МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
"САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО"
Институт компьютерных наук и технологий
Направление **02.03.01** : Математика и компьютерные науки

Literature Review:

«Modern Scheduling Algorithms in Cloud-Fog and High-Performance Computing Environments»

Student: Darya Yashnova 5130201/20002

Teacher: Ph.D, Motorin D. E.

Saint-Petersburg, 2024

1 Analysis

All four articles are devoted to the topic of optimizing algorithms and improving their efficiency in various areas related to information technology. However, each article focuses on its own specific problem and offers unique approaches to solving it.

First article [1] discusses the problem of optimizing task scheduling algorithms in cloud computing. The authors propose an improved Shortest Job First(SJF) algorithm that takes into account the dynamic nature of cloud workloads. The article focuses on the problems faced by traditional task scheduling algorithms and offers solutions to overcome them.

Second article [2] is devoted to the problem of optimization of task scheduling algorithms for cloud and fog computing. The authors propose a new metaheuristic Hybrid Particle Whale Optimization Algorithm (PWOA), which combines the advantages of two other algorithms (Particle Swarm Optimization and Whale Optimization Algorithm) to improve the performance of task scheduling.

Third article [3] addresses the problem of optimizing task scheduling algorithms for simulating job generation similar to jobs in Google ' s data centers. The authors propose a new approach to job generation that takes into account the dynamic nature of workloads, and propose the Particle Swarm Search algorithm for configuring task scheduling parameters.

Fourth article [4] discusses the GARLSched algorithm, which is aimed at optimizing task scheduling in computing systems. The algorithm involves multiple networks - policy network, RL value network, GA value network, and discriminator network - that work together to improve scheduling performance. The algorithm utilizes expert guidance and reinforcement learning to train the networks and make scheduling decisions based on the workload trace provided as input.

General information in articles:

- All articles are devoted to optimization of task scheduling algorithms.
- The authors propose new approaches to solving existing problems in the field of task planning.
- The articles use simulation methods to test the proposed algorithms.

Differences in articles:

- 1. Target areas of application of algorithms:

  Article [1] focuses on cloud computing.

  Article [2] deals with cloud and fog computing.

  Article [3] is devoted to the simulation of job generation in Google data centers.

  Article [4] also examines task generation simulation, but with an emphasis on three different swarms that are trained based on task reconfiguration conditions.

- 2. The proposed algorithms:

  Article [1] improves the Shortest Job First (SJF) algorithm.

  Article [2] proposes a new metaheuristic algorithm Hybrid Particle Whale Optimization Algorithm (PWOA).

  Article [3] presents a new approach to task generation and the Particle Swarm Search algorithm for configuring task scheduling parameters.

Article [4] uses the Particle Swarm Search method and three different swarms for learning based on task reconfiguration conditions.

- 3. Focus on the dynamic nature of workloads:

  Articles [1], [2] and [4] focus on the dynamic nature of workloads and propose appropriate solutions.

  Article [3] also takes into account the dynamic nature of jobs, but focuses on simulating job generation in Google data centers.

2 Table 1

Research Paper	Heuristic methods	Prioritization	Math methods	Comparative studies	Optimization Parameters
Optimizing Task Scheduling in Cloud Computing: An Enhanced Shortest Job First Algorithm		Dynamic prioritization	SJF	+	Task completion time, Resource utilization
A Hybrid Particle Whale Optimization Algorithm with application to workflow scheduling in cloud–fog environment	PWOA (PSO+WOA)		Fitness Function	+	TET, TEC
Energy-aware scheduling of malleable HPC applications using a Particle Swarm optimised greedy algorithm	PSO	FIFO	PSO, Greedy algotithm, Friedman's test	+	Task Reorganization parameters, Server Shutdown Options
GARLSched: Generative adversarial deep reinforcement learning task scheduling optimization for large-scale high performance computing systems	Expert policies: F1, F2, F3, F4, WPT, UNICEF	Expert policies	DRL, GAN, MDP	+	AVGwt, AVGbsld

