

Hybrid Whale-Bee Optimization (HWBO) based Optimal Task Offloading Scheme in MCC

M. S. Premalatha, B. Ramakrishnan

Abstract: Transferring the tasks from portable gadgets to public cloud is one of the important processes in Mobile Cloud Computing (MCC). Subsequently, offloading differed errands in the meantime will build the 'cloudlets' load and enlarges the basic finish time of the offloaded assignments. Storing of tasks in the cloud storage is energy consumed process. The optimal position is to be identified for offloading the tasks from portable gadgets. In order to solve the issue, an optimal task offloading technique is proposed. A hybrid optimization method based on Hybrid Whale Optimization algorithm (WOA) and Artificial Bee colony optimization algorithm (ABC). Dual task assignment process incorporated with queuing models offloads the task in the optimal place of the cloud to reduce the drop rate. The efficiency of the proposed scheme is evaluated with the conventional methods on the basis of energy consumption, drop rate etc.

Keywords: Whale optimization, Average response time, energy consumption, Mobile Cloud Computing, Queuing model.

I. INTRODUCTION

Cloud Computing offers of the processing resources relying on the need of clients or users [1]. In this way with the help of web, it conveys the administrations to clients. Assignments are given to users by the cloud service provider. The monstrous extent of uses is bolstered by the portable gadgets that request substantial use of intensity. It implies that function codes square measure dead part in some obscure places as opposed to the death penalty in included mobile gadgets. This outcome in the decline of assignments process speed, impediment of assets and maximized power utilization [2,3]. In MCC, 3 unmistakable The errand offloading might be a typical action of mobile gadgets inside the climate. To strengthen the execution few errands square measure transmitted by the portable applications to the cloud. layers explicitly, 'cloudlets', 'open cloud' and 'impromptu gadgets' (for example Bluetooth) are incorporated [4, 5]. 'People in general cloud' is set removed from the clients. In this way, the clients will discuss exclusively with the help of web (or get to reason (AP)). Due to the heavy distance, completely fledged by the clients and 'open clouds' it expends extra energy on account of a moderate affiliation. These issues might be settled by the arrangements anticipated in [6-8].

Revised Manuscript Received on February 05, 2019.

M. S. Premalatha, Research Scholar, Manonmaniam Sundaranar University, Abishekapatti, Thirunelveli – 12, Tamil Nadu, India.

Dr. B. Ramakrishnan, Associate Professor, Department of Computer Science and Research Centre, S.T. Hindu College, Nagercoil, Tamil Nadu, India.

Through the 'cloudlets' will associate with the mobile gadgets. Towards the unmistakable geological position, these data focuses square measure meagerly dispersed. Precedents for a couple of the topographical conditions square measure theater, periodic retailers so on. From such condition, the clients choose the nearby cloudlet with respect to distance to dump the tasks. From such condition, the client picks the closer 'cloudlet' to dump the mobile errands. On the off chance that the approaching client asks for square measure in huge interest, the 'cloudlets' can confront the insufficiency of assets. So the client should hold up inside the line of the cloudlet to strategy the assignments. It's fundamental to detect an additional cloudlet having separate servers to allot the spare undertakings that square measure holding up inside the line to finish the processing. To recuperate this client in passable holding up time the portable gadgets make utilization of the cloud administrations with, and coincidental gadgets. These administrations can cause abundance utilization of intensity and incredibly beyond all doubt won [14], [15]. In this way, in order to recuperate these issues, it's fundamental to spot relate degree right consign to dump the assignments by bearing in mind the requirements of assets inside the public cloud.

In this research, an optimal queuing based offloading of tasks from mobile devices to cloud is proposed. The offloading can be done in public cloud, private cloud and cloudlets. The optimal position to store the jobs into the cloud is determined using the queuing theory based algorithm.

The primary objectives of this research are,

- To develop a dual stage offloading technique, this transfers the task via cloudlet to minimize the completion and response time.
- To combine Whale Optimization with ABC and queuing Decision maker to form a choice regarding the task allocation in cloudlets by equalization the load.

II. LITERATURE SURVEY

The errand task in MCC setting has pulled in assortment of analysts inside the movement field of 'cloud' processing. A few seeable reviews were directed amid this field as unequivocal in [2,3,4]. Inside the exacting works, difficulties, and plan of MCC were set up.

The MCC situations asset designation, security, and protection square measure acutely explored in [2], [21]. the conventional programming calculations upheld errand task square measure wide arranged into 2 shapes. they're assignment based programming algorithmic guidelines and framework basically based programming calculation.

These calculations might be divided off on the introduce of load balance, use of memory, vitality utilization then on. As of late, the cell phones may experience the ill effects of decreased power life at the period of offloading versatile functions because of the utilization of a great deal of intensity. Pedram et al. [11] set up one among the programming approaches to decrease of the vitality use by a path referenced as adaptive power issue. So as to save the vitality Wang et al. [12] arranged the versatile transmission procedure to decrease of the correspondence that corrupts the processing sum comparatively as expanding the battery life. Zomaya et al. [13] built up one among the dynamic programming calculations referenced as voltage scaling to lessen the undertaking fruition time and subsequently the utilization of vitality. Additionally, Xie et al. [14] arranged the algorithmic standard to relocate of the assignments in the middle of the 'mists' local centers. in order to decrease the cost and to amplify the data focus use Bossche et al. [15] arranged the detailing of 'paired number program'. along these lines to adjust the 'cloud' structures stack Tayal et al. [16] set up the Fuzzy principally based GA algorithmic guideline at the season of utilization programming. also, Feller et al. [17] arranged the assignment of work by a novel approach. Here it uses of the ACO algorithmic principle to determine the allotment issues. Xu et al. [18] take a shot at the initiate of IMOPSO algorithmic guideline to examine of the implicit technology active ways.

The 'community oriented' undertaking based programming calculations neglect to execute its new approaching assignment while not the finish of the aforesaid allocated errands. The commitments of 'collective' basically based programming algorithmic standard were dole out by [19]. The creators in [19] arranged the "LARAC" algorithmic guideline with 'one-climb' procedure for programming the undertakings that square measure requested inside the straight design. The universal process for the task of the undertaking is contemplated in [20] for the 'communitarian' execution of the portable applications. These were considered for the lessening of vitality utilization. Likewise, the algorithmic standard is employed to determine the enhancement issue inside the task of assignments. In [21] the utilization outline is planned to imbed the applying routes with a standardized slice type to downsize the correspondence sum and in this manner the exchange time at the trading of learning. One among the process is arranged by Xie et al. [22] to figure the undertaking offloading inside the dimension oriented helpful process.

III. PROBLEM DEFINITION

This process contains user request from mobile devices, cloud (public) and cloudlets. The cloudlet is scattered into the entire urban environment. During offloading services, the huge amount of time and energy is consumed. So as to limit the total task's drop rate with expanded clients demands, to limit the avg.completion time and the utilization of power the choice capacity is completed. The user requests are processed by the internet to the cloud via cloudlet. The offloading of task from mobile devices to cloud via cloudlet is presented in figure 1. But if the number of task or jobs enlarged means cloudlet is busy and it is become a time consumed process. The process is delayed for a long time if the cloudlet doesn't meet the QoS requirements.

