A COMPREHENSIVE STUDY OF VARIOUS SECURITY-SENSITIVE JOB SCHEDULING TECHNIQUES

[1] Amanpreet Kaur, [2]Anil Kumar
[1]Student Mtech 4th sem. , [2] Assistant Professor GNDU(Amritsar)
[1] Department of Computer Engineering & Technology
GNDU,Amritsar,India

Abstract: Cloud computing has become popular due to its wide range of applications via internet. The service composition based techniques that are conscious from the server selection from the cloud can progress to the cost and efficiency of cloud computing. This paper has focused on the comprehensive review on security-sensitive scheduling techniques. The comparative analysis indicates that existing techniques suffer from various issues. The overall objective of this paper is to compare and evaluate the various limitations in existing job scheduling techniques. In the end of this paper, suitable future challenges have also been presented.

***IndexTerms* – Job Duplication, Security, Task Scheduling.**

I. INTRODUCTION

1.1 PARALLEL COMPUTING

In the easiest sense, parallel computing is the simultaneous use of multiple compute resources to solve a computational problem:

- A problem is broken into discrete parts that can be solved concurrently.
- Each part is further broken down to a series of instructions.
- Instructions from each part execute simultaneously on different processors.
- An overall control/coordination mechanism is employed.

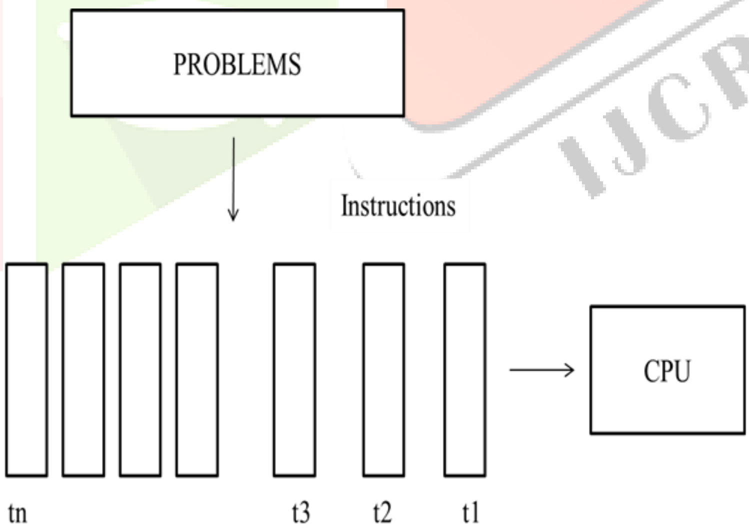

Figure 1.1: Parallel Computing

The computational problem should be able to:

- Be broken apart into discrete pieces of work that can be solved simultaneously;
- Execute multiple program instructions at any moment in time;
- Be solved in less time with multiple compute resources than with a single compute resource.

Parallel computing is a form of computing architecture where many processors execute or process an application or computation simultaneously. Parallel computing helps in performing large computations by dividing the workload between more than one processor, which works through the computation at the exact same time. Parallel computing can also referred as parallel processing. Parallel computing is closely linked to concurrent computing, they are commonly used together, and often conflated, although the two are distinct: it's possible to possess parallelism without concurrency (such as bit-level parallelism), and concurrency without parallelism (such as multitasking by time-sharing on a single-core CPU). In parallel computing, a computational task is normally broken down in several, often many, very similar subtasks which can be processed independently and whose results are combined afterwards, upon completion.

