ISSN:0975-9646

Task Scheduling in Cloud Computing: Review

Raja Manish Singh[#1], Sanchita Paul[*2], Abhishek Kumar [#3]

[#]*M.E, Research Scholar,*
Department of Computer Science and Engineering,
Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India.
[*]*Assistant Professor,*
Department of Computer Science and Engineering,
Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India.

***Abstract*— Task scheduling plays a key role in cloud computing systems. Scheduling of tasks cannot be done on the basis of single criteria but under a lot of rules and regulations that we can term as an agreement between users and providers of cloud. This agreement is nothing but the quality of service that the user wants from the providers. Providing good quality of services to the users according to the agreement is a decisive task for the providers as at the same time there are a large number of tasks running at the provider's side.**

The task scheduling problem can be viewed as the finding or searching an optimal mapping/assignment of set of subtasks of different tasks over the available set of resources (processors/computer machines) so that we can achieve the desired goals for tasks. In this paper we are performing comparative study of the different algorithms for their suitability, feasibility, adaptability in the context of cloud scenario, after that we try to propose the hybrid approach that can be adopted to enhance the existing platform further. So that it can facilitate cloud-providers to provide better quality of services.

***Keywords*— Cloud Computing, Cloud Architecture, Task Scheduling, Scheduling Types, GA, PSO**

I. INTRODUCTION

There is no exact definition of cloud but we can define cloud in various ways and by considering various means. Cloud computing is Internet-connected mode of supercomputing. It is a type of shared infrastructure, which simply puts the huge system pools together by using various means; distributed, virtualization etc. It gives users a variety of storage, networking and computing resources in the cloud computing environment via Internet, users put a lot of information and accesses a lot of computing power with the help of its own computer.

According to R.Buyya that defines the cloud as " Cloud is a parallel and distributed computing system which basically consist of a collection of inter-connected and virtualized computers that are provisioned dynamically and presented as one or more than one unified computing resources based on service-level agreement (SLA) established through negotiation between the service providers of cloud and users [1].

Cloud computing is a large-scale distributed computing model, which depends on the economic size of the operator of cloud that is abstract, virtualized and dynamic. The main content of cloud computing is to manage computing power, storage, various kind of platforms and services which assigned to the external users on demand through the internet. Cloud computing is a rapidly emerging computation paradigm with the goal of freeing up users of cloud from the management of hardware, software, networks and data resources and shifting these burdens to cloud service providers[2].

Clouds provide a very large number of resources, including platforms for computation, data centers, storages, Networks, firewalls and software in form of services. At the same time it also provides the ways of managing these resources such that users of cloud can access them without facing any kind of performance related problems.

Cloud Computing Services are divided into three classes, according to the abstraction level and the service model of providers, namely: (1) Infrastructure as a Service, (2) Platform as a Service, (3) Software as a Service.

Distribution, virtualization and elasticity are the basic characteristics of cloud computing. Virtualization is one of the main features of cloud. Most of the software and hardware have provided support to virtualization. We can perform virtualization on many factors such as hardware, software, storage and operating system, and manage them in cloud platform.

II. RELATED WORK

Previous research on task scheduling problem includes as the NP-hard [3] problem. As the scheduling is NP-hard we can say that it is basically handled by various known heuristic methods which give solutions for problem instances basically in restricted manner. A heuristic is nothing but an algorithm which guarantees for finding an approx optimal solution in fewer than the polynomial time. It searches for finding a path in the available solution space by ignoring and eliminating the other paths that may seem possible [4]. The algorithms which are based on heuristic can also be classified into: 1.Cluster Scheduling 2.List Scheduling.

Clustering algorithms [5], [6], [7], assumes that for the execution of sub-tasks there are large numbers of processors available for doing the work. As it assumes that a large number of processors are available for work so, it uses as many processors as possible for reducing the makespan of the generated schedules.

Most of the heuristics comes under the list scheduling class. The searching under list scheduling algorithm can be divided under two separate phases: first phase, each subtask gets a priority according to some criteria and after setting

priority they are added to the list of task which is waiting for their turns in order to decreasing priority. Second phase, when the different processors becomes available for serving more, the highest priority subtask is selected from the list and assigned to the processor which is most suitable. So, here searching is narrowed down to small portions of solution space by using some other heuristics; greedy heuristics, thus it gives a better makespan. A drawback is that is not suitable for giving consistent results for heterogeneous computing systems.