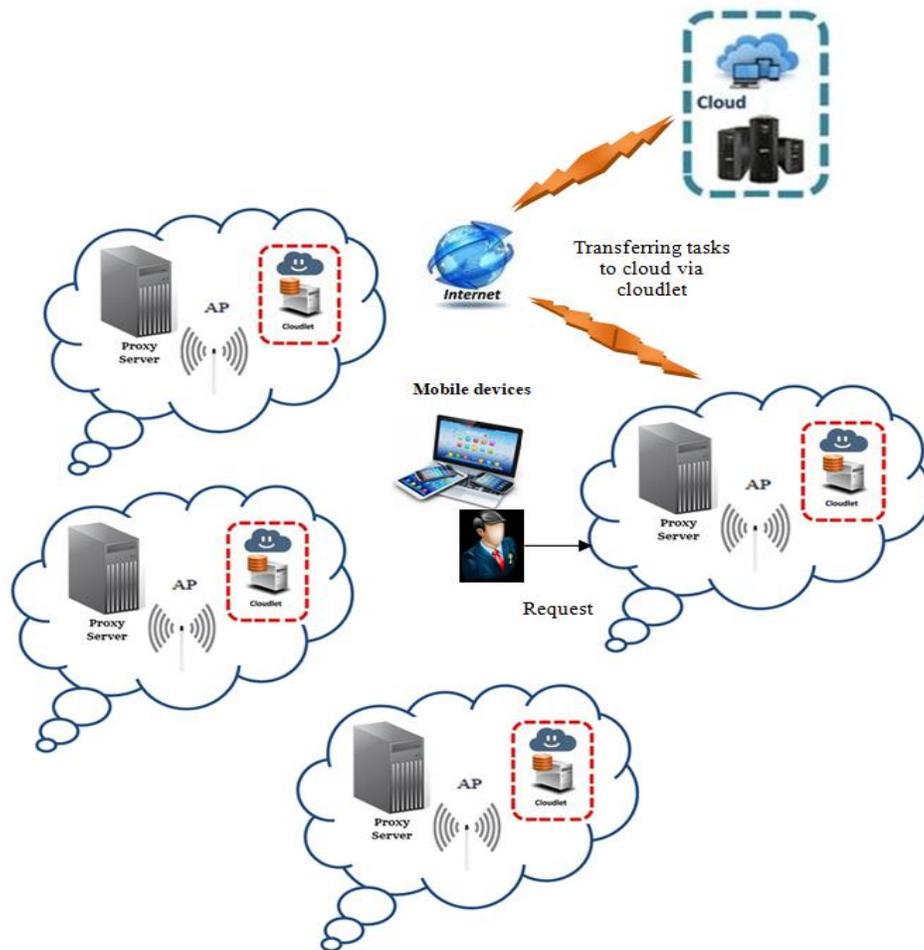


Fig.1 ask offloading from mobile devices to cloud

In this research problem is formulated using Queuing Theory in order to reduce the delay time if the number of task increased. The basic principle of Queue is First Come First Serve (FCFS). Using this principle, maximum resources are utilized without buffer. The time for completing the process and communication time are directly proportional to each other. When transferring of process through web, the average time taken is denoted as λ_{web} undertakings to 'mists' through a web association, the consummation time of the whole errands is about seconds. While, alludes to the transmission time of the assignments through by means of (second walk). The 'cloudlets' are disposed of by methods for the likelihood in the event that it decays to give an errand (third walk). The errand finish time is assessed by including both the λ_{web} and Q_{3G} .

The solicitations which are transmitted to compare the clients are kept up by methods for assignments. The errands are portrayed in the frame ts_j , where $j = 1, \dots, M_{ts}$. Processing of each undertaking is finished by the 'cloudlets' or 'open mists' servers. One of the invaluable behind 'people in general cloud' is it incorporates of adequate assets when differentiated to the accessibility of assets in 'cloudlets'. The

expression vS_k is alluded to as the 'virtual' servers which are incorporated into the cloudlet Ct_k , where $k = 1, \dots, M_{Ct}$. The lattice is appointed to one for the component R_{jk} but the cloudlet is allocated with the undertaking ts_j , else the grid is relegated to zero. The real worry of the framework is to limit the whole undertakings consummation time t_c .

IV. PROPOSED OPTIMAL TASK SCHEDULING PROCESS

The Optimal position to store the jobs from the portable gadgets to public cloud or cloudlets is calculated by the proposed scheme. A hybrid Whale-bee optimization (HWBO) along with queuing model utilizes the maximum resources, user requests, the communication time between users and cloudlets etc. Here queuing theory estimates the delay time and process time of the users for their request on the account of resources and services available.

Hybrid Whale-Bee Optimization (HWBO) based Optimal Task Offloading Scheme in MCC

The Whale optimization hybridized with queuing model to offloading jobs. The search space is elaborated by mining of tasks from all the public cloud and cloudlets. In order to reduce the cause of dimensionality, the tasks are optimally allocated to the cloud server using “Artificial Bee Colony Optimization (ABC)”. The proposed calculation likewise made an endeavor to adjust the cloudlet stack for the decrease of denied demands rate.

Dual stage offloading technique

There are 2 modules in this process, one is user request and another one is offloading ie, transferring the tasks. Initially, the user having any of the mobile devices, disseminate the demand call to the nearest cloudlets. Inside the cloudlet, the demand request received by the proxy server. The Proxy is maintaining a buffer to store the request as in the order they received. Here queuing principle FCFS is incorporated along with optimization. The task dispatcher role is to calculate the existence of user request.

The storehouse is related with the “Decision Maker” section and makes utilization of the accessible checking information as its contribution; by this ensuing impact, it can take viable proficient choices. The ongoing status with respect to the use of 'cloudlets' servers just as the normal time taken by the clients to offload the assignments in 'cloudlets' or 'open cloud' are protected in each 'cloudlets' store that keeps up of the 'intermediary' servers in it. On the other hand, the observing unit gathers this fundamental data's and this unit is put in the intermediary's of 'cloudlets'. The synchronization and Updation of storehouse's whole information are done intermittently through the web. In light of the use of servers and the status of the system the open term may fluctuate for every period and furthermore it will encounter a variety dependent on the approaching solicitations. This unit deals with the premise of proposed line based calculation. Identified with the dual stage organize, the data of the 'Choice maker' is used by the transmitter phase. The data is used by the transmitter so as to appoint and to transfer approaching undertakings to the particular public cloud or cloudlet. The fundamental capacity of the transmitter is to elevate the clients to create utilization of the 3G association in the disappointment of errand serving and furthermore in the drop of undertakings dependent on the client's demand. The choice produced based on “Decision Maker” unit transmits the errands to the 'cloudlet' (Ct) in light of the likelihood Q_{Ct} and the assignments are transmitted to 'people in general cloud' through the web dependent on the likelihood Q_{web} . The point by point portrayal of our proposed calculation is expressed underneath.

