

SISTEME DE AȘTEPTARE PENTRU CALCULUL PARALEL ȘI DISTRIBUIT

Asist. drd. Felician Alecu

Academia de Studii Economice București

Rezumat: Performanțele sistemelor de așteptare în condiții de suprasolicitare joacă un rol important în ceea ce privește percepția consumatorilor asupra calității serviciilor. Timpii de așteptare și întârzierile sunt inevitabile în cadrul acelor sisteme de așteptare, care răspund unor cereri aleatoare, a căror apariție în timp și spațiu este guvernată de anumite legi probabilistice, cunoscute sau necunoscute. A oferi, în cadrul unui sistem de așteptare, capacități de servire suficiente, pentru a evita așteptările în absolut orice circumstanțe, implică costuri uriașe. Din acest motiv, scopul teoriei așteptării este acela de a ne asista în proiectarea unor sisteme de servire, în care există un echilibru între costurile de operare și timpii de așteptare ai utilizatorilor sistemului.

Cuvinte cheie: Teoria așteptării, sistem de așteptare, rețea de așteptare, sisteme distribuite geografic, planificarea programelor paralele, sisteme de calcul paralel și distribuit, eficiența programelor paralele.

1. Introducere

Teoria așteptării este principalul instrument folosit în studiul fenomenelor de congestionare, care apar atunci când numărul cererilor depășește capacitatea de servire. Termenul *coadă* provine din cuvântul latin *coda* și semnifică dispunerea sub forma unui șir a unor persoane aflate în așteptare.

Teoria așteptării se constituie ca o ramură a cercetărilor operaționale, ce are ca obiectiv abordarea matematică a cozilor sau a firelor de așteptare. Firele de așteptare și teoria așteptării și-au găsit aplicabilitate în diverse domenii cum ar fi telecomunicațiile, controlul traficului, anticiparea performanțelor computerelor, serviciile medicale (planificarea paturilor dintr-un spital), traficul aerian, vânzarea biletelor de avion, industria minieră etc.

În ceea ce privește latura practică, se pleacă de la un sistem din lumea reală, care se analizează, iar în urma analizei se decide asupra modelului matematic, care descrie cel mai bine sistemul. Acest model matematic poate fi unul nou creat sau se poate apela la unul deja existent. Prin analizarea acestui model matematic, se obțin o serie de concluzii care pot fi extinse asupra sistemului original. Activitatea de creare a unui model matematic are la bază simplificarea și aproximarea sistemului inițial. Astfel, detaliile considerate a fi nesemnificative nu sunt incluse în cadrul analizei. De asemenea, pentru a se putea transpune comportamentul sistemului sub forma unui model matematic este necesar ca datele primare să fie transformate în mărimi matematice, folosind aproximarea. De multe ori, în ceea ce privește comportamentul sistemului real, se fac o serie de presupuneri bazate, mai ales, pe intuiția și experiența analistului, decât pe observații efectuate în cadrul sistemului [1].

Atunci când se alege un model matematic, se poate opta pentru un model realist, dar care poate să nu conducă la nici un fel de rezultate, sau se poate opta pentru un model simplificat, care să genereze rezultate aproximative. Marea majoritate a rezultatelor exacte, din cadrul teoriei așteptării, s-au obținut în acele sisteme în care ritmul de sosire al consumatorilor și/sau ritmul de servire al stațiilor sunt guvernate de o lege de probabilitate exponențial negativă. Din fericire, în cazul majorității sistemelor de așteptare, cel puțin ritmul de sosire al consumatorilor în sistem este exponențial negativ deoarece sosirile consumatorilor pot fi modelate după o distribuție de tip Poisson [4].

În practică, teoria așteptării este, în special, folosită pentru a scoate în evidență disfuncționalitățile existente în cadrul unui sistem aflat în funcțiune și pentru a arăta direcțiile de eficientizare a funcționării acestuia prin indicarea valorilor pe care trebuie să le atingă anumite variabile de sistem, pentru a se ajunge la un nivel satisfăcător al performanțelor.

Principalul avantaj al teoriei așteptării este acela că ne pune la dispoziție informații extrem de importante despre timpii de așteptare, care apar în sistem pe baza unor date minimale despre caracteristicile sosirilor în sistem, caracteristicile stațiilor de servire și disciplina sistemului.

