

Distribuirea fluxurilor de interpelari în rețelele locale

Prof.dr. Ion BOLUN, Academia de Studii Economice, Chisinau, bolun@ase.md
Anatol CIUMAC, Accent Electronic S.A., Chisinau, aciumac@accent.md

Multiserver LANs that offer many categories of services with absolute priority discipline are investigated. There is formulated the problem of optimal user request flows distribution among LANs' servers; the goal is to minimize the functional that depends on the mean response time to user requests of different categories. Analytical solutions for particular cases referred to two different criteria are obtained.

Keywords: LAN, server, distribution, request, services with priorities, response time.

Introducere

Asigurarea calitatii necesare a serviciilor informatice oferite de rețele este o problema de prim plan [1], [5]. În rețelele informatice moderne se solicita implementarea din ce în ce mai largă a unor servicii ce presupun folosirea intensă a procesării de grafică, de secvențe video, de video la cerere etc. În asemenea condiții cresc cerințele față de procesarea operativă a interpelărilor și în rețele predomină [2], de regulă, fluxurile de interpelări operative. Satisfacerea cerințelor către deservirea unor asemenea interpelări, la costuri rezonabile, necesită folosirea în rețea a unei discipline cu prioritate în deservirea interpelărilor de diferite categorii. În lucrare se cercetează aspectele referitoare la distribuirea fluxurilor de interpelări de diferite categorii între serverele rețelelor locale, urmărind minimizarea unor funcționale dependente de durata medie de răspuns la interpelare. Fluxurile de interpelări generate la stațiile utilizatorilor se consideră de tip Poisson, iar durata de servire a interpelărilor de către servere de distribuție exponențială. Setul de servere al LAN este deja cunoscut. Acesta poate fi determinat, de exemplu, folosind rezultatele [3]. Este formulată problema de optimizare în cauză și sunt obținute soluții analitice pentru două cazuri particulare de criterii. Cazul unei singure categorii de interpelări în rețea este cercetat în [4].

Modelul de funcționare a setului de servere

Rețeaua locală este o rețea cu difuzare ce constă din N stații de lucru ale utilizatorilor și un set M_j ,

$j = \overline{1, m}$ de servere; aici j este tipul, iar m – numărul total de servere în rețea. Stațiile utilizatorilor necesită oferirea de către servere a $i = \overline{1, n}$ categorii de servicii. Un server poate oferi unul sau mai multe din cele n categorii de servicii. Fiecare interpelare către serviciul i va fi referită la categoria i de interpelări pentru $i = \overline{1, n}$. Serviciile se oferă cu priorități. Se folosește disciplina absolută cu continuarea deservirii; mai înaltă este prioritatea interpelărilor de categoriile cu valoarea indicelui i mai mică. Durata servirii interpelărilor de către servere se consideră de repartiție exponențială, deoarece repartiția exponențială corespunde în mare parte realității [1, 5]. Fluxurile de interpelări ale utilizatorilor generate în rețea sunt de tip Poisson staționare, ceea ce are loc, de regulă, în rețele [1, 5]. De asemenea, se consideră că durata transferului datelor este mult mai mică, comparativ cu durata procesării acestora în rețea, de aceea aceasta nu se ia în considerare.

Notatii folosite:

t_{ji} – durata medie (rata $m_{ji} = 1/t_{ji}$) de deservire a unei interpelări de categoria i de către un server de categoria j ;

T_{ji} – durata medie de răspuns de către un server de tip j la o interpelare de categoria i ;

T_i – durata medie de răspuns la o interpelare de categoria i ;

g_{li} – rata fluxului de interpelări de categoria i generate de stația l a rețelei;

Λ_i – rata fluxului sumăr de interpelări de categoria i în rețea;

I_{ji} – rata fluxului sumăr de interpelări de categoria i ce se deservesc de un server de tip j .

Evident, rata Λ_i se determina ca

$$\Lambda_i = \sum_{l=1}^N g_{li}, i = \overline{1, n}. \quad (1)$$

$$T_{ji} = \frac{t_{ji}}{1 - \sum_{k=1}^{i-1} I_{jk} t_{jk}} + \frac{\sum_{k=1}^i I_{jk} t_{jk}^2}{\left(1 - \sum_{k=1}^{i-1} I_{jk} t_{jk}\right) \left(1 - \sum_{k=1}^i I_{jk} t_{jk}\right)}, j = \overline{1, m}, i = \overline{1, n}. \quad (2)$$

La formularea si solutionarea diverselor probleme, se va tine cont de necesitatea asigurarii functionarii sistemului cercetat în regim stationar. Pentru un server de tip j , regimul stationar de functionare exista, daca încarcarea serverelor este mai joasa decât 1, adica

$$\sum_{l=1}^N I_{li} t_{ji} < 1, j = \overline{1, m}. \quad (3)$$

Formularea matematica a problemei

La prioritatea absoluta cu continuarea deservirii, deservirea interpelarilor de prioritate mai înalta nu depinde de deservirea interpelarilor de prioritate mai joasa. De aceea ca criteriu de optimizare este oportuna folosirea minimului duratei medii de raspuns la interpelarile de categoria $i = 1$, apoi la cele de categoria $i = 2$ si tot asa pâna la cele de categoria $i = n$.