Hybrid Whale-bee optimization (HWBO)

The hybrid Whale-Bee algorithm is described for the ideal task of the undertaking in with the expect to limit the whole errands normal fruition time. It limits the transmitted errands fulfillment time by adjusting the 'cloudlets' stack. Here “Decision Maker” engineering in which it incorporates the

reduce the delay time, energy consumption and drop rate of half breed calculation is planned to develop. The whale behaviour based enhancement calculation is employed along with queuing theory to compromise the shortage in the accessibility of assets to cloud and the collaboration postpone skilled in the middle of the cloudlets and versatile clients. The WOA calculation is joined with the queuing decision making to diminish the clients holding up period in accepting the transferred assignments, utilization of intensity. It can likewise adjust the cloudlets stack by allowing the line replica of cloudlet. In this manner it checks the cloudlet's status. In the event that it discovers every one of the servers are in the bustling state, the offloaded undertakings can remain in the line until the point when the server enters an inert state. This joined calculation WOA_{QDM} goes for two fundamental concern specifically, reduction of the correspondence span and reaction time.

The principal target goes for the minimization of reaction time. The second target goes for the minimization of the open time frame among the cloudlet and portable clients, this can be viably accomplished by giving the yield of the queuing models to the WOA algorithm. This unit takes out the data in regards to the whole demands requested by the client, starting position of the line and 'cloudlets/mists' stockpiling limit. This unit fundamentally works totally relying on the model of the 'lining' hypothesis. Inferring the 'lining' demonstrate with various particular undertakings and assessing the probabilities of best errands allocated to every single 'cloudlet' or 'cloud' will prompt the development of enormous hunt space. Henceforth to determine of this dimensionality problem it is proposed to use the whale algorithm that elevates us to distinguish best arrangements. The line show issue swings to be difficult to assess with the common arrangements because of the dimensionality issue (enormous hunt space). Alternately, the pursuit space is relative to the number of cloudlet. Therefore to determine this problem the perfect decision is to utilize the WOA. The aces behind the WOA are

that it can get the ideal arrangement with limited cost, limited exertion and with less time. To recognize the ideal arrangement the WOA method forces of the randomized neighborhood look capacities and because of its probabilistic principles of determination it kills out the restricted ideal answer for be caught in the solution space.

The number of task available in the cloudlet is referred as t_1, t_2, \dots, t_{Ct} and which is given as the input for WOA. The solution space of WOA contains the task and resources available in cloudlet. WOA is the biological characteristics of humpback whales in which the solution updation is reflected by its behaviour. The fitness of WOA is

$$T_{fit} = Q_{3G}(\lambda_{3G} + \lambda_{web}) + Q_{web}\lambda_{web} + Q_{ct}t_{total}$$

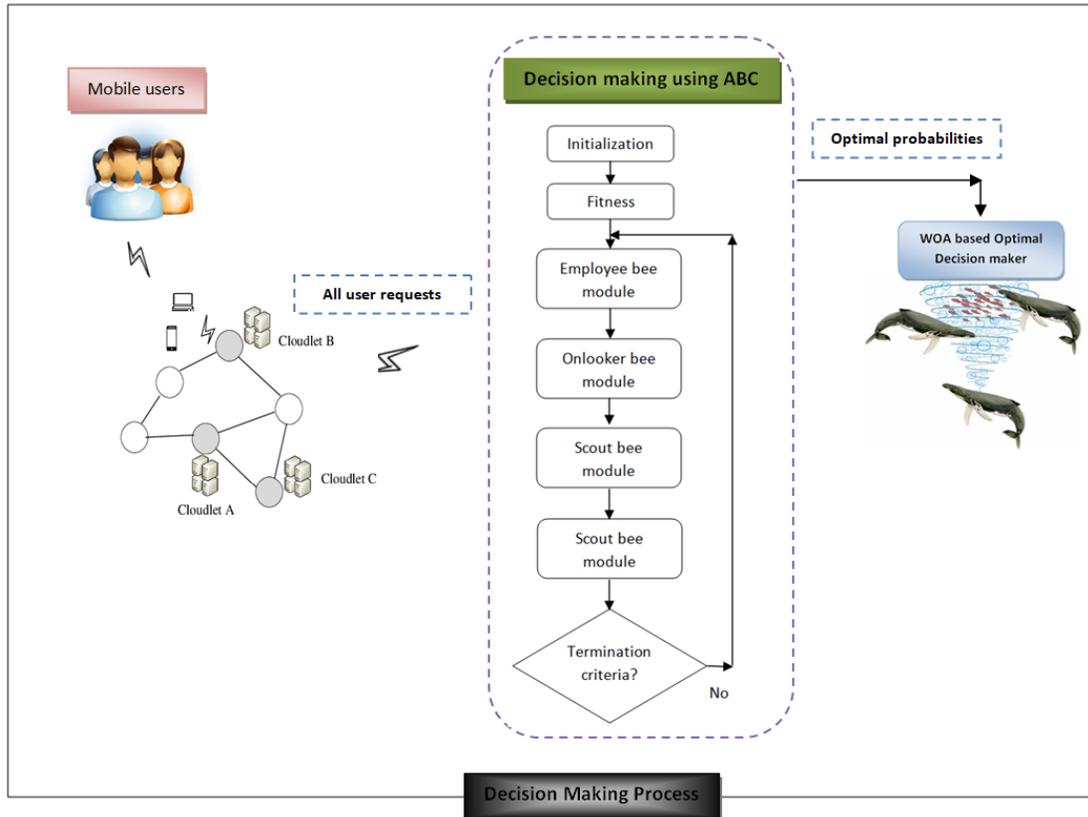


Fig. 2 Decision maker process using Whale-Bee Optimization

The cost function, t_{Total}^r represents the complete response time for the entire tasks. The results from queuing model are given as the input for Whale optimization along with the time taken for dissemination among user and cloudlets. The major intension of queuing models is to assign probabilities with minimum response time. WOA_{QDM} used this probabilities in to the working process of WOA. The optimal result is estimated by WOA_{QDM} based on the time taken to communicate among cloudlets and mobile users. The service time of MCC is reduces by this optimal process. The Combinatorial Problems in the optimization is resolved by using the WOA along with QDM. The whales are the most emotional animals in the world. Among different types of whales, humpback whales are having typical hunting behaviour. Using their, bubble net nourishing characteristics they chase the victim in the deep sea. In this chasing method, a circular shaped bubble net is created in the region of victim. This chasing behaviour is used to assign the tasks optimally. The process of WOA is explained as follows.