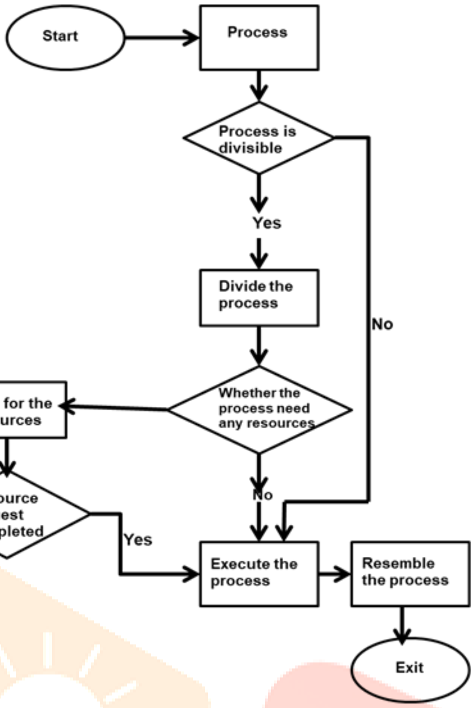

Figure 1.2: Flow Diagram of a Process

1.2 BASIC SCHEDULING IN PARALLEL COMPUTING

Scheduling [2] process in the cloud computing scenario has several fundamental components:

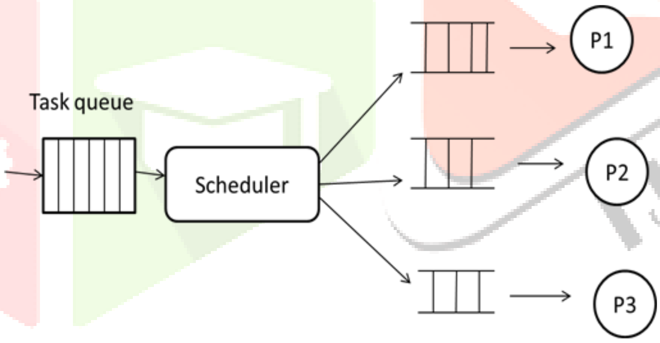

Figure 1.3: Basic Scheduling Diagram

1.3 TASK SCHEDULING

Scheduling is the group of strategies that manage an order of execution of multiple tasks within the processors in an effort to decrease the enough time and the expense required to carry out all of these tasks. While in the cloud environment, task scheduler plays vital role of allocating cloud provider's resources among the list of large number of users. Task scheduling [1] refers to distribution of the tasks among the list of cloud servers which process or execute these tasks for user (or client). A simple yet effective task scheduling policy provides proper usage of resources, load balancing [6] and optimization [16] of execution cost and time.

In Static Scheduling, for every single task, communication cost and computation expenditure is considered beforehand. In Dynamic Scheduling [7], decisions are taken in advance at run time and without cost details are available earlier. Static Scheduling further comprise Heuristic and Guided Random algorithms [10]. Heuristics involves 3 subcategories namely clustering, list and duplication algorithms. In Clustering based algorithms for instance Clustering for Heterogeneous Processors [10] (CHP), clusters of tasks are assigned to appropriate processors. In contrast, duplication algorithms try to copy tasks to reduce makespan. Algorithms such as for instance Contention aware (CA-D) duplication algorithms [12] get rid of the communication cost by placing tasks on same processor. A large time complexity and more processor usage limit the utilization of duplication based algorithms in cloud environment. List Scheduling algorithms such as for example Critical Path on Processor (CPOP) gives minimum makespan along side efficient time complexity. The list Scheduling Algorithms are most practical.

1.4 TASK DUPLICATION

Task duplication is a common technique to lessen the mandatory communication between processors. In this technique certain crucial tasks are executed on more than one processor. The information they procedure is then locally on different processors and less communication must be sent between the processors. Task duplication is just a relatively new approach for DAG scheduling. This type of scheduling cans also an NP-complete problem. Performance of the duplication based algorithms is better than non-duplication based ones when it comes to generating smaller schedule lengths. However, that is usually achieved at the cost of higher time complexity and larger amount of processors.

The interprocessor communication [1] overhead occurs when two tasks of a similar program assigned to different processors have dependencies and they have to exchange data among them. Task duplication is method of reducing the interprocessor communication overhead which often can improve the total execution time. Task duplication [2] means scheduling a simultaneous program by redundantly allocating a few of its tasks on which other tasks critically depend. This reduces the start times of waiting tasks that may eventually improve the overall execution time of the entire program. Duplication based scheduling may be particularly useful for systems with high communication overhead like a network of workstations.

