Curs 12: Procese Poisson

1.1 Procese Poisson: definiţie, proprietăţi

Sistemele coadă sau liniile (firele) de aşteptare sunt modele conceptuale ale unor sisteme ce constau din entităţi ce fac coadă pentru a fi servite.

Un sistem coadă conţine unul sau mai multe servere, adică staţii care execută servirea clienţilor. Clienţii reprezintă entităţi omogene, solicitând toate acelaşi tip de serviciu de la o staţie. Clienţii sosiţi aşteaptă momentul începerii servirii lor. Dacă momentul sosirii nu coincide cu cel al servirii, atunci se aşează la coadă în aşteptarea serviciului. Se presupune că instantaneu, în momentul creerii unei disponibilităţi în sistemul de servire, un client din coadă, dacă există, este servit (intră în serviciu).

Există o teorie matematică a cozilor (aşteptării) care se bazează pe teoria proceselor stochastice. Simularea sistemelor de aşteptare este însă subiect de studiu al disciplinei Sisteme de evenimente discrete (*Discrete Event Systems* (DES)), care dezvoltă bazele teoretice ale tehnologiei controlate de calculator. Exemple de DES: sistemele de calcul, reţelele de comunicaţii, fabricaţia automată, sistemele trafic etc.

Exemple de clienţi şi servere:

• sistemul PC sau sistemul multiprocesor: clienţii sunt procesele şi job-urile, iar serverul este procesorul (procesoarele).

• sistemul de comunicaţii de date: clienţii sunt pachetele de informaţii, iar server este procesorul nod de reţea.

• sistemul bază de date: clienţii sunt cererile de tranzacţii (inserare, modificare, ştergere), serverul este serverul bază de date.

Noi studiem doar fluxul sosirii clienţilor la coadă, acesta fiind modelat de un proces Poisson.

Intuitiv, un proces stochastic este un fenomen aleator care evoluează în timp. Formal, un proces stochastic este o familie de variabile aleatoare (X_t), $t \in T$, definite pe acelaşi câmp de probabilitate $(\Omega, \mathcal{K}, P)$, cu valori într-o mulţime S ($S = \{1, 2, \dots, n\}$, $S = \mathbb{N}$ sau $S \subseteq \mathbb{R}$), numită spaţiul stărilor. Un proces stochastic în timp continuu $(N_t), t \geq 0$, ce ia valori în mulţimea numerelor naturale se numeşte *proces de numărare*. Variabila aleatoare N_t indică numărul de evenimente rare ce se produc în intervalul de timp $(0, t]$. Pentru $s < t$, $N_t - N_s$ este numărul de evenimente ce se produc în intervalul $(s, t]$.

Definiţia 1.1.1 Un *proces Poisson* este un proces stochastic de numărare (N_t), $t \geq 0$, ce verifică proprietăţile:

1) $N_0 = 0$;

2) Evenimentele rare ce se produc în intervale disjuncte sunt independente, adică pentru orice alegere de momente de timp $t_0 < t_1 < \cdots < t_n$, variabilele aleatoare $N_{t_1} - N_{t_0}$, $N_{t_2} - N_{t_1}$, ..., $N_{t_n} - N_{t_{n-1}}$ sunt independente;

3) Distribuţia de probabilitate a variabilelor $N_{t+s} - N_s$, $s \geq 0$, ce dau numărul de evenimente rare (numărul de clienţi ce intră în sistemul coadă) ce se produc în intervalul $(s, s+t]$, este

$$P(N_{t+s} - N_s = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!},$$

unde $\lambda > 0$ este un parametru fixat, numit *rata procesului Poisson*.

Implicaţiile acestei definiţii sunt următoarele:

a) Din condiţia 3) rezultă că distribuţia de probabilitatea a variabilelor aleatoare $N_{t+s} - N_s$ nu depinde decât de lungimea t a intervalului $(s, s+t]$, nu şi de extremităţi. Deci, cum intervalul $(0, t]$ are aceeaşi lungime ca şi intervalul $(s, s+t]$, avem că

$$P(N_{t+s} - N_s = k) = P(N_{t+s-s} - \underbrace{N_0}_{=0} = k) = P(N_t = k).$$

Prin urmare, pentru fiecare $t > 0$, variabila aleatoare N_t are distribuţia Poisson de parametru λt.

b) Se poate arăta, folosind condiţia 3) din definiţie, că probabilitatea ca într-un interval de timp h foarte mic (apropiat de zero) să se producă un singur eveniment numărat de procesul Poisson este

$$Q(h) = P(N_{t+h} - N_t = 1) = \lambda h + \mathfrak{o}(h),$$

unde $\mathfrak{o}(h)$ este o funcţie cu proprietatea că $\lim_{h \to 0} \frac{\mathfrak{o}(h)}{h} = 0$, adică o funcţie care tinde la zero mai repede decât h (o astfel de funcţie este de exemplu h^2).

