

ՀՀ ԳԱԱ ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ ՊՐՈԲԼԵՄՆԵՐԻ
ԻՆՍՏԻՏՈՒՏ

Վարդանյան Արթուր Պապինի

ԲԱԶՄԱՊՐՈՑԵՍՈՐԱՅԻՆ ՀԱՄԱԿԱՐԳԵՐՈՒՄ ՀԵՐԹԵՐԻ
ՀԵՏԱԶՈՏՈՒՄ ՍՊԱՍՄԱՆ ԺԱՄԱՆԱԿԻ ՍԱՀՄԱՆԱՓԱԿՄԱՄԲ

Ե.13.05 - «Մաթեմատիկական մոդելավորում, թվային մեթոդներ և ծրագրերի
համալիրներ» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի
գիտական աստիճանի համար

ՍԵՂՄԱԳԻՐ

Երևան 2025

INSTITUTE FOR INFORMATICS AND AUTOMATION PROBLEMS OF THE NAS RA

Vardanyan Artur

STUDY OF QUEUES IN A MULTIPROCESSOR SYSTEMS
WITH A WAITING TIME RESTRICTION

ABSTRACT

of the dissertation for obtaining a Ph.D. degree in Technical Sciences on specialty
05.13.05 “Mathematical modelling, numerical methods and program complexes”

Yerevan 2025

Ատենախոսության թեման հաստատվել է ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում:

Գիտական ղեկավար՝ ֆիզ. մաթ. գիտ. թեկնածու Վ. Գ. Սահակյան
Պաշտոնական ընդդիմախոսներ՝ տեխ. գիտ. դոկտոր Գ. Տ. Կիրակոսյան
ֆիզ. մաթ. գիտ. թեկնածու Ա. Ն. Հարությունյան
Առաջատար կազմակերպություն՝ Երևանի Պետական Համալսարան

Ատենախոսության պաշտպանությունը տեղի կունենա 2025թ. փետրվարի 14-ին ժամը 14:00-ին ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում գործող 037 «Ինֆորմատիկա» մասնագիտական խորհրդի նիստում հետևյալ հասցեով՝ Երևան, 0014, Պ. Սևակի 1:

Ատենախոսությանը կարելի է ծանոթանալ ՀՀ ԳԱԱ ԻԱՊԻ գրադարանում:

Սեղմագիրն առաքված է 2025 թ.-ի հունվարի 13-ին:

Մասնագիտական խորհրդի գիտական

քարտուղար ֆիզ. մաթ. գիտ. դոկտոր՝



Մ. Ե. Հարությունյան

The topic of the dissertation was approved at the Institute of Informatics and Automation Problems of NAS RA.

Scientific supervisor: V. G. Sahakyan, Ph.D.
Official opponents: G. T. Kirakosyan, D.Tech.S.
A. N. Harutyunyan, Ph.D.
Leading organization: Yerevan State University

The Defense will take place on 14 February 2025; at 14:00, at the Specialized Council 037 “Informatics” at the Institute of Informatics and Automation Problems of NAS RA.
Address: Yerevan, 0014, P. Sevak 1.

The Dissertation is available in the library of IIAP NAS RA.

The abstract is delivered on 13 January, 2025.

Scientific Secretary of the Specialized Council, D.Ph.M.S.



M. E. Haroutunian

Աշխատանքի ընդհանուր նկարագիրը

Թեմայի արդիականությունը: Տեղեկատվական տեխնոլոգիաների ոլորտում ինտենսիվորեն կատարվում են հետազոտական աշխատանքներ, որոնցում անդրադառնում են բաշխված և զուգահեռ արագագործ հաշվողական համակարգերին: Ներկայումս այդպիսի համակարգերի հաճախակի օգտագործումը պայմանավորված է մի շարք ուղղությունների զարգացմամբ՝ մասնավորապես դրանք օգտագործվում են բազմագործակալական խելացի տեղեկության որոնման մեթոդներում, տվյալագիտությունում՝ մեծ տվյալների մշակման և պահպանման համար, իսկ գիտական ոլորտներում՝ բարդ հետազոտություններում խնդիրների մոդելավորման համար: Բացի այդ, գիտության և տեխնիկայի բնագավառի առաջադրանքների մոդելավորումները պահանջում են ավելի մեծ ճշգրտություն, արագ և լայնածավալ հաշվարկներ, որոնց առաջացրած բարդությունները հնարավոր է դառնում շրջանցել հենց նման հաշվողական համակարգերի օգտագործմամբ:

Բազմապրոցեսորային համակարգերի օգտագործմամբ սպասարկվող առաջադրանքները պահանջում են մեծածավալ հաշվարկներ, որոնք կատարվում են մեկից ավելի պրոցեսորների զուգահեռ օգտագործման միջոցով: Ինչի արդյունքում էլ առաջանում են որոշակի խնդիրներ, թե ինչպես կազմակերպել մի քանի առաջադրանքների միաժամանակ կատարումը հաշվողական համակարգի ռեսուրսների արդյունավետ օգտագործմամբ: Ինչն էլ էական նշանակության խնդիր է բազմապրոցեսորային համակարգերի պլանավորման (scheduling) ալգորիթմների՝ մշակման կոնտեքստում:

Արդյունավետ պլանավորման ալգորիթմները բազմապրոցեսորային համակարգերի համար ներկայացնում են համակարգչային գիտության և զուգահեռ հաշվարկների հետազոտության կարևորագույն ոլորտ: Տեխնոլոգիաների և բազմապրոցեսորային համակարգերի զարգացմանը զուգահեռ արդյունավետ պլանավորման ալգորիթմների թե՛ անհրաժեշտությունն է մեծանում, և թե՛ դրանց մշակումն է բարդանում²: Ավելին, դրանք պահանջում են ինչ որ իմաստով հարմարվողական պլանավորման ռազմավարություններ և համակարգի ռեսուրսների³ արդյունավետորեն բաշխում՝ տարբեր ցուցանիշների օպտիմալացման համար, ինչպիսիք են՝ համակարգի թողունակությունը, սպասարկման ուշացումը, էներգաարդյունավետությունը և այլն:

Այսպիսով, պլանավորման ալգորիթմների կատարելագործման գործընթացը

¹ Adaptive Computing Enterprises, “Maui Scheduler™ Administrator’s Guide”, 2014. Available online at: <http://docs.adaptivecomputing.com/maui/index.php>, Accessed on: July 27, 2024.

² Varela C. A., “Programming Distributed Computing Systems: A Foundational Approach”, The MIT Press, London, England, pp. 43-256, 2013, ISBN: 978-0-262-01898-2.

³ Guo W., Liu F., “Optimized Grid Resource Allocation and Task Scheduling Strategy”, Applications and Techniques in Cyber Intelligence: ATCI 2019, Springer International Publishing, pp. 142-149, 2020, DOI: https://doi.org/10.1007/978-3-030-25128-4_19.

