View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by TELKOMNIKA (Telecommunication Computing Electronics and Control)

TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 18, No. 3, June 2020, pp. 1252~1258
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v18i3.11915

Improved quality of service-based cloud service ranking and recommendation model

Sirisha Potluri, Katta Subba Rao
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, India

Article Info

Article history:

Received Nov 27, 2018
Revised Jan 31, 2020
Accepted Feb 24, 2020

Keywords:

Cloud computing
Quality of service
Service ranking

ABSTRACT

One of the ongoing technologies which are used by large number of companies and users is cloud computing environment. This computing technology has proved that it provides certainly a different level of efficiency, security, privacy, flexibility and availability to its users. Cloud computing delivers on demand services to the users by using various service-based models. All these models work on utility-based computing such that users pay for their used services. Along with the various advantages of the cloud computing environment, it has its own limitations and problems such as efficient resource identification or discovery, security, task scheduling, compliance and sustainability. Among these resource identification and scheduling plays an important role because users always submits their jobs and expects responses in least possible time. Research is happening all around the world to optimize the response time, make span so as to reduce the burden on the cloud resources. In this paper, QoS based service ranking model is proposed for cloud computing environment to find the essential top ranked services. Proposed model is implemented in two phases. In the first phase, similarity computation between the users and their services is considered. In the second phase, computing the missing values based on the computed similarity measures is calculated. The efficiency of the proposed ranking is measured and the average precision correlation of the proposed ranking measure is showing better results than the existing measures.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Sirisha Potluri,
Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation,
Green Fields, Vaddeswaram, Andhra Pradesh 522502, India.
Email: sirisha.vegunta@gmail.com

1. INTRODUCTION

On demand service selection is providing wide selection of facilities and satisfying the changing needs of IT industries. There are many advantages and issues which are associated with cloud computing environment and ongoing research is helping it improve in all its regards. Cloud computing provides all the advantages to the cloud users to migrate to next level of computing. This computing environment is still in an evolving phase and delivers the services in various ways using various deployment models. The relationship between cloud service provider and customer completely lies on the on demand availability of the resources, efficient management of the resources and competent use of the resources. Along with the various advantages of the cloud computing environment, it has its own limitations and problems such as efficient resource identification or discovery, security, task scheduling, compliance and sustainability.

Among these resource identification and scheduling plays an important role because users always submits their jobs and expects response in least possible time. Research is happening all around the world to optimize the response time, make span so as to reduce the burden on the cloud resources [1, 2].

2. RESEARCH ISSUES AND CHALLENGES IN CLOUD COMPUTING

2.1. Privacy and security

Yunchuan Sun Et al. stated that cloud computing environment make use of large amounts of datasets for storage and processing. On premise data storage, handling, processing is completely risk free. But if the data is getting saved in other places there is a greater risk involved in that. Cloud service providers, intermediaries or the companies involved in hosting of cloud platforms may disclose the data which is being in access to them. Due to which data may be tampered or lost. In either of the cases it leads to security and privacy issues [3]. Due to these reasons of security and privacy, cloud computing is taking slightly a back step apart from its wider adoption [4, 5].

2.2. Performance

Weizhong Qiang Et al. stated that in cloud environment, this issue is very much important as it is one of the measuring technique in computing environment. Performance is a major concern in cloud system which affects the delivery of various services, revenue generated due to service delivery and the number of customers invlovled in the process [6].

2.3. Reliability and availability

Mohammad Reza Mesbahi Et al. stated that reliability defines the certain expected functioning of the system under selected time interval and with given conditions. Availability of any system can be defined as degree in which the set of resources are available and these available resources are used whenever and wherever there is a need. These two factors comes together and defines the strength and potency of the system [7].

2.4. Scalability and elasticity

Emanuel Coutinho Et al. stated that the scalability can be defined as its ability to adopt to the changes. When there is a need of resources by various cloud users, the system should take care of workloads efficiently. Elasticity of the system can be defined as the factor up to which it can handle the workloads efficiently. Elasticity ensures allocation and deallocation of resources to handle the needs of on demand users [8].