1.5 SECURITY

In today's world internet has been utilized by almost everyone. Numerous file exchanges take place online including many official documents. These files require some kind of security mechanisms while being transmitted within the internet. Cryptography [23] is one particular means through which an individual can encrypt the information and send it through the internet. In this way the information is safe and unreadable for the intruders.

1.6 MOTIVATION

Task duplication is effective in preventing both the information transmission time and encryption time from delaying workflow tasks start. The primary reason is that whenever workflow tasks and their successor tasks are running on the single virtual machines (VMs), the output data of those tasks duplicating execution may be used by their successor tasks directly without the data transmission and encryption. To guarantee data security, output data of tasks original execution is likely to be encrypted before being stored. We study to schedule workflow tasks to VMs with selective duplicating tasks into idle time slots, caused by data dependencies between tasks, to minimize the start time of workflow tasks that may further minimize both makespan and the monetary cost for workflow and guarantee data security.

II. LITERATURE SURVEY

Huangke Chen et al. (2016) [1] investigates that the issue of scheduling workflows with security-sensitive intermediate data. We first present four theorems to minimize both makespans and monetary costs for executing workflow applications. From then on, we produce a security-aware scheduling approach, namely SOLID, striving to enhance the makespans and costs of executing workflows, and resource efficiency of VMs, while guaranteeing the security requirements of workflows. To judge the potency of SOLID, we conduct extensive simulation experiments in the context of randomly generated workflows and real-world workflow traces. The outcomes reveal that the proposed SOLID approach prevails over existing algorithms with regards to makespan, monetary costs and resource efficiency.

Arani Bhattacharya et al. (2015) [2] proposed that that by allowing duplicate execution of a couple of selected tasks results in a polynomial time scheduling algorithm that minimizes the sum total completion time of an application. The algorithm ATOM (Algorithm for Time Optimization on Mobiles) determines a schedule to execute tasks of a concurrent application with duplication in a way that makespan is minimized. The simulation and trace-driven experiments reveals that ATOM significantly reduces makespan and energy consumption while executing in polynomial time. The trace-driven simulation on benchmark applications shows that our algorithm reduces the scheduling time by 8 times in comparison to a typical optimization solver while guaranteeing minimum makespan.

Safwat A. Hamad et al. (2016) [3] proposes an better Genetic Algorithm for task scheduling problem in the Cloud computing environment. The proposed algorithm targets to minimize completion time and cost, and maximize resource utilization. The completion time for the proposed TS-GA algorithm is reduced by (41.83%) and (39.26%) about the default GA, and RR algorithms, respectively. The expense of the proposed TS-GA algorithm is reduced by (3.6%) and (6.07%) regards to the default GA and RR algorithms respectively. The resource utilization of the proposed TS-GA algorithm is improved by (47%) and (30.04%) in regards to the default GA and RR algorithms, respectively. The speedup of the proposed TS-GA algorithm is improved by (34.03%) and (33.65%) about the default GA and RR algorithms, respectively. The efficiency of the proposed TS-GA algorithm is improved by (34.06%) and (33.66%) concerning the default GA and RR algorithms respectively.

Indrajeet Gupta et al. (2016) [4] propose a task duplication-based workflow scheduling algorithm
for heterogeneous cloud environment which is centered on task duplication realization. The proposed algorithm has two phases. The first phase computes the priority of all the tasks and second phase goes through the scheduling with task duplication by calculating data arrival time from one task to another task. Proposed algorithm aims to minimize workflow execution time and to maximize the resource utilization. The performance evaluation of the proposed algorithm is done on benchmark scientific workflow applications with different task-cloud heterogeneity. Comparisons of the simulated results with the some existing workflow scheduling algorithms noticeably show that the proposed algorithm outpaces in terms of makespan and average cloud

utilization. He propose a task duplication-based workflow scheduling (TDWS) algorithm that is proficient to schedule large DAG of dependent tasks in heterogeneous cloud environment by duplicating parent task nodes before the child task on the same cloud server.