In addition of these heuristics based algorithms, meta-heuristics based algorithms apply a combinatoric process in finding the solution. These algorithms require sufficient sampling of solutions in given search space and have shown best performance over a number of scheduling problems. Particle Swarm optimization [8], Simulated Annealing [9], Genetic Algorithms [10], and Ant Colony Optimization [11] have been successfully and effectively applied to various scheduling problems. Only drawbacks with these algorithms than those whose based on heuristics is that, these generate higher computational cost and less efficient.

III THE ARCHITECTURE OF CLOUD

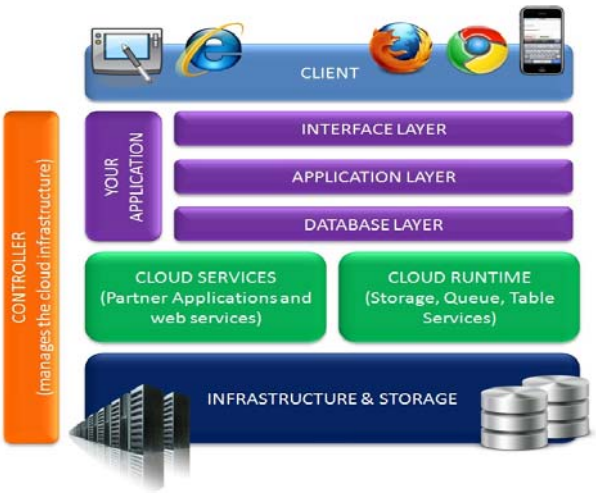

Fig.1 Layered View of Cloud Architecture

Cloud computing architecture has two most significant components that are basically known as front and back end. Front end is that part of cloud which is visible to the user of cloud. This include applications and computer that user uses to access the cloud. Back end of cloud computing is nothing but the cloud itself, which mainly comprises of computers, storage devices.

Cloud Environment basically consists of software applications that are accessed via internet as services when wishes to use them. Applications that are built on Cloud Architectures are such that the underlying infrastructure of computing is used only when it is actually needed, draw the necessary resources when someone make a demand for those, perform a particular job, then relinquish the unneeded resources and often dispose them after the job is completed. During their operation the applications scale up or down elastically based on need of resources.

Cloud Architectures address key difficulties that mainly related to processing of large amount of data. In traditional way of processing the data it is quite difficult to get as many machines that an application needs for its operation to complete. It is one of the most difficult things to get the machines when one/particular needs them. It is really a tough work to distribute and co-ordinate a large-scale job on different-different machines, run processes on them, and provisions another machine to recover if one machine fails during the operation. It is also difficult to auto-scale up and down based on changing nature of workloads. It is difficult to get rid of all those machines when the job is completed.

Applications that built over Cloud Architectures run in-the-cloud where the actual physical location of the infrastructure is determined by the cloud providers. They simply take advantage of simple APIs of Internet-accessible services that are able to scale on-demand, where the complex logic of scalability and reliability of the services remains implemented and hidden inside the vast cloud.

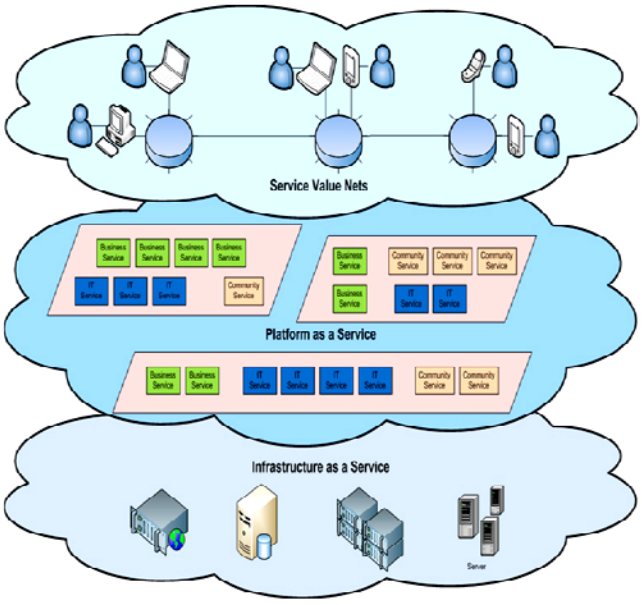

Fig.2. Service Layers of Cloud

IaaS is the lower most layers of the cloud computing systems and it basically provides virtualized resources ex: computation, communication and storage that are available on demand. PaaS makes the cloud easily programmable. SaaS can be term as the software delivery model.

