Optimization of Computation and Communication Driven Resource Allocation in Mobile Cloud

R. Shankar
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
shankarrajendran75@gmail.com

Tharani Vimal
Department of Information Technology, St. Peter’s College of Engineering and Technology, Avadi, Chennai, Tamil Nadu, India
tharanivimal2010@gmail.com

Received: 29 January 2022 / Revised: 09 March 2022 / Accepted: 13 March 2022 / Published: 30 April 2022

Abstract – With the emergence of accessing Smartphones in day-to-day life, Mobile Cloud Computing (MCC) technology has become popular with the advantage of resolving the resource constraints in mobile devices through the offloading method. The existing models have presented the different resource allocation solutions to ensure the seamless execution of the applications for the resource-constrained mobile devices with the Quality of Service (QoS). The optimization of resource allocation is the process of potentially allocating remote resources to mobile users without violating the Service Level Agreements (SLAs). However, resource allocation is still becoming a major constraint in the Mobile Cloud (MC) data centers due to higher consumption of energy and time factors during the execution of mobile requests on the remote cloud. The consumption of the energy and response time of the offloaded tasks or applications heavily relies on the cloud resource allocation for the mobile users. Hence, Resource Allocation Optimization (RAO) emerged as the significant objective to select the appropriate cloud resources for the requested tasks to increase the lifetime of the devices with improved time efficiency. Thus, this work focuses on optimizing MC resource allocation by optimizing the allocation of both the computation and communication resources. The proposed RAO model considers two potential factors, such as the energy and response time while allocating the computational and communicational resources. Initially, the Energy and Response time-driven RAO (EARO) approach prioritizes the request generated from the mobile users based on the estimated execution time. Modeling the Estimated Communication and Execution Time (ECET) algorithm tends to allocate the cloud resources and accomplish the minimal response time of the application requests. The EARO approach intends to minimize the execution time as well as the response time towards the target of alleviating the energy consumption during the resource allocation. Moreover, it selects the resources for the inter-VM communication with the knowledge of the minimal migration time ensuring bandwidth resources. Thus, EARO preserves the device's energy with minimal application completion time. The experimental results illustrate that the time efficiency of the proposed EARO model outperforms the existing resource allocation model in the MC environment.

Index Terms – MCC, Resource Allocation, Computation, Communication, Optimization, Energy Consumption, Bandwidth, Response Time.

1. INTRODUCTION

Mobile Cloud Computing (MCC) [1] paradigm enables the execution of resource-constrained mobile requests remotely on the large-scale and on-demand cloud to preserve the device energy through the offloading method. Numerous scientific mobile applications utilize cloud resources, involving gaming, finance, linguistics, economics, social networks, engineering, geophysics, and mathematics. Moreover, mobile applications are emerging in various fields of computing. The MCC enables improved reliability, scalability, data storage capacity, processing power, and battery lifetime with the advantages of the multi-tenancy and dynamic service provisioning characteristics. Even though offloading method relieves the pressure of the device by enabling the moving of computation energy from the mobile device to the cloud [2], the energy consumption of the cloud server increases with the computations in the data center. From the perspective of carbon emission control and environmental impact control, energy consumption heavily relies on the task or computation instead of the system. Hence, reducing the energy consumption is crucial to potentially satisfy the mobile users. The energy consumption becomes a key factor in demanding the operation cost for the cloud services from the mobile users. Traditional cloud and Mobile Cloud (MC) resource allocation models focus on the different factors, including scalability, data confidentiality, customer satisfaction, battery consumption, number of clients per server, and minimum SLA violations. Several resource allocation works have
developed different strategies from the aspect of together allocation of Virtual Machine (VM) resources on the available host to balance the resource utilization and SLA violations. Furthermore, resource allocation with live migration techniques have been emerged to modify the allocated VM resources during the execution, moving the VMs from one host or Physical Machine (PM) to another PM. Although the number of migrations increases the overall computation time and energy consumption, optimally allocating the resources is vital with the limited number of migrations in the MC environment.

Several existing researchers have targeted ensuring the Quality of Service (QoS) while providing resources to the mobile users in the distributed computing environment [3]. Hence, Resource Allocation Optimization (RAO) plays a vital role in significantly improving the performance of the MCC. Several conventional RAO models have been focused on the potential factors of energy consumption, response time, profit, and latency when dealing with the unlimited resource distribution to more mobile users [4]. In an MC environment, the optimization of cloud resource provisioning is prominent to meet the QoS requirements of the mobile user. In the MC environment, an optimal resource allocation is essential to reduce the energy consumption on mobile devices, execution time, and latency [5].