Akhila Ka et al. (2016) [5] discuss various methods are discussed where deduplication methods are carried from encrypted data in a sizeable storage area. The majority of the methods studied here focus on the foundation of convergent encryption, which is really a simple approach that produces deduplication appropriate for encrypted data. This record dense world, we cannot compromise on both security and duplication of data across storage areas. A technique must be formulated that may enhance storage optimization without negotiating on encryption method; by giving deduplication technique in data storage servers where in fact the available data is encrypted.

Mokhtar A. Alworafi et al. (2017) [6] focused on Scheduling Cost Approach (SCA) that calculates the cost of CPU, RAM, bandwidth, storage available. In this approach, the tasks will be distributed among the VMs based on the priority given by user. The priority depends on the user budget satisfaction. The proposed SCA will try to improve the load balance by selecting suitable VM for each task. The results of SCA are compared with the results of FCFS and SJF algorithms which proves that, the proposed SCA approach significantly reduces the cost of CPU, RAM, bandwidth, storage compared to FCFS and SJF algorithms. He presented the cost priority to schedule tasks on cloud resources that meet the user budget satisfaction. The Scheduling Cost Approach (SCA) calculates the cost of all resources. Each task is assigned based on task priority taking into consideration suitable resources for execution and distribution of load balancing between the VMs in clusters. He conducted six experiments to test the performance of our approach. The comparison of SCA was done with FCFS and SJF algorithms under same task priority and resource cost processing.

A. Kumaravel et al. (2014) [7] proposed the grid task scheduling algorithm that is designed to achieve high throughput computing in a grid environment. Fault Tolerance is likely to be attained by the Task Duplication Method. The Refinement Task improves the performance by neglecting when the rest of jobs when any among the replica gets completed successfully. He creates the scheduler which improves the performance. The Duplication-based Dynamic Scheduling Algorithm (DDSA) which involves Task duplication and Refinement combined with the dynamic threshold could be the core of the technology.

GuanWang et al. (2016) [8] proposed a novel task scheduling algorithm for heterogeneous computing named HSIP (heterogeneous scheduling algorithm with improved task priority) whose functionality depend on three pillars: (1) an improved task priority strategy centered on standard deviation with improved magnitude as computation weight and communication cost weight to produce scheduling priority more reasonable; (2) an access task duplication selection policy to help make the makespan shorter; and (3) an better idle time slots (ITS) insertion-based optimizing policy to really make the task scheduling more efficient. He evaluates the proposed algorithm on randomly generated DAGs, with a couple real application DAGs in comparison with some classical scheduling algorithms. Based on the experimental results, our proposed algorithm appears to execute much better than other algorithms with regards to schedule length ratio, efficiency, and frequency of best results.

Elhossiny Ibrahim et al. (2016) [9] possesses an enhancement task scheduling algorithm on the Cloud Computing environment has been introduced to lessen the make-span, along with, decrease the cost of executing the independent tasks on the cloud resources. The principles of the algorithm is dependent on calculating the sum total processing power of the available resources (i.e., VMs) and the total requested processing power by the users' tasks, then allocating a small grouping of users' tasks to each VM on the basis of ratio of its needed power in accordance with the total processing power of most VMs. He proves that the efficiency of the enhancement algorithm by minimizing make-span by 26.06 % additionally, decreasing the cost by 67.52 %.

Aida A Nasr et al. (2016) [10] focused on a new task duplication scheduling algorithm has been presented for heterogeneous distributed computing systems (HDCS) to enhancement scheduling performance. This algorithm uses new attribute called Rank to assign a priority for every single task. Additionally it uses task duplication technique to decrease the communication overhead. The performance analysis indicated that the proposed MCND algorithm has better performance than CPOP with duplication algorithm. According to the simulation results, it is found that the MCND algorithm is better than the other algorithm in terms of SLR, speedup and execution time. The new algorithm applies new job duplication algorithm to lessen the schedule amount of DAG. It applies the task duplication on VIT only not on every task in DAG. So, it takes low execution time and energy to schedule the tasks.

III. COMPARISON TABLE:

Table 3.1: Description of different Techniques.