Această relaţie indică faptul că probabilitatea ca într-un interval de timp h foarte mic să se producă un singur eveniment este proporţională cu $\lambda > 0$. Calculând

$$\lim_{h \to 0} \frac{Q(h)}{h} = \lim_{h \to 0} \frac{\lambda h + \mathfrak{o}(h)}{h} = \lambda,$$

obţinem intensitatea producerii evenimentelor rare contorizate de procesul Poisson sau rata procesului Poisson, λ.

c) Se mai poate arăta că

$$P(N_{t+h} - N_t \geq 2) = \mathfrak{o}(h).$$

Această relaţie ilustrează că probabilitatea ca în intervalul $(t, t+h]$, de lungime h, să se producă mai mult de o intrare este $0+\mathfrak{o}(h)$, adică pentru h foarte mic această probabilitate este aproape 0, iar viteza de variaţie a probabilităţii $Q(h) = P(N_{t+h} - N_t \geq 2)$ este

$\lim_{h \to 0} Q(h)/h = 0$. Cu alte cuvinte, nu se produc simultan două evenimente rare (în sistemul coadă nu intră simultan doi sau mai mulţi clienţi).

Exemplul 1. Clienţii sosesc la o coadă conform unui proces Poisson cu rata de 6 clienţi pe oră. Fie N_t numărul de clienţi sosiţi până la momentul t, inclusiv. Să se calculeze, exploatând proprietăţile procesului Poisson, următoarele probabilităţi:

a) Probabilitatea ca până la momentul $t = 5$ să sosească 3 clienţi, adică $P(N_5 = 3)$;

b) $P(N_2 = 3, N_5 = 7)$;

c) $P(N_5 = 7 | N_2 = 3)$.

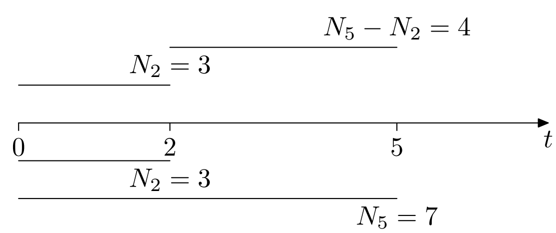

Fig.1: Echivalenţa dintre $(N_2 = 3, N_5 = 7)$ şi $(N_2 = 3, N_5 - N_2 = 4)$.

Rezolvare: Variabila aleatoare N_t are distribuţia Poisson de parametru $6t$, deci

$$P(N_t = k) = e^{-6t} \frac{(6t)^k}{k!}, \; k \in \mathbb{N}.$$

În particular, pentru a) avem

$$P(N_5 = 3) = e^{-6 \cdot 5} \frac{(6 \cdot 5)^3}{3!}.$$

b) Evenimentul $(N_2 = 3, N_5 = 7)$ coincide cu evenimentul $(N_2 - N_0 = 3, N_5 - N_2 = 4)$ (vezi Fig.1). Conform proprietăţii 2) din definiţie, pentru diviziunea

$$t_0 = 0 < t_1 = 2 < t_2 = 5$$

a intervalului de timp $[0, 5]$ variabilele aleatoare $N_2 - N_0$, $N_5 - N_2$ sunt independente, de unde rezultă

$$\begin{aligned} P(N_2 - N_0 = 3, N_5 - N_2 = 4) &= P(N_2 - N_0 = 3) P(N_5 - N_2 = 4) \\ &= P(N_2 = 3) P(N_3 = 4) \\ &= e^{-6 \cdot 2} \frac{12^3}{3!} e^{-6 \cdot 3} \frac{18^4}{4!}. \end{aligned}$$