3 Research results

Research Paper	Results
Optimizing Task Scheduling in Cloud Computing: An Enhanced Shortest Job First Algorithm	Experiments in CloudSim tested the modified SJF algorithm. Resource allocation, task completion time, and fair distribution of resources were evaluated. The comparison was made with different metrics, including Make Span, average waiting time, and the level of resource usage. Optimization included comparison of VM configurations and scheduling algorithms such as Traditional SJF, Round Robin, and First - Come - First - Serve. The results showed that the modified SJF has a lower average waiting time, which indicates its effectiveness in the operational allocation of resources.
A Hybrid Particle Whale Optimization Algorithm with application to workflow scheduling in cloud–fog environment	To test the algorithm, a series of experiments were performed in various workflows, including Inspiral, Montage, Epigenomics, Sipht, and Cybershake. The proposed PWOA algorithm was compared with the standard PSO and WOA algorithms. According to the simulation results, the proposed PWOA algorithm surpassed both the standard PSO and WOA algorithms for TET and TEC. For more accurate results, further studies plan to take into account additional parameters, such as total execution energy, latency, offlooding, and delivery times.
Energy-aware scheduling of malleable HPC applications using a Particle Swarm optimised greedy algorithm	To test the algorithm, experiments were performed using ten different experimental settings under identical operating conditions. In each of these settings, jobs were generated so that there is only one large set of job sets G_1, consisting of 100 subsets S_i, where each S_i contains 50 different jobs to schedule on ten identical servers. The main difference between the first four settings and the last six is the decision-making process. For each setting, different steps of the FIFO (FIFO), FIFO-Rcfg (FIFO-Rcfg), FIFO-Poff (FIFO-Poff), FIFO-Rcfg-Poff (FIFO-Rcfg-Poff), Rand algorithms are shown- Param1, Rand - Param2, Rand - Param3, Swarm1, Swarm2, and Swarm3, which were used to evaluate the scheduler's performance.
GARLSched: Generative adversarial deep reinforcement learning task scheduling optimization for large-scale high performance computing systems	The proposed algorithm, called GARLSched, is evaluated using various workloads and optimization goals, such as average task waiting time and slow down. The experiment shows that GARLSched outperforms basic DRL methods and algorithms in terms of scheduling performance for various workloads and optimization purposes. A model trained on specific workload data(SDSC-SP2) demonstrates a good generalization ability when applied to invisible workloads such as Lublin-256, HPC2N, and SDSC - BLUE. The results show that GARLSched can effectively learn high-quality planning policies and perform well in practical scenarios with a high load on medium-and large-scale computing resources, solving problems of service degradation.

References

[1] Yellamma Pachipala, Kavya Sri Sureddy, A.B.S. Sriya Kaitepalli, Nagalakshmi Pagadala, Sai Satwik Nalabothu, Mihir Iniganti (2024)
Optimizing Task Scheduling in Cloud Computing: An Enhanced Shortest Job First Algorithm, doi: https://doi.org/10.1016/j.procs.2024.03.250, (ICIDCA 2024).

[2] Sumit Bansal, Himanshu Aggarwal (2023)
A Hybrid Particle Whale Optimization Algorithm with application to workflow scheduling in cloud–fog environment,
doi: https://doi.org/10.1016/j.dajour.2023.100361, Decision Analytics Journal.

[3] Briag Dupont, Nesryne Mejri, Georges Da Costa (2020)
Energy-aware scheduling of malleable HPC applications using a Particle Swarm optimised greedy algorithm, doi: https://doi.org/10.1016/j.suscom.2020.100447, Sustainable Computing: Informatics and Systems.

[4] Jingbo Li, Xingjun Zhang, Jia Wei, Zeyu Ji, Zheng Wei (2022)
GARLSched: Generative adversarial deep reinforcement learning task scheduling optimization for large-scale high performance computing systems,
doi: https://doi.org/10.1016/j.future.2022.04.032, Future Generation Computer Systems.