Chapter 8

Conclusions and Future Directions

In this research study, we have investigated scheduling of scientific workflow applications on cloud computing infrastructure, a high-performance computing system. This chapter concludes the dissertation with emphasis on the contributions made by this study. Some of the future research directions are also laid down at the end.

8.1 Contributions

The algorithms presented in this study aim to satisfy several QoS requirements such as minimizing makespan, execution cost, failure probability and so on. This section summarizes the devised strategies to achieve the above goal.

A comprehensive survey of existing literature in the domain of scheduling in distributed computing systems is presented in Chapter 2. The scheduling techniques are classified based on application models, allocation techniques and objective functions. In particular, the scheduling techniques employing different optimization methods viz. heuristic, metaheuristic and hybrid are explored. This chapter facilitates in understanding of prevailing research, identification of research gaps and outlining the need for additional research.

To achieve our target of minimizing makespan, we have presented a scheduling strategy utilizing a novel metaheuristic technique, Intelligent Water Drops (IWD) algorithm in Chapter 3. Numerous IWDs are employed to search diverse solutions where an IWD builds a solution by moving from one task to another in precedence order and selects a resource for each task. This is done by using a probabilistic function which relies on the information about soil between task and resource, and completion time of the task on resource. Soil and velocity of an IWD as well as soil between task and resource is updated at each step to promote the selection of promising resources. After numerous iterations, the best solution in terms of makespan is selected as the final schedule. The proposed technique is compared with PSO and SGA based scheduling strategies considering three scientific workflow applications namely Montage, LIGO and Epigenomics from diverse domains. Experiments are performed in cloud simulation environment. by varying the size of each
workflow application. In case of LIGO workflows, the proposed strategy is 13-54% better than PSO and 3-25% better than SGA. The results for Epigenomics workflows show that the reduction in average makespan of IWD over PSO and SGA is 5-42% and 2-35% respectively. In case of Montage workflows, the average makespan obtained by PSO is 29-54% more than IWD whereas SGA performs slightly better than IWD.

Chapter 4 extended previous approach to form Load Balanced-IWD (LB-IWD) algorithm which not only reduces the makespan of workflows but also balances the load of VMs in clouds. Load balancing of VMs results in fair and maximum utilization of resources in cloud data center. The probabilistic function is modified to incorporate the resource utilization of resource in addition to soil between task and resource, and completion time of the task on resource. Some new ideas are integrated into the algorithm that contribute in achieving solutions of higher quality. Tasks are prioritized in the descending order of upward rank and delta soil is constrained to control the convergence rate. LB-IWD is compared with five scheduling techniques: S-IWD, S-PSO, Min-Min, Max-Min and MCT considering three scientific workflows. Evaluation results demonstrate that LB-IWD achieves 30% lower execution time than these algorithms and has less CV than other algorithms which means the system is more balanced in LB-IWD as compared to other algorithms.

The cost-optimized scheduling strategy based on IWD, which keeps makespan under user stated deadlines, is suggested in Chapter 5. It considers key features of IaaS cloud such as pay-per-use model and diverse configuration of the resources along with issues of performance variation and delay in allocation of VMs. Empirical analysis based on four scientific workflows demonstrated that the proposed strategy outperforms three state-of-the-art algorithms, DOGA, CEDS and PSO in context of meeting deadlines and cost optimization. Mostly our approach produces lowest cost schedules followed by CEDS, DOGA and PSO. Though these schedules have comparatively higher average makespan at moderate and loose deadlines, they are always below deadline values. PSO often does not comply with tight deadline constraints and generates schedules with higher execution costs. The solutions produced by DOGA have lesser costs than PSO for most of the cases, still the costs are comparatively much higher than other algorithms. CEDS performed better than PSO and DOGA with respect to execution costs, however its performance is inferior to IWD in majority of the cases.

A multi-objective scheduling strategy is presented in Chapter 6 which focused on optimization of execution time, cost and reliability with limitations on time and budget. The
strategy combines IWD and GA algorithms along with non-dominance sorting to achieve trade off solutions. The initial population of GA is seeded with IWD schedule obtained from previous chapter to achieve the optimal solution in lesser number of generations. Single point crossover and uniform mutation operations are carried out to extensively search the problem space which further improves the performance of the algorithm. Fitness of each chromosome is evaluated based on non-dominance sorting and its perimeter value. A set of Pareto-optimal solutions are provided to users so that they can select a solution based on individual preference. Its potential is demonstrated using four scientific workflows from different domains having varying resource requirements. IWD-GA achieves 32-47% and 5-47% higher hypervolume than NSGA-II and HPSO respectively. Further IWD-GA also accomplishes promising results in terms of two set-coverage. It implies the achievement of a higher quality Pareto-optimal set by IWD-GA with appropriate diversity of solutions.

Another multi-objective scheduling strategy, which focused on reduction of makespan, cost and energy consumption in data center under deadline constraints, is implemented in Chapter 7. The simulation results based on performance metrics of hypervolume and two-set coverage reveals its significant potential over existing techniques.

### 8.2 Future Directions

Enhancement in technology is an unending process which provides scope for future research directions. The research study undertaken in this dissertation focused on makespan, execution cost, failure probability and energy consumption. The presented work can be extended as follows.

1. **(i)** As security issues in cloud have fueled concerns about its adoption by users, the scheduling problem in cloud can be further investigated by focusing on the aspect of security.

2. **(ii)** It would be stimulating to employ alternative techniques based on machine learning concepts to improve the efficiency of scheduling algorithms.

3. **(iii)** A novel pricing model based on spot instances has come into existence. It allows provisioning of resources at much lesser prices than their on-demand prices but are subject to interruption at any time. The challenge is to devise scheduling techniques which wisely choose spot instances and on-demand instances for executing user tasks to reduce execution costs while dealing with volatility of spot instances.