2. Structura unui sistem de așteptare

Un sistem de așteptare reprezintă un model generic, care se compune din următoarele trei elemente (figura 1):

- *clienții (consumatorii)* care solicită un serviciu;
- *stația de servire* ce are ca menire satisfacerea cererilor clienților; într-un sistem de așteptare, stația de servire poate avea un singur post sau pot exista mai multe posturi (număr finit sau infinit) identice, care lucrează în paralel;
- *firul de așteptare sau coada* care se formează în cazul în care consumatorii trebuie să aștepte.

Modelele din teoria așteptării se diferențiază între ele în ceea ce privește:

- legile de probabilitate ce guvernează sosirea clienților și servirea acestora;
- numărul posturilor din stația de servire;
- disciplina firului de așteptare;
- structura populației consumatorilor (număr finit sau infinit de consumatori).

Din acest motiv, este evident faptul că există nenumărate modele de sisteme și rețele de așteptare. Pentru descrierea celor mai utilizate, în literatura de specialitate s-a impus o notație de forma:

$$A/B/m \text{ sau } A|B|m$$

unde A și B sunt simboluri de litere, iar m reprezintă o constantă numerică. Această notație a fost folosită pentru prima dată de către Kendall, în anul 1953.

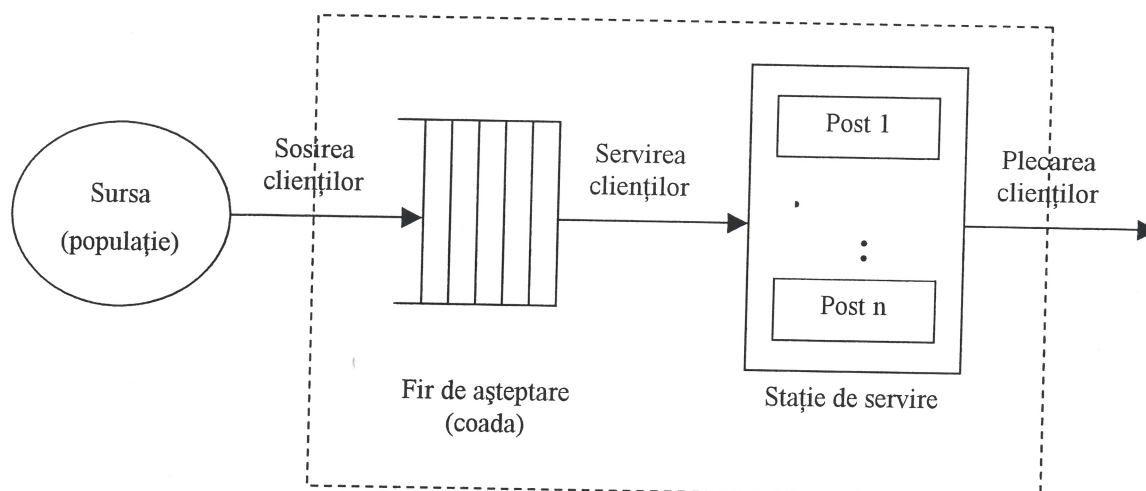


Figura 1. Structura unui sistem de așteptare

Semnificația celor trei simboluri este următoarea:

- prima literă indică legea de probabilitate care guvernează sosirile consumatorilor în sistem;
- cea de-a doua literă simbolizează legea de probabilitate care descrie procesul de servire;
- constanta numerică desemnează numărul de posturi identice, care funcționează în paralel, în cadrul stației de servire.

Literele A și B pot lua una din valorile prezentate în Tabelul 1, în timp ce constanta numerică poate lua valori de la 1 la ∞ .

Desigur că poate părea straniu faptul că, dintre toate distribuțiile de probabilitate posibile, doar patru (M , D , E , H) au asociate simboluri speciale, toate celelalte fiind incluse în categoria generică G care, evident, include și categoriile anterioare. Motivul pentru care doar cele patru distribuții de probabilitate au fost evidențiate este acela că numai aceste distribuții oferă avantaje semnificative în ceea ce privește analiza sistemului de așteptare [1].

Tabelul 1. Distribuții de probabilitate care au simboluri asociate

SIMBOL	SEMNIFICAȚIE
M	Distribuție Poisson (distribuție exponențial negativă)
D	Distribuție deterministică (la intervale constante)
E_k	Distribuție Erlang de ordinul k
H_k	Distribuție hyperexponențială de ordinul k
G	Distribuție generică (orice fel de distribuție)

Alte abrevieri care s-au impus sunt cele referitoare la cele mai întâlnite discipline de servire:

- FIFO indică faptul că primul client sosit este și primul servit (*first-in, first-out*);
- LIFO presupune servirea ultimului consumator dispus în coada de așteptare (*last-in, first-out*);
- SIRO semnifică faptul că servirea se face în mod aleator (*service in random order*).