բավականին դինամիկ է, այսինքն շարունակական անհրաժեշտություն կա, որ դրանց մշակման հիմքում ավելի ու ավելի շատ ցուցիչներ հաշվի առնվեն: Վերջին սերնդի մի շարք պլանավորման ալգորիթմների արդյունավետության բարձրացման նպատակով դրանց մշակման մեջ օգտագործում են օպտիմալացման տեսության և մեքենայական ուսուցման գործիքներ: Բայց նաև արժեքավոր պատկերացումներ են ստացվում, երբ այդ համակարգերը դիտարկվում են զանգվածային սպասարկման տեսության⁴ շրջանակներում:

Հաշվի առնելով վերոնշյալ դիտարկումները՝ անհրաժեշտ է բազմապրոցեսորային համակարգը մոդելավորել որպես զանգվածային սպասարկման մոդել, որում դրված է առաջադրանքների զուգահեռ սպասարկման պայման և հերթում առաջադրանքների սպասման ժամանակային սահմանափակում:

Աշխատանքի նպատակը և դիտարկված խնդիրները: Աշխատանքի հիմնական նպատակն է բազմապրոցեսորային համակարգի հետազոտումը որպես ընդլայնված զանգվածային սպասարկման մոդել՝ առաջադրանքների մուտքի հոսքի, դրանց սպասարման տևողության, սպասարկման համար պահանջվող հաշվողական ռեսուրսների քանակի և հերթում սպասման ժամանակի պատահական մեծություններով սահմանափակումների դեպքում: Այս նպատակին հասնելու համար դիտարկվել են հետևյալ խնդիրները.

- Հետազոտել, նկարագրել և մոդելավորել համակարգի վիճակները դիտարկելով առաջադրանքների սպասարկման համար պատահական թվով հաշվողական ռեսուրսների քանակի և հերթում սպասման ժամանակի սահմանափակումները:
- Համակարգի տրված պարամետրերի և բաշխումների համար հաշվել համակարգի վիճակների հավանականությունների արժեքները: Դրանց օգտագործմամբ կատարել համակարգի կատարողականության չափումներ և գտնել սպասարկման օպտիմալ տարբերակը: Արդյունքները ներկայացնել անալիտիկ կամ թվային հաշվարկների տեսքով:

Հետազոտության օբյեկտները: Այս աշխատանքում հետազոտման օբյեկտը բազմապրոցեսորային համակարգն է, նրա հիման վրա ընդլայնված զանգվածային սպասարկման մոդելը, դրա դինամիկ վիճակը, պայմանավորված համակարգում սպասարկվող և հերթում դրված առաջադրանքների քանակներով, ընդլայնված զանգվածային մոդելի պարամետրերի արժեքները և բաշխումները, առաջադրանքների հերթագրման և սպասարկման եղանակը:

Հետազոտության մեթոդներ: Աշխատանքի հետազոտության տեսական հիմքը հավանականության տեսության, զանգվածային սպասարկման և դիսկրետ մաթեմատիկայի մեթոդներն են: Կիրառական արդյունքները հենված են թվային հաշվարկների մեթոդների վրա՝ օգտագործելով Python ծրագրավորման լեզուն և իր

⁴ Shortle J. F., Thompson J. M., Gross D., Harris C. M., “Fundamentals of Queueing Theory”, John Wiley and Sons, New York, pp. 35-475, 2018, ISBN: 111894352X.

որոշ ընդլայնված գրադարաններ (NumPy, SciPy, math, matplotlib)⁵:

Աշխատանքի գիտական նորոյթը: Այս աշխատանքի համատեքստում ներկայացվում են հետևյալ գիտական արդյունքները.

- Բազմապրոցեստորային համակարգի համար մշակվել է զանգվածային սպասարկման մոդել: Դասական զանգվածային սպասարկման մոդելներն օգտագործում են մեկ սպասարկող սարք յուրաքանչյուր առաջադրանքի համար, իսկ այս աշխատանքում ընդլայնվում է հետազոտման ուղղությունը դեպի բազմապրոցեստորային համակարգեր, որոնցում յուրաքանչյուր առաջադրանք կարող է սպասարկվել մեկից ավելի սարքերի միաժամանակ օգտագործմամբ: Ընդլայնված սպասարկման մոդելում յուրաքանչյուր առաջադրանք իր սպասարկման համար պահանջում է պատահական քանակի պրոցեսորներ և ունի հերթում սպասման ժամանակի սահմանափակում:
- Բազմապրոցեստորային սպասարկման մոդելի միջոցով արտաձվել է համակարգի վիճակների հավանականությունների համար հավասարումներ: Ինչպես նաև մշակվել և իրականացվել է այդ հավասարումներով համակարգի լուծման ալգորիթ՝ հավանականությունների հաշվման համար: Սպասարկման մոդելի մասնավոր դեպքի համար արտաձվել են համակարգի վիճակների հավանականությունների հաշվման անալիտիկ բանաձևեր:
- Համակարգի տրված պարամետրերի և բաշխումների համար մշակվել և իրականացվել է սպասարկման օպտիմալ տարբերակը որոշող ալգորիթ:

Ստացված արդյունքների կիրառական նշանակությունը: Աշխատանքում ստացված արդյունքները կարող են օգտագործվել բազմապրոցեստորային հաշվողական համակարգերի պլանավորման փաթեթների մշակման մեջ՝ համակարգի աշխատանքի արդյունավետ կազմակերպման, ռեսուրսների օպտիմալ օգտագործման և ընդհանուր առմամբ համակարգի արտադրողականության բարձրացման համար:

Պաշտպանության ներկայացվող հիմնական դրույթները:

- Բազմապրոցեստորային համակարգի մոդելավորման եղանակ, որում առաջադրանքները սպասարկվում են զուգահեռ, յուրաքանչյուր առաջադրանք իր սպասարկման համար ունի պատահական քանակով պրոցեսորների պահանջ և իր հերթում սպասման ժամանական ունի պատահական մեծությամբ բնութագրվող սահմանափակում:
- Բազմապրոցեստորային զանգվածային սպասարկման համակարգի համար սպասարկվող և հերթում գտնվող առաջադրանքներով պայմանավորված համակարգի վիճակների հավանականությունների անալիտիկ

⁵ NumPy Contributors, “NumPy Documentation”, Available online at: <https://numpy.org/doc/stable/>, Accessed on: July 26, 2024.

նկարագրություն, տրված պարամետրերի և բաշխումների համար սպասարկման օպտիմալ տարբերակի որոշման ալգորիթմ և թվային հաշվարկներ:

Ստացված արդյունքների գրաքննությունը և փորձարկումը: Ստացված արդյունքները զեկուցվել են միջազգային CSIT գիտաժողովում.

1. 12th International Conference on Computer Science and Information Technologies (CSIT 2019), September 23 - 27, 2019, Yerevan, Armenia,
2. 13th International Conference on Computer Science and Information Technologies (CSIT 2021), Yerevan, Armenia, September 27 - October 1, 2021,
3. 14th International Conference on Computer Science and Information Technologies (CSIT 2023), Yerevan, Armenia, September 25-30, 2023.

