum_{j=0}^{n} \lambda_i \cdot r_{ji}.$$

Rezultă că fluxul de unități prelucrate (cereri) ce traversează fiecare sistem $SA_i$ al rețelei *RSA* se poate calcula cu ajutorul ecuației matriceale în care:

$$\Lambda = \Lambda \cdot R,\ \Lambda \cdot D = 0\ .$$

Aici $\Lambda$ = vectorul-linie a traficelor caracteristice nodurilor rețelei, considerate ca sisteme $SA_i$ elementare:

$$\Lambda = (\lambda_0,\lambda_1,\lambda_2,...,\lambda_N).$$

- $D$ = matricea dinamică a rețelei:

$$D=R-I=\{d_{ij}\};$$

pentru $i, j = \overline{0},\overline{n}$ cu $\sum_{j=0}^{n} d_{ij} = 0$, ceea ce înseamnă că:

$$D = \begin{pmatrix} r_{00}-1 & r_{01} & \cdots & r_{0N} \\ r_{10} & r_{11}-1 & \cdots & r_{1N} \\ \vdots & \vdots & \vdots & \vdots \\ r_{N0} & r_{N1} & \cdots & r_{NN}-1 \end{pmatrix}$$

Relația $\Lambda.D$=0 permite scrierea unui sistem de ecuații lineare care admite totdeauna o infinitate de soluții, dependente de un $\lambda_i$, ales arbitrar pe lângă soluția banală ($\Lambda$=0). De regulă, dacă se precizează o anumită valoare pentru $\lambda_0$, atunci se obține relația de dependență:

$$\lambda_i = a_i \cdot \lambda_0, \qquad i = \overline{1,n}.$$

Se realizează chiar o clasificare a rețelelor în raport cu valoarea lui $\lambda_0$ şi anume:

- rețele deschise, pentru care $\lambda_0 \neq 0$;

- rețele închise, pentru care $\lambda_0 = 0$. În acest ultim caz toate fluxurile se stabilesc în raport cu un alt $\lambda_i$ arbitrar ales.

### 3.4.1.2. Aplicație

Fie o rețea stochastică *RSA*1 (fig. 3.5) în care nodul 0 este sursă şi totodată destinație finală. Rețeaua are două noduri principale de transfer 1 şi 2 şi două noduri intermediare 3 şi 4, care ar putea fi eventual înglobate în 1, respectiv în 2 ca etaje suplimentare de aşteptare.

Matricea *R* de transfer este precizată în fig. 3.6, elementele sale reprezentând de fapt *factorii de cointeresare* între abonații conectați ca terminale într-o asemenea rețea (se poate uşor verifica faptul că suma elementelor unei linii din matricea *R* este egală cu unitatea):