In the cloud or MC environment, optimal resource allocation satisfies the mobile users’ anticipations without SLA violations. The existing RAO models [6] have addressed the energy and response time constraints for optimizing the MC environment. However, several existing RAO models lack to jointly optimize the computation and communication resources in the MC environment [7]. In addition, optimizing the MC resource allocation with the only awareness of computational resources is inadequate because the communication time greatly impacts the device energy and response time [8]. Hence, it is necessary to examine both the computation and communication resources to execute the mobile application requests in the MC environment optimally. Thus, this work targets optimizing the resource allocation through prioritized request-based cloud resource allocation and the bandwidth resource-based task migration in the MC. The proposed computation and communication-driven resource allocation model ensures improvements in the performance of the device energy and response time in the MC environment. It incorporates three major processes such as prioritization, ECET based resource allocation, and optimization through migration.

1.1. Objectives of the Research Work

- To optimize the resource allocation model for the computation and communication resources in the MC environment
- To alleviate the energy consumption and response time by optimizing the resource allocation in the MC.

1.2. Contributions of the Research Work

The main contributions of the proposed Energy and Response time-driven RAO (EARO) methodology are presented as follows.

- This work presents the RAO model incorporating the computation and communication resource allocation in the MC environment by developing the ECET algorithm.
- The EARO model modifies the allocation order based on the request size of the mobile applications to achieve the minimum response time. The EARO approach selects the time-aware cloud resources for the compute-intensive mobile applications to mitigate the energy and response time from considering the computational and communicational resources.
- Effectively selecting the target hosts in the cloud minimizes the migration time among the inter-VM resources considering the communication resources and thus, ensures the optimized resource allocation in the MC environment.

1.3. Paper Organization

This section outlines the organization of this work, which is structured as follows. Section 2 reviews the existing research works in the resource allocation and RAO model in the MC environment. The optimization problem in the MC resource allocation is stated in Section 3. Section 4 discusses the system model and overall process of the proposed RAO methodology in detail. The experimental results of the EARO approach and the existing approach are illustrated in Section 5. Finally, Section 6 concludes this work.

2. RELATED WORKS

This section reviews the existing cloud, MCC resource allocation, and RAO approaches.

2.1. Resource Allocation Approaches

eTime model [9] leverages the energy-efficient data transmission between the cloud and mobile devices. Lyapunov optimization minimizes the device energy during the application transmission through the wireless network in an MC environment. Optimal and fair service allocation [10] enables a tiered MC approach to ensure QoS requirements and improves the scalability and device performance. The fair service allocation model considers the Location-time workflows (LTW) considering power consumption, delay, and cost of mobile application execution. ThinkAir framework [11] enhances the method-level offloading of application parallelization, reducing mobile device energy consumption. The recursive algorithm allocates the resources dynamically in terms of on-demand resource provisioning.
using multiple VM images. An energy-efficient cooperative offloading model (E2COM) schedules the tasks to minimize energy consumption of mobile user and internet data traffic redundancy in the MC environment. The task scheduling algorithm applies the pricing method and Lyapunov optimization to minimize WLAN energy [12]. Context-aware multi-objective resource allocation model [13] mitigates the energy and the completion time while offloading the tasks in the MC environment. It acquires the contextual information from the execution environment through the Non-dominated Sorting Genetic Algorithm II and the entropy weight to handle the offloading process. Consequently, it minimizes the energy along with completion time based on the deadline, budget, and residual energy constraints. Energy and time-efficient task offloading based resource allocation model [14] addresses the time cost and energy minimization problem by utilizing the fog and cloud architecture advantages. It improves the energy consumption completion time by performing the computation offloading and transmission power allocation. The research work [15] performs the joint resource allocation to minimize the energy in the Cloud Radio Access Network (C-RAN). Incentive-Compatible Auction Mechanism (ICAM) [16] effectively allocates the resources between the service users, the buyers, the service providers, and sellers. It satisfies the service demands and analyzes the pricing by allocating the cloudlet resources. The joint multi-resource allocation method [17] applies the Semi-Markov Decision Processing (SMDP) in the cloudlet system to maximize the system performance and improve the QoS. It enhances MCC performance through optimal wireless resource allocation and computation decision-making. The joint offloading and resource allocation model [18] presents the three-step algorithm such as alternating optimization, semidefinite relaxation, and sequential tuning. It allocates the processing and network resources in the MC with the reference of the computing access point. The research work [19] presents the Energy Efficient dynamic Resource Allocation (EERA) model to optimally locate the virtual machine, which tends to the modeling of the green cloud with the reduced energy in the MC data center. By consolidating the resources in the MC, it effectively manages the energy consumption based on selecting the target resources.