Աշխատանքի արդյունքները քննարկվել են ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի, ինչպես նաև CSIT “Data Analytics and Mathematical Modeling” (Թբիլիսի, հունիսի 26-28, 2024 թ.) և “High Performance Artificial Intelligence” (Երևան, սեպտեմբերի 22-26, 2024 թ.) սեմինարների ընթացքում:

Հրատարակումներ: Ատենախոսության հիմնական արդյունքները հրատարակվել են ութ (8) գիտական աշխատություններում (4-ը WoS/Scopus-ում), որոնց ցանկը բերված է սեղմագրի վերջում:

Աշխատանքի ծավալը և կառուցվածքը: Ատենախոսության ծավալը կազմում է 104 էջ, ներառում է 82 գրականության հղումներ և բաղկացած է ներածությունից, 3 գլուխներից, եզրակացությունից, օգտագործված գրականության ցանկից:

Աշխատանքի բովանդակությունը

Ներածություն բաժնում հիմնավորվում է ատենախոսության արդիականությունը, ձևակերպված է աշխատանքի նպատակը, դիտարկված խնդիրները, գիտական նորոյթը, կիրառական նշանակությունը և պաշտպանության ներկայացված հիմնական դրույթները:

Առաջին գլխում ներկայացված է բազմապրոցետորային հաշվողական համակարգերը, նրանց աշխատանքի սկզբունքը, առաջադրանքների ընդունման, հերթագրման և կատարման կարգերը: Բազմապրոցետորային հաշվողական համակարգերի մոդելավորման եղանակները զանգվածային սպասարկման համակարգերի միջոցով⁶:

1.1 ենթագլխում ներկայացված են բազմապրոցետորային հաշվողական համակարգերի ներածությունը, դրանցում աշխատանքի կազմակերպման ձևերը և տեխնոլոգիաները:

1.2 ենթագլխում ներկայացված են բազմապրոցետորային հաշվողական

⁶ Bocharov P.P., D'Apice C., Pechinkin A.V., Salerno S., "Queueing Theory", VSP, Utrecht, Boston, pp. 34-98, 2004. ISBN: 90-6764-398-X.

համակարգերի հետազոտման հիմնական ուղղությունները: Հաշվի առնվելով դրանք, ընդգծվել է աշխատանքում դիտարկվող բազմապրոցեստորային համակարգը որպես զանգվածային սպասարկման մոդել հետազման կարևորություն, ինչի արդյունքում հնարավոր է դառնում ստանալ արժեքավոր պատկերացումներ համակարգի կատարողականության արդյունավետության, հաշվողական ռեսուրսների արդյունավետ օգտագործման, բեռնվածության բաշխման և աշխատանքի արդյունավետության չափանիշների գնահատման մասին:

1.3 ենթագլխում ներկայացված են բազմապրոցեստորային հաշվողական համակարգերում հերթերի ներկայացման և սպասարկման առանձնահատկությունները, դրանցում առաջադրանքների հերթագրման արդյունավետությանը նպաստող մի շարք տեխնոլոգիաները և հերթերի արդյունավետ սպասարկման մեթոդները՝ կախված սպասարկման տարբեր սխեմաների և քաղաքականությունների ընտրություններից:

1.4 ենթագլխում ներկայացված է զանգվածային սպասարկման համակարգերի ներածությունը: Ենթագլխում բերված են զանգվածային սպասարկման համակարգերի հետազոտության համապարփակ վերլուծություն, ըստ դիտարկված խնդիրների առանձնահատկությունների և սույն աշխատանքում դիտարկված բազմապրոցեստորային սպասարկող համակարգի հետազոտման կարևորություն ու առանձնահատկություն հանդիսացող դրույթները:

1.5 ենթագլխում ներկայացված են բազմապրոցեստորային հաշվողական համակարգերը զանգվածային սպասարկման⁷ համատեքստում, ընդգծելով դրանցում հերթերի սպասարկման ուսումնասիրությունն ուղղված զուգահեռ մշակման, ռեսուրսների բաշխման, ծանրաբեռնվածության հավասարակշռման և կատարողականության օպտիմալացման մարտահրավերների

Երկրորդ գլխում ներկայացված է առաջարկվող բազմապրոցեստորային սպասարկող համակարգի մոդելավորման⁸ եղանակը:

2.1 ենթագլխում ներկայացված է առաջարկվող բազմապրոցեստորային սպասարկող համակարգի նկարագրությունը: Ենթագլխում նկարագրված է բազմապրոցեստորային համակարգում առաջադրանքների մանրակրկիտ ներկայացման նշանակումները, որով հստակ ընդգծվում է թե՛ առաջադրանքների բազմապրոցեստորային սպասարկման, և թե՛ հերթում սպասման ժամանակի սահմանափակման առանձնահատկությունը:

Ենթադրվում է, որ հաշվողական համակարգ մուտք գործող առաջադրանքների

⁷ Anamika J., Jain M., Bhardwaj D., “Controllable multiprocessor queueing system”, Applications of Mathematical Modeling, Machine Learning, and Intelligent Computing for Industrial Development, pp. 61-76, 2023. ISBN: 9781003386599.

⁸ Ghimire S., Thapa G. B., Ghimire R. P., Silvestrov S., "A Survey on Queueing Systems with Mathematical Models and Applications", American Journal of Operational Research, vol. 7.1, pp. 1-14, 2017. DOI: <https://doi.org/10.5923/j.ajor.20170701.01>.

հոսքում, յուրաքանչյուր առաջադրանք բնութագրվում է $(v, \beta, \omega, \gamma)$ չորս պատահական պարամետրերով, որտեղ v -ն հաշվողական ռեսուրսների (պրոցեսորներ, միջուկներ, կլաստերային հանգույցներ և այլն) քանակն է, որոնք անհրաժեշտ են առաջադրանքը սպասարկելու համար, β -ն առաջադրանքը սպասարկելու համար պահանջվող առավելագույն ժամանակն է, ω -ն այն հնարավոր ժամանակն է, որ առաջադրանքը կարող է սպասել մինչև սպասարկման առաջադրվելը, որից հետո այն թողնում է համակարգն առանց սպասարկվելու, իսկ γ -ն առաջադրանքի համակարգ մուտք գործելու պահից այն ժամանակային միջակայքն է, որից հետո թույլատրվում է սկսել առաջադրանքի սպասարկումը: Երբ $\gamma = 0$, այս պարամետրի արժեքը կարող է բաց թողնվել, իսկ սպասման ժամանակի սահմանափակում չունեցող համակարգերում ω պարամետրի արժեքը նույնպես բաց է թողնվում:

Համակարգի պարամետրերը նկարագրվում են.

m - համակարգի հաշվողական ռեսուրսների քանակը ($m \geq 1$),

n - առաջադրանքների առավելագույն թույլատրելի քանակը հերթում ($n \geq 1$),

α - հարևան մուտքերի միջև ժամանակային միջակայքի պատահական արժեք, որն ունի հետևյալ հավանականության բաշխումը.

$$P(\alpha < t) = A(t),$$

β - առաջադրանքի սպասարկման ժամանակի պատահական արժեք, որն ունի հետևյալ հավանականության բաշխումը.

$$P(\beta < t) = B(t),$$

ω - հերթում առաջադրանքի թույլատրելի սպասման ժամանակի պատահական արժեք, որն ունի հետևյալ հավանականության բաշխումը.

$$P(\omega < t) = W(t),$$

v - առաջադրանքը սպասարկելու համար անհրաժեշտ հաշվողական ռեսուրսների քանակի պատահական արժեք, որն ունի հետևյալ հավանականության բաշխումը.

$$P(v \leq k) = V(t), \quad k = 1, 2, \dots, m:$$

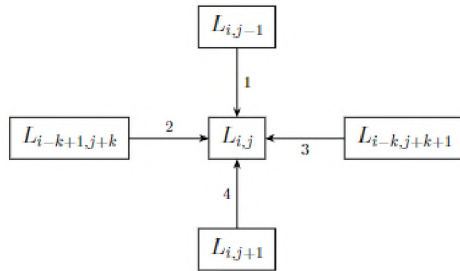
Ենթազլխում նաև մանրակրկիտ քննարկված է առաջադրանքների սպասարկման մերժում ստանալու սցենարները և դրանց սպասարկման կարգապահության ընտրությունը:

Առաջադրանքները սպասարկվելու են համակարգ մուտք գործելու հերթականությամբ, այսինքն՝ օգտագործվում է FIFO կարգապահությունը: Այն առաջադրանքները, որոնք համակարգը ստանում է հերթի լրիվ զբաղեցված լինելու դեպքում (այսինքն հերթում արդեն n հատ առաջադրանքներ սպասում են սպասարկման առաջադրվելու) ստանում են համակարգի մուտքի և ծառայության մերժում:

2.2 Ենթագլխում ներկայացված է առաջարկվող բազմապրոցեսորային սպասարկող համակարգի մոդելավորման հիմնական նշանակումները: Ենթագլխում նկարագրվում է համակարգի հնարավոր վիճակների վերջավոր քանակի պատճառով համակարգի աշխատանքի կայուն ռեժիմի⁹ անցնելու փաստը:

Դասական զանգվածային սպասարկման մոդելներում դիտարկվում է սպասարկող համակարգի վիճակ հասկացությունը, որը բնութագրում է այդ պահին համակարգի աշխատանքային պրոցեսը: Համակարգի վիճակը որոշվում է այդ պահին համակարգում սպասարկվող և հերթում դրված առաջադրանքների թվով: Քանի որ դասական զանգվածային սպասարկման մոդելներում յուրաքանչյուր առաջադրանք օգտագործում է մեկ սպասարկող սարք, ապա այդ մոդելներում հիմնականում համակարգի վիճակը որոշվում է հերթագրված առաջադրանքների թվով:

Այս աշխատանքում, դիտարկվող ընդլայնված սպասարկման մոդելի դեպքում համակարգի վիճակը որոշվում է այդ պահին համակարգում սպասարկվող և հերթագրված առաջադրանքների թվով: Համակարգի վիճակը, երբ i առաջադրանքները սպասարկվում են, իսկ j առաջադրանքները սպասում են հերթում նշանակվում է $L_{i,j}$ -ով: Բազմապրոցեսորային հաշվողական համակարգի պրոցեսորների և հերթում առաջադրանքների վերջավոր թվերով պայմանավորված համակարգի հնարավոր վիճակների թիվը սահմանափակ է և դեպքերը, երբ համակարգը կարող է անցնել $L_{i,j}$ վիճակի այլ վիճակից, ներկայացված է հետևյալ սխեմայում.



Սխեմայում պատկերված անցումները համապատասխանաբար նկարագրվում են հետևյալ կերպ.

1. համակարգի վիճակը $L_{i,j-1}$ էր, և համակարգը մեկ առաջադրանք ստացավ, որն էլ միացավ հերթին;
2. համակարգի վիճակը $L_{i-k+1,j+k}$ էր, որտեղ $k = 1, 2, \dots, \min(i, n - j)$ և մեկ առաջադրանք ավարտեց սպասարկումը և հեռացավ համակարգից, որի արդյունքում հերթի առաջին k առաջադրանքները ընդունվեցին

⁹ Kleinrock L., “Queueing Systems: Vol. I - Theory”, John Wiley and Sons, New York, 1975, ISBN: 0-471-49110-1.

սպասարկման;

3. համակարգի վիճակը եղել է $L_{i-k, j+k+1}$, որտեղ $k = 1, 2, \dots, \min(i-1, n-j-1)$ և հերթի առաջին առաջադրանքը լքեց հերթը (նրա սպասման ժամանակը սպառվեց) և հերթից առաջին k առաջադրանքները ընդունվեցին սպասարկման.
4. համակարգի վիճակը $L_{i, j+1}$ էր, և հերթից մեկ առաջադրանք, ոչ առաջին առաջադրանքը լքեց հերթը (նրա սպասելու ժամանակը սպառվեց):

2.3 ենթազբյուրում ներկայացված են այն սահմանափակումները որոնք դրվում են համակարգ առաջադրանքների ստացման, դրանց սպասարկման և հերթում ձախողման պատահական մեծությունների բաշխումների վրա բազմապրոցետորային սպասարկող համակարգի կայուն վիճակի համար հավասարումների ստացումը կատարելու համար, որն արվում է նախորդիվ համակարգի վիճակների անցումները նկարագրող սխեմայով բոլոր հնարավոր անցումները դիտարկելու միջոցով:

Բազմապրոցետորային սպասարկող համակարգի կայուն վիճակի համար հավասարումների ստացումը կատարվում է, երբ որպես $A(t)$, $B(t)$, $W(t)$, և $V(k)$ վերցվում են հետևյալ բաշխման ֆունկցիաները.

$$A(t) = 1 - e^{-at},$$

$$B(t) = 1 - e^{-bt},$$

$$W(t) = 1 - e^{-wt},$$

$$V(k) = \frac{k}{m}, \quad k = 1, 2, \dots, m,$$

որտեղ a -ն ($a > 0$) մուտքային հոսքի ինտենսիվությունն է, b -ն ($b > 0$) առաջադրանքների սպասարկման ինտենսիվությունն է, իսկ w -ն ($w > 0$)՝ հերթից առաջադրանքի սպասարկման ձախողման ինտենսիվությունը:

Նախորդ ենթազբյուրում ներկայացված սխեմայում դեպքերը, երբ համակարգը կարող է անցնել $L_{i, j}$ վիճակի կախված i , j և k փոփոխականների ընդունած արժեքներից, կարող են պակասել: Այդ փաստը հաշվի առնելով, համակարգի $L_{i, j}$ վիճակում լինելու հավանականությունը գտնելու համար նախ դիտարկվում են այդ փոփոխականների ընդունած եզրային արժեքներով որոշվող դեպքերը, իսկ սխեմայի համապատասխան անցումներն արտացոլում են $\delta_i^{(1)}$, $\delta_{i, j, k}^{(2)}$ և $\delta_{i, j, k}^{(3)}$ հավանականությունները:

Ակնհայտ է, որ երբ $i = 0$ ու $j = 0$, ապա

$$P_{0,0} = \frac{b}{a} P_{1,0}, \tag{1}$$

իսկ երբ $i = 0$ ու $1 \leq j \leq n$, ապա

$$P_{0,j} = 0: \quad (2)$$

Այն դեպքում, երբ $1 \leq i \leq m$ և $j = 0$, ապա

$$P_{i,0} = \frac{1}{a+ib} \left[a\delta_{i-1}^{(1)} + b \sum_{k=k_0}^{k_1} \left((i-k+1)\delta_{i,0,k}^{(2)} P_{i-k+1,k} \right) + w \sum_{k=0}^{k_1} \left(\delta_{i,0,k}^{(3)} P_{i-k,k+1} \right) \right], \quad (3)$$

որտեղ

$$k_0 = \begin{cases} 0, & \text{երթև } 1 \leq i < m \\ 1, & \text{երթև } i = m \end{cases}$$

$$k_1 = \min(i, n),$$

$$k_2 = \min(i-1, n-1):$$

Պարզ է, որ այս դեպքում $P_{i,0}$ -ն որոշող հավասարումներում միայն այն դեպքն է հանդես գալիս, երբ հերթը լքում է հերթի առաջին առաջադրանքը:

Այն դեպքում, երբ $1 \leq i \leq m$ և $1 \leq j \leq n$, ապա

$$P_{i,j} = \frac{1}{a+ib+jw} \left[a(1-\delta_{i-1}^{(1)})P_{i,j-1} + b \sum_{k=k_0}^{k_1} \left((i-k+1)\delta_{i,j,k}^{(2)} P_{i-k+1,k+j} \right) + w \sum_{k=0}^{k_1} \left(\delta_{i,j,k}^{(3)} P_{i-k,k+j+1} \right) + wjP_{i,j+1} \right], \quad (4)$$

որտեղ k_0 -ն որոշվում է ինչպես նախորդ դեպքում, իսկ

$$k_1 = \min(i, n-j),$$

$$k_2 = \min(i-1, n-j-1):$$

Պետք է նկատել, որ այս դեպքում $P_{i,j}$ -ն որոշող հավասարումների նախավերջին գումարելիում ներառված է այն դեպքը, երբ հերթը լքում է հերթի առաջին առաջադրանքը, իսկ վերջին գումարելիում ներառված է այն դեպքը, երբ հերթը լքում է հերթի ոչ առաջին առաջադրանքը: Մնաց դիտարկելու վերջին երկու եզրային դեպքերը՝ նախ երբ $1 \leq i < m$ և $j = n$, ապա

$$P_{i,n} = \frac{1}{a+ib+nw} \left[a(1-\delta_{i-1}^{(1)})P_{i,n-1} + b(i+1)\delta_{i,n,0}^{(2)}P_{i+1,n} \right], \quad (5)$$

և վերջին եզրային դեպքը, երբ $i = m$ և $j = n$, ապա

$$P_{m,n} = \frac{a(1 - \delta_{m-1}^{(1)})P_{m,n-1}}{a + mb + nw} \quad (6)$$

Արդյունքում նկարագրած բազմապրոցետորային սպասարկող համակարգի համար (1)-ից (6) հավասարումներով $P_{i,j}$ ($0 \leq i \leq m$ և $0 \leq j \leq n$) անհայտ փոփոխականներով, ստացվում է գծային հավասարումների համակարգ, որի լուծումով կորոշվեն սպասարկող համակարգի յուրաքանչյուր $L_{i,j}$ ($0 \leq i \leq m$ և $0 \leq j \leq n$) վիճակում համակարգի գտնվելու հավանականությունները:

Ավելին, քանի որ $P_{i,j}$ ($0 \leq i \leq m$ և $0 \leq j \leq n$) անհայտ հավանականությունները միևնույն պատահույթի բոլոր հնարավոր ելքերին համապատասխանող հավանականություններն են, ապա դրանք բավարարում են նաև հետևյալ նորմալության պայմանին.

$$\sum_{i=0}^m \sum_{j=0}^n P_{i,j} = 1: \quad (7)$$

2.4 ենթազիլում ներկայացված են բազմապրոցետորային սպասարկող համակարգի մոդելավորման համար որոշ հավանականությունների նկարագրությունները և որոշ պատահույթների սահմանումներ:

2.5 ենթազիլում ներկայացված են մի քանի օժանդակ հավանականությունների և նախորդիվ սահմանված որոշ հավանականությունների արժեքների հաշվարկման բանաձևերի արտածումները:

2.6 ենթազիլում ներկայացված է բազմապրոցետորային սպասարկող համակարգի աշխատանքի կայուն ռեժիմի անցնելու փաստով պայմանավորված համակարգի կայուն վիճակի հավանականությունների ստացումը, որը ձևակերպվում է հետևյալ թեորեմի տեսքով՝

Թեորեմ. Պարամետրերի տրված արժեքների դեպքում, (1)-ից (6) հավասարումներով գծային հավասարումների համակարգի լուծումը որոշում է սպասարկման համակարգի $P_{i,j}$ կայուն վիճակի հավանականությունները ($0 \leq i \leq m$ և $0 \leq j \leq n$):

2.7 ենթազիլում դիտարկվել է աշխատանքում նկարագրված բազմապրոցետորային սպասարկման համակարգի մոդելը մասնավոր դեպքում՝ մասնավորապես երբ համակարգում հերթ չկա: Ինչի արդյունքում համակարգը նկարագրող որոշ պարամետրեր բաց են թողնվում և համակարգ ժամանելուն պես առաջադրանքները ենթակա են կամ սպասարկման համար ընդունման, կամ մերժման: Առաջադրանքները ստանում են ծառայության մերժում, եթե համակարգ մուտք գործելիս ակնհայտ է դառնում, որ իրենց համապատասխան պարամետրերով սպասարկումը չի կարող բավարարվել: Դա տեղի է ունենում, երբ համակարգը չունի

անհրաժեշտ ազատ պրոցեսորներ առաջադրանքի սպասարկումը սկսելու համար:

Այս դեպքում համակարգի կայուն վիճակի հավանականությունների ստացման համար կատարվել է անալիտիկ արտածում, ինչի արդյունքում էլ ստացվել է համակարգի կայուն վիճակի հավանականությունների գնահատելու համար բանաձև.

$$P_k = \frac{\alpha^k}{k! m^k} \binom{m}{k} P_0, \quad k \geq 1$$

որտեղ P_k նշանակումը ներկայացնում է հավանականությունը, այն բանի, որ k հատ առաջադրանքներ սպասարկվում են համակարգում, α -ն սահմանվում է որպես առաջադրանքների մուտքային հոսքի և սպասարկման ինտենսիվությունների հարաբերություն, իսկ m -ը համակարգի պրոցեսորների թիվն է:

Համակարգի հզորությունը վերջավոր է՝ դիցուք առաջադրանքների ընդհանուր թիվը, որոնք կարող են սպասարկվել n է, ապա նորմալության պայմանի օգտագործմամբ P_0 -ն կարող է որոշվել հետևյալ կերպ.

$$P_0 = \frac{1}{1 + \sum_{k=1}^n \frac{\alpha^k}{k! m^k} \binom{m}{k}}:$$

2.8 ենթագլխում ներկայացված է աշխատանքում դիտարկված բազմապրոցեսորային համակարգում առավել օպտիմալ սպասարկման մոդել՝ դիտարկվում է համակարգում առաջադրանքների սպասարկման ընդհատումները՝ անցակետեր(checkpoints) ստեղծելու մեխանիզմի կիրառմամբ ընդհատված առաջադրանքները հերթում պահելով: Նկարագրվում է առաջադրանքների սպասարկման ընտրության և հաջորդ ընդհատման ժամանակը որոշելու ալգորիթմ:

Առաջարկվող ալգորիթմը կարող է ծառայել որպես անհրաժեշտ պայման առաջադրանքների սպասարկման համար առանց ընդհատման, այսինքն եթե հերթը չի կարող սպասարկվել ընդհատմամբ, ապա այն չի կարող սպասարկվել նաև առանց ընդհատման: Առաջարկվող ալգորիթմը նաև կատարում է հերթի վիճակի նախնական գնահատում և տալիս է առաջարկներ սպասարկման հնարավոր հերթականության մասին:

Երրորդ գլխում ներկայացված է աշխատանքում կատարված հետազոտությունների հիման վրա կիրառական արդյունքներ՝ նկարագրված է բազմապրոցեսորային սպասարկող համակարգի կայուն վիճակի հավանականությունների ստացման համար թվային ալգորիթմի մշակման, համակարգի աշխատանքի ընթացքում ժամանակի որոշակի պահին առաջադրանքի մուտքի մերժման և առաջադրանքի ձախողման հավանականությունների գնահատման ընթացքները:

3.1 ենթագլխում ներկայացված է բազմապրոցեսորային սպասարկող համակարգի կայուն վիճակի հավանականությունների ստացման համար (1)-ից (6) հավասարումներով հավասարումների համակարգի լուծման գործընթացը:

Վերլուծելով հավասարումների համակարգի կառուցվածքը և հատկությունները նկատվում է, որ հավասարումների համակարգը գծային է, քանի որ անհայտ փոփոխական հանդիսացող հավանականությունները հավասարումներում ներկայացված են գծային կոմբինացիաներով: Երկրորդը, որ կարևոր է նկատել, այն է, որ (1) -ից (6) հավասարումներով ներկայացված հավասարումների համակարգը համասեռ է: Հավասարումների համակարգի համասեռությունը նկատելով՝ լուծմանը կարելի է մոտենալ՝ փոխելով համակարգի հավասարումներից մեկը (7) պայմանով, որի արդյունքում կստացվի գծային հավասարումների ոչ համասեռ համակարգ: Իսկ հավասարումների համակարգի գծայինությունը թույլ է տալիս լուծման համար կիրառել տարբեր մաթեմատիկական մեթոդներ գծային համակարգերի լուծման համար:

Համակարգերի լուծման համար կիրառվում է գծային հավասարումների համակարգերի լուծման ամենաընդհանուր մոտեցում՝ օգտագործելով դրա մատրիցային ներկայացումը հավասարումների համակարգը լուծելու համար: Այսպիսով նախ կառուցվում է A մատրիցը, որը ներկայացնում է $P_{i,j}$ ($0 \leq i \leq m$ և $0 \leq j \leq n$) անհայտ հավանականությունների գծային կոմբինացիաներում դրանց գործակիցների մատրիցը, իսկ b վեկտորը, որը ներկայացնում է այդ գծային կոմբինացիաների արժեքների վեկտորը: A մատրիցը $(m + 1)(n + 1) \times (m + 1)(n + 1)$ չափի քառակուսի մատրից է, իսկ b վեկտորը $(m + 1)(n + 1) \times 1$ չափի վեկտոր սյուն է:

Վերանայելով կատարված նշանակումները հավասարումների համակարգը կգրվի հետևյալ կերպ.

$$Ax = b,$$

որտեղ x -ը վեկտոր սյուն է, որը ներկայացնում է $P_{i,j}$ ($0 \leq i \leq m$ և $0 \leq j \leq n$) անհայտ հավանականությունները: Հավասարումների համակարգի այսպիսի ներկայացումով ավելի պարզ է դառնում համակարգի կառուցվածքը և լուծման գործընթացը: Հավասարումների համակարգը լուծելու համար մշակվել է ալգորիթմ, որն իրականացվել է Python ծրագրավորման լեզվով՝ օգտագործելով NumPy գրադարանի որոշ գործիքներ: Մշակված ալգորիթմն ապահովում է թվային եղանակ բազմապարոցետորային սպասարկման համակարգի կայուն վիճակի հավանականությունները հաշվարկելու համար:

3.2 ենթազգլխում ներկայացված են բազմապարոցետորային սպասարկման համակարգի կատարողականության որոշ չափանիշների սահմանումներ և վերլուծություն, ժամանակի որոշակի պահին առաջադրանքի մուտքի մերժման և առաջադրանքի ձախողման (եթե առաջադրանքը թողնում է հերթը իր սպասման ժամանակի ավարտի պատճառով) և առաջադրանքի հերթագրման հավանականությունների գնահատումը:

Առաջադրանքի մուտքի ժամանակ, համակարգը մերժում է դրա մուտքը, եթե հերթում սպասող առաջադրանքների թիվը հասել է իր առավելագույն քանակին՝ n -

ին: Այլ կերպ ասած, առաջադրանքի մուտքի մերժման հավանականությունը համակարգի աշխատանքի ընթացքում կարելի է հաշվարկել՝ հաշվի առնելով համակարգի այն վիճակների հավանականությունները, որտեղ հերթը լիովին զբաղված է: Մուտքի մերժման հավանականությունը, որը նշանակվում է P_r -ով, արտահայտվում է հետևյալ կերպ.

$$P_r = \sum_{i=0}^m \frac{aP_{i,n}}{a + ib + nw}:$$

Հերթում առաջադրանքի ձախողման հավանականությունը, որը նշանակվում է P_f -ով, կարելի է հաշվել համակարգի վիճակների հավանականությունների միջոցով հետևյալ բանաձևով.

$$P_f = \sum_{i=0}^m \sum_{j=1}^n \frac{jwP_{i,j}}{a + ib + jw},$$

որտեղ, ինպես վերևում, a -ն ներկայացնում է առաջադրանքների մուտքի ինտենսիվությունը, b -ն՝ սպասարկման ինտենսիվությունը, w -ն՝ հերթում սպասման ժամանակի սպառման պատճառով առաջադրանքների ձախողման ինտենսիվությունը, իսկ $P_{i,j}$ ($0 \leq i \leq m$ և $0 \leq j \leq n$) վիճակների հավանականություններն են, երբ համակարգում սպասարկվում են i առաջադրանքներ և j առաջադրանքներ սպասում են հերթում:

Այստեղ սահմանվում է նաև համակարգի օգտագործման գործակիցը, որը հանդիսանում է սպասարկման հզորության պահանջի և համակարգի ընդհանուր սպասարկման հզորության հարաբերությունը: Համակարգի օգտագործման գործակիցը նշանակվում է ρ -ով և արտահայտվում է հետևյալ կերպ.

$$\rho = \frac{a(m+1)(1-P_r)(1-P_f)}{2mb}:$$

3.3 ենթազվիտում ներկայացված է մոդելավորվող համակարգի տրված պարամետրերի և բաշխումների համար սպասարկման օպտիմալ տարբերակի որոշման ընթացքը: Համակարգի պարամետրերի տրված բաշխումների համար համակարգի կողմից առաջադրանքի մուտքի մերժման P_r և հերթագրված առաջադրանքի ձախողման P_f հավանականությունների միջոցով հնարավոր է դառնում գնահատել P_i հավանականությունը, որը հանդիսանում է համակարգի կողմից չսպասարկվող առաջադրանքների քանակի հիմնական ցուցիչը.

$$P_i = P_r + (1 - P_r)P_f:$$

Ակնհայտ է, որ երբ $n = 0$, ապա P_r մերժման հավանականությունը ինչ-որ արժեք ունի, իսկ P_f -ը զրո է, և երբ համակարգում հերթ կա ու հերթի երկարությունը մեծացվի, ապա P_r մերժման հավանականության արժեքը կծգտի զրոյի, իսկ P_f հավանականության արժեքը՝ մեկի, այսպիսով.

$$\begin{aligned} n = 0 &\Rightarrow 0 < P_r < 1, & P_f = 0 \\ n \rightarrow \infty &\Rightarrow P_r \rightarrow 0, & P_f \rightarrow 1 \end{aligned}$$

Հաշվի առնելով այս փաստը՝ խնդիր է դրվել P_i հավանականության մինիմիզացիան կատարել ըստ հերթի երկարության փոփոխման, գտնել հերթի երկարության այն արժեքը, որի դեպքում P_i հավանականությունը ընդունում է իր փոքրագույն արժեքը համակարգի տրված պարամետրերի դեպքում և միննույն ժամանակ համակարգի օգտագործման գործակիցը փոքր է լինում մեկից ($\rho < 1$):