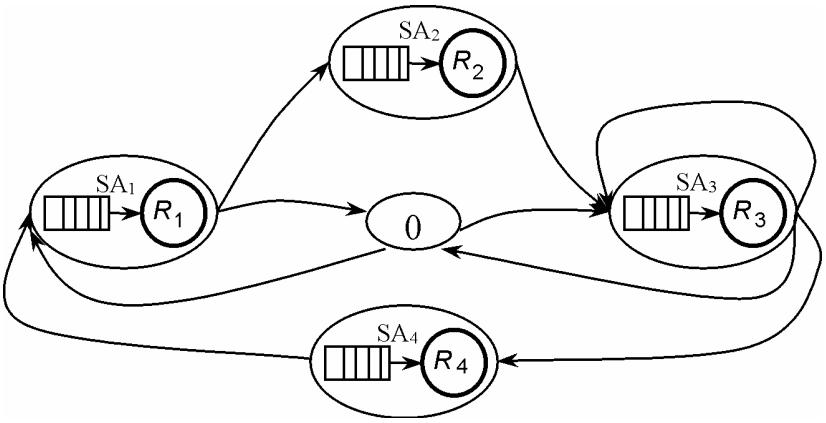



Figura 3.5. Structura rețelei RSA1

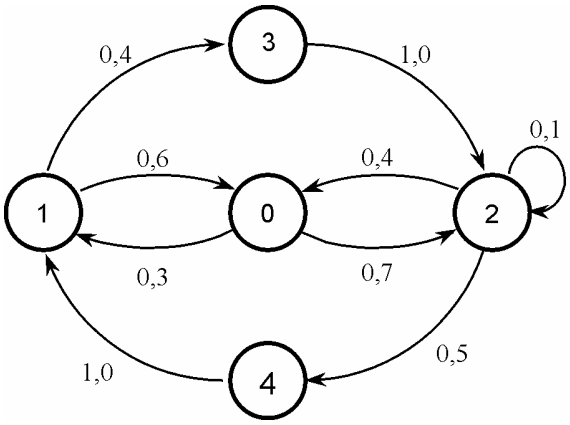



Figura 3.6. Graful tranziției între starile rețelei

$$R = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{pmatrix} 0 & 0.3 & 0.7 & 0 & 0 \\ 0.6 & 0 & 0 & 0.4 & 0 \\ 0.4 & 0 & 0.1 & 0 & 0.5 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \\ \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array}$$

Se cere a fi calculată matricea $\Lambda$ caracteristică acestei structuri de rețea.

*Rezolvare*. Pe baza datelor propuse matricea dinamică *D*, ce corespunde transferurilor prin rețea este:

$$D = \begin{pmatrix} -1 & 0.3 & 0.7 & 0 & 0 \\ 0.6 & -1 & 0 & 0.4 & 0 \\ 0.4 & 0 & -0.9 & 0 & 0.5 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{pmatrix}$$

Fluxurile de intrare în nodurile rețelei propuse se determină prin rezolvarea următorului sistem de ecuații liniare, sistem întocmit pe baza matricei dinamice *D*:

$$\begin{cases} -\lambda_0 + 0.6\cdot\lambda_1 + 0.4\cdot\lambda_2 + 0\cdot\lambda_3 + 0\cdot\lambda_4 = 0 \\ 0.3\cdot\lambda_0 - \lambda_1 + 0\cdot\lambda_2 + 0\cdot\lambda_3 + \lambda_4 = 0 \\ 0.7\cdot\lambda_0 + 0\cdot\lambda_1 - 0.9\cdot\lambda_2 + \lambda_3 + 0\cdot\lambda_4 = 0 \\ 0\cdot\lambda_0 + 0.4\cdot\lambda_1 - 0\cdot\lambda_2 - \lambda_3 + 0\cdot\lambda_4 = 0 \\ 0\cdot\lambda_0 + 0\cdot\lambda_1 + 0.5\cdot\lambda_2 + 0\cdot\lambda_3 - \lambda_4 = 0 \end{cases}$$

Se poate uşor calcula soluția:

$$\Lambda = \left[ \lambda_0,\ \frac{31}{35}\cdot\lambda_0,\ \frac{41}{35}\cdot\lambda_0,\ \frac{62}{175}\cdot\lambda_0,\ \frac{41}{70}\cdot\lambda_0 \right]$$

*Observație*. În situația tuturor sistemelor $SA_i$ în echilibru static s-a stabilit că trebuie să se folosească un raport subunitar între traficul oferit $\lambda_i$ şi mărimea ratei $\mu_i$ a ratei grupului resurselor de prelucrare (serveri) [2,7,8].

Aceasta înseamnă că pentru oricare dintre sistemele $SA_i$ ale rețelei trebuie să fie îndeplinită inegalitatea $0 < \rho_i = a_i\lambda_0 / \mu_i < 1$, ceea ce conduce la impunerea unei condiții obligatorii pentru stabilirea valorii fluxului sursei.

Ținând seama de expresia $\lambda_i = a_i.\lambda_0$ care precizează dependența generală a fluxurilor $\lambda_i$ din noduri față de fluxul $\lambda_0$ al sursei, înseamnă că relația de obligativitate pentru a se obține o rețea staționară este:

$$0 < \lambda_0 < \lambda_0^{\max} = \min_{(i)} \left\{ \frac{\mu_i}{a_i} \right\}.$$

În cazul în care această relație nu este verificată, adică $\lambda_0 > \lambda_0^{\max}$, atunci este necesar de a modifica numărul de serveri $k_j$ în stația $SA_j:M/M/k_j$ astfel încât va fi verificată relația

$$\lambda_0 \le \min_{(j)} \left( \frac{\mu_j}{a_j} \right)$$

Pentru o rețea *RSA* staționară caracteristicile numerice de performanță sunt următoarele:

- $\overline{n}_{SAi} = \frac{\rho_i}{1-\rho_i}, i = 1,...,n$ - numărul mediu de cereri în sistemul $SA_i$;
- $\overline{n}_{şi} = \frac{\rho_i^2}{1-\rho_i}, \ i = 1,...,n$ - numărul mediu de cereri în şirul de aşteptare al $SA_i$;
- $\overline{\tau}_{SAi} = \overline{n}_{SAi} / \lambda_i$, $\overline{\tau}_{şi} = \overline{n}_{:i} / \lambda_i$ (respectiv ) - durata medie de aşteptare a unei cereri în $SA_i$ (respectiv în şirul de aşteptare);
- $\overline{N}_{RSA} = \sum_{i=1}^{n} \overline{n}_{SAi}$ - numărul mediu de cereri în rețeaua *RSA*;
- $\overline{\tau}_{RSA} = \overline{N}_{RSA} / \lambda_0$ - durata medie de aşteptare a unei cereri în rețeaua *RSA*;
- $\pi_i(m_i) = (1-\rho_i) \cdot \rho_i^{\ m}, i = 1,...,n; j \ge 0$ - probabilitatea staționară că în momentul considerat în sistemul $SA_i$ se află $m_i$ cereri;

♦ $\pi_{RSA}(0) = \prod_{i=1}^{n}(1-\rho_i)$ - probabilitatea staționară că în momentul considerat în rețea RSA nu este nici o cerere;

♦ $\pi_{RSA}(\vec{m}) = \pi_{RSA}(0) \cdot \prod_{i=1}^{n} \rho_i^{m_i}$ - probabilitatea staționară că în momentul considerat în rețeaua *RSA* se află cereri, adică în acelaşi moment în fiecare $SA_i$ sunt $m_i$, $i$=1,...,$n$, cereri.

Deseori la analiza *RSA* se introduc unele restricții temporale față de durata medie de aflare a unei cereri (mesaj) în rețea, ceea ce poate, de exemplu, fi redată de relația $\bar{\tau}_{RSA} \le \tau^*_{RSA}$. Aici $\bar{\tau}_{RSA} = \bar{N}_{RSA} / \lambda_0$, iar $\bar{\tau}^*_{RSA}$ este specificată în prealabil de beneficiar.

În cazul în care această relație nu este verificată (adică $\bar{\tau}_{RSA} > \bar{\tau}^*_{RSA}$) este necesar de a introduce unele modificări în ce constă numărul de serveri folosiți în sistemul $SA_j$ şi anume:

- determinăm $N^*_{\max} = \lambda_0 \cdot \bar{\tau}^*_{RSA}$;
- calculăm $\hat{n}_{SA_j} = \max_{(i)}\{\bar{n}_{SAi}\}$ şi pentru fiecare $SA_j$, $j$=1,...,$n$ modificăm $\bar{n}^*_{SAj} = \hat{n}_{SAj} / l^*_j,\ \ l^*_j \ge 1$ (număr întreg) până când vom obține $N^*_{\max} \ge \sum_{j=1}^{n} \bar{n}^*_{SAj}$.

$k^*_j = \left]\frac{\lambda_j}{\mu_j}\left(1+\frac{1}{\bar{n}^*_{SAj}}\right)\right[$ (aici $]x[$ - este număr întreg al lui $x$) numărul de serveri (canale) necesare pentru a obține caracteristica temporală specificată de beneficiar.

### 3.4.2. Scopul lucrării

Studierea metodelor de descriere, elaborare a algoritmilor de funcționare şi de evaluare a rețelelor stochastice deschise cu sisteme de aşteptare model Jackson.