2.2. Resource Allocation Optimization Approaches

A joint optimization algorithm solves the RAO problem considering the energy consumption, latency, and execution time in MCC. Low complexity backward induction optimizes the integer part overall performance of the resource allocation and partitioning [20]. An agent-based optimization framework [21] optimizes the request execution time to minimize the processing time and energy consumption in the MC environment using a mobile-agent-based partition. A novel framework introduces the joint allocation approach that optimizes the allocation of the network and processing resources in the MC environment. Joint optimization considers the multi-user application transmit power minimization, latency reduction, and energy of the devices [22]. The optimal joint multiple resource allocation model minimizes the request loss probability and overall resource utilization in the cloud [23]. An Optimal Cloud resource provisioning (OCRP) [24] algorithm minimizes the resource provisioning total cost in cloud computing. OCRP algorithm includes Benders decomposition, deterministic equivalent formulation, and sample-average approximation which formulate the multi-stage resource management of VMs optimization efficiently. The energy-optimal framework [25] minimizes the device energy consumption while transmitting and executing the mobile application requests in the cloud. Task tolerance-based energy consumption optimization approach increases resource utilization and minimizes energy consumption using the resource scheduling optimization algorithm.

Optimization depends on the increased task tolerance of resource execution in cloud computing [26]. Resource allocation framework [27] solves the optimization problem of mixed-integer programming (MIP) problem in a cloud environment. Resource mapping procedure offers QoS-aware resource provisioning and cost-efficient computing resources. Off-the-Cloud service optimization (OCSO) [28] discusses the heuristic-based off-the-cloud to optimize resource consumption using an intelligent resource scaling algorithm. Bandwidth along with energy-aware resource allocation model [29] optimizes the energy utilization in the C-RAN through the bandwidth power allocation and the energy-aware resource allocation. It satisfies the QoS requirements based on allocating the feasible bandwidth resources with the reduced number of active virtual machines. The RAO model [30] selects the cloudlet resources by applying mixed-integer linear programming, optimizing the execution latency. It allocates the resources based on the optimized reward and resource usage in the selected cloudlet in subsequence. The research work [31] jointly optimizes the computation and the wireless bandwidth resources to enhance the QoS of MCC. The cloud-assisted mobile edge computing model [32] performs the computation offloading and the optimal resource allocation. It presents the Collaborative Computation Offloading and RAO (CCORAO) scheme to address the NP-hard and the non-convex constraints in the vehicular networks. Joint Computation Offloading and RAO (JCORAO) model [33] develops the optimal offloading strategy, allocation of uplink subchannel and the uplink transmission power, and scheduling of computational resources. It jointly allocates the computational and communicational resources in the edge computing environment by proving the Nash equilibrium.

The research work [34] develops the Properly Identifying Optimized Cloudlet in Mobile cloud computing (PIOCM) model to offload the mobile application requests based on the
waiting time and energy. By applying the Support Vector Machine (SVM), it predicts the resourceful cloudlet and selects the optimized cloudlet considering the internal and external conditions of the network profile. Energy Efficient Mobility Management in MCC (E2M2MC2) model [35] employs the Elective Repeat Multi-Objective Optimization (ERMO2) algorithm to select the best cloud resource in the perspective of the packet loss, delay, jitter, communication cost, bit error rate, network load, and response time. It ensures the energy efficiency for the network resources over the increased user mobility in the 5G heterogeneous networks. Advanced Q-MAC (AQ-MAC) model [36] performs the mobile data offloading among the cloudlets and leverages the resource allocation for the mobile devices in the cloud environment. By considering the position and velocity of the users, it improves the Quality of Service (QoS) along with higher reliability and efficiency.

3. PROBLEM STATEMENT
MCC often confronts the inadequate processing capability and battery life of the resource-constrained mobile device. Resource allocation and request dispatching have gained significant attention among researchers due to the limited resource availabilities of the device. Performance optimization is an important aspect during application execution on heterogeneous VMs. Hence, RAO necessitates enhancing the resource allocation in the cloud server while executing the applications to save energy consumption. Lack of analyzing the utility during the resource allocation affects the RAO in terms of execution time, energy, and response time. In MCC, the mobile user meets congestion problems due to network delay and bandwidth limitations. Bandwidth resource allocation is the significant optimization constraint among different resource characteristics in the MC environment. In addition, considering or optimizing the communication resource while executing the mobile application requests is a challenging task rather than initially selecting the communication resources in the MC. Hence, it is essential to mitigate or save energy and minimize the response time by allocating the computation and communication resources.