Այս դեպքում նույնպես մշակվել է այգորիթ, որը հնարավորություն է տալիս տրված քանակի պրոցեսորներով բազմապրոցեսորային սպասարկման համակարգի՝ առաջադրանքների մուտքային հոսքի, սպասարկման և հերթագրված առաջադրանքի ձախողման տրված ինտենսիվությունների համար պարզել հերթի երկարության օպտիմալ արժեքը: Ինչի արդյունքում տրված պարամետրերով և այդ երկարությամբ հերթով համակարգում կլինի սպասարկման օպտիմալ տարբերակ:

3.4 Ենթազվիսում ներկայացված են փորձարարական հաշվարկներ, որոնք կատարվում են դիտարկված բազմապրոցեսորային սպասարկման համակարգի պարամետրեր, ինչպիսիք են՝ մուտքային հոսքի, սպասարկման և հերթում գտնվող առաջադրանքների սպասարկման խափանումների ինտենսիվությունների համար ընտրված որոշակի արժեքների դեպքում:

Հաշվարկների համար ենթադրվել է, որ հաշվողական համակարգը ունի 256 ($m = 256$) հաշվողական պրոցեսորներ և 10 ($n = 10$) առաջադրանքների առավելագույն տեղերով հերթի երկարություն: Առաջադրանքների մուտքային հոսքի ինտենսիվության, ծառայության ինտենսիվության և հերթում առաջադրանքի ծառայության ձախողման ինտենսիվության համար վերցված են այս արժեքները՝ $a = 240$, $b = 190$ և $w = 2$:

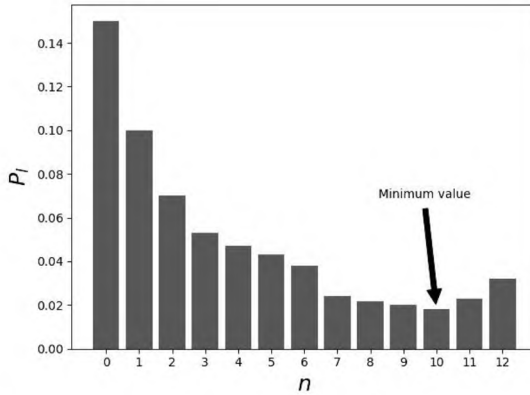
Օգտագործելով հաշվումների համար համակարգի ընտրված պարամետրերը, հաշվարկվել են առաջադրանքի մերժման և հերթագրված առաջադրանքի ձախողման հավանականությունները, համակարգի օգտագործման գործակիցը.

$$P_r = 0.00763,$$

$$P_f = 0.0137,$$

$$\rho = 0.621:$$

Հաշվարկների համար համակարգի ընտրված պարամետրերով սպասարկման պարամետրերի օպտիմալ ընտրություն կատարելով ծրագրային կոդով հաշվարկների արդյունքում ստացվել է, որ համակարգի այդպիսի մուտքային պարամետրերի դեպքում սպասարկման համակարգի հերթի երկարության օպտիմալ արժեքը 10 է: Այդ հաշվարկների արդյունքների հիման վրա կառուցվել է հետևյալ դիագրամը.



Նկար 1: Հերթի երկարության օպտիմալ արժեքի գնահատման ընթացքը

Հաշվարկներն արվել են օգտագործելով աշխատանքում ստացված արդյունքների հիման վրա մշակված և իրականացված ծրագրային փաթեթը¹⁰:

Աշխատանքի հիմնական արդյունքները.

Այս աշխատանքի շրջանակներում բազմապրոցետորային համակարգը մոդելավորվել է կիրառելով զանգվածային սպասարկման մոդելների տեսությունը, ի տարբերություն դասական զանգվածային սպասարկման մոդելների, որոնք օգտագործում են մեկ սպասարկող սարք յուրաքանչյուր առաջադրանքի համար, այստեղ ընդլայնվել է հետազոտման ուղղությունը դեպի բազմապրոցետորային համակարգեր, որոնցում յուրաքանչյուր առաջադրանք կարող է սպասարկվել մեկից ավելի սարքերի միաժամանակ օգտագործմամբ:

- Առաջարկվում է բազմապրոցետորային համակարգերի ուսումնասիրության մոտեցում, որպես ընդլայնված զանգվածային սպասարկման մոդել, որտեղ ներառվել են առաջադրանքների իրենց սպասարկման համար անհրաժեշտ պրոցեսորների քանակի և հերթում դրանց սպասման ժամանակների պատահական մեծություններով որոշվող սահմանափակումներ [1, 2, 3, 4]:
- Ընդլայնված զանգվածային սպասարկման մոդելի միջոցով նկարագրվել է բազմապրոցետորային համակարգի վիճակները: Վիճակների հավանականությունների ստացման համար արտաձվել է հավասարումների համակարգ, որի լուծման համար մշակվել և իրականացվել է ալգորիթմ: Մասնավոր դեպքում համակարգի վիճակների հավանականությունների ստացման համար անալիտիկ բանաձևեր են արտաձվել [5, 6, 7]:

¹⁰ Computational codes for the multiprocessor queueing system model, Available online at: <https://github.com/a17ur/MQSM>.

- Սպասարկման համակարգի տրված պարամետրերի և բաշխումների համար սպասարկման պարամետրերի օպտիմալ ընտրություն կատարելու համար արտածվել է համակարգի կողմից առաջադրանքների չսպասարկման որոշվող հավանականության բանաձև: Այդ հավանականության արժեքի նվազեցման համար մշակվել և իրականացվել է օպտիմիզացիայի ալգորիթմ[8]:

Հրապարակված աշխատանքների ցանկ

1. Vardanyan A., “Improved Model of Scheduling Algorithm”, *Mathematical Problems of Computer Science*, Volume 50, pp. 104–106, 2018. DOI: 10.51408/1963-0027.
2. Sahakyan V., Vardanyan A., “The Queue Distribution in Multiprocessor Systems with the Waiting Time Restriction”, *Mathematical Problems of Computer Science*, Volume 51, pp. 82–89, 2019. DOI: 10.51408/1963-0035.
3. Sahakyan V., Vardanyan A., “The Queue State for Multiprocessor System with Waiting Time Restriction”, *IEEE Xplore, Computer Science and Information Technologies (CSIT 2019)*, Yerevan, Armenia, pp. 116-119, 2019. DOI: 10.1109/CSITechnol.2019.8895093.
4. Sahakyan V., Vardanyan A., “The Steady State Distribution for M|M|m|n Model with the Waiting Time Restriction”, *Mathematical Problems of Computer Science*, Volume 54, pp. 34–40, 2020. DOI: 10.51408/1963-0057.
5. Sahakyan V., Vardanyan A., “About the Possibility of Executing Tasks with a Waiting Time Restriction in a Multiprocessor System”, *AIP Conference Proceeding, Computer Science and Information Technologies (CSIT 2021)*, 2757 (1), 030003, 2023. DOI: 10.1063/5.0135784.
6. Sahakyan V., Vardanyan A., “A Computational Approach for Evaluating Steady-State Probabilities and Virtual Waiting Time of a Multiprocessor Queuing System”, *Programming and Computer Software*, Volume 49, pp. S16–S23, 2023. DOI: 10.1134/S0361768823090098.
7. Sahakyan V., Tadevosyan R. and Vardanyan A., “Model of a System for Servicing Multiprocessor Tasks without a Queue”, *Programming and Computer Software*, Volume 50, Suppl. 1, pp. S47–S50, 2024. DOI: 10.1134/S0361768824700403.
8. Vardanyan A., “Advanced Queuing Model of a Multiprocessor Computing System”, *Mathematical Problems of Computer Science*, Volume 62, pp. 43-51, 2024. DOI: 10.51408/1963-0119.