In cloudlets, the task are allocated using WOA_{QDM} part, therefore it can reduce the total process completion time and dissemination time by the usage of WOA. The intentional

After shrinking and encircling the prey, the spiral position is updated. The primary intension of Queuing Decision Maker part is to reduce the Process reaction time. arguments are employed by the WOA with the help of Queuing Decision Maker unit.

To update the delay time and process time, the reaction time of every cloudlets is calculated as

$$ts_k^r = ts_k^w + ts_k^s = \frac{U_{avg} + U_{ini}}{\mu_k} + \frac{1}{\lambda}$$

The avg. no. of incoming tasks inside cloudlet Ct_k presented as U_{avg} and the summation of all the reation time of task gives the total response time t_{Total}^r

$$ts_{Total}^r = \sum_{k=1}^{M_{Ct}} Q_k ts_k^r = \frac{U_{avg} + U_{ini}}{\mu_k} + \frac{1}{\lambda}$$

Based on the above eqn. the function $\mu_k, k = 1, \dots, M_{Ct}$ is developed to describe the total time to respond

$$\min(ts_{Total}^r) w.r.t \mu_1 + \mu_2 + \dots + \mu_{M_C} = \mu_T$$

Thus, by adjusting the cloud service provider heap the all out reaction period is decreased. The *QDM* limits the reaction time of framework just as to adjust the heap of cloudlets. This research, the assignments effort rate of the cloudlet *j* meant by the inconsistent μ_k is viewed as the sustenance basis and the above equation is used as the cost function. Also the above equation ought to be used to diminish of the frameworks complete reaction time. Consequently, undertaking task probabilities are evaluated with $\mu_k = Q_k \mu_T$. The process of this method is based on the likelihood Q_k . The transmission undertakings normal execution time over the 'cloudlets' is limited by utilizing these likelihood ratio. The user in the nearby location is more benefitted than the distanced user. Accordingly, the WOA method considers of both the correspondence time of the client and cloudlet just as the reaction time to go before its last procedure. Subsequently, the WOA decision is to dole out each undertaking ts_j to the 'cloudlet' Ct_k . It is conceivable to dole out more than one errands to each of the 'cloudlets'. Whale behaviour favor the high bubble net to reduce the correspondence time + reaction time.

Design of Whale-Bee with *WOA_{QDM}* :

In this research, Whale-Bee algorithm, *A* the matrix is described with users and cloudlets as rows and columns respectively and A_{jk} represents the time duration between user task ts_j and cloudlet Ct_k .

Initialization: In WOA, the input is the probabilities of task allocation from *WOA_{QDM}*. The whales are randomly selects the chunk of allocated tasks R_{jk} iteratively. The cost function is defined using dissemination duration, delay time of cloudlets. In the initial solution, the random values of the threshold are chosen and the search agent parameters are initialized and the population size is indicated by using the number of items.

Updating Process of MWO

Encircling Prey: In this, Humpback whale surrounds the prey (little fishes) at that point, refreshes its position towards the ideal arrangement throughout expanding number of iteration from beginning to the greatest no.of iteration, it's described by the

$$A = |\bar{y} \cdot \bar{C}^*(t) - \bar{C}(t)|$$

$$C(t+1) = C^{best}(t) - H \cdot y \bar{1}$$

Where *t* current iteration \bar{H} and \bar{y} is coefficient vectors,

\bar{C} is the position on a vector,

The coefficient values are calculated by

$$y \bar{1} = 2y \bar{2} \cdot \bar{r} - y \bar{2} \quad \text{and } y = \bar{2} \cdot r$$

Where $y \bar{2} \rightarrow$ reduces from 2 to 0

Bubble net attacking Model: It's included the two approaches to updated the new CH in vehicle clustering model

Shrinking encircling process: This process employed by decreasing linearly the value of $y \bar{2}$ from 0 to 2, the Random value of vector $k \bar{1}$ range between (-1 to 1).

Spiral updating position: To replicate the helix formation of whales, a spiral shape constraint is created in the gap of whale and prey.

$$\bar{C}(t+1) = \bar{H} \cdot e^{bt} \cdot \cos(2\Pi s) + \bar{C}(t)$$

The above enq describes the separation process of *i*th prey and the *b* as the steady state logarithmic spiral. The model followed by the

$$\bar{C}(t+1) = \begin{cases} \bar{C}(t) - y \bar{1} \cdot \bar{H} & \text{if } i < 0.5 \\ \bar{H}' \cdot e^{bs} \cdot \cos(2\pi s) + \bar{C}(t) & \text{if } i \geq 0.5 \end{cases}$$

Where *s* is a random value between - 1 to 1, $b \rightarrow$ constant and $i \rightarrow [0, 1]$. To show this synchronous behavior, we expect that there is a probability of half to pick between both the contracting encompassing system and the spiral model to refresh the position of whales amid optimization.

(C)Search for prey exploration phases: For searching a prey, an approach based on the variety of the $y \bar{1}$ vector can be used. Indeed, humpback whales look haphazardly as indicated by the position of each other.

$$\bar{H} = |y \cdot \bar{C}_{rand} - \bar{C}| \quad \text{And}$$

$$C(t+1) = C_{rand} - y \bar{1} \cdot \bar{H} / \text{Correlation}$$

Where, C_{rand} is an arbitrary value, *rand* is a random number and its selection is based on the velocity updating of the PSO algorithm.

Velocity updating of the PSO algorithm: In the typical form of PSO, the velocity vector for a particle is updated according to Gbest and Pbest value. The formulation for

updating the velocity as well as position of the particles in the PSO is given as:

The velocity and the position of particle changed by means following eqn

$$v_i(t+1) = v_i(t) + b_1 rand(Pbest(t) - r_i(t)) + b_2 rand(Gbest - r_i(t))$$

$$r_i(t+1) = r_i(t) + v_i(t+1)$$



Where, V_i is the particle velocity, r_i is the current particle, rand is a random number between (0, 1), b_1, b_2 are learning factor, usually $b_1 = b_2 = 2$.

This mechanism and $|y\bar{1}| > 1$ emphasize exploration and allow the MWO algorithm to perform a global optimum and

$|y\bar{1}| < 1$ for updating the position of the search agents. To spread the whale towards the prey effectively the updation parameter is used.