2.5. Interoperability and portability

Beniamino Di Martino Et al. stated that interoperability of cloud computing environment can be defined as the power or ability which is adopted by the system to run various services which are from various vendors. Its associated factor interoperability can be defined as the way in which different platforms are effectively used among different cloud service providers. Portability is an important factor which defines the ability to transfer the related data applications from one CSP to another CSP without much dependency involved in it [9].

2.6. Resource management and scheduling

Harshil Mehta Et al. stated that resource management can be defined as a factor which represents the efficient usage and maintenance of the resources. Various types of resources are mapped to the submitted jobs so as schedule them in a more efficient manner. A better resource scheduling or provision can be achieved by using a better scheduling policy. By using these scheduling algorithms, we can optimize cloud services to achieve high level of quality of service. There are many categories of users, tasks and resources using which task mapping and execution is achieved [10].

2.7. Energy consumption

Nazmul Hossain Et al. stated that any cloud data center or storage consists of infrastructure in large and voluminous amounts. Cloud computing environment is delivering various types of services to enormous users by using its massive infrastructure. Cooling infrastructure is needed to compensate the heat which is produced by these servers. This Expensive infrastructure should be developed, operated and maintained very carefully as it produces greenhouse gases which are adverse to environmental balance. Proper cooling system and mechanisms are maintained by cloud centers to maintain the cloud infrastructure as environment friendly [11-13].

2.8. Virtualization

Yuping Xing Et al. stated that in distributed and cloud computingenvironement, the concept of virtualization plays a very important role. This refers to the process of creating virtual machines for the various physical resources. This concept is closely related to task scheduling process where task queue is mapped to various virtual machines for scheduling. The main advantages of virtualiztion are cost reduction, highly scalable and provide elastic behavior to the cloud computing environment [14, 15].

2.9. Bandwidth cost

Phuong Ha Et al. stated that according to cloud computing, bandwidth referes to the rate in which bits are transferred. This measurement is mainly useful when we are calculating throughput of the system. Since cloud computing environement is providing high speed transfer of data among systems it requires high bandwidth [16].

3. QUALITY OF SERVICE IN CLOUD COMPUTING

3.1. SLA issues in cloud computing environment

Service level agreement commonly includes various segments like measuring the performance, managing the problems, customer satisfaction, recovery from failures and finally termination of the agreement. SLA's are consistently checked and validated using which it creates a better communication channel between consumer and the provider. SLA establishes the quality of service (QoS) agreement between service-based system providers and users. With a violation of SLA, the provider must pay penalties. Continuous performance check is required in order to ensure that these CSPs are providing services as mentioned in SLAs. SLA is consistently met, these agreements are frequently designed with specific lines of demarcation and the parties involved are required to meet regularly to create an open forum for communication. Cloud computing is using collection of resources and all these are supported by the interface and infrastructure concept of distributed and cloud computing environment. Serive level agreement can be viewed as a document between customer and provider and is maily based on services and not on customers [17].

3.2. QoS metrics and classification

As we have discussed earlier cloud computing environment is in demand of continuous performance measurement. The two mechanisms such as measurement monitor and exposure of cloud computing environment is done by using these metrics. The main issue and drawback with this computing environement is its performance. The perforamce is measured by using various factors related to user experience and ability. The main drawback and problem of cloud computing is having difficultly in identifying base or root cause for service problems such as interruptions because the structure of the environement is very complex. The dynamic environement of the cloud makes use of various metrics to measure the performance. All these attributes are listed as QoS metric to enforce the promised level of bahaviour in cloud [18, 19]. There are many QoS metrics using which the performance of CSP can be measured. Mainly there are three aspects of QoS metric namely performance, dependability and configuration. Performance can be measured by using response time, processing time, service throughput, data transfer rate and latency. Dependability can be measured by using availability, elasticity, reliability, timeliness, resilience and scalability. Similarly configuration can be measured by using virtual systems and location. The QoS metrics classification [20] is given in Table 1.

3.3. Existing QoS prediction approaches and models for service ranking

Existing QoS prediction approaches are CStrust, COFILL, LoNMF, OLMF, SCMF, GNMF, TaSR and CSMF etc. Existing ranking prediction approaches are Cloud rank QoS ranking prediction framework- CloudRank1 and CloudRank2, PSO based QoS ranking prediction algorithm, Time aware trust worthiness ranking prediction approach, Multirank1 and MultiRank2 and Ranking oriented prediction method etc. The objective of the proposed mechanism to achieve better recommendation compared to conventional approaches.