ИССЛЕДОВАНИЕ ОЧЕРЕДЕЙ В МНОГОПРОЦЕССОРНЫХ СИСТЕМАХ С ОГРАНИЧЕНИЕМ НА ВРЕМЯ ОЖИДАНИЯ

Абстракт

В области информационных технологий интенсивно ведутся исследовательские работы, в которых идет речь о распределенных и параллельных высокоскоростных вычислительных системах. В настоящее время использование таких систем обусловлено развитием ряда направлений, в частности, они используются в многоагентных интеллектуальных методах поиска информации, в науке о данных для обработки и хранения больших объемов данных, а также в научных областях для решения сложных исследовательских задач. Кроме того, задачи моделирования науки и техники требуют большей точности, быстрых и масштабных расчетов, сложности которых можно обойти за счет использования таких вычислительных систем.

В современных многопроцессорных системах приложения становятся все более разнообразными и динамичными. Задачи, решаемые с помощью многопроцессорных систем, требуют крупномасштабных вычислений, выполняемых с использованием более чем одного процессора. В результате возникают определенные проблемы, как организовать одновременное выполнение нескольких задач с эффективным использованием ресурсов многопроцессорной вычислительной системы. Это важный вопрос в контексте разработки алгоритмов планирования для таких систем.

Эффективные алгоритмы планирования для многопроцессорных систем представляют собой важную область исследований в области информатики и параллельных вычислений. Они требуют определенного понимания стратегий адаптивного планирования и эффективного распределения системных ресурсов для оптимизации различных параметров, таких как пропускная способность системы, задержка обслуживания, энергоэффективность и т. д.

Таким образом, процесс совершенствования алгоритмов планирования достаточно динамичен, то есть существует постоянная необходимость учитывать все больше показателей в основе их разработки. С целью повышения эффективности ряда алгоритмов планирования последнего поколения при их разработке используются инструменты теории оптимизации и машинного обучения. Ценную информацию можно получить и тогда, когда эти системы рассматриваются в рамках моделей массового обслуживания.

Учитывая вышеизложенное, в данной работе рассматривается многопроцессорная система как модель массового обслуживания, в которой задаются условие параллельного обслуживания задач и ограничение времени ожидания задач в очереди.

Цель и рассматриваемые задачи

В данной работе исследуется многопроцессорная система с применением теории массового обслуживания. Состояние вычислительной системы носит случайный характер. Входной поток задач, продолжительность их обслуживания, количество требуемых вычислительных ресурсов, предел допустимого времени ожидания в очереди — все это случайные величины.

- Поставлена задача исследовать, описать и смоделировать состояния системы, учитывая случайное количество требуемых вычислительных ресурсов для задач и ограничения на время ожидания в очереди.
- Для заданных параметров системы и распределений рассчитать значения вероятностей состояний и найти оптимальный вариант обслуживания. Результаты представить в виде аналитических или численных решений.

Теоретическая и практическая значимость работы

Предложенные в результате исследования методы подхода к исследованию состояния многопроцессорных систем позволят расширить их исследование при различных начальных параметрах. Полученные результаты могут быть использованы в алгоритмах планировщиков очередей многопроцессорных систем.

Структура и объем работы

Диссертация состоит из введения, 3 глав, списка использованной литературы. Диссертация написана на 104 страницах и имеет 82 ссылки на литературу.

Основные результаты работы

1. Предлагается подход к исследованию многопроцессорных систем как расширенной модели массового обслуживания, включающей случайное количество процессоров для задач и ограничения на время их ожидания в очереди [1, 2, 3, 4].
2. С помощью расширенной модели массового обслуживания описаны состояния многопроцессорной системы, и для вероятностей этих состояний выведена система уравнений, для решения которой разработан численный алгоритм. В частных случаях получены аналитические формулы для вероятностей состояний системы [5, 6, 7].
3. Для заданных распределений выведена формула для вероятности, определяемой несервисными задачами в системе, чтобы оптимально выбрать параметры обслуживания. Для уменьшения значения данной вероятности разработан и реализован алгоритм оптимизации [8].

STUDY OF QUEUES IN A MULTIPROCESSOR SYSTEMS WITH A WAITING TIME RESTRICTION

Abstract

In the field of information technology, significant research is being conducted on distributed and parallel high-speed computing systems. These systems are essential in areas such as multi-agent intelligent information retrieval, data science for processing and storing large volumes of data, and scientific research addressing complex problems. The demands for increased accuracy, faster computation, and large-scale calculations in scientific and technological modeling also necessitate the use of such computing systems.

Problems that require large-scale computations often rely on multiprocessor systems, which involve more than one processor. This introduces challenges in efficiently managing the simultaneous execution of tasks while optimizing the use of resources within the multiprocessor system. Addressing these challenges is crucial in developing effective scheduling algorithms for such systems.

The development of efficient scheduling algorithms for multiprocessor systems is a key area of research in computer science and parallel computing. They require some understanding of adaptive scheduling strategies and efficient allocation of system resources to optimize various parameters such as system throughput, service latency, energy efficiency, etc.

The dynamic nature of scheduling algorithm development requires the continual incorporation of additional metrics. Optimization theory and machine learning tools are often employed to enhance the efficiency of next-generation scheduling algorithms. However, valuable insights can also be gained by analyzing these systems using queueing theory.

This study examines a multiprocessor system as a queueing model, where tasks are processed in parallel, and constraints on task waiting times in the queue are taken into account.

The purpose and problems of the work

This work investigates a multiprocessor system using queueing theory. The system's state is random, as the flow of incoming tasks, their service duration, the required amount of computational resources, and the limitation on waiting time in the queue are all random variables.

- The aim is to study, describe, and model the system's states by examining the random number of computational resources required by tasks and the limitations on waiting time in the queue.
- For the given system parameters and distributions, calculate the probabilities of

the states and determine the optimal service option. The results should be presented in the form of analytical or numerical solutions.

The practical significance of the work

The proposed approach to studying the state of multiprocessor systems allows for the exploration of these systems under various initial parameters. The results obtained can be applied to the queue scheduling algorithms in multiprocessor systems.

Structure and scope of work

The dissertation consists of an introduction, 3 chapters, a list of used literature. The thesis spans 104 pages and references 82 literature sources.

The main results of the work

1. A research approach is proposed for studying multiprocessor systems as an extended queueing model that incorporates random numbers of processors for tasks and limitations on their waiting times in the queue [1, 2, 3, 4].
2. Through the extended queueing model, the states of the multiprocessor system are described, and a system of equations for the probabilities of these states is derived. An algorithm has been developed for solving this system. In specific cases, analytical formulas for the probabilities of system states have been derived [5, 6, 7].
3. For the given distributions, a formula for the probability determined by the non-service tasks in the system has been derived to make optimal choices for service parameters. An optimization algorithm has been developed and implemented to minimize this probability [8].