IV. OVERVIEW: TASK SCHEDULING

Cloud consists of a number of resources that are different with one other via some means and cost of performing tasks in cloud using resources of cloud is different so scheduling of tasks in cloud is different from the traditional methods of scheduling and so scheduling of tasks in cloud need better attention to be paid because services of cloud depends on them.

Task scheduling plays a key role to improve flexibility and reliability of systems in cloud. The main reason behind scheduling tasks to the resources in accordance with the given time bound, which involves finding out a complete and best sequence in which various tasks can be executed to give the best and satisfactory result to the user. In cloud

computing, resources in any form i.e. cups, firewall, network are always dynamically allocated according to the sequence and requirements of the task, subtasks. So, this leads task scheduling in cloud to be a dynamic problem means no earlier defined sequence may be useful during processing of task. The reason behind the scheduling to be dynamic is that because flow of task is uncertain, execution paths are also uncertain and at the same time resources available are also uncertain because there is a number of tasks are present that are sharing them simultaneously at the same time.

The scheduling of tasks in cloud means choose the best suitable resource available for execution of tasks or to allocate computer machines to tasks in such a manner that the completion time is minimized as possible. In scheduling algorithms list of tasks is created by giving priority to each and every tasks where setting of priority to different-different tasks can be based on various parameters. Tasks are then chooses according to their priorities and assigned to available processors and computer machines which satisfy a predefined objective function [12].

A. Scheduling Types

1) Static scheduling schedule tasks in known environment i.e. it already has the information about complete structure of tasks and mapping of resources before execution, estimates of task execution/running time.

2) Dynamic scheduling must depend on not only the submitted tasks to cloud environment but also the current states of system and computer machines to make scheduling decision.

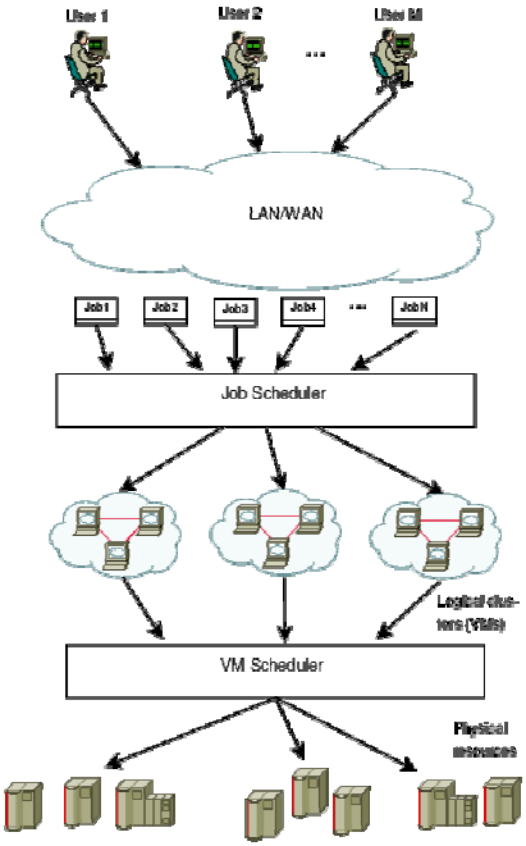

Fig.3. General View of Task Scheduling

Cloud computing uses virtualization technique for mapping the resources of cloud to the virtual machine layer, implement the user's task, so the task scheduling of cloud computing environment achieve at the applications layer and the virtual layer of resources. Scheduling is nothing but the mapping of tasks and resources in accordance with some certain principles for achieving the desired goal. Cloud computing paradigm simplifies the mapping of tasks to resources; the required resources together form to be virtual machines (VMs), the process of search the desired resource package is same as the process of searching the various VMs.