Prin interconectarea a două sau mai multe sisteme de așteptare obținem o rețea de așteptare (rețea de sisteme de așteptare – figura 2). În cadrul unei rețele de așteptare, populația sursă a unui sistem poate proveni din clienții care au părăsit un alt sistem de așteptare, aparținând rețelei.

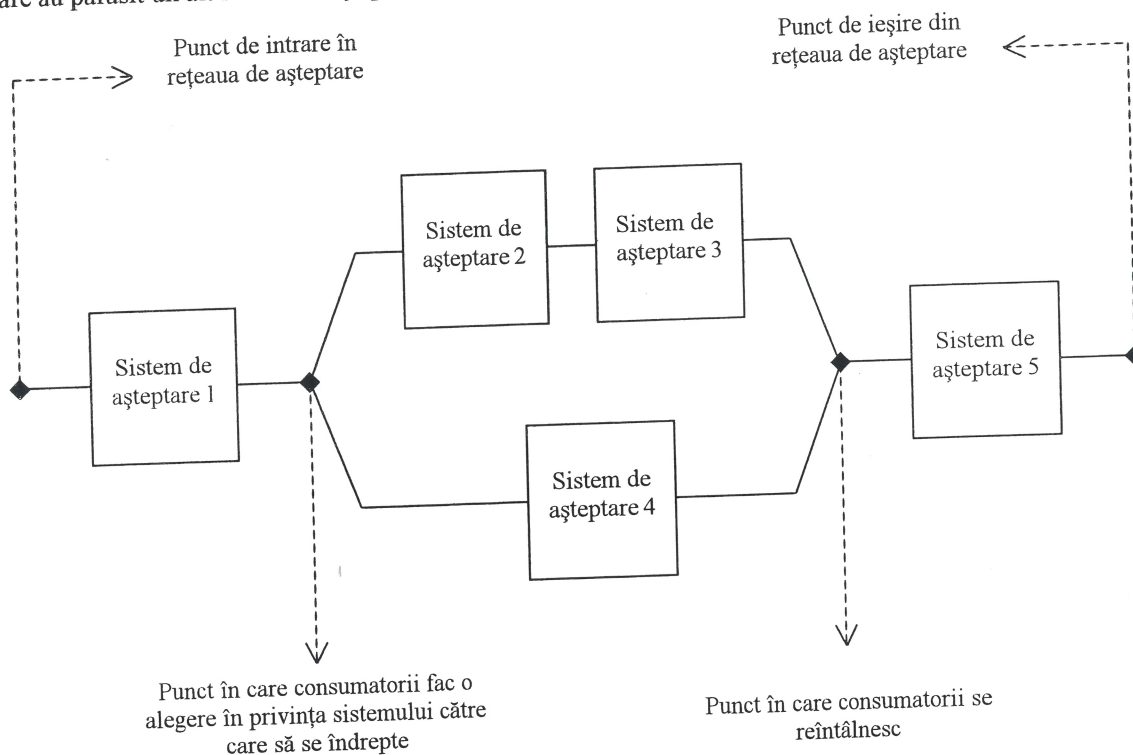


Figura 2. Rețea de așteptare

Capacitatea sistemului reprezintă un alt parametru important în descrierea unui sistem de așteptare și semnifică numărul maxim de consumatori ce pot fi, la un moment dat, în sistem, unde, prin numărul de clienți din sistem, înțelegem atât clienții care sunt în cadrul stației de servire, cât și pe cei care se află în coada de așteptare. Prin capacitatea cozii de așteptare, înțelegem numărul maxim de consumatori din care poate fi format firul de așteptare.

Pe baza legilor de probabilitate și a caracteristicilor stației de servire, se pot calcula duratele de așteptare atât pentru consumatori, cât și pentru stații într-un anumit interval de timp. În cazul în care se asociază costuri timpilor de așteptare, se obține o funcție economică dorită a fi minimizată, acționând fie asupra ritmului de sosire fie asupra capacității stației.