Table 1. Classification of QoS metric in cloud computing environment

Aspect	Metric		Description
Performance	Response Time		Elapsed time between user submitted a task as a service request and final response is received
	Processing Time		It is represented as average response time
	Service Throughput		It is represented as the number of tasks completed by cloud computing environment per unit amount of time.
	Data Transfer Rate		It is represented as the speed at which data can be transferred quickly and securely
	Latency	Packet Length Network Latency Through of a CSP's network	It is represented as the delay between when first packet bit passed the input check point and the last packet bit passed the output check point
Dependability	Availability		It is represented as the percentage of time that the resource is ready for the immediate use
	Elasticity		It is represented as the ability by which the CSP can expand and contract his services
	Reliability		It is represented as the ability to ensure continuous process of the program without loss
	Timeliness		It is represented as the ability to supply information in time when users are in need to access it
	Resilience		It is represented as the ability to resist and recover in an effective and timely manner, including the preservation and restoration of its essential basic structures and functions
	Scalability	Horizontal Vertical	It is represented as the ability to increase cloud resources of the same type by using VMs
Configuration	Virtual Systems		It is represented to identify cloud virtual infrastructure
	Location		Based on QoS location affinity

4. QOS BASED SERVICE RANKING MECHANISM

QoS is a measurement to check for the guaranteed service is achieved from the cloud providers or not [21-24]. This mechanism contains two phases for optimal service selection in cloud computing environment. In the first phase, the computation of similarity among the users and their services is considered. In the second phase, computation of the missing values based on the values which are using computed similarity measures is calculated. The mechanism of prediction and ranking based on QoS is validated using WSDream#1 QoS dataset and Web Service QoS dataset using statistical measures mean squared error and root mean square error. QoS matrix which contains QoS values of various services for various users acts as essential data source for service evaluation and selection. But this is a sparse matrix with missing values. Proposed method is useful for QoS value prediction and subsequently helps in optimal ranking sequence of services. Cosine based similarity measure helps for QoS value prediction and improved binary gravitational search is useful for service rankingmechanism. Similarity computation and missing QoS value prediction are performed on U usersand S services and the output is optimal service ranking of the resources in cloud computing environment [25].

5. PROPOSED METHOD

Algorithm steps:

Step 1: Computing the task rating similarity requested by the active and normal user is given by $L\bar{r}_a$ and $L\bar{r}_n$ represents the average log like lihood estimator of all services invoked by active and normal users to service. The contextual task similarity requested by the active cloud and normal cloud users is given by:

$$sim\ R(r_a, r_n) = \frac{\sum_{s\in S}(r_{a,s}-\overline{Lr}_a)(r_{n,s}-\overline{Lr}_n)}{\sqrt{\sum_{s\in S}(r_{a,s}-\overline{Lr}_a)^2 \sum_{s\in S}(r_{n,s}-\overline{Lr}_n)^2}} * (mas(T_a, T_n))$$

Step 2: In this step, the similarity computation of the users based on the QoS values is given by the similarity between the users interacting with the service is

$$sim\ U(v_a, v_n) = SimU(Qos(v_a, v_n) = \frac{\sum_{s\in S}(v_{a,s}-\bar{v}_{a,s})\,(v_{n,s}-\bar{v}_{n,s})}{\sqrt{\sum_{s\in S}(v_{a,s}-\bar{v}_{a,s})^2 \sum_{s\in S}(v_{n,s}-\bar{v}_{n,s})^2}}$$

Step 3: QoS missing value prediction using the weighted similarity measures as

Step 4: Normalized confidence score for users is computed as

$$Normalized\ confidence\ score\ for\ users = \phi_U = NCWU = N(simU(v_a, v_n).(\sum_{UeU} \frac{simU(v_a, v_n)}{\sum_{ueU} simU(v_a, v_n)}$$

Step 5: Normalized confidence score for tasks is computed as:

$$Normalized\ confidence\ weighted\ score\ for\ tasks = \phi_t = NCWS = N(simR(r_a, r_n).(\sum_{seS} \frac{simR(r_a, r_n)}{\sum_{seS} simR(r, r_n)}$$