Reference	Technique	Year	Task Duplication	Objective	Security	Platform
[7]	Duplication-based Dynamic Scheduling Algorithm(DDSA)	2014	✓	To attain high throughput computing in a grid environment , Fault Tolerance,improves performances	×	Grid Computing
[11]	Genetic					Parallel

	based technique	2014	✓	Solving task scheduling in multi-processor systems.	×	Computing
[18]	DBD-CTO algorithm, PSO& Genetic Simulated Annealing Algorithm	2014	×	Study of the different algorithms for their suitability, feasibility, adaptability in the context of cloud scenario	✓	Cloud computing
[20]	Batch mode Heuristic Algorithm ,On- line mode Heuristic Algorithm ,Dependency Mode heuristic Algorithm	2014	✓	Aims to map the appropriate tasks, jobs to the resources that are available by keeping in mind the QoS (Quality of Service) requirements of the user.	✓	Cloud computing
[17]	Cuckoo Search Algorithm	2015	×	Minimize total execution time& CSA to schedule the tasks in Cloud computing.	✓	Cloud computing
[3]	Genetic based task scheduling algorithm	2016	×	Minimize completion time and cost of tasks& maximize resource utilization.	×	Cloud Computing
[4]	duplication-based workflow scheduling algorithm TDWS	2016	✓	Makespan & Cloud utilization	×	Heterogeneous Cloud Environment
[5]	Storage optimization techniques	2016	✓	To makes deduplication compatible with encrypted data.	✓	Cloud Computing
[8]	Heterogeneous Scheduling with Improved Task Priority (HSIP)	2016	×	our proposed algorithm appears to execute a lot better than other algorithms when it comes to schedule length ratio, efficiency, and frequency of best results	✓	Heterogeneous Computing
[9]	Enhancement task scheduling algorithm	2016	✓	minimized make span and decreased in cost	×	Cloud Computing
[10]	Mean Communication Node with Duplication (MCND)algorithm	2016	✓	low execution time to schedule the tasks	×	Heterogeneous Distributed Computing Systems

[13]	Load balancing task scheduling algorithm based on weighted random and feedback mechanisms	2016	×	Achieved balanced load as well as self-adaptability	×	Cloud Computing
[14]	Green clonal scheduling optimization algorithm	2016	×	Reduce the execution time and energy consumption, and can achieve resource load balancing, thus effectively improve the resource utilization and scheduling efficiency	×	Green cloud computing
[15]	Genetic algorithm	2016	×	produces the optimal solution of the tasks and optimize the waiting time of overall system	✓	Cloud computing
[16]	Improved ant colony algorithm; Improved particle swarm optimization algorithm	2016	×	Improve the operating efficiency &Improve the scheduling ability of the algorithm.	×	Cloud computing
[19]	Tasks scheduling model based on the original priority calculation method	2016	✓	Subtasks execution time can be advanced and the complete time of the whole task-set can be cut down to a certain extent.	×	Cloud computing
[1]	SOLID	2017	✓	Reduce Communication Overheads,Makespan,Mon-etary cost	✓	Cloud Computing
[2]	ATOM (Algorithm for Time Optimizatio n Mobiles)	2017	✓	Reduce Scheduling Time	×	Mobile Cloud
[6]	Scheduling Cost Approach (SCA)	2017	×	efficient and capable of achieving the cost and improving the task scheduling	×	Cloud Computing

IV. COMPARISON OF PARAMETERS

Table 4.1: Different Parameters Used.

	Makespan	Speedup	Utilization	Overheads	Security	Efficiency
Duplication-based Dynamic Scheduling Algorithm(DDSA) [7]	✓	×	✓	×	×	×
Genetic based technique [11]	×	✓	✓	×	×	✓
DBD-CTO algorithm, PSO, Genetic Simulated Annealing Algorithm[18]	✓	×	×	×	✓	×
Batch mode Heuristic Algorithm ,On- line mode Heuristic Algorithm ,Dependency Mode heuristicAlgorithm[20]	×	✓	✓	×	✓	✓
Cuckoo Search Algorithm[17]	×	×	✓	✓	✓	✓
Genetic based task scheduling algorithm [3]	×	✓	✓	×	×	✓
Task duplication-based workflow scheduling algorithm(TDWS)[4]	✓	×	✓	✓	×	×
Storage optimization techniques [5]	✓	×	×	×	✓	✓
Heterogeneous Scheduling with Improved Task Priority (HSIP)[8]	×	✓	✓	✓	✓	✓
Enhancement task scheduling algorithm [9]	✓	✓	✓	×	×	×
Mean Communication Node with Duplication (MCND)algorithm [10]	×	✓	✓	×	×	×