V. STUDY OF EXISTING TASK SCHEDULING ALGORITHMS

A. Deadline and Budget Distribution based Cost-Time Optimization Algorithm [13]

This approach considered two constraints: deadline of executing the tasks and budget. This paper proposed (DBD-CTO) scheduling algorithm to schedule tasks in cloud computing environment. This algorithm achieve its goal or target by finishing the execution of tasks before the given deadline as well as at the same time minimizes the cost of computation that occur during execution.

B. Improved cost-based algorithm for task scheduling in Cloud Computing Environment[14]

The author of this paper proposed the approach which is known as improved cost-based scheduling algorithm. The main objective of his work is to schedule groups of task in cloud computing platform, where resources are having different resource costs and different computation performance. When grouping of jobs is done, communication between jobs and resources optimizes computation/communication ratio. This algorithm measured performance of computation and cost of resources. This also increased the execution of tasks / transfer of data between tasks ratio by combining various tasks during execution. The process of combining task is usually done by after analysing the capability of different available resource and its processing. CloudSim has been used for performing the simulation and the inputs of the simulation are: average MI of tasks, granularity size of tasks, total number of tasks and task overhead time. Result of his work shows that for this particular algorithm time taken to complete tasks after grouping of tasks is very less as compared to when grouping is not done.

C. A PSO-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments [15]

The paper gives us idea about particle swarm optimization based algorithm. Here scheduling of tasks considers both computation cost and data transmission cost. The author mainly focuses on minimizing the total cost of executing the applications on resources that is provided by the cloud service providers. Work that is performed in this paper can be divided into two parts: 1. Formulation of a model for task-resource mapping to minimize the overall execution cost. 2. Designing a heuristics that uses PSO for solving task-resource mappings based on the above model. The proposed algorithm considers communications costs of

all the tasks, including various dependencies between them. So when a comparison is made in terms of the cost savings with the 'Best Resource Selection' (BRS) algorithm, this PSO based algorithm achieved better performance over BRS with three times savings of cost as well as good distribution of workload onto resources. The heuristics that is proposed in this paper is as generic that it can be used for any number of tasks by increasing the particle dimensions and the resources number.

D. Multi-Objective Task Assignment in Cloud Computing by Particle Swarm Optimization [16]

Here author of this paper performs scheduling of tasks by keeping in mind not only minimizing the total cost of transferring and executing, but also minimizing the total transferring time and the time of execution. In this paper the problem of task scheduling described as the assigning all the data of task to all resources in cloud computing environment makes the total cost and time of processing and communication to minimize. Here task scheduling can be regarded as the mapping which maps all the tasks to a Directed Acyclic Graph (DAG) G (V, E). Each node has a weight which mainly denotes the amount of data to be performed by a task on a special node. Each edge has a weight which denotes the amount of information from one task generating to another task to be dealt with. This algorithm objective is not the one but includes processing and transferring time, processing and transferring cost as well. Experiment result shows that this algorithm is efficient in decreasing all types of cost and time in cloud computing environment.

E. Bi-Criteria Priority based Particle Swarm Optimization Workflow Scheduling Algorithm for Cloud (BPSO) [17]

This paper develops an approach for scheduling workflow tasks over the available resources of cloud that minimized the execution time and execution cost under the given deadline and budget constraint. In, BPSO the workflow tasks are executed in order of their priority which is basically computed using bottom level which is same as that is defined in HEFT [20]. The assigned priority is then used to initialize the PSO. After assigning the priority the tasks are sorted according to the descending order of bottom level. The tasks are then sending to different processors according to their order of execution for completing the workflow application. For performing the experiment author developed a simulation program in java which consists of a data center which includes six resources with different processing speed. Experiment results shows that this algorithm has a promising performance when compared with PSO.

