Workflow Scheduling in Cloud Computing Environment Using Firefly Algorithm

[1]Amanpreet Kaur, [2]Sonia Sharma
[1,2]Dept. CSE, Guru Nanak Dev University, Amritsar, Punajb, India

Abstract
Cloud computing is the new generation of networks that uses remote servers hosted on the Internet for various uses such as data storage, data management, software usage etc. There are huge amount of resources provided and users can make use of the resources in any way they want to. Workflow scheduling is the most important task in cloud computing field and users have to pay for resources that were used based in a pay-per-usage scheme. Today, researchers attempt to find newer ways for Workflow scheduling which could work well in the cloud environment. Another important element to be considered about cloud computing is Load balancing. Here in this document, we are recommending Firefly criteria for effective fill controlling in reasoning processing. This criterion is based on the travel behavior of the fireflies which go looking for the closest possible maximum alternatives. We employ Firefly algorithm to schedule the jobs and thereby evenly distribute the load and in turn reduce the overall completion time (makespan).

Keywords
Firefly, Cloud Computing, Scheduling

I. Introduction

Now a day's Cloud Computing is an evolving field that provides resource as services to the customers. Cloud computing provide services, shared resources or common infrastructure on demand through internet. We are having a huge number of cloud service providers who actually provide their customers with resources as services. They don't actually provide products but only services. Service provider provides the facilities on pay per use policy. Cloud computing provides an ease of access for the remote sites via internet. This would allow each user to sign in into web based services and hosts all the programs remotely. Cloud computing is providing a different environment using virtual machines. In other words 'Cloud' is defined as the virtual flexible execution environment of resources from multiple remote servers to provide metered service to users based on their preferences. Virtual machines helps user to fulfill their task without sacrificing the quality of service. To provide services to the customers and best utilizing resources, service providers must have to follow some criteria or scheduling algorithms to manage all these resources optimally. Scheduling is a difficult task in cloud computing environment because a cloud provider has to take care of many users according to their different QoS needs. Scheduling in cloud computing is of various types: task scheduling, workflow scheduling, resource scheduling, job scheduling etc. Every task could have varied parameters like needed information, desired completion time, expected execution time, job priority etc..,. An algorithmic program, which is typically associates in nursing abstract model that defines tasks while not specifying the physical location of valuable provides on that the tasks area unit executed. The population based mostly improvement algorithms that area unit used Genetic algorithms (GA), Particle Swarm improvement (PSO), microorganism search (searching) improvement (BFO) and Bat algorithmic program (BA) that uses a population of individuals to unravel the issue.

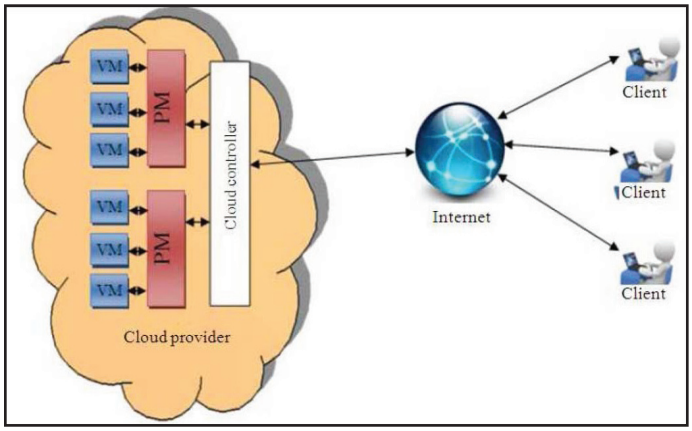


Fig. 1: Cloud Environment Model

A. Scheduling Problems in Cloud System

Scheduling in cloud system is not a single problem, it has many problems. Because of many QoS parameters like cost, efficiency, time, resources come between scheduling. Scheduling becomes a critical issue in cloud environment. In following, concepts related to cloud scheduling and various scheduling types are given.

B. Types of Scheduling

Task scheduling in different cloud environment is of many types-static and dynamic scheduling workflow scheduling, user level scheduling, real time scheduling and heuristic scheduling.

User level scheduling- In this, things like Load Balance, user requirement and other factors that effects user consumption rate of resources are considered.

Dynamic level task scheduling–According to the user's demand, an internet-based computing provides dynamically-scalable, efficient and optimized services, platforms and resources.

Real time task scheduling- Tasks should complete before deadline. Good throughput is the primary objective of real time scheduling.