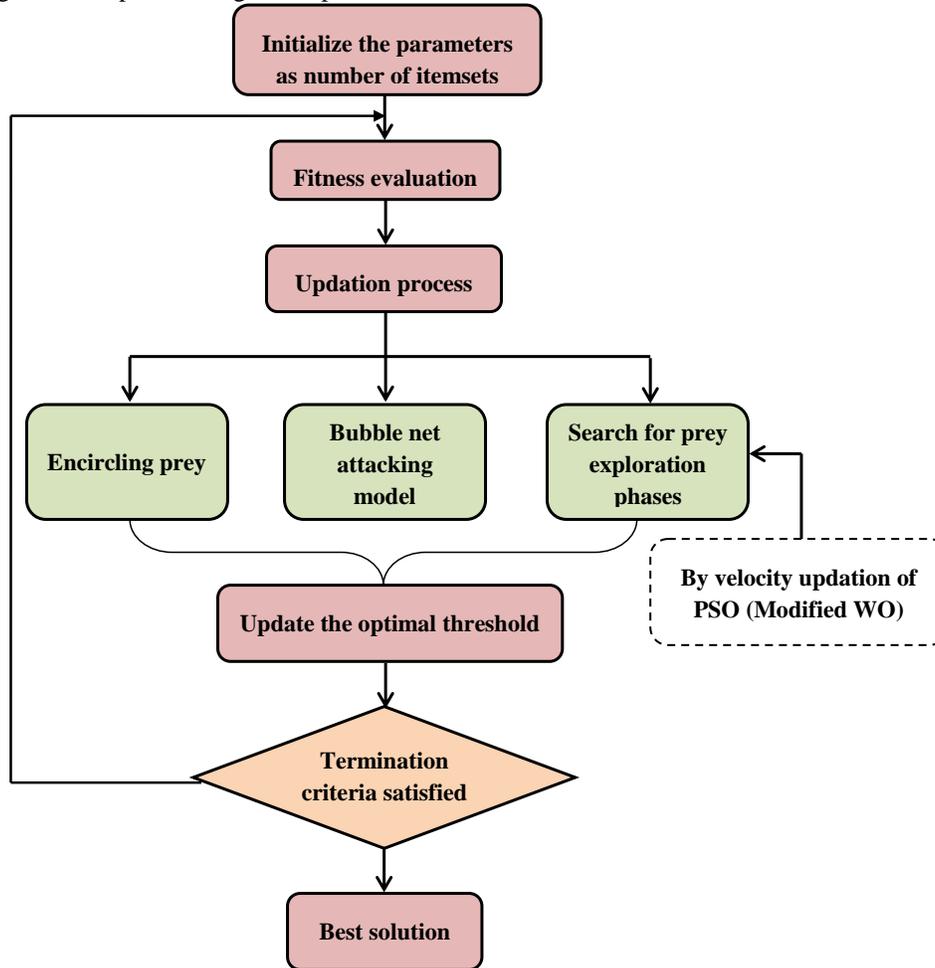


Fig. 3 Flowchart of MWO Algorithm

Termination: The average reaction time is minimized using the proposed algorithm with the help of equation. The task are optimally allocated by using filtered probabilities

$$t_{Total} = t^* + t^r$$

In addition, offloading time and reaction time required to stable the load is reduced.

V.EXPERIMENTAL RESULTS

The proposed HWBO technique is analyzed with some of the performance metrics which is described in the following session.

Evaluation metrics

“Average Completion Time”: An instance needed for finishing the appealed task with cloudlet of minimum obstruction

“Average Transmission Time”: An instance needed for transferring the demand tasks without any hindrance. It is done with more number of cloudlets.

“Average energy Consumption”: complete energy usage during the offloading of tasks from mobile devices.

“Drop Rate”: Total no.of tasks dropped based on mobile users needs.

Here the “NASA” workload with 10 and 15 cloudlets is taken for consideration. The cloudlets are spread over the urban area. It is described in fig.4

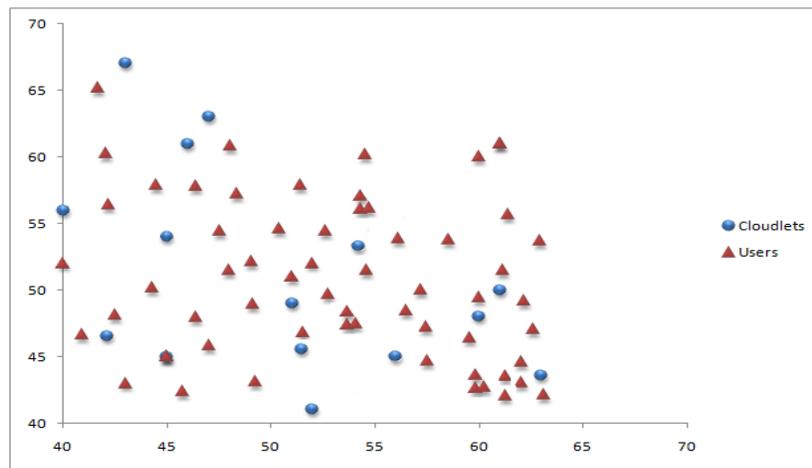
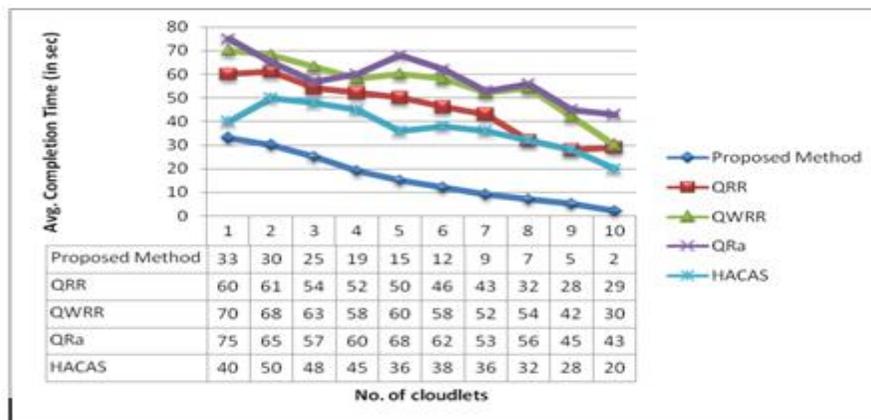


Fig.4 15 cloudlet representation in NASA

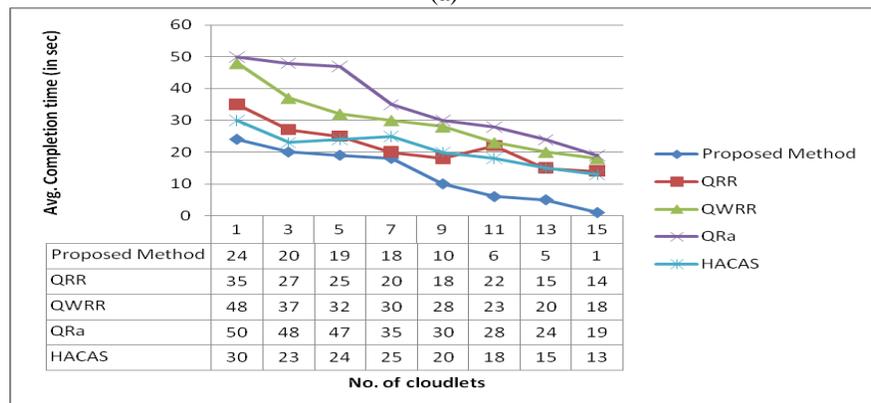
In this research, the performance is analyzed and compared with other existing techniques [10].

