Review on Meta-heuristic Scheduling Optimization Techniques in Heterogeneous Clouds

Shubhdeep Kaur Sandhu, Anil Kumar (Asst. Professor)

CET Department, GNDU, Amritsar,

sandhu.shubh92@gmail.com

Abstract— As cloud computing is turning out to be evident that the eventual fate of the cloud industry relies on interconnected cloud systems where the resources are probably going to be provided by various cloud service suppliers. Clouds are also seen as being multifaceted; if the user requires only computing capacity and wishes to personalize it as per his requirements, the infrastructure cloud suppliers are able to provide this convenience as virtual machines. Many optimized meta-heuristic scheduling techniques are introduced for scheduling of bag-of-tasks applications in heterogeneous framework of clouds. The overall analysis demonstrates that, utilizing different meta-heuristic techniques can offer noteworthy benefits in the terms of speed and performance.

Keywords— Bag-of-tasks, Heterogeneous clouds, Meta-scheduling, Meta-heuristics, Simulated Annealing, Tabu Search, Multi-criteria Decision Making

1. INTRODUCTION

1.1 Cloud Computing

Within the course of the past couple of years, cloud computing has come forth as a standout amongst other solutions for delivering IT oriented services to the clients. It is the novel concept with the help of which services are distributed amongst consumers and providers after identifying the customer demands and sandboxing their requirement in virtualized settings [12]. From the infrastructure point of view, Cloud Computing is propitious resolution that extends the resource capacity of independent computing systems dynamically. Cloud computing is analogous to Grid computing in the manner that it also deploys the distributed resources to attain application-level targets [8]. Its proficiency to leverage virtual technologies at the hardware level as well as application level in order to recognize the properties of sharing the resources, providing dynamic resource scaling "on-demand" while offering a flexible price framework in conjunction with ease of modification and high availability makes it superior to the Grids. On the other hand, with the help of utility based price frameworks and on-demand resource as well as service provisioning, service suppliers can maximize the resource utilization along with minimization of operational cost. A service provider does not need to offer capacities in accordance with the peak load anymore, which results in magnificent savings when the resources are set free to save operational costs in case service request is reduced [8].

1.2 Inter-cloud Systems

The term "inter-cloud" implies an interoperable environment in which multiple criteria collude to satisfy QoS levels [12]. Once the multiple clouds are interlinked together, different clouds provide dissimilar architectures and varying resources which are consolidated into a single entity in a transparent manner [16]. Inter-cloud intends to expand the service elasticity of cloud and scalability while minimizing the performance and service cost overheads [15]. Inter-cloud systems support dynamic workload supervision to initiate decision making for job distribution at meta-brokering level. Inter-cloud meta-broker is built to be decentralized and dynamic by improving the way choices are made for service distribution [12]. This can be carried out through the use of heuristic criterion and algorithms to achieve improved meta-scheduling in inter-cloud environments. In each scheduling decision, percentage of required resources is ought to be reconfigured, displacing them to an alternate cloud region. This course of action causes some virtual machines to be paused for a short time period, which in turn can cause performance degradation temporarily [10].

1.3 Meta-scheduling Paradigms

We pay attention to performance optimization using meta-scheduling paradigm to attain a much better job scheduling across multiple clouds. When numerous distinct clouds are merged, a multi-layered technique is needed that ought to have a universal scheduler, which manages the allocation of jobs amongst the clouds in addition to the ones that are local cloud schedulers [16]. The meta-broker invokes the scheduler sporadically that allows optimization of entire infrastructure cost dynamically by placing some VMs to the most inexpensive cloud [6]. The nature of jobs being processed is a crucial aspect of multi-layered model [16]. The conventional parallel and distributed systems could capture only a single characteristic of jobs to be scheduled in the real workload. But in the realistic workload of the modern parallel systems, apart from the fact that they are distributed identically and independently, the workload is identified by other significant features like burstiness (temporal as well as spatial), long range dependence in the method of job arrival and bag-of-tasks behavior [5].