3. Extensii ale modelelor clasice

Modelele de așteptare clasice au la bază o serie de ipoteze legate de comportamentul consumatorilor cum ar fi faptul că nici un consumator nu se descurajează, părăsind astfel coada, pe parcursul așteptării pentru a fi servit. În plus, atunci când un consumator nu este acceptat în cadrul unui sistem de așteptare, datorită faptului că acesta funcționează la capacitate maximă, se presupune că utilizatorul nu va reveni mai târziu, ci se va îndrepta către un alt sistem care oferă servicii similare.

Modelele pot fi extinse prin exprimarea ritmurilor de sosire și de servire sub următoarea formă:

$$\lambda_n = c_n \lambda \quad n = 0, 1, 2, \dots$$

$$\mu_n = d_n \mu \quad n = 1, 2, 3, \dots$$

unde λ și μ sunt constante, iar c și d sunt coeficienți ai căror valori depind de starea sistemului de așteptare.

Cu ajutorul extensiilor aduse acestor modele putem lua în considerare o serie de comportamente ale consumatorilor care, până acum, erau considerate imposibile.

Un prim caz ar fi acela al utilizatorilor care prospectează starea sistemului de așteptare și pe baza celor observate decid să nu se alăture firului de așteptare, ci să se îndrepte către alte sisteme care oferă servicii similare.

O a doua situație ar fi reprezentată de cazul unui consumator care, deși se află în așteptare în coadă, decide să o părăsească, chiar dacă nu a beneficiat încă de serviciul pentru care a așteptat.

Cele mai utilizate forme ale coeficienților c_n și d_n sunt:

$$c_n = (n + j)^{-a} \quad n = 0, 1, 2, \dots$$

$$d_n = n^b \quad n = 1, 2, 3, \dots$$

unde j , a și b sunt constante pozitive.

Prin ajustarea valorilor lui a și/sau b pot fi modelate tipurile comportamentale ale utilizatorilor, indiferent de numărul posturilor de servire prezente în sistem sau de capacitatea acestora.

O extensie deosebit de importantă a modelelor de așteptare clasice o reprezintă acele sisteme în care consumatorii au atașate priorități care sunt luate în considerare de către stația de servire [4]. În cazul tuturor modelelor clasice, ordinea în care utilizatorii au acces la stația de servire este determinată de disciplina firului de așteptare. Din acest motiv, consumatorii sunt diferențiați doar pe baza ordinii în care au sosit în cadrul sistemului de așteptare fără să se țină cont de tipul și de durata serviciului pe care îl solicită.

Uzual, consumatorii potențiali ai unui sistem de așteptare sunt împărțiți în clase, iar fiecărei astfel de clase i se asociază o prioritate. Ori de câte ori un client părăsește sistemul, următorul consumator servit va face parte din clasa cu cea mai mare prioritate asociată. Alegerea unui anumit consumator din cadrul clasei cu prioritatea cea mai ridicată se efectuează în concordanță cu disciplina asociată clasei respective.

Figura 3 este reprezentarea schematică a unui sistem de așteptare cu r clase de priorități. Fiecare utilizator al sistemului este încadrat în una din cele r clase, iar fiecare clasă are asociat un număr k ($k = 1, \dots, r$), care reprezintă prioritatea clasei respective. Prin convenție, se consideră că, cu cât valoarea este mai mică, cu atât prioritatea clasei este mai mare.