### 3.4.3. Ordinea îndeplinirii lucrării

Îndeplinirea lucrării prevede următoarele acțiuni:

- pentru o rețea *RSA* deschisă, model Jackson (conform variantei date), redată de matricea de transfer *R* şi parametrii fluxurilor de intrare şi de servire a cererilor (clienților) de construit graful lanțului *LMC* în timp continuu, ce descrie comportarea stabilă a rețelei date;

- de scris ecuațiile de echilibru ale ratelor de probabilități pentru fiecare macrostare a lanțului *LMC* subiacent rețelei *RSA* date;

- de elaborat algoritmii şi programul de calcul numeric pentru determinarea repartizării probabilităților staționare de macrostare;

- pentru varianta stabilită de profesor, de evaluat indicatorii numerici de performanță ai *RSA*, astfel încât $\bar{\tau}_{RSA} \le \bar{\tau}^*_{RSA}$, unde $\bar{\tau}^*_{RSA}$ este durata medie de aflare a unei cereri în rețea, specificată în prealabil.

### 3.4.4. Prezentarea şi susținerea lucrării de laborator

Lucrarea se prezintă în formă de referat şi se susține profesorului la calculator în mod practic.

*Conținutul referatului:*

- foaia de titlu cu denumirea lucrării;

- obiectivele lucrării de laborator, scurte date teoretice;

- schema-bloc, programul şi rezultatele numerice ale indicatorilor de performanță a rețeli *RSA* date.

- concluzii.

### 3.4.5. Întrebări şi teme

- ecuațiile Chapman-Kolmogorov ce descriu comportamentul rețelei *RSA*;

- echilibrul static local şi global al fluxurilor de cereri în *RSA*;

- legea lui Little pentru o rețea *RSA*;

- funcții de repartiții ale duratei de servire şi clasificarea lor;

- rețele *RSA* închise model *Norton*;

- teorema *BCMP* pentru rețele *RSA*.

# Bibliografie


1. Артамонов Г.Т., Брехов О.М. *Аналитические вероятностные модели функционирования ЭВМ*. М.: Энергия, 1978.
2. Башарин Г.П., Бочаров П.П., Коган Я.А. *Анализ очередей в вычислительных сетях*. М.: Наука, 1989.
3. Бужор П.Ф., Гуцуляк Е.Н. *Анализ производительности информационно-вычислительной системы с мультидоступом*. Известия Академии наук МССР, Серия физико-технических и математических наук, № 1, с. 105-110, Кишинэу, 1980.
4. Саати Т. *Элементы теории массового обслуживания и ее приложения*. М.: Сов. Радио, 1965.
5. Cinlar E. *Introduction to stochastic processes*. Prentice Hall, 1975.
6. Гнеденко Б.В., Коваленко И.Н. *Введение в теорию массового обслуживания* М.: Наука, 1987.
7. Gelenbe E. et al. *Reseaux de files d'attente. Modelisation et traitement numerique.* France, Edition Homes et Techniques, 1980.
8. Феллер В. *Введение в теорию вероятностей и ее приложения*. Т.1, пер. с англ. М.: Мир, 1967.
9. Кемени Дж., Снелл Дж. *Конечные цепи Маркова.* М.: Наука, 1970.
10. Kleinrock L. *Queueing systems, vol. 1 "Theory", vol. 2 "Computer aplications"*. Wiley&Sons, 1976.
11. Кофман А., Крюон Р. *Массовое обслуживание. Теория и практика*. М.: Мир, 1965.
12. Neuts M.F. *Matrix-geometryc solutions in stochastic models - an algorithmic approach*. The John Hopkins University Press, London, 1981.
13. Onicescu O. *Teoria probabilităţilor şi aplicaţii*. Bucureşti, Editura Didactică şi pedagogică, 1978.
14. Plateau B., and Fourneau I.M. *A methodology for solving Markov models of parallel systems*. Journal of parallel and distributed computing, 12:370-378, 1991.
15. Philippe B., Saad Y., and Stewart W.J. *Numerical methods in Markov chain modeling*. Rapport de rechardi 1115, INRIA, Rocquencourt, France, November 1989.
16. Stewart W.J. *Introduction to the numerical solution of Markov chains*. Princeton University Press, USA, 1994.
17. Тихонов В.И., Миронов М.А. *Марковские процессы*. М.: Сов. радио, 1977.
18. Bucholz P., Ciardo G., Donatelli S., Kemper P. Complexity of memory-efficient Kronecker operations with applications to the solutions of the Markov models. *Informs J. Comp*., no. 12(3), Summer 2000, p. 203-222.