Step 6: Weighted score for users is computed as:

$$Weighted\ score\ of\ users = w_u = \frac{\tau.\phi_U}{\tau(\phi_U) + (1-\tau)\phi_T}$$

Step 7: Weighted score for tasks is computed as:

$$Weighted\ score\ of\ tasks = w_T = \frac{(1-\tau).\phi_T}{\tau(\phi_T) + (1-\tau)\phi_U}$$

Step 8: Weigted predicted score for Qos missing value is computed as

$$Predicted\ value\ for\ Qos\ missing\ values = \frac{2(w_U.w_T)}{(w_U + w_T)} Max\ \{\phi_{U[i]}, \phi_{T[j]}\}$$

Step 9: Apply PSO algorithm for users to tasks ranking.

6. RESULTS AND ANALYSIS

The efficiency of the proposed scheme is compared with six existing competing methods namely IMEAN, UMEAN, UPCC, IPCC, UIPCC and IBGSSQPred. IMEAN: item mean: average QoS value observed from the used service as the predicted QoS value of the user for the unused service. User mean (UMEAN)–average QoS value known by the user for the used service as the predicted QoS value of the user for the unused service. User-based collaborative filtering method using PCC (UPCC)-Top-K neighbors of the users are found using PCC similarity measures. Item-based collaborative filtering method using PCC (IPCC)-Top-K neighbors of the services are found using PCC similarity measures. User and item-based collaborative filtering method using PCC (UIPCC)–combines UPCC and IPCC; top-K neighbors of the users and services are found using PCC similarity measures for QoS prediction. The comparison of mean squared error of proposed QoS measure to the traditional QoS measures on the cloud dataset is shown in Figure 1.

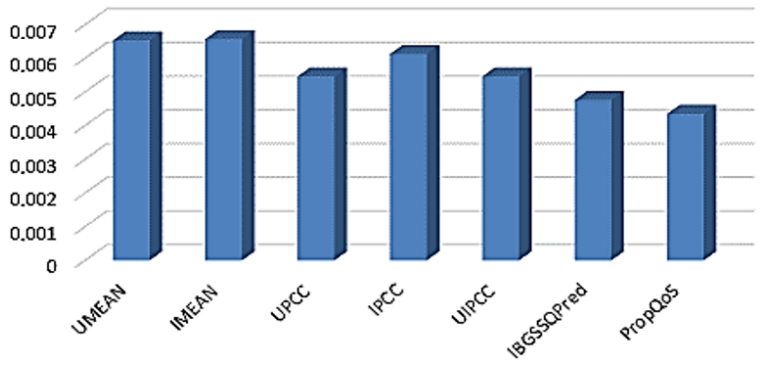

Figure 1. Comparison of mean squared error of proposed model to the traditional QoS models on the training cloud dataset

Figure 1 represents the comparative analysis of proposed QoS model to the conventional QoS models on the training cloud dataset. In the figure, the mean square error rate of the proposed model and existing models are computed on the training dataset. From the Figure 1 it is noted that the present model has less error rate than the conventional models. Figure 2 represents the comparative analysis of proposed QoS model to the conventional QoS models on the training cloud dataset. In the figure, the root mean square error rate of the proposed model and existing models are computed on the training dataset. From the Figure 2 it is

noted that the present model has less RMSE rate than the conventional models. The comparison of QoS runtime of proposed QoS measure to the existing QoS runtime measure on the cloud dataset is shown in Figure 3. Figure 4 represents the efficiency of the proposed ranking measure on cloud users to services and it is showing that the average value of precision correlation of the proposed algorithm is better than the previously existing measures for users to services interaction.

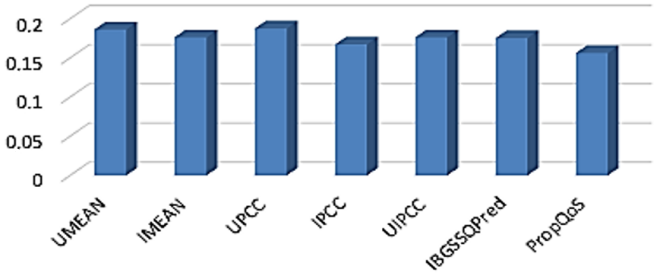

Figure 2. Comparison of root mean squared error (RMSR) of proposed model to the traditional QoS models on the training cloud dataset