Asadar, fie sunt cunoscute marimile: $\Lambda_i; M_j, t_{ji}, j = \overline{1, m}, i = \overline{1, n}$. Se cere determinarea distribuirii $I_{ji}, j = \overline{1, m}, i = \overline{1, n}$ a fluxurilor de interpelari $\Lambda_i, i = \overline{1, n}$ între serverele $M_j, j = \overline{1, m}$ ale retelei, care ar asigura minimumul duratei medii T_i de raspuns la interpelarile de categoria i consecutiv pentru $i = 1$, apoi pentru $i = 2, \dots$ si, în sfârșit, pentru $i = n$

$$T_i = \frac{1}{\Lambda_i} \sum_{j=1}^m M_j \sum_{i=1}^n I_{ji} T_{ji} \rightarrow \min, \quad (4)$$

$i = \overline{1, n}$

la satisfacerea restrictiilor

$$\sum_{j=1}^m M_j I_{ji} = \Lambda_i, i = \overline{1, n}, \quad (5)$$

unde T_{ji} se determina conform formulei (2). De mentionat ca desi în cazul (4) se urmareste minimizarea T_i , solutia obtinuta poate avea o dispersie relativ mare a duratei

De asemenea, la deservirea exponentiala (repartitia exponentiala a duratei servirii) si disciplina absoluta cu continuarea deservirii interpelarilor utilizatorilor, au loc relatiile [5]:

servirii interpelarilor. Aceasta dispersie este cu atât mai mare, cu cât este mai mare diferenta între productivitatea serverelor. Deseori, însa, este importanta reducerea dispersiei duratei de raspuns la interpelari. În asemenea cazuri, ca functie scop ar putea fi folosita functia

$$T_{1i} = T_{2i} = \dots = T_{mi} = T_i ? \min. \quad (6)$$

Atât problema (4), (5), cât si problema (6), (5), se refera, luând în considerare relatiile (2), la probleme de programare matematica neliniara. Mai jos se va cerceta mai întâi cazul folosirii functiei scop (4) (p. 4), iar apoi si cel al folosirii functiei scop (6).

Funcția scop se determina de relatia (4)

Pentru acest caz, functia scop (4), tinând cont de relatia (2), ia forma (7):

$$T_i = \frac{1}{\Lambda_i} \sum_{j=1}^m M_j \left[\frac{\lambda_{ji} \tau_{ji}}{S} + \frac{\lambda_{ji}^2 \tau_{ji}^2 + H_e \lambda_{ji}}{S(S - \lambda_{ji} \tau_{ji})} \right] \rightarrow \min,$$

unde

$$S = 1 - \sum_{k=1}^{i-1} I_{jk} t_{jk}; H_e = \sum_{k=1}^{i-1} I_{jk} t_{jk}^2. \quad (8), \quad (9)$$

Problema (7), (5) poate fi rezolvata folosind metoda multiplicatorilor Lagrange. Functia Lagrange L corespunzatoare este

$$L = T_i + c \left(\Lambda_i - \sum_{j=1}^m M_j I_{ji} \right) \rightarrow \min, \quad (10)$$

unde c este multiplicatorul Lagrange, iar T_i se determina conform formulei (7).

Luând derivatele partiale de la L fata de variabilele $I_{ji}, j = \overline{1, m}$ si c , avem

$$\begin{cases} \frac{\partial L}{\partial \lambda_{ji}} = \frac{\partial T_i}{\partial \lambda_{ji}} + \chi(-M_j) = 0, j = \overline{1, m} \\ \frac{\partial L}{\partial \chi} = \Lambda - \sum_{j=1}^m M_j \lambda_{ji} = 0. \end{cases} \quad (11)$$

Din ecuatiile primei linii a sistemului (11), în rezultatul unor transformari obtinem

$$\mathbf{I}_{ji} = \mathbf{m}_{ji} \left(S - \sqrt{\frac{S \mathbf{t}_{ji} + H_e}{c \Lambda_i}} \right), \quad j = \overline{1, m}. \quad (12)$$

Înlocuind în ecuatia din linia a doua a sistemului (11) variabilele \mathbf{I}_{ji} , $j = \overline{1, m}$ cu expresiile pentru acestea din (12), obtinem

$$\frac{1}{\sqrt{c \Lambda_i}} = \frac{S \sum_{j=1}^m M_j \mathbf{m}_{ji} - \Lambda_i}{\sum_{j=1}^m M_j \mathbf{m}_{ji} \sqrt{S \mathbf{t}_{ji} + H_e}}, \quad (13)$$

si, înlocuind în (12), avem

$$\lambda_{ji} = S \mu_{ji} - \sqrt{\mu_{ji} (S + H_e \mu_{ji})} \frac{S \sum_{k=1}^m M_k \mu_{ki} - \Lambda_i}{\sum_{k=1}^m M_k \sqrt{\mu_{ki} (S + H_e \mu_{ki})}}, \quad j = \overline{1, m}, i = \overline{1, n}. \quad (14)$$