1.4 Bags-of-tasks
The inherent extensive dissemination of heterogeneous and dynamic nature of clouds induces them to be more suited to execute the loosely coupled parallel applications like BoTs. These embarrassingly parallel tasks can be executed on any processor and have the ability of scaling out, but do not facilitate the inter-task communication [11]. According to the definition According to the definition proposed in [5], each of the jobs within a BoT can have the identical credentials like group name, queue name, user name, user approximate runtime, which makes it evident to assume that all the jobs within same BoT are considered to have comparable runtime. Due to environmental heterogeneity, tasks belonging to same BoT can have different completion times [13]. A part of jobs arrived at the local level are also crucial and are required to be scheduled with precedence much higher than the remaining jobs. The distinct permutations of a respective schedule delegating BoT tasks to various virtual machines assist to form the search area of the problem [16].

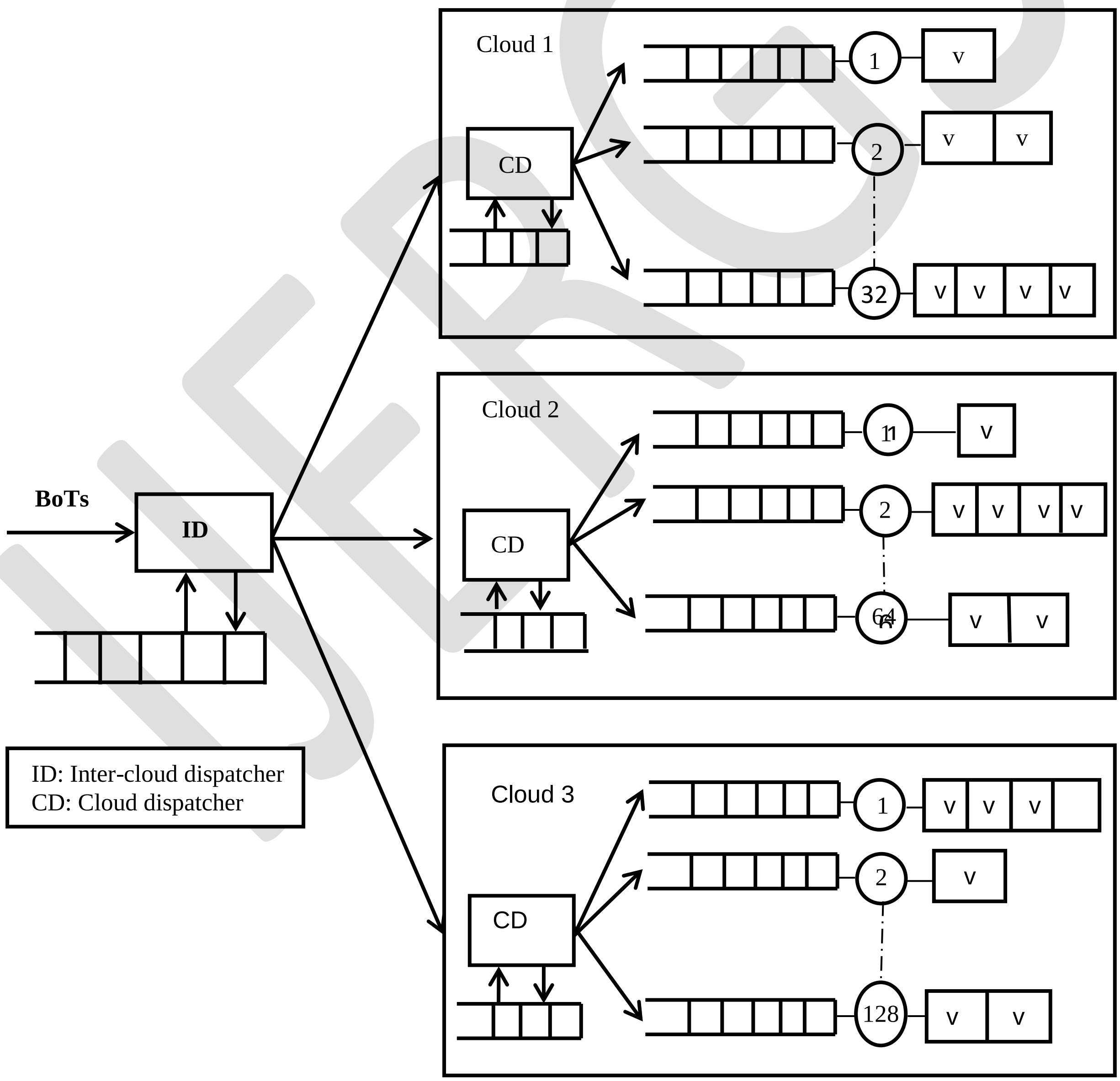

Figure 1. Inter-cloud System

1.4 Meta-heuristic Techniques

Meta-heuristics based techniques can be easily hybrid with any auxiliary approach available to solve non-linear optimization problems. The main aim behind the hybridization of different algorithmic ideas is to fully exploit and combine the advantages of individual pure strategies to get the better performing systems. Multi-criteria decision making problems are special case of vector optimization problems. Scheduling is conceived as a multi-objective task because we use multiple criteria for the evaluation of the quality solutions by minimizing two or more conflicting objectives instead of taking only one objective into the account. Herein, the main motive is to choose a trade-off among all the feasible solutions. In order to select the best suited alternative amongst the available ones, every solution is measured according to more than a single objective function, each of which must be maximized or minimized.

2. TECHNOLOGIES USED

2.1 Meta-heuristic Techniques

Meta-heuristics are considered to be the generic methods that provide good solutions, global optimum within a genuine computation time [7]. They mimic the natural metaphors to solve complex optimization problems such as annealing process, particle swarm, bee colony, artificial bee colony. In other words, meta-heuristic is the upper level approach that is used to guide the underlying heuristics to solve specific problems [7]. They direct the search through the solution space, using substitute algorithms as some form of