Load balancing task scheduling algorithm based on weighted random and feedback mechanisms[13]	✓	✗	✓	✗	✗	✓
Green clonal scheduling optimization algorithm[14]	✗	✓	✓	✓	✗	✓
Genetic algorithm[15]	✗	✓	✓	✗	✓	✓
Improved ant colony algorithm; Improved particle swarm optimization algorithm[16]	✗	✓	✓	✗	✗	✗
Tasks scheduling model based on the original priority calculation method[19]	✓	✓	✗	✓	✗	✓
SOLID [1]	✓	✗	✓	✗	✓	✓
ATOM (Algorithm for Time OptimizationMobiles) [2]	✓	✗	✓	✗	✗	✗
Scheduling Cost Approach (SCA)[6]	✗	✓	✓	✓	✗	✓

V. CONCLUSION

In this review paper, the existing job scheduling techniques have been reviewed which have designed so far by various researchers. The benefits and limitations of existing job scheduling techniques have also been presented. In order to overcome the issues associated with existing job scheduling techniques, in this paper an efficient tiger optimization based multi-objective scheduling technique will be proposed in near future. The multi-objective fitness function will also be designed which will consider security and energy consumption in joules, simultaneously. Further, job duplications will also be considered to reduce the inter-communication overheads.

REFRENCES

[1] H. Chen, X. Zhu, D. Qiu, L. Liu, and Z. Du, "Scheduling for workflows with security-sensitive intermediate data by selective tasks duplication in clouds," IEEE Transactions on Parallel and Distributed Systems, 2017.

[2] A. Bhattacharya, A. Banerjee, and P. De, "Scheduling with task duplication for application offloading," in Consumer Communications & Networking Conference (CCNC), 2017 14th IEEE Annual, pp. 678{683, IEEE, 2017.

[3]S. A. Hamad and F. A. Omara, "Genetic-based task scheduling algorithm in cloud Computing environment," International Journal of Advanced computer Science and Applications, vol. 7, no. 4, pp. 550-556, 2016.

[4]I. Gupta, M. S. Kumar, and P. K. Jana, "Task duplication-based workflow scheduling for heterogeneous cloud environment," in Contemporary Computing (IC3), 2016 Ninth International Conference on, pp. 1-7, IEEE, 2016.
[5] Akhila K, Amal Ganesh, Sunitha C," A Study on Deduplication Techniques over Encrypted Data" Fourth International Conference on Recent Trends in Computer Science & Engineering Chennai, Tamil Nadu, India, doi: 10.1016/j.procs.2016.05.123, 1877-0509 © 2016 The Authors. Published by Elsevier B.V.

[6] Mokhtar A. Alworafi, Atyaf Dhari, Asma A. Al-Hashmi, Suresha, A. Basit Darem ,” Cost-Aware Task Scheduling in Cloud Computing Environment” I.J. Computer Network and Information Security, 2017, 5, 52-59.

[7] A. Kumaravel,” Review on a Dynamic Scheduling Algorithm for Grid with Task Duplication” Middle-East Journal of Scientific Research 20 (1): 94-99, 2014 ISSN 1990-9233 © IDOSI Publications, 2014 DOI: 10.5829/idosi.mejsr.2014.20.01.11255

[8] GuanWang,YuxinWang, Hui Liu and He Guo,” HSIP: A Novel Task Scheduling Algorithm for Heterogeneous Computing” Hindawi Publishing Corporation Scientific Programming Volume 2016, Article ID 3676149, 11 pages http://dx.doi.org/10.1155/2016/3676149.