Substituind în (7) \mathbf{I}_{ji} , $j = \overline{1, m}$ cu expresiile pentru acestea din (14), durata minima T_i^* de raspuns la interelariile de categoria i în retea este

$$T_i^* = \frac{1}{\Lambda_i} \left[\frac{\left(\sum_{j=1}^m M_j \sqrt{\mu_{ji} (S + H_e \mu_{ji})} \right)^2}{S \sum_{j=1}^m M_j \mu_{ji} - \Lambda_i} - \frac{1}{S} \sum_{j=1}^m M_j (S + H_e \mu_{ji}) \right], \quad i = \overline{1, n}. \quad (15)$$

5. Functia scop

În acest caz, din (2) si (6) avem

$$T_i = T_{ji} = \frac{1}{S} \left[\mathbf{t}_{ji} + \frac{\mathbf{I}_{ji} \mathbf{t}_{ji}^2 + H_e}{S - \mathbf{I}_{ji} \mathbf{t}_{ji}} \right], \quad j = \overline{1, m}, \quad (16)$$

$$\text{de unde } \mathbf{I}_{ji} = S \mathbf{m}_{ji} - \frac{1}{T_i} - \frac{H \mathbf{m}_{ji}}{S T_i}, \quad j = \overline{1, m}. \quad (17)$$

Înlocuind în restrictia (5) \mathbf{I}_{ji} , $j = \overline{1, m}$ cu expresia pentru acestea din (17), obtinem

$$\left(1 - \frac{H}{S T_i} \right) \sum_{j=1}^m M_j \mathbf{m}_{ji} - \frac{1}{T_i} \sum_{j=1}^m M_j = \Lambda_i, \quad (18)$$

$$\text{de unde } T_i^* = T_i = \frac{\sum_{j=1}^m M_j (S + H \mu_{ji})}{S \left(\sum_{j=1}^m M_j \mu_{ji} - \Lambda_i \right)}, \quad i = \overline{1, n}. \quad (19)$$

si, înlocuind în (17),

$$\lambda_{ji} = S \mu_{ji} - (S + H \mu_{ji}) \frac{S \sum_{k=1}^m M_k \mu_{ki} - \Lambda_i}{\sum_{k=1}^m M_k (S + H \mu_{ki})}, \quad j = \overline{1, m}, i = \overline{1, n} \quad (20)$$

sau într-o alta forma

$$\lambda_{ji} = \frac{S^2 \left(\mu_{ji} \sum_{k=1}^m M_k - \sum_{k=1}^m M_k \mu_{ki} \right) + \Lambda_i (S + H \mu_{ji})}{\sum_{k=1}^m M_k (S + H \mu_{ki})}, \quad j = \overline{1, m}, i = \overline{1, n}. \quad (21)$$

Prezinta interes diferenta $\Delta T_i'$ dintre durata T_i pentru cazul din p. 4 (expresia (15)) si cea calculata conform formulei (19). În rezultatul unor transformari, avem

$$\Delta T_i' = \frac{\sum_{k=1}^m M_k \mu_{ki} \sum_{j=1}^m M_j (S + H\mu_{ji}) - \left(\sum_{j=1}^m M_j \sqrt{\mu_{ji} (S + H\mu_{ji})} \right)^2}{\Lambda_i \left(S \sum_{j=1}^m M_j \mu_{ji} - \Lambda_i \right)}, \quad i = \overline{1, n}. \quad (22)$$

Este usor de verificat ca are loc relatia

$$\Delta T_i' \Big|_{m_j=m, j=1, m} = 0, \quad i = \overline{1, n}. \quad (23)$$

Concluzii

Pentru cazul servirii exponentiale, sunt obtinute solutii analitice privind distribuirea I_{ji} , $j = \overline{1, m}, i = \overline{1, n}$ a interpelarilor utilizatorilor între serverele retelei. Se elaboreaza un produs program destinat distribuirii rationale a fluxurilor de interpelari ale utilizatorilor între serverele retelelor locale.

Bibliografie

1. G. Bolch, St. Greiner, H. Meer, K. S. Trivedi. *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*. 1998.

2. I.Bolun. *Macrosinteza retelelor de calculatoare*. – Chisinau: Editura ASEM, 1999.
3. I.Bolun, A.Ciumac. *Configuraiton of local area network set of servers*, Computer Science Journal of Moldova, vol. 10, no. 1.
4. I.Bolun, A.Ciumac. *Distribution of user requests among LAN's servers*. Proceedings of the International Symposium "Innovative applications of information technologies in business and management", May 2002, Iasi, Romania. – pp. 74-79.
5. T.G. Robertazzi. *Computer networks and systems: queuing theory and performance evaluation*. – New York: Springer-Verlag, 1994.