heuristic, usually local search that can formulate the problems to find a solution maximizing a criterion among a number of candidate solutions [2]. Meta-heuristics customize the operations of supporting heuristics to generate higher quality results efficiently, optimizing both performance and cost while considering heterogeneity of virtual machines [14]. The different meta-heuristic algorithms comply separate procedures for multi-criteria scheduling of loosely coupled parallel jobs named, BoTs in multiple clouds. They are known to be the iterative master processes that improve the solutions at each step until a forbidden criteria is met. Two contrasting criteria must be taken into consideration while designing a meta-heuristic exploration of the search space which is referred to as diversification, and exploitation of the best solutions found, termed as intensification. Meta-heuristics are extended to hybridized versions of variant algorithms [7]. Hybridization of different algorithmic abstractions aims at obtaining more effective systems that exploit the merits of respective classic strategies.

(a) Simulated Annealing

It is one of the earliest meta-heuristic techniques and is motivated by the physical annealing process that establishes the link between its thermodynamics and hunts for global minima in discrete optimization problem [16]. The fundamental characteristic of Simulated Annealing is that it allows an effective approach to escape local optima with sharp probability and time variations by permitting the hill climbing moves hoping to discover global optimum [1]. Simulated annealing refers to the process used in metallurgy in which physical substances are elevated to a higher degree of energy and after that they are gradually cooled until metal alloys are typically in solid state. At each step, a neighbor state is determined by using a neighboring function. The choice of relevant neighborhood ends up being significant for the quality of the outcomes and has probably enormous effect on the quality of SA algorithm [14]. The system can either remain at the current state or move to the next one. Right here, simulated annealing makes the usage of virtual cooling schedule that defines the temperature drop. It figures out if a “worse” move to a favorable machine be accepted, searching for a global optimal solution. As the temperature falls, it becomes hard for the “worse moves” or moves towards high energy states, to be accepted, but the system always accepts the moves to the neighbors having lesser energy. In due course, when the temperature becomes very low, the algorithms being greedy, starts carrying out down-hill moves [16].