[9] Elhossiny Ibrahim, Nirmeen A. El-Bahnasawy, Fatma A. Omara,” Task Scheduling Algorithm in Cloud Computing Environment Based on Cloud Pricing Models”, 2016 World Symposium on Computer Applications & Research, 978-0-7695-5832-5/16 $31.00 © 2016 IEEE DOI 10.1109/WSCAR.2016.20

[10] Aida A Nasr, Nirmeen A EL-Bahnasawy, Ayman EL-Sayed,” A New Duplication Task Scheduling Algorithm in Heterogeneous Distributed Computing Systems” ISSN: 2302-9285 Vol. 5, No. 3, September 2016, pp. 373~382, DOI: 10.11591/eei.v5i3.664.

[11] Ahmad, Ishfaq, and Yu-Kwong Kwok. "On exploiting task duplication in parallel program scheduling." IEEE Transactions on Parallel and Distributed Systems 9, no. 9 : 872-892.

[12] Sinnen, Oliver, Andrea To, and Manpreet Kaur. "Contention-aware scheduling with task duplication." Journal of Parallel and Distributed Computing 71, no. 1 (2011): 77-86.

[13] Qian, Zhang, Ge Yufei, Liang Hong, and Shi Jin. "A load balancing task scheduling algorithm based on feedback mechanism for cloud computing." International Journal of Grid and Distributed Computing 9, no. 4 (2016): 41-52.

[14] Liu, Yang, Wanneng Shu, and Chrish Zhang. "A Parallel Task Scheduling Optimization Algorithm Based on Clonal Operator in Green Cloud Computing." Journal of Communications 11, no. 2 (2016).

[15] Lakshmi, R. Durga, and N. Srinivasu. "A dynamic approach to task scheduling in cloud computing using genetic algorithm." Journal of Theoretical and Applied Information Technology 85, no. 2 (2016): 124.

[16] He, XiaoLi, Yu Song, and Ralf Volker Binsack. "The intelligent task scheduling algorithm in cloud computing with multistage optimization." International Journal of Grid and Distributed Computing 9, no. 4 (2016): 313-324.

[17] Navimipour, Nima Jafari, and Farnaz Sharifi Milani. "Task scheduling in the cloud computing based on the cuckoo search algorithm." International Journal of Modeling and Optimization 5, no. 1 (2015): 44.

[18] Singh, Raja Manish, Sanchita Paul, and Abhishek Kumar. "Task Scheduling in Cloud Computing." International Journal of Computer Science and Information Technologies (IJCSIT) 5, no. 6 (2014): 7940-7944.

[19] Chen, Qing-Yi, Zhi-Hong Liang, Hong-Wei Kang, Yu-Ming Ma, and Dong Wang. "Research of Dependent Tasks Scheduling Algorithm in Cloud Computing Environments." In ITM Web of Conferences, vol. 7, p. 08001. EDP Sciences, 2016.

[20] Nallakumar, R., N. Sengottaiyan, and KS Sruthi Priya. "A Survey on Scheduling and the Attributes of Task Scheduling in the Cloud." *Int. J. Adv. Res. Comput. Commun. Eng* 3 (2014): 8167-8171.

[21] S. Pearson and A. Benameur, “Privacy, security and trust issues arising from cloud computing,” in Proceedings of the Second International Conference on Cloud Computing Technology and Science(CloudCom). IEEE, 2010, pp. 693–702.

[22] Z. Zhu, G. Zhang, M. Li, and X. Liu, “Evolutionary multi-objective workflow scheduling in cloud,” IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 5, pp. 1344–1357, 2016.

[23] X. Zhu, Y. Zha, P. Jiao, and H. Chen, “Security-aware workflow scheduling with selective task duplication in clouds,” in Proceeding of the 2016 Spring Simulation Multi-Conference (SPRINGSIM). SCS,2016, pp. 407–414.

[24] R.Kaur and S. Kinger, "Enhanced Genetic Algorithm based Task Scheduling in Cloud Computing," International Journal of Computer Applications, vol. 101, 2014.

[25] R. N. Calheiros and R. Buyya, “Meeting deadlines of scientific workflows in public clouds with tasks replication,” IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 7, pp. 1787–1796, 2014.