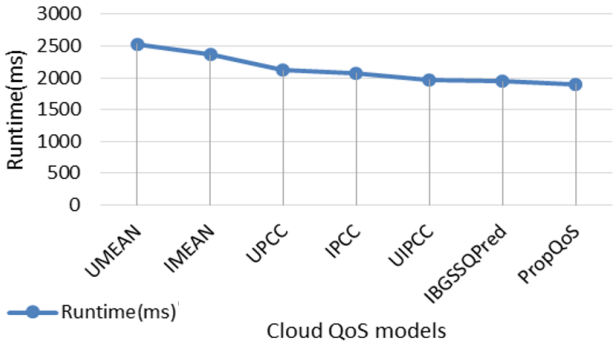

Figure 3. Comparative analysis of proposed QoS runtime to the existing QoS runtime measures on the input cloud service dataset

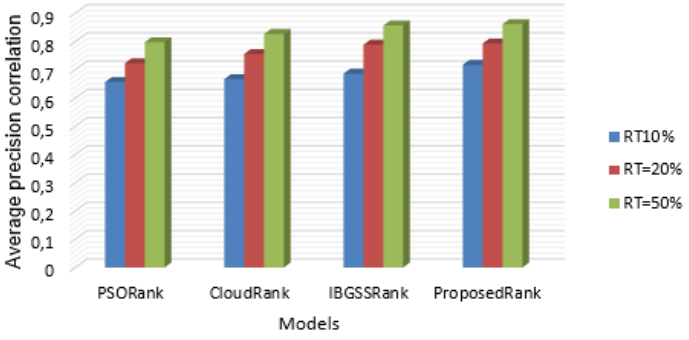

Figure 4. Contextual ranking comparison of proposed ranking measure to the existing ranking measure on cloud web service dataset

7. CONCLUSION AND FUTURE WORK

QoS based service ranking mechanism using service similarity based PSO in cloud computing environment is proposed in this paper. This mechanism contains two phases for optimal service selection in cloud computing environment. In the first phase, the computation of similarity among the users and their services is considered. In the second phase, computation of the missing values based on the values which are using computed similarity measures is calculated. The mechanism of prediction and ranking based on QoS is validated using WSDream#1 QoS dataset and web service QoS dataset using statistical measures mean squared error and root mean square error. The efficiency of the ranking and recommendation model which is

proposed in this paper is measured and the value of average precision correlation is giving better results than the existing measures. This ranking mechanism can be used in task scheduling to efficiently schedule the tasks to resources.In the futurework, a hybrid PSO optimization model is integrated with the proposed recommendation and ranking model for efficient task scheduling in cloud computing environment.

REFERENCES

[1] N. Kratzke, P. Christian Quint, "Understanding cloud-native applications after 10 years of cloud computing-A systematic mapping study," *Journal of Systems and Software*, vol. 126, pp. 1-16, April 2017.

[2] T. Lynn et al., "A Preliminary Systematic Review of Computer Science Literature on Cloud Computing Research using Open Source Simulation Platforms," *Proceedings of the 7th International Conference on Cloud Computing and Services Science (CLOSER 2017)*, vol. 1, pp. 537-545, 2017.

[3] Y. Sun, J. Zhang, Y. Xiong, G. Zhu, "Data Security and Privacy in Cloud Computing," *International Journal of Distributed Sensor Networks*, vol. 2014, pp. 1-9, July 2014.

[4] A. Albugmi, M. O. Alassafi, R. J. Walters, G. Wills, "Data Security in Cloud Computing,"*2016 Fifth International Conference on Future Generation Communication Technologies (FGCT)*, Luton, pp. 55-59, 2016.

[5] I. Ahmed, "A brief review: security issues in cloud computing and their solutions," *TELKOMNIKA Telecommunication Computing Electronics Control*, vol. 17, no. 6, pp. 2812-2817, December 2019.

[6] W. Qiang, "Performance and security in cloud computing," *The Journal of Super Computing*, vol. 75, no. 1, pp. 1-3, January 2019.

[7] M. R. Mesbahi, A. M. Rahmani, M. Hosseinzadeh, "Reliability and high availability in cloud computing environments: a reference roadmap," *Human-centric Computing and Information Sciences*, vol. 8, no. 20, December 2018.