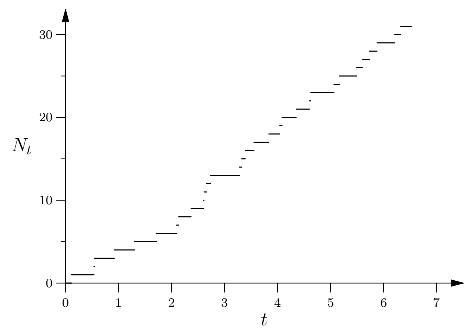

Fig.2: Vizualizarea procesului Poisson de rată $\lambda = 5$ pe intervalul [0,7].

c) Se aplică definiţia probabilităţii condiţionate şi b). Astfel, se obţine:

$$P(N_5 = 7|N_2 = 3) = \frac{P(N_2 = 3, N_5 = 7)}{P(N_2 = 3)} = e^{-18}\frac{18^4}{4!}.$$

Relaţia unui proces Poisson cu distribuţia exponenţială

Deoarece nu intră la coadă simultan doi sau mai mulţi clienţi (nu se produc simultan două evenimente rare), momentele aleatoare ale sosirilor/producerilor evenimentelor rare în intervalul de timp $[0, T]$ sunt distincte, $0 < t_1 < t_2 < \cdots < t_{n-1} < t_n \leq T$, iar variabila aleatoare N_t, ce dă numărul clienţilor sosiţi până la momentul T, ia valorile:

$$\begin{aligned}
N_t &= 0, && \text{pentru } t \in [0, t_1),\\
N_t &= 1, && \text{pentru } t \in [t_1, t_2),\\
&\vdots\\
N_t &= n-1, && \text{pentru } t \in [t_{n-1}, t_n).
\end{aligned}$$

Momentele de timp ale intrării clienţilor în sistem sunt aleatoare, deci şi lungimile intervalelor dintre două sosiri consecutive sunt aleatoare. Notăm cu X variabila aleatoare ce dă lungimea intervalelor inter–sosiri.

Propoziţia 1.1.1 *Variabila aleatoare inter–sosiri asociată unui proces Poisson de rată λ are distribuţia exponenţială de parametru $\theta = 1/\lambda$, $X \sim Exp(\theta = 1/\lambda)$.*

Demonstraţie: Pentru a simplifica şi întelege mai bine demonstraţia vom presupune că variabila X dă lungimea intervalului de timp până la sosirea primului client, inclusiv. Să calculăm funcţia de repartiţie a variabilei aleatoare X în acest caz:

$$F_X(t) = P(X \leq t) = 1 - P(X > t).$$

Dar evenimentul $(X > t)$ (lungimea intervalului intersosiri este mai mare decât $t > 0$) este echivalent cu evenimentul $(N_t = 0)$, adică până la momentul t nu a sosit nici un client. Astfel, cele două evenimente au aceeaşi probabilitate:

$$P(X > t) = P(N_t = 0) = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t},$$

deci funcţia de repartiţie a variabilă aleatore X este

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

şi 0 în rest, deoarece pentru $t < 0$, $P(X \leq t) = 0$. Prin urmare, funcţia de repartiţie a variabilei aleatoare X coincide cu funcţia de repartiţie a distribuţiei exponenţiale de parametru $\theta = 1/\lambda$. □

Consecinţă: Timpul mediu între două sosiri consecutive, cu rata sosirilor λ, este

$$M(X) = 1/\lambda,$$

deoarece pentru $X \sim \text{Exp}(\theta)$, $M(X) = \theta$.

Notăm cu X_k variabila aleatoare ce dă lungimea intervalului de timp dintre sosirea clientului $k-1$ şi k. Variabilele aleatoare X_k sunt independente şi identic distribuite, $X_k \sim \text{Exp}(1/\lambda)$. Deci, $M(X_k) = 1/\lambda$ şi $\sigma^2(X_k) = 1/\lambda^2$.

Variabila aleatoare

$$T_n = X_1 + X_2 + \cdots + X_n$$

dă momentul sosirii celui de-al n-lea client în sistem.

Suma a n variabile aleatoare independente şi identic distribuite $Exp(\theta)$ este o variabilă aleatoare ce are distribuţia de probabilitate n–Erlang, deci momentul intrării în sistem a clientului n, T_n, este o variabilă aleatoare n–Erlang. Momentul mediu al sosirii este

$$M(T_n) = M(X_1) + M(X_2) + \cdots + M(X_n) = n/\lambda,$$

iar dispersia acestuia este

$$\sigma^2(T_n) = \sigma^2(X_1) + \sigma^2(X_2) + \cdots + \sigma^2(X_n) = n/\lambda^2,$$

deoarece variabilele aleatoare $X_1, X_2, \ldots, X_n$ sunt independente.