4. THE PROPOSED RESOURCE ALLOCATION OPTIMIZATION MODEL
The proposed methodology employs the ThinkAir architecture [11, 37] to implement the RAO model. With the enormous advantages in the ThinkAir architecture, the EARO approach adopts this architecture for RAO. Dynamic adaptation, ease of use, cloud-based performance improvement, and dynamic scaling are the potential benefits of executing compute-intensive mobile application requests. In essence, the rapidly changing characteristics of the mobile computing environment require efficient and quick adaptive architecture to ensure seamless execution even when there is a loss of connection. ThinkAir architecture avoids framework misuse through its simple interface and enables novice and less skilled developers to cope with the ever-increasing mobile application market. Moreover, it fulfills the performance requirements of the customers by dynamically scaling the computational power in the remote server. Figure 1 shows the proposed RAO in ThinkAir Architecture.

![Figure 1 Proposed RAO in ThinkAir Architecture](image-url)
establishing the connection between Smartphone and the remote Cloud. Profiler contains device, program, and network for monitoring the status of the mobile device. Moreover, the profiler is associated with the energy model of the smartphone. ThinkAir comprises the application server, server handler, and dynamic object input stream in the cloud server. The server handler receives the mobile user offloaded requests from the smartphone and reports executed requests to the smartphone. The application server manages application offloaded code from the smartphone. Cloud server consists of a dynamic object input stream for managing the exceptions during mobile user request execution. The EARO approach performs the RAO on the cloud execution controller that executes the requests offloaded from the mobile device based on the application server offloaded code and server handler resources.

4.1. System Model

MCC system comprises ‘N’ number of mobile devices and cloud in which smartphones consider mobile devices through the connection of Wi-Fi network. In MCC system, each mobile device (M_i) has unique ID for identification where, i = \{1,2,.....,N\}. Cloud consists of multiple servers with different capabilities in a data center. The resource pool contains unlimited resources, including the CPU, memory, and network resources in the cloud. Cloud resource pool consists of \(j^\text{th}\) Physical machines (PMs) and Virtual machines (VMs) resources for allocating the resources to the mobile user. In this proposed system, the data center is modeled as the cloud provider in which the servers are the heterogeneous resources for providing the service to heterogeneous requests. Cloud provider has the pool of resources \(R_j\) where, \(j = \{1,2,....., M\}\) for providing resources to the mobile user.

Table 1 Notations Used in the Proposed Methodology

| Symbols/Notations | Description |
|-------------------|-------------|
| \(RP\)            | Request Pool |
| \(MU_k\)          | Mobile user request |
| \(VM_k\)          | Cloud VM resources |
| \(E_{ET}\)        | Estimated execution time |
| \(W_T\)           | Waiting time |
| \(E_{CT}\)        | Estimated completion time |
| \(VM_k\)          | VM to be migrated |
| \(T^M_{VM}\)      | Tasks to be migrated from VM_k |
| \(VM_i^{U}\)      | Utilization of VM_i |
| \(VM_i^{H}\)      | Host VM |
| \(VM_i^{S}\)      | Specification of VM_i |
| \(VM_i^{C}\)      | Specification of VM_i |
| \(P_S\)           | Processing speed |
| \(\beta_T\)       | Threshold transmission time |
| \(\beta_S\)       | Threshold speed |
| \(BW_k\)          | Bandwidth |
| \(\alpha_e\)      | Threshold time |

In cloud storage, resources involve computing power, storage, and network bandwidth. The energy of the mobile device and RF components are inter-related to each other. MCC optimization framework depends on the computation and communication components. The bandwidth between primary VM and host VM is known as inter-VM bandwidth. Mobile device application code offloading method and remote server execution overcome resource limitation and long spending time in high power condition of the mobile devices. Each server has a different processing speed (\(P_S\)) and energy level in the cloud data center. Hence, the EARO approach selects the resources in the cloud server based on the application requests. In the MC environment, energy consumption depends on the computation and communication requirements. Table 1 depicts the details of the notations or symbols used in this work.

4.2. The Proposed EARO Methodology

The EARO approach formulates computation and communication-guided RAO models in an MC environment. In the MC environment, the EARO approach considers the computation and communication constraints during the resource allocation for achieving high performance. In the EARO, the optimization phase contains energy consumption optimization and response time optimization. Firstly, the optimization of the energy consumption is responsible for the resource selection and task migration during the computation process. The proposed task migration is based on selecting the minimum transmission time of inter-VM bandwidth for migrating the tasks to the host VM in the remote server. Secondly, the response time optimization concentrates on resource selection at the communication level. It selects the high processing speed of RF component bandwidth for transmitting the application requests in the MC environment. The proposed EARO approach optimizes the processing time to reduce the energy consumption by considering the inter-VM bandwidth-based task migration. With the advantage of task migration in addressing the data loss, time, and cost consumption during the execution, the EARO approach adopts the migration process for RAO. Thus, the EARO approach ensures the optimal resource allocation by the computation and communication constraints-based task migration. Figure 2 illustrates the optimization system for the energy consumption and response time in MC.