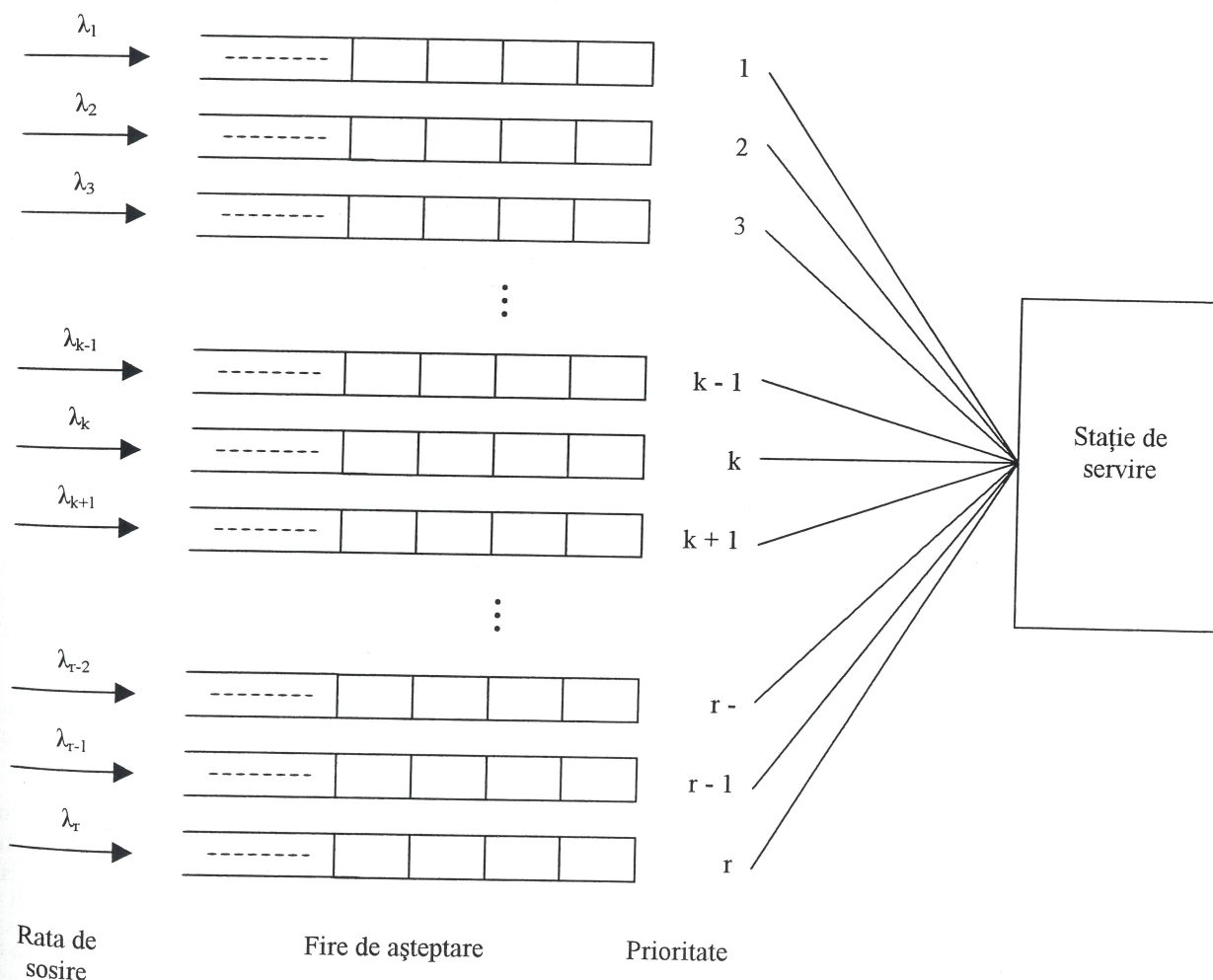


Figura 3. Sistem de așteptare cu r clase de priorități

Sosirile în cadrul fiecărei clase sunt Poisson de parametru λ_k pentru clasa k . Se observă că, prin utilizarea claselor de prioritați, se obțin de fapt r fire de așteptare paralele, care concurează pentru obținerea accesului la serviciul oferit de stația de servire.

Prioritățile care diferențiază clasele de utilizatori pot fi clasificate în două categorii:

- *preemptive* – în sistemele de așteptare care utilizează astfel de prioritați, utilizatorii cu prioritate ridicată nu vor fi niciodată nevoiți să aștepte în fir datorită consumatorilor cu prioritate mai scăzută. Astfel, dacă la coadă se așează un consumator care face parte dintr-o clasă cu prioritate mai ridicată decât cea a clientului care tocmai este servit, atunci servirea acestuia din urmă este întreruptă, urmând a fi reluată integral sau continuată din punctul din care s-a rămas. De asemenea, este posibil ca, în compensație pentru faptul că servirea i-a fost întreruptă, utilizatorului în cauză să-i fie asociată o nouă prioritate mai ridicată decât cea anterioară;
- *nonpreemptive* – servirea unui consumator din cadrul unui sistem de așteptare, care utilizează prioritați nonpreemptive, nu va fi niciodată întreruptă datorită sosirii în cadrul sistemului a unui consumator care face parte dintr-o clasă cu prioritate mai ridicată. Servirea acestuia din urmă se va face doar după ce clientul din stația de servire părăsește sistemul.