Workflow scheduling- plays an important role. Workflow scheduling is used so that workload can be managed and profit can be increased. The algorithm that we propose is the firefly algorithm which works based on the swarm behavior of fireflies. This algorithm was first designed by a person named Yang who was attracted by a swarm of fireflies.
The firefly algorithm is known to be the best among all the other swarm based algorithms due to its rapid convergence properties, flexibility and error tolerance capacities. This firefly algorithm is a heuristic kind of algorithm and thus the precision of the remedy relies on the parameter places in the criteria. Meta-heuristic techniques can efficiently tackle these issues and can

find near-optimal solutions in comparatively shorter period of time. Meta-heuristic refers to a high level procedure which aims at finding, generating or selecting heuristic that can provide an approximate solution to an optimization problem [1]. This technique can efficiently solve problems that are larger in size and are complex.

II. Literature Survey

Scheduling of workflow is an essential conception in the field of cloud computing. There is plentiful number of algorithms shaped especially for the purpose of workflow scheduling.

A. Cat Swarm Optimization is a Heuristic Optimization Algorithm

In this CSO heuristic optimization algorithm, created based on the inspiration towards the swarm behavior of cats. Cats that generally have swarm behavior are said to have two modes of behavior namely

1. seeking mode and
2. Tracking mode

- **Seeking mode**

In seeking mode the cat stays idle and only has position whereas they do not have velocity.

- **Tracking Mode**

In tracing mode the cat is in motion and is said to possess both position and velocity.

This algorithm wholly lies on two modes of operation. The fitness factor for each cat is calculated and the best one is picked out. The best cat is stored in memory and it is updated with the next best cat. Here in the Cat swarm algorithm the virtual machines are disguised as cats. Another such implementation was done in the field based on the similar swarm behavior of where particles which was known as the particle swarm optimization algorithm.

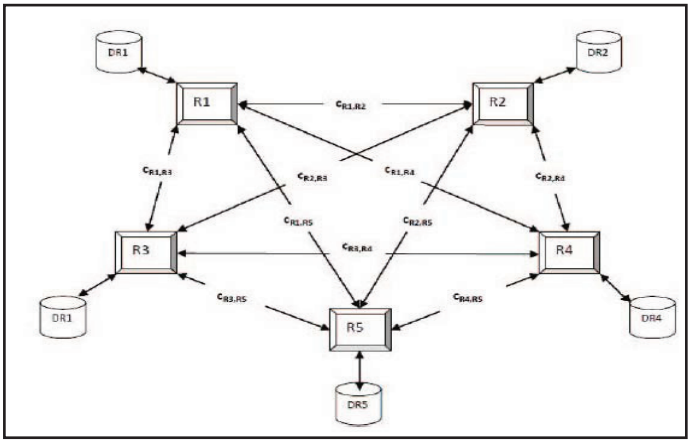


Fig. 2: Sample Architecture of Computing Nodes with Shortage

Where,

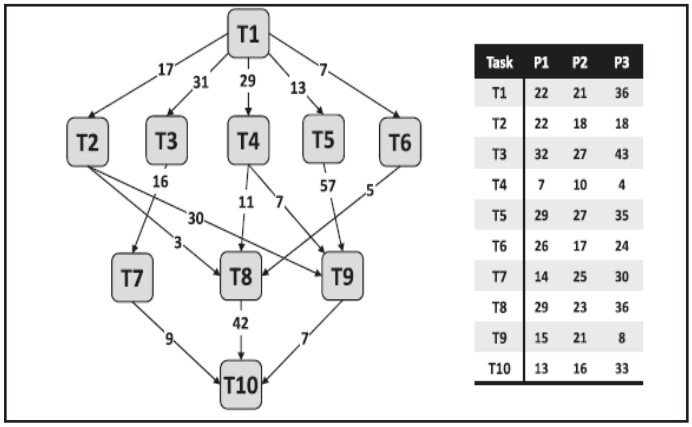
d_{ij} represents data flow from T_i and T_j which is assumed constant in size for all tasks.

C_{Ri} and C_{Rj} represents cost per unit data transferred between two resources R_i and R_j.

R1, R2, R3, R4, R5 are available 5 resources.

B. Particle Swarm Optimization

A compound in PSO is much like a fowl or even seafood travelling through a search area. The movement of each compound is associated with a speed which is said to have both scale and route. Place of each compound at any type of time is impacted by its best position and also within the best compound in an issue area. There is a particular property called as fitness value for each particle which is used to measure the performance ability of each particle in the space. Using this, the global best solution is determined by iterating to find local best solution with respect to fitness function. Each compound place at any type of time is affected by its best place and within the best compound in an issue space. The ability of a particle is calculated by a fitness value, which is a concern specific strategy.