In the MC environment, the RAO problem depends on the QoS. The cloud server receives heterogeneous mobile user requests from a mobile device for remote processing in the cloud in the proposed optimization system. Request pool comprises more number of requests with a unique ID and required resource capabilities for processing the same
application. Request pool dispatches the service requests (MU'R') to corresponding server resources (R_j), relying on the capabilities of the request and resources. The EARO approach concentrates on both the VM and PM resources for optimizing resource allocation. The proposed framework focuses on both the execution time and transfer time of the tasks. In the proposed methodology, task execution time is computed from the task processing ability on the corresponding VM characteristics in the MC. Task transfer time is computed from the time taken by the inter-VM bandwidth in the remote cloud and the bandwidth of RF components.

4.2.1. Time-Aware Resource Allocation

In the MC environment, the RAO problem depends on the QoS. The cloud server receives heterogeneous mobile user requests from a mobile device for remote processing in the cloud in the proposed optimization system. Request pool comprises more number of requests with a unique ID and required resource capabilities for processing the same application. Request pool dispatches the service requests (MU'R') to corresponding server resources (R_j), relying on the capabilities of the request and resources. The EARO approach concentrates on both the VM and PM resources for optimizing resource allocation.

The proposed framework focuses on both the execution time and transfer time of the tasks. In the proposed methodology, task execution time is computed from the task processing ability on the corresponding VM characteristics in the MC. Task transfer time is computed from the time taken by the inter-VM bandwidth in the remote cloud and the bandwidth of RF components.
ThinkAir addresses the problem of scalability using VM instances creation of smartphones in the cloud. It supports on-demand resource allocation for various requirements of mobile user requests. ThinkAir architecture dynamically allocates the resources to mobile user requests depending on the requirements using execution parallelization. The EARO approach follows the initially $E_{CT}$-based $MU_{R_i}$, VM selection, and $E_{ET}$ based VM allocation. As modeled in Equation (1), the proposed estimated completion time ($E_{CT}$) depends on the estimated execution time ($E_{ET}$) on VM resource, waiting time in queue ($W_T$), migration time on inter-VM bandwidth ($M_T$), and transmission time on channel bandwidth ($BW_T$).

$$E_{CT} = E_{ET} + W_T + M_T + BW_T$$  \hspace{1cm} (1)

The proposed algorithm selects the mobile user request ($MU_{R_i}$) and $VM_i$ resource that executes the corresponding mobile user request at a minimum completion time ($E_{CT}$) than other VMs. If the proposed algorithm selects more than one $VM_i$ resource, it selects the first initiated $VM_i$ for executing the request. Furthermore, the EARO approach calculates the mean ($\mu$) and standard deviation ($\sigma$) of the estimated completion time ($E_{CT}$) to allocate the resource with a minimum completion time using Equations (2) and (3).

$$\mu = \frac{\sum_{i=1}^{N} E_{CT}^i}{N}$$ \hspace{1cm} (2)

$$\sigma = \sqrt{\frac{\sum(E_{CT}^i-\mu)^2}{N}}$$ \hspace{1cm} (3)

The proposed algorithm allocates the $VM_i$ using the higher priority of minimum estimated execution time ($E_{ET}$) if the mean time of estimated completion time is greater than the standard deviation. Suppose the request list satisfies the condition ($\sigma \mu$), it comprises the tasks with small size. In other cases, the proposed EARO model exploits the allocation order of maximum estimated execution time ($E_{ET}$) and minimum estimated execution time ($E_{CT}$). Hence, it reduces the makespan of the application due to task size-based allocation priority.

4.2.2. Resource Allocation Optimization

The EARO approach optimizes the two constraints such as the energy and time, particularly response time through the computation and communication resource-aware algorithms. The proposed energy consumption optimization model aims to reduce the request completion time in the remote server and to minimize the energy consumed by the mobile device. The proposed optimization technique-based estimated completion time ($E_{CT}$) reduction minimizes the energy consumption of each request due to the selection of resources with minimum $E_{CT}$. The proposed optimization technique concentrates more on selecting and placing the VMs in the cloud server. The proposed algorithm selects the $PM_j$ for migrating the resource if $PM_j$ has the maximum estimated execution time ($E_{ET}$). It chooses the $VM_k$ (VM to be migrated) resource among all allocated VMs ($VM_{k^i}$) from that $PM_j$. The EARO approach selects the $VM_k$ if $VM_j$ ($VM_{k^j}$) is greater than the threshold utilization level ($\alpha_i$). It selects the set of migratable tasks from $VM_k$ to achieve an earlier task completion time. Migration time reduction alleviates the completion time as well as the energy consumption in the MC. The proposed algorithm identifies the host $VM_j$ if it satisfies a few conditions. The conditions are 1) Host (target) VM must contain the similar characteristics of $VM_i$ to execute tasks, 2) target VM is not placed in idle $PM_j$, 3) target VM utilization level is not more than threshold level after the placement of migrating tasks, 4) $E_{ET}$ is not greater than the primary $PM_j$ while executing the migrating tasks on that $PM_j$, 5) inter-VM bandwidth has the minimum transmission time while migrating the tasks from primary VM to target VM. The resource allocator selects the target VM from the allocated VMs for the similar characteristics of source VM and minimum migration time on bandwidth channel during resource allocation.