The average completion time performance is compared with the existing methods which are presented in Fig 5. If the number of task increases the total completion time is

decreased for proposed HWBO algorithm because of queuing based decision maker process. The conventional methods perform less when the task increases. The optimal assignment of task with minimum time is done using the proposed method.



(a)

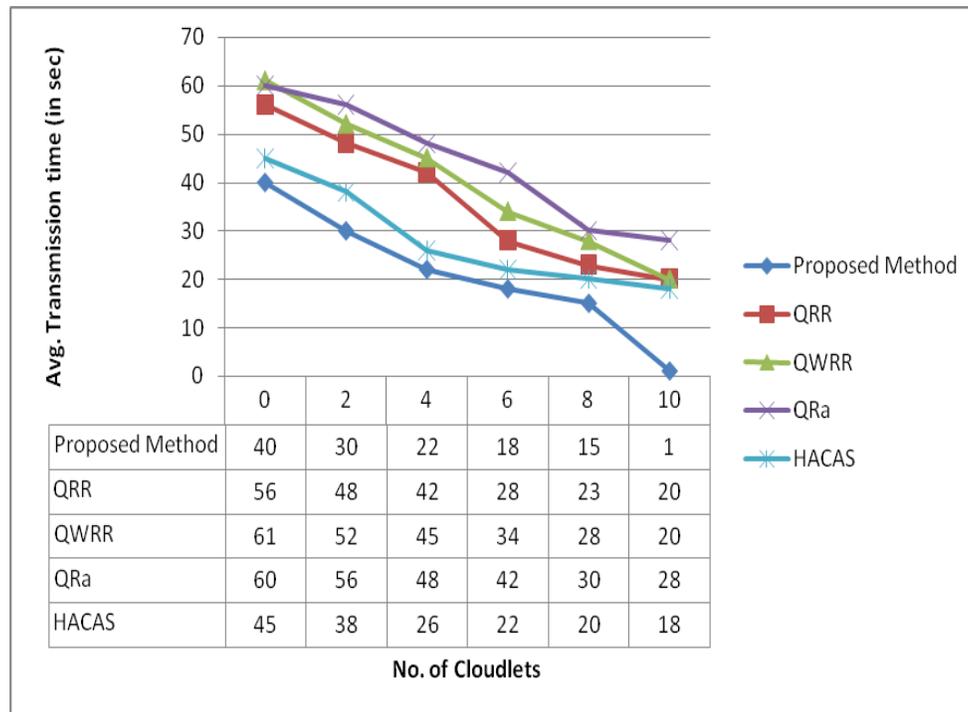


(b)

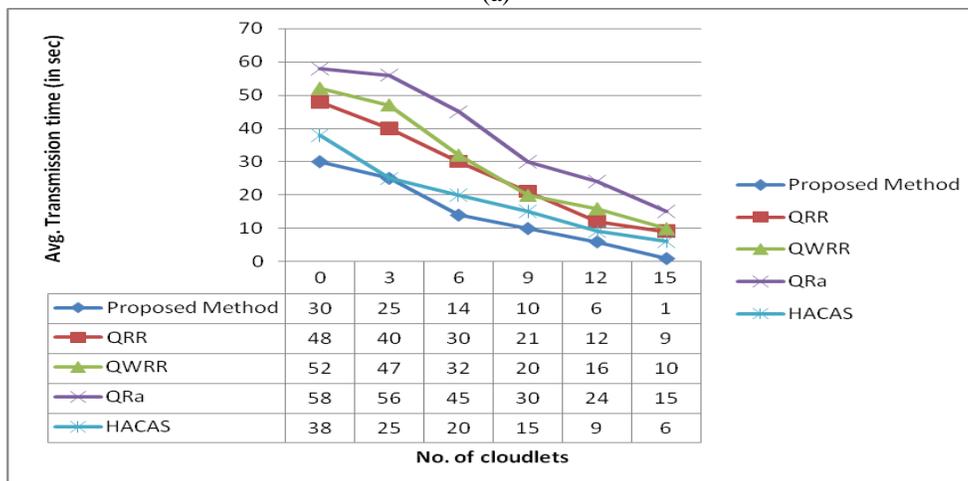
Fig. 5 comparison of avg. completion time

The WOA technique selects the task optimally with minimum response time with maximum cloudlets. The transmission time is analyzed in the following fig. 6 which is compared with existing methods, QRa, HACAS and QWRR etc. The conventional method doesn't consider the distance

amount the users and cloudlets. So the dissemination time is maximum when compared to the proposed research. The tasks are stored in to the nearest cloudlets by considering the distance.



(a)



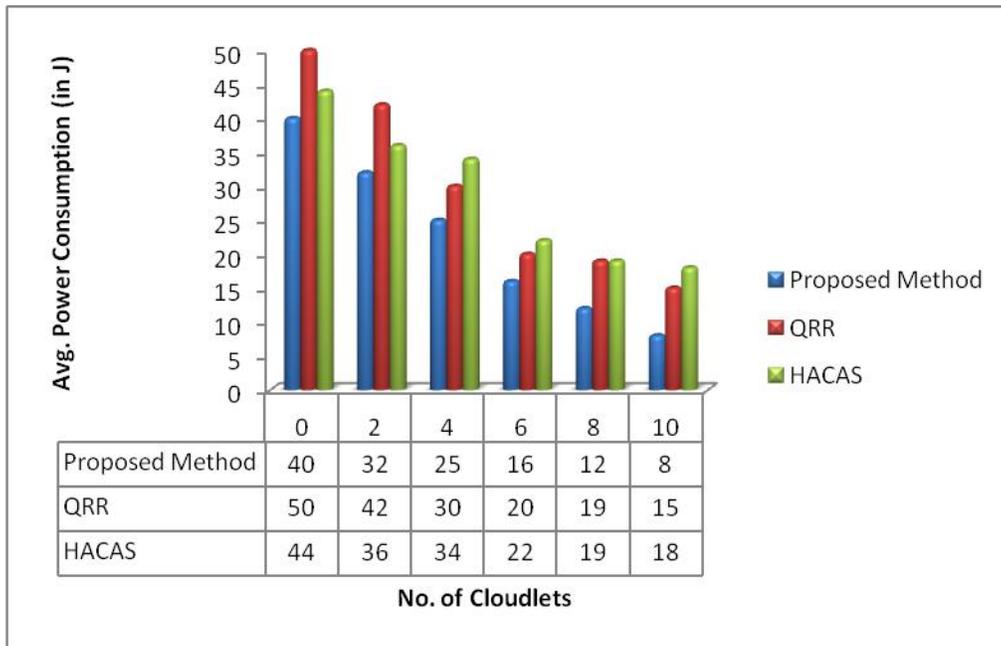
(b)