Multe sisteme de așteptare operează cu sisteme de prioritați diferite pentru clase diferite. Astfel, utilizatorii care fac parte dintr-o clasă care are asociată o prioritate foarte ridicată pot obține servicii preemptive, în timp ce consumatorii unei clase cu prioritate medie pot beneficia doar de servicii nonpreemptive, față de cei din clasele cu prioritați inferioare. În interiorul claselor aceleiași sistem de așteptare, pot fi folosite discipline diferite [1].

Astăzi, sistemele de așteptare, în care sosirea clienților din diverse locații geografice se poate face prin intermediul liniilor telefonice, sunt întâlnite la tot pasul. Apelurile sunt preluate de către un computer care memorează data, ora, originea și natura apelului. Succesiunea acestor apeluri care solicită un anumit serviciu reprezintă, de fapt, coada de așteptare numai că, de data aceasta, ea nu mai poate fi localizată fizic deoarece există doar în memoria unui computer.

Stația de servire, fie că este o mașină sau o persoană, urmează să se deplaseze la domiciliul solicitanților pentru efectuarea serviciului solicitat. Un astfel de sistem poartă numele de *sistem de așteptare distribuit geografic*.

Consumatorii și/sau posturile stației de servire se caracterizează prin faptul că au o distribuție geografică. Există și o incintă a stației de servire, care este localizată static și din care pleacă în teritoriu posturile componente. Fie consumatorii de deplasează către stația de servire, fie posturile acestora sunt cele care merg la locațiile consumatorilor pentru completarea serviciului.

Servirea poate fi făcută acolo unde se află localizat consumatorul, în incinta stației de servire sau pe parcursul traseului de la „domiciliul” clientului către stația de servire. În practică, pot fi întâlnite cazuri în care servirea consumatorilor se realizează prin combinarea celor trei variante anterioare. Un exemplu în acest sens îl reprezintă serviciile de ambulanță.

Momentul, locația și durata serviciului solicitat sunt elemente care variază de la un consumator la altul. Structura spațiului geografic din care provin clienții joacă un rol foarte important pentru sistemul de așteptare în ansamblu. Distribuție spațială a consumatorilor și caracteristicile drumului de la stația de servire până la consumator reprezintă factori critici în ceea ce privește calitatea serviciilor oferite.

4. Aplicații ale firelor de așteptare în procesarea paralelă și distribuită

În ultimii ani, sistemele paralele și distribuite au devenit din ce în ce mai atractive pentru aplicații cu cerințe de calcul intensive cum ar fi cele destinate simulării unor sisteme complexe (aerodinamică, meteorologie). Principalul avantaj al unor astfel de sisteme este raportul mai mult decât atractiv dintre preț și performanțele care pot fi obținute [3].

Pentru a putea beneficia din plin de multitudinea resurselor, programele paralele sunt divizate în task-uri independente, care intră în competiție, pentru fi executate pe procesoarele sistemului. Independența a două task-uri înseamnă obținerea aceluiași rezultat indiferent dacă ele sunt executate secvențial, în orice ordine sau în paralel [5].

Funcționarea sistemelor de calcul, fie ele secvențiale sau paralele, are la bază organizarea de fire de așteptare pentru accesul la resursele partajate ale sistemului (procesor, memorie, dispozitive periferice). Astfel, pentru fiecare resursă, există un sistem de așteptare în care resursa respectivă reprezintă stația de servire, iar task-urile din sistem formează clienții [2].

În cazul procesorului, coada de așteptare poartă numele de fir de execuție. Firele de execuție fac trecerea de la programarea secvențială la programarea concurrentă. Un fir de execuție reprezintă, de fapt, o listă în care se găsesc acele procese care sunt gata de execuție.

Dacă această listă este menținută în memoria partajată a sistemului, accesul trebuie să se efectueze într-o secțiune critică pentru prevenirea situației în care două procesoare ar putea să aleagă același proces pe care să-l planifice în execuție simultan. Excluderea mutuală pentru accesul la resursele critice se asigură folosind mecanisme standard din sistemele uniprocessor (semafoare, bariere, mutex-uri, monitoare etc.). Astfel, atunci când un procesor intră în secțiunea critică, acesta are acces exclusiv la lista de procese gata de execuție. În funcție de disciplina de servire și de modul de implementare a priorităților în sistemul de așteptare, procesorul extrage un proces din listă după care părăsește secțiunea critică și începe să execute procesul.