Energy consumption of the application request (i,j) depends on the transmission energy ($E_{T}^i$), receiving energy ($E_{R}^i$), and transmission rate ($R_{T}^i$) of the channel at a time. The device energy consumption ($E_{D}^i$($MU_{R_i}$)) contains transmitting and receiving energy on the channel bandwidth and transmitting energy consumption ($E_{T}^i$) from ‘i’ to ‘j’ calculated as ($F_{in}/R_{T}^i$).

$$E_{D}^i(MU_{R_i})=E_{T}^i + E_{R}^i$$ \hspace{1cm} (4)

In Equation (4), $F_{in}$ and $F_{out}$ refer to the transmitting and receiving file sizes. Cloud server energy consumption is the sum of the energy consumed during the computation and the communication, expressed in Equation (6). In an MC environment, offloading energy consumption is modeled in Equation (5) over the ‘n’ servers.

$$E_{D}^i(MU_{k_j})=E_{T}^i + E_{C}^i(MU_{R_i}) + E_{R}^i$$ \hspace{1cm} (5)

$$E_{D}^i(MU_{R_i})=E_{T}^i + E_{C}^i(MU_{R_i}) + E_{R}^i$$ \hspace{1cm} (6)

Equations (5) and (6) compute the energy consumed among the cloud resources and the energy consumed during the offloading, respectively, in which $E_c^i$ ($MU_{R_i}$) denotes the energy consumption of the computation in the MC environment. According to the proposed EARO model, the energy of inter-VM bandwidth relies on the transmission speed and data transmission time. The proposed algorithm predicts the inter-VM bandwidth with high speed and minimum transmission time for migrating the resources to the cloud server. The EARO approach reduces the overall energy.
consumption when selecting a high processing speed bandwidth. Energy consumption of mobile devices spends more energy for computing and communicating resources in the cloud server. Computation and communication resource-induced optimization algorithms minimize the cloud server response time of each request. Transmission time (\(T_i\)) of each application request application to the \(i^{th}\) resource is based on the average length of mobile user request \(b_i\) bits, delay time (\(d_i\)) between the device and the cloud, communication channel bandwidth (\(b_0\)) of a particular request (\(R_i\)) on the transmission channel.

\[
T_i = d_i + b \times \frac{R_i}{b_0}
\]  

Equation (7) finds the data transfer time for remote processing in the cloud. Mobile users request transmission time, and the number of bits transmitted per second identifies the transmission speed of bandwidth. Response time minimization is targeted to minimize the completion time of each request. The transmission channel bandwidth allocation depends on the minimum estimated transfer time (\(E_T\)) on the communication channel over the network traffic. Each bandwidth has a different processing speed for transmitting the request or data. The proposed optimization technique finds the minimum estimated transfer time of bandwidth among multi-channel bandwidths. Response time optimization depends on the estimated delay of the server during the selection of the server in a cloud environment. The proposed algorithm minimizes the latency in terms of transmission delay, Queuing delay, and migration delay during computation and the communication in the MC. Hence, the proposed EARO approach mitigates the response time while optimizing the energy. Response time optimization mostly concentrates on reducing queuing delay and migration delay during resource allocation. Transmission delay (\(d_i\)) measures the ratio between the request size and the transmission rate. The transmission time of each request is computed from the factors of the processing speed and file size of each transmitting and receiving data between cloud provider and mobile user. It focuses on high transmission speed with a minimum transfer time of channel bandwidth between mobile devices and cloud providers. Transmission time (\(BW_T\) and \(M_T\)) reduction at the communication stage and execution time (\(E_T\)) reduction at the computation stage improves the response time performance in the MC environment.