Fig. 6 Average transmission time analysis

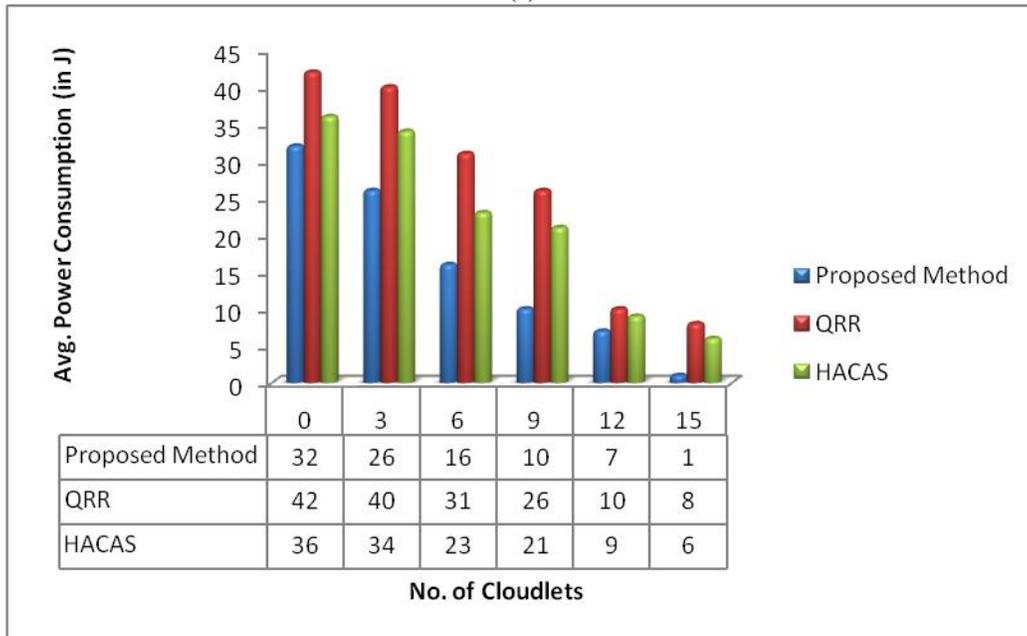
The energy consumed to offload the task is analyzed in the following figure fig. 7. The mobile devices have very less power consumption. So here the task is offloaded using Wifi connections which minimize the energy utilization. The energy consumed is estimated using the following formula

$$Egy = Q_{WiFi} \times \sum_j \sum_k ts_{jk}^{tr}$$

From the figure, the proposed method, consumed very minimum energy when compared to other existing methods.



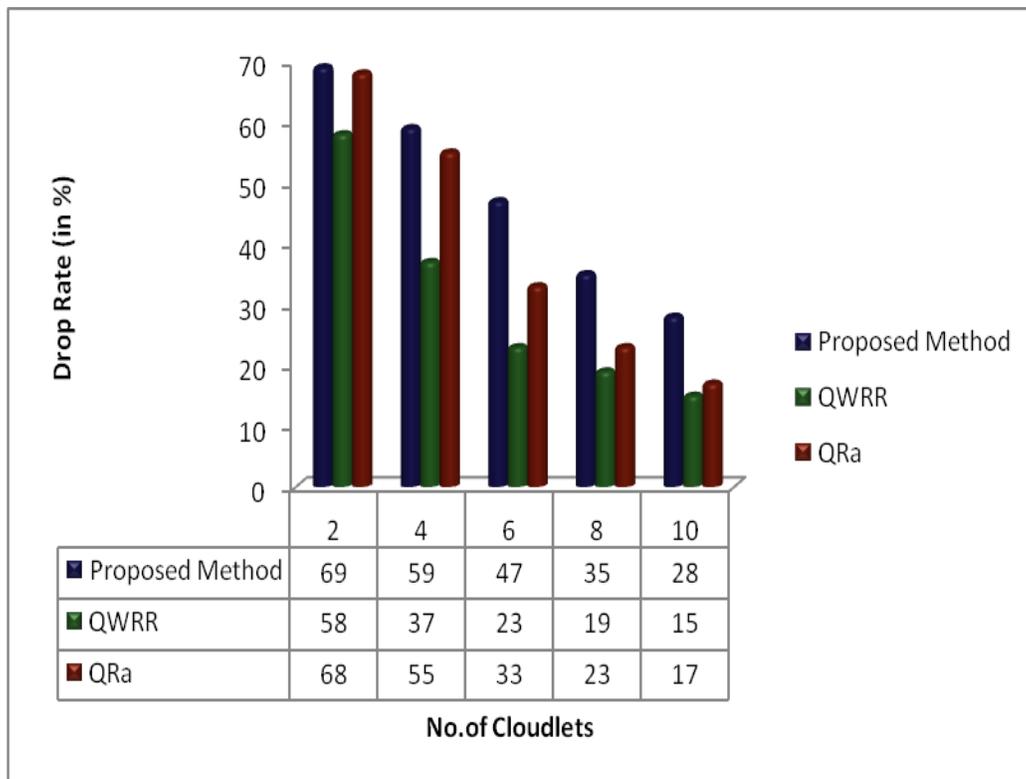
(a)



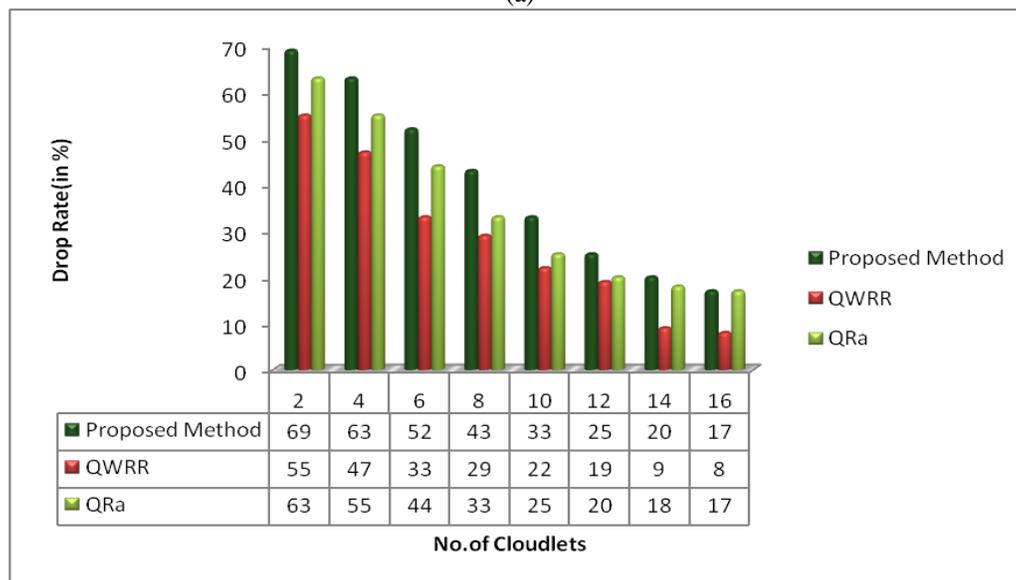
(b)

Fig. 7 Comparison of energy consumption in NASA

The queuing model incorporation with proposed scheme leads to less drop rate during offloading process. It is presented in Fig. 8. In the existing method HACAS provides less drop rate with light load servers in cloud. But if the tasks increases by the incoming demands, the HACAS doesn't works. The proposed scheme's queuing model works well with decision maker unit. It outperforms the existing algorithms



(a)



(b)

Fig. 8 Avg. drop rate comparison

VI.CONCLUSION

On this research, the hybrid Whale-Bee algorithm for most reliable assignment of obligations in MCC surroundings is proposed. For this motive to sort out the performance of 'cloudlets' and the 'public cloud' incorporated with queuing models. The Whale-Bee scheme reduces the processing time of Cloudlets in addition to the verbal exchange phase on the duration of supplying a selection approximately the task of

obligations among the 'cloudlets' and the cellular customers. Based on the performance analysis, the proposed scheme reduces the response time, energy consumption, drop rate etc with the help of HWBO and queuing models. Balancing the load of cloudlets when the number of task increase with more number of user request is the major intension.