Algorithm:

Inputs = x_0, $d_{max.}$

Outputs = x_{best}

$x = x_0$, $g = G(x)$

$x_{best} = x$, $g_{best} = g$

$d = 0$

while $d < d_{max}$ do

    T = temperature (d/d_{max})

    x' = nbr (x)

    $g' = G(x')$

if $P(g, g', T) >$ uniform (0,1) then
x = x', g = g'
if $g < g_{best}$ then
$x_{best} = x$, $g_{best} = g$
end if
end if
d+ = 1
end while
return x_{best}

Although the main loop of the given algorithm is moderately enough to be applied, still there are some other functions which can be modified according to each problem. Such functions are:

i. G = It represents the energy function that computes the energy of the given state.
ii. P ( ) = It is the probability function which figures out if the moves ought to be acknowledged.
iii. nbr ( ) = It gives the neighbors of a given state.
iv. temperature ( ) = It evaluates the cooling schedule.

(b) Tabu Search

Tabu search makes the use of memory constructs to prohibit those states of search space which have already been visited [16]. TS algorithm uses a mathematical function that analyses how much a chosen solution satisfies the desired measures. This function considers a set of numerous possible moves at each stage that are neighbors of the current state. Each time TS is implemented, it may use one or multiple number of memory structures which uphold the lists of states that have either already been visited or are forbidden on the basis of criteria defined by the user [3]. This list of forbidden moves is known as "tabu list" and cannot be expanded beyond the given maximum size. So, it's considered to be expired when they reach the maximum size. Once it expires, tabus are removed in First in First out (FIFO) manner [16]. TS has capability to get high quality solutions with modest computational efforts. Sometimes, tabus may forbid fair moves even if there is no danger of cycling and in addition they may result in the stagnation of the search process [2].

Algorithm:
Inputs = x_0, d_{max}, n_{max}
Outputs = x_{best}
tblist = [ ]
$x = x_0$, g = G (x)
$x_{best} = x$, $g_{best} = g$
d = 0
while $d < d_{max}$ do
ngbrs = [ ]
while nbrs.size $< n_{max}$ do
x_{temp} = nbr (x)
if not x_{temp} in tblist then
nbrs += x_{temp}
end if
end while
x = choose_best (nbrs)
g = G (x)
if $g < g_{bes}$t then
$x_{best} = x$, $g_{best} = g$
tblist+ = x
expire (tblist)
end if
d + = 1
end while

Apart from the functions used in Simulated Annealing, Tabu Search deploys two more functions which are as follows:

i. choose_best: It is used to select the best solution from the respective candidates.

ii. tblist and expiration function: The tabus gets expired when the tabu list hits the maximum size.