F. Independent tasks scheduling based on Genetic Algorithm [18]

In this paper author introduces GA in scheduling tasks to adapt to memory constraints and request of high performance in cloud environment. The author made various assumptions about tasks: 1. Tasks are a periodic. 2. Tasks are non-preemptive. 3. Each task has two ways of access of a process unit i.e. shared and exclusive. This paper also assumes the centralized scheduling scheme [21, 22]. The algorithm that had been developed in this paper synthetically considers both recourse and time utilization so the result that is obtained consists high satisfaction

G. Genetic Simulated Annealing Algorithm for Tasks Scheduling [19]

The paper considered the value of both Genetic Algorithm and Simulated Annealing and develops an algorithm. The algorithm considered the QOS requirements of different type of tasks, corresponding to the nature of tasks of user in cloud. QOS is nothing but a standard of user's satisfaction towards the services. This paper says that as the parameters' dimensions are different and orders of magnitude are very different so, in order to evaluate the implementation of tasks in cloud resources by these parameters, the various parameters must be dealt with dimensionless. According to type of the tasks, different weight of parameters can be given to find resources that satisfy the QOS of the tasks and their expectations. Algorithm steps first execute steps of GA after then annealing comes which helps improving local search ability of GA. This algorithm efficiently completes the searching of resources and allocation process in cloud computing.

TABLE I SUMMARY OF ALGORITHMS

METHOD USED IN ALGORITHM	FACTOR CONSIDERED	ADVANTAGES	TOOL USED
DBD-CTO algorithm [13]	Cost, Time	It lowers the cost of computation and completes task in given time boundary.	Java Environment
Improved Cost-Based Task Scheduling Algorithm [14]	Performance, Cost	It measures resource cost as well as computational performance also improves (computation /communication) ratio.	Cloud Sim
A PSO-based Heuristic for Scheduling Workflow Applications [15]	Cost of computation , Cost of data transmission	It gives three times cost saving as compare to BRS and also balances the load on resources by distributing tasks to available resources.	JSwarm package
Multi-Objective Task Assignment in Cloud Computing by Particle Swarm Optimization [16]	Processing and Transferring time, Processing and Transferring cost	It not only optimizes the time, but at the same time optimizes the cost also.	Matlab R2009b
Bi-Criteria Priority based Particle Swarm Optimization [17]	Execution time and Execution cost	It minimizes the execution cost while meeting the budget and deadline constraint.	Java Environment
Independent Task Scheduling Based on GA [18]	Execution Time	Consider resource and time utilization.	CloudSim
Genetic Simulated Annealing Algorithm [19]	QOS Parameters, Cost	Considers the QOS requirements of different user tasks.	Java Environment

VI. FUTURE SCOPE

Task scheduling is one of the most famous problems in cloud computing so; there is always a chance of modification of previously completed work in this particular field. The researchers at their own time performed their work according to their knowledge space and after some time their work had been carried out some other people. During scheduling they had considered various techniques and applied constraints but as the cloud computing is too vast that they had not been able to capture all aspects at the same time but they mentioned these facts that there is a chance of modification of algorithms and which part has to be modified.

TABLE III.
ALGORITHMS AND FUTURE SCOPE

Algorithm	Future Enhancement
Deadline and Budget Distribution based Cost-Time Optimization Algorithm	Need to be extended for real-time applications.
Improved cost-based algorithm	Extend to handle more complex scenario involving dynamic factors such as dynamically changing cloud and other various QoS attributes.
A PSO-based Heuristic for Scheduling Workflow Applications	Extend to support scheduling workflows of real applications such as brain imaging analysis.
Bi-Criteria Priority based Particle Swarm Optimization	Can be extended by considering existing load of resources during mapping of task to a particular resource.
Independent tasks scheduling based on Genetic Algorithm	Work can be done in future for reducing the solution space in GA to make it more effective.
Genetic Simulated Annealing Algorithm for Tasks Scheduling	Can be modified to support QoS parameters that having different dimensions and also having different orders of magnitude.

REFERENCES

[1] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, Cloud Computing and emerging IT Platforms: Vision, hype, and reality for delivering computing as the 5th utility, Future Generation Computer Systems, 25:599-616, 2009.

[2] Brian Hayes, Cloud computing, Communications of the ACM, Volume 51, Issue 7, July 2008.

[3] M. R. Garey and D. S. Johnson, Computer and intractability: a guide to the theory of NP-Completeness. New York: W. H. Freeman, 1979.

[4] A. Zomaya, C. Ward, and B. Macey, "Genetic scheduling for parallel processor systems: comparative studies and performance issues," Parallel and Distributed Systems, IEEE Transactions on, vol. 10, no. 8, pp. 795-812, aug 1999.