Task	P1	P2	P3
T1	22	21	36
T2	22	18	18
T3	32	27	43
T4	7	10	4
T5	29	27	35
T6	26	17	24
T7	14	25	30
T8	29	23	36
T9	15	21	8
T10	13	16	33

Fig. 3: Workflow Scheduling

A similar firefly algorithm was employed in the field of grid computing for the purpose of job scheduling. In grid, this involves combining the firefly algorithm with the max- min algorithm for the purpose of scheduling jobs on the grid [1]. The implementation of this algorithm was essential in a wide complicated field such as grid computing. One of the well-known features of this algorithm is that it automatically maintains load balancing while scheduling jobs in the grid computing environment.

III. Algorithm

The recent development in the field of cloud computing is the occurrence of distributed systems mostly. A distributed system is a collection of components positioned in differently located systems networked together all of which have the necessity of communicating with each other. In distributed systems, the scheduling of workflow gets more and more complex as the sum of jobs needed to be scheduled goes on increasing. As scheduling of workflow is more significant to process jobs, there need to be some kind of algorithm to schedule the workflow. Also the workload has to be evenly distributed, so that the execution time of the processes is shortened as much as possible. There have been number of algorithms designed for the purpose of scheduling workflow in various fields of computing.

Here in we represent a meta-heuristic algorithm called as the firefly algorithm for the purpose of workflow scheduling. This algorithm enfolds all possible techniques to schedule the workflow in the most efficient manner possible. Prior to the explanation of the algorithm it is necessary to get to know how the algorithm evolved.

This firefly algorithm has been designed based on the inspiration on the swarm behavior of fireflies. Fireflies are generally known to exist as groups and they are said to have a swarm kind of behavior. The blinking light in the fireflies is their attribute of attractiveness mainly used for the purpose of attracting mates and to defend themselves from other predators. The swarm of fireflies usually moves in the direction of the brightest one. All the other fireflies

with lower light intensities move toward the ones with higher light intensities. So as the distance between the fireflies goes on increasing, the light intensity also increases.

A. Firefly Algorithm: Proposed Algorithm

```
Begin
   1) Objective function: f(x),   x = (x_1, x_2, ..., x_d);
   2) Generate an initial population of fireflies x_i   (i = 1,2,...,n);.
   3) Formulate light intensity I so that it is associated with f(x)
      (for example, for maximization problems, I ∝ f(x) or simply I = f(x);)
   4) Define absorption coefficient γ

   While (t < MaxGeneration)
      for i = 1 : n (all n fireflies)
         for j = 1 : n (n fireflies)
            if (I_j > I_i),
               move firefly i towards j;
               Vary attractiveness with distance r via exp(−γ r);
               Evaluate new solutions and update light intensity;
            end if
         end for j
      end for i
      Rank fireflies and find the current best;
   end while

   Post-processing the results and visualization;

end
```

Following characteristics of the fireflies are used for the development of FA.

- All fireflies are of same sex, so that firefly will be interested in all other fireflies;
- Brighter is the flash, more will be the attractiveness and with distance, the attractiveness and brightness both decreases.
- If there are no fireflies less large than a given firefly, it will shift arbitrarily. In other words brightness of a firefly is affected or determined by the landscape of the objective function.

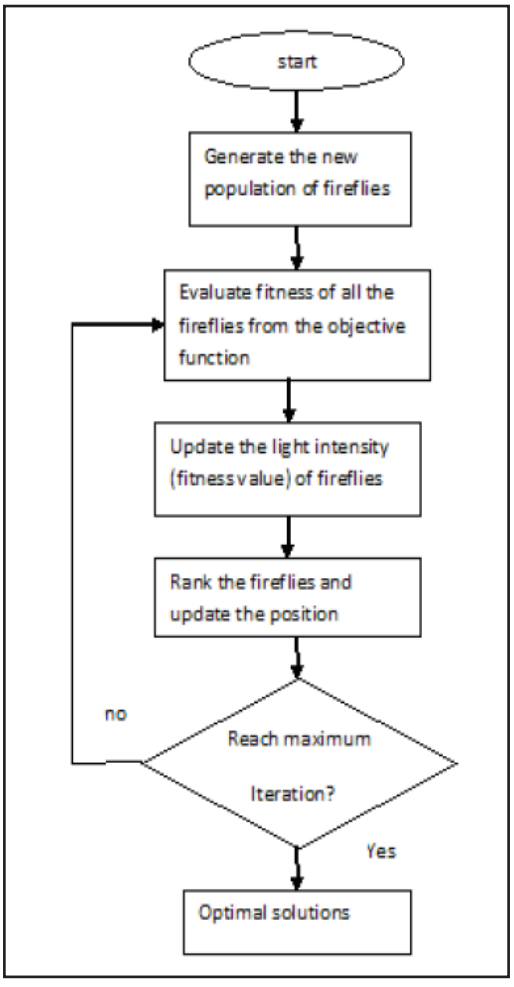


Fig. 4: Flow Chart for Firefly Algorithm

According to the given flow chart first we initialize the number of data centers, virtual machines and cloudlets. After initialization define the population of firefly to generate the new solution. Evaluate the fitness value of the new solution. Choose the virtual machine and compare it with the fitness value. Make rank for a best solution for the precise fitness value and find the current best value. It will give the optimal solution.