Conform definiției date de Tanenbaum [7], un sistem distribuit reprezintă o colecție de calculatoare independente și interconectate între, ele care apar utilizatorului ca un singur calculator. Putem afirma că sistemele distribuite sunt alcătuite din mai multe calculatoare autonome, împreună cu componentele hardware și software de interconectare, care asigură mecanismele de comunicație și sincronizare. În categoria sistemelor distribuite, sunt incluse multiprocesoarele, multicalculatoarele conectate prin rețele de interconectare statice sau dinamice, precum și rețelele locale de stații. Se consideră că noțiunea de sistem distribuit este o noțiune mai cuprinzătoare, care o înglobează și pe aceea de sistem paralel.

Un sistem de distribuit este gestionat de un sistem de operare distribuit, care asigură utilizarea partajată a resurselor sistemului, distribuirea proceselor de calcul și comunicația între procese. Sistemele de operare distribuite reprezintă extensia pentru arhitecturi cu procesoare multiple a sistemelor de operare multitasking, în care se pot defini și executa, concurrent, mai multe procese.

În funcție de tipul sistemului de operare distribuit, care rulează pe sistemul respectiv, pot exista unul sau mai multe fire de execuție. Sistemele de operare cu multiprocesare, care operează asemănător cu un sistem UNIX multitasking, dar pe calculatoare cu procesoare multiple, sunt caracterizate prin existența unei singure cozi de execuție pentru întreg sistemul. Ori de câte ori un procesor devine liber, acesta extrage un proces din firul de așteptare pentru a-l executa. Următorul proces care va fi executat se stabilește în funcție de disciplina firului de așteptare și de prioritățile implementate în cadrul sistemului. Datorită faptului că timpul de execuție al programului paralel este finit, putem considera că numărul proceselor în care poate fi divizat algoritmul reprezintă, de asemenea, o mărime finită. Din acest motiv, unui astfel de sistem de calcul îi putem asocia un model de așteptare, în care stația de servire este compusă din m posturi care lucrează în paralel, unde m reprezintă numărul de procesoare din sistem, iar sosirile au loc dintr-o populație finită (vezi modelul matematic al unui sistem de așteptare, cu posturi de servire multiple și cu sosiri dintr-o populație finită).

La polul opus, se află sistemele de operare în rețea și sistemele de operare real distribuite, în cazul cărora există câte o copie a sistemului de operare, în fiecare element de procesare. Aceste sisteme de operare sunt caracterizate prin existența unui număr de fire de execuție egal cu numărul de procesoare din sistem, fiecare procesor având asociat propriul fir de execuție. La nivelul fiecărui element de procesare, există un sistem de operare care administrează firul de execuție asociat procesorului respectiv. Datorită faptului că timpul de execuție al programului paralel se presupune a fi finit, putem considera că numărul proceselor ce vor fi executate la nivelul întregului sistem este tot un număr finit. Din acest motiv, putem asocia fiecărui nod un model de așteptare, în care stația de servire este formată dintr-un singur post, iar sosirile se realizează dintr-o populație finită (vezi modelul matematic al unui sistem de așteptare cu un singur post de servire și cu sosiri dintr-o populație finită). Sistemele de așteptare din nodurile de procesare formează, la nivelul întregului sistem paralel, o rețea de așteptare, deoarece procesoarele comunică și se sincronizează între ele, iar execuția unui proces poate fi condiționată de rezultatele generate de execuția altor procese pe alte procesoare.

Sistemele de așteptare, despre care s-a discutat anterior, se pot transforma în sisteme ale căror intrări au loc dintr-o populație infinită, în cazul în care timpii de execuție ai programelor paralele devin foarte mari, iar numărul proceselor ce sunt executate în cadrul sistemului crește simțitor (vezi modelul matematic al unui sistem de așteptare cu un singur post de servire și cu sosiri dintr-o populație infinită, modelul matematic al unui sistem de așteptare cu posturi de servire multiple și cu sosiri dintr-o populație infinită).