[5] H. Cheng, "A high efficient task scheduling algorithm based on heterogeneous multi-core processor," in Database Tech-nology and Applications (DBTA), 2010 2nd International Workshop on, Nov. 2010, pp. 1–4.

[6] A. Amini, T. Y. Wah, M. Saybani, and S. Yazdi, "A study of density-grid based clustering algorithms on data streams," in Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on, vol. 3, july 2011, pp. 1652–1656.

[7] A. Ebaid, R. Ammar, S. Rajasekaran, and T. Fergany, "Task clustering amp; scheduling with duplication using recursive critical path approach (rcpa)," in Signal Processing and Information Technology (ISSPIT), 2010 IEEE International Symposium on, dec. 2010, pp. 34–41.

[8] T. Chen, B. Zhang, X. Hao, and Y. Dai, "Task scheduling in grid based on particle swarm optimization," in Parallel and Distributed Computing, 2006. ISPDC '06. The Fifth International Symposium on, july 2006, pp. 238–245.

[9] M. Kashani and M. Jahanshahi, "Using simulated annealing for task scheduling in distributed systems," in Computational Intelligence, Modelling and Simulation, 2009. CSSim '09. International Conference on, sept. 2009, pp. 265–269.

[10] S. Song, K. Hwang, and Y.-K. Kwok, "Risk-resilient heuris-tics and genetic algorithms for security-assured grid job scheduling," Computers, IEEE Transactions on, vol. 55, no. 6, pp. 703–719, june 2006.

[11] K. Li, G. Xu, G. Zhao, Y. Dong, and D. Wang, "Cloud task scheduling based on load balancing ant colony optimization," in Chinagrid Conference (ChinaGrid), 2011 Sixth Annual, aug. 2011, pp. 3–9.

[12] A. Radulescu, A. Gemund, "Fast and effective task scheduling in heterogeneous systems," Proceedings of the 9th heterogeneous computing workshop (HCW 2000), pp. 229-238, 2000.

[13] A. Verma, S. Kaushal, "Deadline constraint heuristic based genetic algorithm for workflow scheduling in cloud," Forthcoming article in international journal of grid and utility computing.

[14] S. Selvarani, G.S. Sadhasivam, "Improved cost-based algorithm for task scheduling in cloud computing," computational intelligence and computing research, pp.1-5, 2010.

[15] S. Pandey, L. Wu, S. Mayura Guru, R. Buyya, "A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments," 24th IEEE international conference on advanced information networking and applications, PP 400-407, 2010.

[16] Lizheng Guo, Guojin Shao, Shuguang Zhao, "Multi-Objective Task Assignment in Cloud Computing by Particle Swarm Optimization", 978-1-61284-683-5/12/$31.00 2012 IEEE.

[17] A. Verma, S. Kaushal, "Bi-Criteria Priority based Particle Swarm Optimization Workflow Scheduling Algorithm for Cloud" Proceedings of 2014 RAECS UIET Panjab University Chandigarh, 06-08 March, 2014.

[18] Chenhong Zhao, Shanshan Zhang, Qingfeng Liu, Jian Xie, Jicheng Hu, "Independent tasks scheduling based on Genetic Algorithm in Cloud Computing", 978-1-4244-3693-4/09/$25.00 2009 IEEE.

[19] GAN Guo-ning, HUANG Ting-lei, GAO Shuai, "Genetic Simulated Annealing Algorithm for task scheduling based on Cloud Computing Environment", 978-1-4244-6837-9/10/$26.00 2010 IEEE

[20] Haluk, T., Salim, H., and Wu, M.Y, "Performance-effective and low complexity task scheduling for heterogeneous computing", IEEE Transaction on Parallel and Distributed Systems, vol. 13, no.3 pp: 260-274, Mar 2002.

[21] Ramamritham K. , Stankovic A. J. ,"Effective scheduling algorithm for real time multiprocessor systems. IEEE Transaction on Parallel and Distributed Systems" 1990 , 1(2): 184~194

[22] Qiao Ying. Wang Hong An, Dai Guo Zhong. "Developing a new Dynamic Scheduling Algorithm for real time multiprocessor systems". Journal of Software, 2002, 3(1):51 ~ 58.