IV. Implementation

The firefly algorithm is implemented on the necessity of scheduling of workflow where the execution time is needed to be reduced in the most possible effective manner. Here in the algorithm, the fireflies are disguised as Virtual machines. Each firefly is compared to a virtual machine. The swarm of fireflies move towards the brightest one and in the very similar way, the jobs is assigned to the best virtual machine.

The fitness factor of each firefly is calculated and also the brightness is determined. The attractiveness of each firefly depends on the brightness of that firefly. The distance between each firefly is determined and update after each iteration. Our implementation is that all the fireflies' move toward the brightest one and each iteration the position of each firefly changes which in turn changes the distance between each firefly and also from the brightest one. So the distance is update after each iteration. The fireflies are ranked based on the fitness function and the over-all global best is determined.

Table 1: Typical Fireflies

	T1	T2	T3	T4	T5
Firefly1	R2	R5	R3	R2	R1
Firefly2	R1	R4	R5	R3	R2
Firefly3	R1	R3	R5	R4	R2

When it comes to scheduling jobs on the virtual machines, the best 957 virtual machine at the time of execution is selected and jobs are allocated to that selected virtual machine. At the time of searching the best one, the one with the best specification is selected and the queued jobs are allocated to the virtual machine.

V. Working Methodology

A. Initialize Number of Data Centers, Virtual Machines and Cloudlets

As mentioned in the past area inhabitants, variety nests, fitness and health operate, place of the nodes, finding possibility. The first step in firefly is interpreting inhabitants and information reflection. The inhabitants are initialized by a vector, in which the duration of vector indicates the number of sources.

B. Generation of New Solution

The objective is to the way to apply the fresh and possibly improved alternatives headed for substitute a not so excellent remedy in the light intensity.

C. Fitness Value

Health and fitness or high quality, value reveals how fit the remedy is, i.e. how well it adjusts to its atmosphere. For a maximization issue, the fitness of a remedy can be proportionate to the value of

the purpose operate. For convenience, we assume that each egg in a home symbolizes a remedy, and a cuckoo egg symbolizes new remedy.

Total Fitness value= sum of VMn (1)

Individual

$VM1, VM2 \ldots VMn = \sum_{k=1}^{n} VM1/VMn$ (2)

Distance = maximum VM –individual VM (3)

Where,

Sum of VMn =total number of VM-RAM

D. Choose the Virtual Machine and compare Fitness

The task, arranging in reasoning processing is a well known NP-hard issue. The issue is even more complex and challenging when the fertilized groups are used to perform a huge number of projects in the reasoning processing system.

E. Ranking Fireflies and Update the Best Solution

Rank the best solution with the precise fitness value from the above process and update its best solution according to the virtual position.

VI. Conclusion

As computing systems are becoming more complex and more distributed systems are evolving, the need to schedule workflow is also required mostly. This algorithm is expected to possibly reduce the execution time by efficient scheduling of workflow and thus allocation of jobs even in complicated systems can be effectively done. This algorithm better serves the purpose of workflow scheduling than other swarm intelligence algorithms. It stands unique in its way that it schedules mainly based on execution time while other such algorithms care only about execution cost which may not be efficient as much as the Firefly algorithm.

References

[1] Existing computational models and heuristic techniques for cloud scheduling problems 2015
[2] Analytical review of three nature inspired algorithms for scheduling in cloud computing 2016
[3] "A Particle Swarm Optimization-based Heuristic for Scheduling Work flow Applications in Cloud Computing Environments", September 2012. Suraj Pandey, LinlinWu, Siddeswara Mayura Guru, Rajkumar Buyya
[4] "Workflow Scheduling in Cloud Computing Environment Using Cat Swarm Optimization" Das, 2014 IEEE. Saurabh Bilgaiyan, Santwana Sagnika, Madhabananda
[5] X.-S.Yang,Nature-inspired Metaheuristic Algorithms, Frome, UK: Luniver Press, 2008.
[6] X.-S. Yang, "Firefly algorithms for multimodal optimization", in: Stochastic Algorithms: Foundations and Applications, SAGA 2009, Lecture Notes in Computer Sciences, Vol. 5792, pp. 169-178 (2009).
[7] Sadasivam, D.G.S., "An Efficient Approach to Task Scheduling in Computational Grids." International Journal of Computer Science and Applications, 2009
[8] "Load balancing of real-time services for cloud Computing", May 2014