În ceea ce privește execuția unui program paralel, pe procesoarele disponibile în cadrul unui sistem multiprocessor, sistemul de așteptare asociat depinde de legea de probabilitate, care descrie procesul de servire și de modul de partiționare a sarcinii totale de calcul. În cazul unui program de tip data-paralel, aceeași succesiune de instrucțiuni se va executa asupra unor seturi de date distincte, iar rezultatul final se va obține prin combinarea rezultatelor intermediare. În această situație, putem considera că durata serviciului este o variabilă aleatoare, repartizată exponențial negativ, iar pentru evaluarea performanțelor sistemului de așteptare și a algoritmului de planificare, cât și pentru îmbunătățirea acestora, se poate folosi un model de tip $M/M/m$. Dacă, însă, partiționarea se realizează în spațiul instrucțiunilor, legea de probabilitate ce guvernează timpii de execuție ai proceselor poate fi una exponențial negativă sau generică, iar modelul asociat sistemului va fi unul de tip $M/M/m$ sau $M/G/m$. Se poate ajunge la un model de tip $M/G/m$, care este foarte dificil de analizat, atunci când un proces poate beneficia de o cantă nelimitată de timp pentru a-și completa execuția.

Din dorința de a partaja echitabil resursele sistemului, este posibil să se impună anumite limite în ceea ce privește caracteristicile proceselor din firul de execuție [6]. Astfel, dacă se stabilește un prag maxim al timpului procesor de care poate beneficia un proces, toate procesele al căror timp de execuție estimat depășește valoarea prag vor fi eliminate din sistem, și nu vor ajunge în firul de execuție. De asemenea, procesele al căror timp de execuție devine mai mare decât limita maximă impusă vor fi întrerupte și eliminate din sistem. În acest fel, se garantează faptul că un proces nu va monopoliza, de unul singur, resursele sistemului.

5. Concluzii

Teoria așteptării se constituie într-un instrument foarte util pentru previzionarea performanțelor sistemelor de calcul, în general, și a celor paralele, în particular. Astfel, mărimile medii de interes (numărul mediu de consumatori din coadă și din sistem, timpul mediu de așteptare în fir și în sistem) ne pot oferi o imagine clară asupra performanțelor de ansamblu ale sistemului și a direcțiilor de îmbunătățire a acestora. Asociind un model de așteptare, unui sistem de calcul, putem scoate în evidență disfuncționalitățile existente, iar teoria așteptării ne va ajuta să eficientizăm funcționarea acestuia prin indicarea valorilor pe care trebuie să le atingă anumiți parametri ai sistemului pentru a se ajunge la un nivel satisfăcător al performanțelor. Sistemul de așteptare, format din procesele aflate în firul de execuție, așteptând să fie rulate de către procesor/procesoare, este gestionat de către un program special, numit planificator, care face alocarea task-urilor pe procesoarele disponibile. Planificatorul implementează, de fapt, disciplina de servire la nivelul sistemului de așteptare și gestionează clasele de priorități asociate proceselor. Felul în care se face servirea clienților afectează performanțele sistemului, în ansamblu. Din acest motiv, se dorește ca algoritmul de planificare folosit să conducă la o creștere a eficienței sistemului. Atunci când se dezvoltă o nouă versiune a acestui program, în primă fază, performanțele acestora sunt estimate folosind suportul oferit de teoria așteptării deoarece utilizarea unei versiuni instabile și netestată ar fi deranjantă pentru utilizatorii sistemului paralel.

Teoria așteptării ne arată că timpul petrecut de un consumator în sistem este format din timpul petrecut în coada de așteptare și din cel necesar servirii. În cazul în care consumatorul din sistemul de așteptare reprezintă un proces care este planificat pentru execuție, timpul petrecut în stația de servire este, de fapt, timpul acestuia de execuție. Pentru ca timpul petrecut de un proces în sistem să fie cât mai mic, nu este suficient să se reducă doar timpul de execuție prin folosirea unor procesoare cât mai rapide, ci este nevoie și de reducerea timpului de așteptare în fir prin folosirea unor algoritmi de planificare eficienți.

Bibliografie

1. ASMUSSEN, S.: Applied Probability and Queues, Springer, 2003.
2. DODESCU, GH., B. OANCEA, M. RACEANU: Procesare Paralelă, Editura Economică, București, 2002.
3. DODESCU, GH.: Sisteme de calcul și operare, Editura ASE, București, 1997.
4. GROSS, D., C. M. HARRIS: Fundamentals of Queuing Theory, Wiley, New York, 2003.
5. SABOT, G. W.: High Performance Computing, Addison-Wesley, 1995.
6. TANENBAUM, A. S.: Organizarea Structurată a Calculatoarelor, Computer Press Agora, 1999.
7. TANENBAUM, A. S.: Distributed Operating Systems, Prentice Hall, 1995.