Exemplul 2. Spamurile sosesc în inbox-ul adresei `student.absent@yahoo.com` cu o rată $\lambda_s = 2/\text{oră}$, iar în mod independent, mailurile uzuale (obişnuite) cu o rată $\lambda_u = 5/\text{oră}$.

a) Care este probabilitatea ca `student.absent` să nu primească nici un spam de la ora 12 la ora 15?

b) La ora 8 dimineaţa, când a accesat inbox-ul, un student nu a găsit nici un mail obişnuit nou. Care este timpul mediu ce trece până la sosirea a 4 mailuri obişnuite?

Rezolvare: Notăm cu (N_t^u), respectiv (N_t^s) procesul Poisson al intrării email-urilor uzuale, respectiv a spam-urilor în inbox.

a) Se cere $P(N_{15}^s - N_{12}^s = 0)$. Deoarece distribuţia de probabilitate a variabilelor $N_t^s - N_\tau^s$, $t > \tau$, nu depinde decât de lungimea intervalului $t - \tau$, adică

$$P(N_t^s - N_\tau^s = 0) = P(N_{t-\tau}^s = 0),$$

rezultă

$$P(N_{15}^s - N_{12}^s = 0) = P(N_3^s = 0) = e^{-\lambda_s 3}(\lambda_s\, 3)^0/0! = e^{-6} \approx 0.0025,$$

deci probabilitatea este extrem de mică.

b) Fie X_1, X_2, X_3, X_4 variabilele aleatoare inter–sosiri pentru procesul Poisson (N_t^u), al intrării mesajelor uzuale în inbox, de rată $\lambda_u = 5$. Variabila aleatoare ce dă momentul sosirii celui de-al patrulea mail uzual (non-spam) este $T_4 = X_1 + X_2 + X_3 + X_4$. Deoarece variabilele aleatoare X_1, X_2, X_3, X_4 sunt identic distribuite, $X_k \sim \text{Exp}(1/\lambda_u)$, rezultă că $M(T_4) = 4/\lambda_u = 4/5$ ore.

1.1.1 Simularea unui proces Poisson

Cunoscând distribuţia de probabilitate a inter–sosirilor putem genera momentele sosirilor astfel:

$$\begin{aligned} t_0 &= 0, \\ t_k &= t_{k-1} + x_k,\ k = 1, 2, \ldots \end{aligned}$$

unde x_k sunt observaţii asupra variabilelor aleatoare $X_k \sim \text{Exp}(1/\lambda)$, $k \in \mathbb{N}^*$. Reamintim că o observaţie asupra variabilei aleatoare $X \sim \text{Exp}(\theta)$ se generează astfel:

`u= urand();`

$\mathtt{x} = -\theta * \ln(\mathtt{1} - \mathtt{u})$;

Avem astfel următorul algoritm de simulare pe un interval fixat $[0, T]$ a momentelor producerii evenimentelor contorizate de un proces Poisson (momentele sosirilor clienţilor în sistemul coadă), de rată λ:

function SimulProcPoisson(λ, T)
    $k = 0$; $t_0 = 0$**;**
    do
        u=urand();
        $\mathbf{x} = -(\mathbf{1}/\lambda) * \ln(\mathbf{1} - \mathbf{u})$**;**
        k=k+1;
        $\mathbf{t_k} = \mathbf{t_{k-1}} + \mathbf{x}$**;**
    while $(\mathbf{t_k} \leq \mathbf{T})$**;**
    return $t_0, t_1, \ldots$
end function

Indicele ultimului $t_n \leq T$ dă numărul de clienţi intraţi în sistemul coadă până la momentul T, iar $t_1, t_2, \ldots, t_n$ sunt momentele intrărilor. $N_t = k$, pentru $t \in [t_k, t_{k+1})$, $k = \overline{1, n-1}$ şi $N_t = n$ pentru $t \in [t_n, T]$.

În Fig 3 sunt vizualizate două fluxuri ale sosirilor, momentele sosirii fiind generate conform algoritmului prezentat.