3. RELATED WORK

Zheng, Shu and Gao et al. 2006 [1] suggested the merging of the merits of simulated annealing and genetic algorithm and came up with a parallel genetic simulated annealing that is employed in order to resolve the crucial challenge of scheduling in grid computing. The algorithm generated the new group of individuals and afterwards simulated annealing normalized all the generated individuals independently. The result provided the overall optimal solution and proposed algorithm is proved to be better than pure Simulated Annealing and Genetic Algorithm. Fayad, M. Garibaldi and Ouelhadj et al. 2007 [2] formulated a scheduling algorithm with an aim to maximize the number of scheduled jobs utilizing Tabu Search to resolve the problem of grid scheduling by determining the optimal solutions. Fuzzy technology became active in this application by supporting the usage of fuzzy sets so that processing times of jobs, patterned with uncertainty could be represented. The algorithm was inspected against robustness while processing times of jobs changed by evaluating its performance in crisp modes as well as fuzzy modes. Moreover, the effect of varying shapes of fuzzy completion times and the average job length on the schedule performance was addressed. Xhafa, Carretero et al. 2009 [3] contrived another variant of Tabu Search to attain high performance by resolving an issue of batch scheduling in grid-based applications. This new form of Tabu Search was considered as a bi-objective algorithm meant for minimizing the flow times and makespan of scheduled jobs. For a classical benchmark, the novel tabu search was formalized against three other algorithms. Furthermore, some more realistic benchmarks were taken into consideration with larger size instances in static and dynamic environments and the results showed us that Tabu search exceeded the compared algorithms to a great extent. Lee, Chun and Karzy et al. 2011 [9] proposed a method to reconsider the resource allocation and job scheduling to comprehend the heterogeneity of cloud-based analytics platforms. They suggested architecture for resource allocation to deploy advanced analytics in heterogeneous clusters with the aim to improve performance and reduce cost overheads. A metric scheme was formulated to achieve better performance and fairness amongst jobs when multiple jobs share the cluster. Sotiriadis, Bessis, Antonpoulos et al. 2012 [12] examined that because of increasing the number of users, supervision of the internal resources in a widely distributed environment is a critical matter that needs to be dealt with. A meta-broker approach was conceptualized for inter-cloud frameworks to arrange them in a decentralized manner, facilitating the coordination of multiple cloud brokers to demonstrate the responsive service mechanization. An inter-cloud system was simulated to evaluate the average execution time required for bulky services and it showed efficient performance with this solution. A.Moschakis and D. Karatza et al. 2014 [14] described another way to get the optimized interlinked cloud systems in the terms of better performance-to-cost ratios and reliability so that cloud clients can acquire high accessibility and quality of service demands. This research involved the resource allocation schemes and distribution of tasks, for which manipulation of Simulated Annealing and Thermodynamic Simulated Annealing were examined with the scheduling of dynamic multi-cloud framework accompanying virtual machines offering heterogeneous performance while executing bags-of-tasks. The simulation results illustrated substantial influence of heuristics in sustaining satisfactory cost-performance trade-off. Sotiriadis, Bessis, Anjum and Buyya et al. 2015 [15] canvassed that the technique of inter-clouds alleviate ascendible resource allocation across multiple cloud infrastructure. A new inter-cloud scheduling paradigm, known as “Inter-Cloud–Meta-Scheduling” was ushered in. The consequences of the above mentioned framework demonstrated better flexibility, robustness and decentralization. To design and enforce various entities of clouds and policies in ICMS, a tool-kit called, “Simulating the inter-cloud” (SimIC) was used. For several arguments such as makespan, turnaround and execution times, this experimental desiccation was proved beneficial as it produced improved performance of individual clouds when imparted together beneath ICMS model.

4. COMPARISON TABLE

Sr No.	Authors	Year	Title	Technique	Heterogeneity	Meta-heuristic	Convergence Speed
1	Zheng, Shijue, Wanneng Shu, and Li Gao	2006	Task scheduling using parallel genetic simulated annealing algorithm	Parallel Genetic Simulated Annealing Algorithm	No	Yes	Higher
2	Fayad, Carole, Jonathan M. Garibaldi,	2007	Fuzzy grid scheduling using tabu search	Tabu search	No	Yes	Average

	and Djamila Ouelhadj.						
3	Xhafa, Fatos	2009	A Tabu Search algorithm for scheduling independent jobs in computational grids	Tabu Search	No	Yes	Average
4	Lee, Gunho, and Randy H. Katz.	2011	Heterogeneity-aware resource allocation and scheduling in the cloud	Hetero-geneous cluster scheduling	Yes	No	Average
5	Sotiriadis, Stelios, Nik Bessis, and Nick Antonpoulos	2012	Decentralized meta-brokers for inter-cloud: modeling brokering coordinators for interoperable resource management	Inter-cloud meta-broker scheduling in decentralized manner	Yes	No	Poor
6	Moschakis, Ioannis A., and Helen D. Karatza	2014	Multi-criteria scheduling of bag-of-tasks applications on heterogeneous interlinked clouds with simulated annealing	Simulated annealing	Yes	Yes	Higher
7	Sotiriadis, Stelios	2015	ICMS simulation framework: architecture and evaluation	Meta - scheduling	Yes	No	Higher
8	Moschakis, Ioannis A., and Helen D. Karatza	2015	A meta-heuristic optimization approach to the scheduling of bag-of-tasks applications on heterogeneous clouds with multi-level arrivals and critical jobs	Simulated Annealing and Tabu Search	Yes	Yes	Higher

5. CONCLUSION

This paper represents the cloud computing has potentially revolutionized a huge portion of IT industry, causing software to be more attractive to a greater extent as a service. It shows the comparison on meta-heuristic techniques based on scheduling of bag-of-tasks applications in heterogeneous environment of clouds. They provide various benefits in speed and performance, but still there are some issues related to them. Simulated Annealing does not determine whether it has found the optimal solution. So, another complementary method is always the utmost need for this purpose. Using Tabu Search, complete solutions can be recorded, but it needs huge storage that makes it highly priced to check if a potential move is tabu. To overcome these issues in the future, we will propose a hybrid technique for parallel scheduling using SA and PSO.