Hybrid Whale-Bee Optimization (HWBO) based Optimal Task Offloading Scheme in MCC

The comparison results show the effectiveness of the proposed scheme with other existing algorithms. In future, the work further extends to optimal resource provisioning and utilization.

REFERENCES

- 1 H. T. Dinh, C. Lee, D. Niyato and P. Wang, "A Survey of Mobile Cloud Computing: Architecture, Applications, and Approaches," *Wireless Communications and Mobile Computing*, Vol. 13, No. 18, p. 1587–1611, 2013.
- 2 M. Chen, Y. Wu and A. V. Vasilakos, "Advances in Mobile Cloud Computing," *Mobile Networks and Applications*, Vol. 19, No. 2, pp. 131–132, 2014.
- 3 S. Binitha and S. S. Sathya, "A survey of bio inspired optimization algorithms," *International Journal of Soft Computing and Engineering (IJSCE) ISSN*, Vol. 2, No. 2, pp. 2231–2307, May 2012.
- 4 E. Ahmed, A. Gani, M. K. Khan, R. Buyya and S. U. Khan, "Seamless application execution in mobile cloud computing: Motivation, taxonomy, and open challenges," *Journal of Network and Computer Applications*, Elsevier, Vol. 52, pp. 154–172, 2015.
- 5 E. Ahmed, A. Gani, M. Sookhak, S. H. A. Hamid and F. Xia, "Application optimization in mobile cloud computing: Motivation, taxonomies, and open challenges," *Journal of Network and Computer Applications*, Elsevier, Vol. 52, pp. 52–68, 2015.
- 6 R. Achary, V. Vityanathan, P. Raj and S. Nagarajan, "Dynamic Job Scheduling Using Ant Colony Optimization for Mobile Cloud Computing," *Intelligent Distributed Computing Advances in Intelligent Systems and Computing*, Vol. 321, pp. 71–82, 2015.
- 7 L. Chunlin and L. LaYuan, "Cost and energy aware service provisioning for mobile client in cloud computing environment," *The Journal of Supercomputing*, Springer, Vol. 71, No. 4, pp. 1196–1223, February 2015.
- 8 J. Vilaplana, F. Solsona, I. Teixidó, J. Mateo, F. Abella and J. Rius, "A queuing theory model for cloud computing," *Journal of Supercomputing*, Springer, Vol. 69, No. 1, pp. 492–507, July 2014.
- 9 M. Altamimi, A. Abdrabou, K. Naik and A. Nayak, "Energy Cost Models of Smartphones for Task," *Emerging Topics in Computing, IEEE Transactions*, Vol. 3, No. 3, pp. 384 - 398, 2015.
- 10 X. Wei, J. Fan, Z. Lu and K. Ding, "Application Scheduling in Mobile Cloud Computing with Load Balancing," *Journal of Applied Mathematics*, pp. 13, 2013.
- 11 Rong P, Pedram M, "Power-aware scheduling and dynamic voltage setting for tasks running on a hard real-time system", In: Asia and South Pacific Conference on Design Automation, 2006. IEEE, pp. 6, 2006
- 12 Wu H, Wang Q, Wolter K, "Tradeoff between performance improvement and energy saving in mobile cloud offloading systems. In: 2013 IEEE International Conference on Communications Workshops (ICC), IEEE, pp 728–732, 2013.
- 13 Lee YC, Zomaya AY, "Minimizing energy consumption for precedence-constrained applications using dynamic voltage scaling", In: 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, 2009. CCGRID'09, IEEE, pp. 92–99, 2009
- 14 Lin X, Wang Y, Xie Q, Pedram M, "Energy and performance-aware task scheduling in a mobile cloud computing environment", In: 2014 IEEE 7th International Conference on Cloud Computing. IEEE, pp. 192–199, 2014.
- 15 Van den Bossche R, Vanmechelen K, Broeckhove J, "Cost optimal scheduling in hybrid iaas clouds for deadline constrained workloads", In: 2010 IEEE 3rd International Conference on Cloud Computing, IEEE, pp. 228–235, 2010.
- 16 Tayal S, "Tasks scheduling optimization for the cloud computing systems. IJAEST-international journal of advanced engineering sciences and technologies, Vol. 1, No.5, pp.111–115, 2011.
- 17 Feller E, Rilling L, Morin C, "Energy-aware ant colony based workload placement in clouds", In: Proceedings of the 2011 IEEE/ACM 12th International Conference on Grid Computing, IEEE Computer Society, pp. 26–33, 2011.
- 18 Xu B, Peng Z, Xiao F, Gates AM, Yu J-P, "Dynamic deployment of virtual machines in cloud computing using multi-objective optimization", *Soft Comput*, Vol. 19, No.8, pp.2265–2273, 2015.
- 19 Zhang W, Wen Y, Wu D, "Energy-efficient scheduling policy for collaborative execution in mobile cloud computing. In: 2013 Proceedings IEEE INFOCOM, IEEE, pp. 190–194, 2013.
- 20 Zhang W, Wen Y, "Cloud-assisted collaborative execution for mobile applications with general task topology", In: 2015 IEEE International Conference on Communications (ICC). IEEE, pp. 6815–6821, 2015.
- 21 Giurgiu I, Riva O, Juric D, Krivulev I, Alonso G, "Calling the cloud: enabling mobile phones as interfaces to cloud applications", In: ACM/IFIP/USENIX International Conference on Distributed Systems Platforms and Open Distributed Processing. Springer, pp. 83–102, 2009
- 22 Xie J, Dan L, Yin L, Sun Z, Xiao Y, "An energy-optimal scheduling for collaborative execution in mobile cloud computing", In: 2015 International Conference and Workshop on Computing and Communication (IEMCON), IEEE, pp. 1–6, 2015.
- 23 M. F. Arlitt and C. L. Williamson, "Web server workload characterization: the search for invariants," in *Proceedings of the 1996 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, New York, NY, USA, 1996.