**Input:** Mobile user request with file size, VM resources with MIPS

**Output:** Optimized allocated resources

Initialize RP

for all \(MU_R^i\) offloaded on cloud do

for all VM resources (\(VM_j\)) do

Find \(E_T\) of \(MU_R^i\) on \(VM_j\)

end for

Prioritize all \(MU_R^i\) in ascending order depends on \(E_T\)

end for

While \(MU_R^i\) is not empty then

for all \(MU_R^i\) offloaded on cloud do

for all VM resources (\(VM_j\)) do

// Waiting time of the \(MU_R^i\) and Allocation
if (\(MU_R^i(W_T) \geq \alpha\))

Select \(MU_R^i\) in FCFS order

end if

Calculate \(E_{CT}\) of \(MU_R^i\) on \(VM_j\) based on \(W_T, BW_T, E_T, M_T\)

end for

Find \(MU_R^i\) and \(VM_j\) based on min(\(E_{CT}\))

if select No.of(min(\(E_{CT}\))(\(VM_j\))>1) then

Select \(VM_j\) on FCFS basis

end if

Calculate \(\sigma\) based on the \(E_{CT}\) of \(MU_R^i\)

if (\(\sigma < \mu\)) then

Allocate \(VM_j\) to \(MU_R^i\) based on Min(\(E_T\))

else

Allocate \(VM_j\) to \(MU_R^i\) based on Max(\(E_T\))

end if

end for

end while

// Energy consumption based Task Migration
for all \(VM_j\) on \(PM_j\) do

Select max(\(E_T\)) among all active servers

if (\(VM_j^{U} \geq \alpha_U\)) from that selected \(PM_j\) then

Select \(VM_k\) <- \(VM_j^A\) and \(T_i^M\) <- \(VM_k\)

Select \(VM_j^{H}\) for \(VM_k\) placement

Allocate \(VM_j^{H}\) to \(T_i^M\) of \(VM_k\)

// Selection of \(VM_j^{H}\) process

if \(VM_j (VM_j^{S} = VM_k^{s})\) then
RESEARCH ARTICLE

Select that VM\textsubscript{j} as VM\textsubscript{j}^H

\hspace{1cm} \textbf{// Select Bandwidth for Migration and Transmission}
\hspace{1cm} \textbf{if} VMBW\textbf{(}P_S \geq \beta_S \& \& E_T \leq \beta_T\textbf{)} \textbf{then}
\hspace{1cm} Select BW\textsubscript{i} for T\textsubscript{i}^M
\hspace{1cm} \textbf{end if}
\hspace{1cm} T\textsubscript{i}^M \rightarrow VM\textsubscript{j}^H
\hspace{1cm} \textbf{end if}
\hspace{1cm} \textbf{end for}
\hspace{1cm} \textbf{end for}

Algorithm 1 Pseudocode of the EARO Model

The proposed EARO approach formulates the resource allocation of the server and network resources to the mobile application requests. The proposed bandwidth allocation depends on the minimum transfer time (E\textsubscript{T}) for transferring the requests at the task migration and transmission stage on RF components. The proposed algorithm selects inter-VM bandwidth and communication bandwidth if processing speed (P\textsubscript{S}) is more than threshold speed (\beta_S) and data transfer time (E\textsubscript{T}) is less than threshold transfer time (\beta_T). Algorithm 1 describes the pseudocode of the proposed EARO approach. Mobile user request transmission channel consumes more energy for optimally allocating the resources under cloud server execution. Transmitting and receiving time delay variation depends on the number of data transmission in bits per second. The proposed optimization algorithm minimizes the migration time using task migration and reduces the transmission and receiving time through the bandwidth resource consideration. Migration time reduction decreases the application processing delay in the cloud. Thus, it reduces energy consumption and improves the response time efficiency by considering the energy consumption of bandwidth and the transmission time and delay.

5. EXPERIMENTAL EVALUATION

To exemplify the performance of the RAO model, the experimental model implements the proposed EARO approach and the existing EERA [19] and AQ-MAC approaches [36].

5.1. Experimental Setup

The experimental model implements the proposed RAO model for the MC environment. It employs the CloudSim tool and simulates the computation and communication-aware RAO model. The experimental model considers the MC environment that consists of 25 mobile devices and 50 tasks to execute the proposed algorithm. It models the cloud environment with 5 PM resources and 30 VM resources with different capabilities. The proposed simulation setup contains five heterogeneous physical machines. Each PM has one CPU with a processing speed of 2000, 4000, 6000, 8000, 10000 MIPS. Also, it considers that each PM has 2 TB storage and 10 GB memory. It evaluates the proposed RAO model by considering 25 mobile users submit the resource requests for allocating 30 heterogeneous VMs. The detailed simulation parameters for implementing the EARO model are presented in Table 2.