REFERENCES:

[1] Zheng, Shijue, Wanneng Shu, and Li Gao. "Task scheduling using parallel genetic simulated annealing algorithm." *2006 IEEE International Conference on Service Operations and Logistics, and Informatics*. IEEE, 2006.

[2] Fayad, Carole, Jonathan M. Garibaldi, and Djamila Ouelhadj. "Fuzzy Grid Scheduling Using Tabu Search." *FUZZ-IEEE*. 2007.

[3] Xhafa, Fatos, et al. "A tabu search algorithm for scheduling independent jobs in computational grids." *Computing and informatics* 28.2 (2009): 237-250.

[4] Buyya, Rajkumar, Rajiv Ranjan, and Rodrigo N. Calheiros. "Intercloud: Utility-oriented federation of cloud computing environments for scaling of application services." *International Conference on Algorithms and Architectures for Parallel Processing*.Springer Berlin Heidelberg, 2010.

[5] Minh, Tran Ngoc, Lex Wolters, and Dick Epema. "A realistic integrated model of parallel system workloads." *Cluster, Cloud and Grid Computing (CCGrid), 2010 10th IEEE/ACM International Conference on*. IEEE,2010.

[6] Papazachos, Zafeirios C., and Helen D. Karatza. "Performance evaluation of bag of gangs scheduling in a heterogeneous distributed system." *Journal of Systems and Software* 83.8 (2010): 1346-1354.

[7] Xhafa, Fatos, and Ajith Abraham. "Computational models and heuristic methods for Grid scheduling problems." *Future generation computer systems*26.4 (2010): 608-621.

[8] Zhang, Qi, Lu Cheng, and Raouf Boutaba. "Cloud computing: state-of-the-art and research challenges." *Journal of internet services and applications* 1.1 (2010): 7-18.

[9] Lee, Gunho, and Randy H. Katz. "Heterogeneity-Aware Resource Allocation and Scheduling in the Cloud." *Hot Cloud*. 2011.

[10] Simarro, Jose Luis Lucas, et al. "Dynamic placement of virtual machines for cost optimization in multi-cloud environments." *High Performance Computing and Simulation (HPCS), 2011 International Conference on*.IEEE, 2011.

[11] Farahabady, M. Hoseiny, Young Choon Lee, and Albert Y. Zomaya. "Non-clairvoyant assignment of bag-of-tasks applications across multiple clouds." *2012 13th International Conference on Parallel and Distributed Computing, Applications and Technologies*. IEEE, 2012.

[12] Sotiriadis, Stelios, Nik Bessis, and Nick Antonpoulos. "Decentralized meta-brokers for inter-cloud: Modeling brokering coordinators for interoperable resource management." *Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on*. IEEE, 2012.

[13] Netto, Marco AS, and Rajkumar Buyya. "Coordinated rescheduling of Bag-of-Tasks for executions on multipleresource providers." *Concurrency and Computation: Practice and Experience* 24.12 (2012): 1362-1376.

[14] Moschakis, Ioannis A., and Helen D. Karatza. "Multi-criteria scheduling of Bag-of-Tasks applications on heterogeneous interlinked clouds with simulated annealing." *Journal of Systems and Software* 101 (2015): 1-14.

[15] Sotiriadis, Stelios, et al. "An inter-cloud meta-scheduling (icms) simulation framework: Architecture and evaluation." (2015).

[16] Moschakis, Ioannis A., and Helen D. Karatza. "A meta-heuristic optimization approach to the scheduling of Bag-of-Tasks applications on heterogeneous Clouds with multi-level arrivals and critical jobs." *Simulation Modelling Practice and Theory* 57 (2015): 1-25.