1.1.2 Operaţii cu procese Poisson

Pentru a enunţa două proprietăţi importante ale proceselor Poisson, dăm mai întâi un exemplu:

Exemplul 3. Mailurile uzuale (non-spam) sosesc în inbox cu o rată $\lambda_u = 3$/oră, iar în mod independent, spam-urile sosesc cu o rată $\lambda_s = 9$/oră.

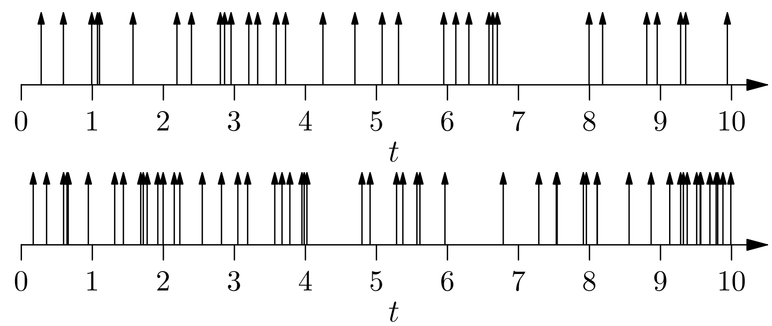

Fig.3: Simularea intrărilor într-un sistem coadă în intervalul $[0, 10]$. Sus, rata sosirilor este $\lambda = 3$, iar jos $\lambda = 5$. Deasupra punctului de coordonate $(t_k, 0)$ este "vizualizat" clientul.

Cu o probabilitate $p_1 = 2/5$ spam-urile sunt oferte de produse, cu o probabilitate $p_2 = 1/10$ sunt de la un moştenitor bogat din Nigeria, care propune să verse o parte din cash-ul moştenit în contul destinatarului, iar cu probabilitatea $p_3 = 1/2$ sunt spam-uri diverse.

Există deci două fluxuri de intrare ale mailurilor, unul cu rata λ_u şi unul cu rata λ_s. Este natural să ne întrebăm care este rata fluxului global al mailurilor de orice fel.

Pe de altă parte, fluxul spam-urilor are trei subfluxuri. De asemenea, prezintă interes determinarea ratei fiecărui subflux cunoscând rata λ_s a fluxului principal al spamurilor şi probabilităţile p_1, p_2, p_3 ca spam-urile să fie de tip 1, 2 sau 3.

În continuare dăm proprietăţi ale procesului Poisson care răspund acestor întrebări.

Suprapunerea (mixarea) mai multor procese Poisson independente.

Pentru a deduce distribuţia de probabilitate a procesului rezultat din suprapunerea sau mixarea mai multor procese Poisson independente demonstrăm:

Propoziţia 1.1.2 *Dacă X, Y sunt variabile aleatoare independente având distribuţia Poisson, $X \sim Poiss(\lambda_1)$, $Y \sim Poiss(\lambda_2)$, atunci suma lor are de asemenea distribuţia Poisson de parametru $\lambda = \lambda_1 + \lambda_2$, $X + Y \sim Poiss(\lambda_1 + \lambda_2)$.*

Demonstraţie: Pentru a arăta că variabila sumă $X + Y$ are distribuţia Poisson de parametru $\lambda_1 + \lambda_2$ demonstrăm că

$$P(X + Y = k) = e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^k}{k!},\ k \in \mathbb{N}.$$

Evenimentul $(X + Y = k)$ se scrie ca o reuniune de evenimente mutual exclusive astfel:

$$(X + Y = k) = (X = 0, Y = k) \cup \cdots \cup (X = i, Y = k - i) \cup \cdots \cup (X = k, Y = 0),$$

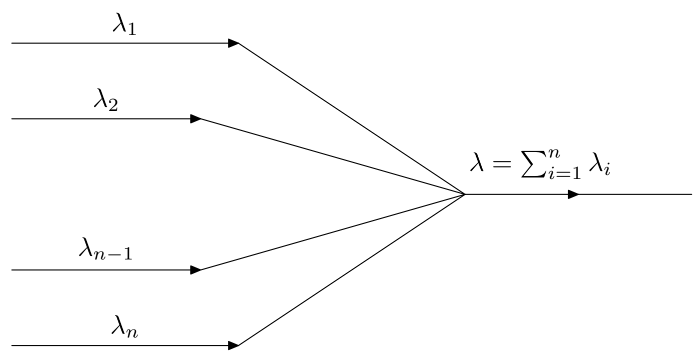

Fig.4:

deci probabilitatea sa este

$$
\begin{aligned}
P(X+Y=k) &= \sum_{i=0}^{k} P(X=i, Y=k-i) \overset{X,Y\ ind.}{=} \sum_{i=0}^{k} P(X=i)P(Y=k-i) \\
&= \sum_{i=0}^{k} e^{-\lambda_1} \frac{\lambda_1^i}{i!} e^{-\lambda_2} \frac{\lambda_2^{k-i}}{(k-i)!} = \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \underbrace{\sum_{i=0}^{k} \frac{k!}{i!(k-i)!} \lambda_1^i \lambda_2^{k-i}}_{\text{binomul lui Newton}} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^k}{k!}.
\end{aligned}
$$

□

Se poate demonstra, mai general, că dacă $X_1, X_2, \ldots, X_n$ sunt variabile aleatoare independente Poisson distribuite, de parametrii $\lambda_1, \lambda_2, \ldots, \lambda_n$, atunci $X_1 + X_2 + \cdots + X_n$ are de asemenea distribuţia Poisson de parametru $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

Ca o consecinţă a proprietăţii precedente avem:

Propoziţia 1.1.3 *Dacă* $(N_t^1), (N_t^2), \ldots, (N_t^n)$ *sunt* n *procese Poisson independente de rate* $\lambda_1, \lambda_2, \ldots, \lambda_n$*, atunci procesul* $(N_t = N_t^1 + N_t^2 + \cdots + N_t^n)$ *este de asemenea un proces Poisson de rată* $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

Se spune că procesul Poisson (N_t), definit mai sus, este rezultatul suprapunerii sau combinării celor n procese. Suprapunerea proceselor Poisson poate fi interpretată ca o coadă constituită din combinarea a n cozi (Fig.4). Intervalul inter–sosirilor în procesul Poisson rezultat din suprapunerea a n procese Poisson independente de rate $\lambda_1, \lambda_2, \ldots, \lambda_n$ este $X \sim \text{Exp}(1/(\lambda_1 + \lambda_2 + \cdots + \lambda_n))$.

Exemplul 4. Fluxul global al mail-urilor din Exemplul 2 se obţine din combinarea fluxului mailurilor uzuale cu spam-urile. Prin urmare, fluxul sosirii mailurilor de orice fel este un proces Poisson de rată $\lambda = \lambda_u + \lambda_s = 3 + 9 = 12$ mailuri/oră.

Ramificarea (descompunerea) unui proces Poisson în n subprocese

Propoziţia 1.1.4 *Dacă* (N_t) *este un proces Poisson de rată* λ *şi* $p_1, p_2, \ldots, p_n \in (0,1)$ *cu* $p_1 + p_2 + \cdots + p_n = 1$*, atunci subfluxurile generate sunt procese Poisson independente având ratele* $p_1\lambda, p_2\lambda, \ldots, p_n\lambda$.

Se spune că fluxul principal se ramifică în n subfluxuri. Evident că suprapunerea subfluxurilor dă fluxul principal de rată $\lambda = p_1\lambda + p_2\lambda + \cdots + p_n\lambda = (p_1 + p_2 + \cdots + p_n)\lambda$.

O astfel de situaţie apare în modelarea şi simularea reţelelor de cozi, în care un client după ce este servit are mai multe opţiuni: alege la întâmplare alte servere din reţea sau iese din reţea .

Exemplul 5. Să se determine rata subfluxurilor spam-urilor constituite din mesaje ce sunt oferte de produse, mesaje din Nigeria, respectiv spam-uri diverse. Care este probabilitatea ca în 12 ore în inbox să intre cel puţin un spam din Nigeria?

Rata fluxului spam-urilor este $\lambda_s = 9$. Ratele celor trei subfluxuri sunt $9p_1, 9p_2, 9p_3$, adică $18/5, 9/10, 9/2$ mailuri pe oră. Notăm cu (N_t^g) procesul Poisson al sosirii spam-urilor din Nigeria. Se cere $P(N_{12}^g \geq 1) = 1 - P(N_{12}^g = 0)$. Deoarece procesul (N_t^g) are rata $9/10 = 0.9$, avem

$$P(N_{12}^g = 0) = e^{-0.9\times 12} = e^{-10.8} = 2.039950341117192e - 005,$$

deci $P(N_{12}^g \geq 1) = 1 - P(N_{12}^g = 0) = 0.99997960049659$.