[8] E. Coutinho, F. R. C Sousa, P. A. L. Rego, J. Souza, "Elasticity in cloud computing: a survey," *Annals of Telecommunications-Annales Des Telecommunications*, vol. 70, no. 7-8, pp. 289-309, August 2015.

[9] B. D. Martino, G. Cretella, A. Esposito, "Cloud Portability and Interoperability," *Encyclopedia of Cloud Computing*, May 2016.

[10] H. Mehta, V. K. Prasad, M. Bhavsar, "Efficient Resource Scheduling in Cloud Computing," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 3, pp. 809-815, April 2017.

[11] N. Hossain, Md. A. Hossain, A. K. M. F. Islam, "Research on Energy Efficiency in Cloud Computing," *International Journal of Scientific and Engineering Research*, vol. 7, no. 8, pp. 358-367, August 2016.

[12] R. A. Hasan et al., "HSO: A Hybrid Swarm Optimization Algorithm for Reducing Energy Consumption in the Cloudlets," *TELKOMNIKA Telecommunication Computing Electronics Control*, vol. 16, no. 5, pp. 2144-2154, October 2018.

[13] A. Ismail, N. A. Jamaludin, S. Zambri, "A Review of Energy-aware Cloud Computing Surveys," *TELKOMNIKA Telecommunication Computing Electronics Control*, vol. 16, no. 6, pp. 2740-2746, December 2018.

[14] Y. Xing, Y. Zhan, "Virtualization and Cloud Computing," *Future Wireless Networks and Information Systems*, vol. 143, pp. 305-312, 2012.

[15] Taskeen Zaidi, Rampratap, "Virtual Machine Allocation Policy in Cloud Computing Environment using CloudSim," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 1, pp. 344–354, February 2018.

[16] H. Phuong, X. Lisong, "Available bandwidth estimation in public clouds," *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Honolulu, HI, pp. 238-243, 2018.

[17] C. Zhang, Y. Wang, Y. Lv, H. Wu, H. Guo, "An Energy and SLA-Aware Resource Management Strategy in Cloud Data Centers," *Hindawi Scientific Programming*, vol. 2019, pp. 16, November 2019.

[18] Swami Das M., Govardhan A., Vijaya Lakshmi D. (2019) "Web Services Classification Across Cloud-Based Applications," *Soft Computing: Theories and Applications*, vol. 742, pp. 245-260, August 2018.

[19] S. Heidari, R. Buyya, "Quality of Service (QoS)-driven resource provisioning for large-scale graph processing in cloud computing environments: Graph Processing-as-a-Service (GPaaS)," Future Generation Computer Systems, Vol. 96, pp. 490-501, 2019

[20] S. Qiping, W. Xiaochao, N. Guihua, C. Donglin, "QoS-aware cloud service composition: A systematic mapping study from the perspective of computational intelligence," *Expert Systems with Applications,* vol. 138, December 2019.

[21] M. H. Ghahramani, M. Zhou, T. H. Chi, "Toward Cloud Computing QoS Architecture: Analysis of Cloud Systems and Cloud Services," in IEEE/CAA *Journal of AutomaticaSinica*, vol. 4, no. 1, pp. 6-18, January 2017.

[22] S. Potluri, K.S. Rao, "Quality of Service based Task Scheduling Algorithms in Cloud Computing," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 2, pp. 1088-1095, April 2017.

[23] Potluri S., Rao K. S., "Simulation of QoS-Based Task Scheduling Policy for Dependent and Independent Tasks in a Cloud Environment," *Smart Intelligent Computing and Applications,* vol. 159, pp. 515-525, September 2019.

[24] Shefali Varshney, Rajinder Sandhu, P. K. Gupta, QoS Based Resource Provisioning in Cloud Computing Environment: A Technical Survey, International Conference on Advances in Computing and Data Sciences *ICACDS 2019: Advances in Computing and Data Sciences*, pp. 711-723, 2019

[25] N. Somu, Gauthama R.M.R., A. Kaveri, Akshay R. K., K. Krithivasan, Shankar S.V.S., "IBGSS: An Improved Binary Gravitational Search Algorithm basedsearch strategy for QoS and ranking prediction in cloud environments," *Applied Soft Computing,* vol. 88, March 2020.