| Entity Type    | Parameter          | Values       |
|----------------|--------------------|--------------|
| User           | Number of Mobile   | 25           |
| User           | Devices            |              |
| Cloudlet       | Number of Tasks    | 50           |
| Cloudlet       |                    |              |
| Host           | Number of PM       | 5            |
| Host           | Resources          |              |
| Host           | CPU                | 2000 to 10000 MIPS |
| Host           | RAM                | 10 GB        |
| Host           | Storage            | 2 TB         |
| Host           | Bandwidth          | 5 GB         |
| Virtual Machine| Number of VM       | 30           |
| Virtual Machine| Resources          |              |
| Virtual Machine| CPU                | 500 to 1000 MIPS |
| Virtual Machine| RAM                | 512 MB to 1 GB |
| Virtual Machine| Bandwidth          | 128 to 512 MB |
| Virtual Machine| Operating System   | Linux        |

5.2. Evaluation Results

The experimental model compares the EARO approach with the existing AQ-MAC approach and the EERA approach using four performance metrics.

5.2.1. Number of Tasks Vs. Execution Time

Figure 3 depicts the execution time of the EARO approach and the existing AQ-MAC and EERA approach while varying the number of tasks from 10 to 50 with an interval of 10. Execution time is the time consumed to compute the mobile user requests on the cloud resources. The execution time of
the offloaded mobile tasks relies on the complexity of the task the number of tasks in the mobile application. The proposed EARO approach accomplishes 11.26% of minimal execution time than the existing EERA approach, even when the number of tasks is 50. It focuses on the request size-based resource allocation prioritization that prioritizes the minimum and maximum estimated execution time. In the same scenario, the AQ-MAC approach consumes only a 6.34% higher execution time than the proposed EARO approach by considering the delay and bandwidth-based cloudlet selection and offloading. Moreover, the EARO approach concentrates on the estimated execution time to minimize the completion time when allocating sorted requests on the cloud resources. The numerical results of the execution time for the EARO, AQ-MAC, and EERA for the different number of tasks are presented in Table 3.

![Graph](image)

Table 3 Performance of Execution Time

| Number of Tasks | Execution Time (s) |
|-----------------|--------------------|
|                 | EARO   | AQ-MAC | EERA   |
| 10              | 3.5    | 3.9    | 4.1    |
| 20              | 3.8    | 4.25   | 4.7    |
| 30              | 4.5    | 4.83   | 5.3    |
| 40              | 6      | 6.3    | 6.8    |
| 50              | 6.3    | 6.7    | 7.1    |

5.2.2. Number of Tasks Vs. Makespan

Figure 4 and Table 4 depict the makespan of the EARO and the AQ-MAC and the EERA approaches for the different number of mobile application tasks with their experimental result values. Makespan combines the transmission and execution time of the mobile application requests in the MC, which results from the execution of the application requests on the mobile device, network resources, and cloud server. The proposed EARO approach greatly minimizes the completion time by 10.84% and 5.37% than the existing EERA and the AQ-MAC approaches when the number of tasks is 40. The existing resource allocation models lack focus on both the communication and computation resource allocation optimization in the mobile cloud environment. The optimized resource allocation reduces the execution time on the cloud and migration time on inter-VM bandwidth, and transmission time on the channel bandwidth. Inter-VM bandwidth-based task migration minimizes the migration time in a cloud server. By implementing the resource allocation model in the ThinkAir architecture, the EARO approach achieves the reduced makespan on multiple VM instances through its parallel execution.

![Graph](image)

Table 4 Performance of Makespan

| Number of Tasks | Makespan (s) |
|-----------------|--------------|
|                 | EARO | AQ-MAC | EERA |
| 10              | 5.9  | 6.25   | 6.8  |
| 20              | 6.5  | 6.9    | 7.2  |
| 30              | 6.8  | 7.35   | 7.8  |
| 40              | 7.4  | 7.82   | 8.3  |
| 50              | 7.9  | 8.67   | 8.9  |
5.2.3. Number of Tasks Vs. Energy Consumption

![Energy Consumption Graph]

Even though delay and bandwidth factors are based on resource allocation in the AQ-MAC approach, it leads to increased energy consumption in the large-scale mobile cloud. Table 5 provides the experimentation data points for the energy consumption.

**Table 5 Performance of Energy Consumption**

| Number of Tasks | Energy Consumption (J) |
|-----------------|-------------------------|
|                 | EARO | AQ-MAC | EERA |
| 10              | 0.7  | 1.42   | 1.3  |
| 20              | 1.2  | 2.08   | 1.9  |
| 30              | 1.6  | 2.34   | 2.2  |
| 40              | 2.1  | 2.91   | 2.8  |
| 50              | 2.5  | 3.29   | 3.2  |

Table 5 provides the experimentation data points for the energy consumption.

**5.2.4. Number of Tasks Vs. Response Time**

![Response Time Graph]

Response time of the proposed EARO and the existing EERA approaches are shown in Figure 6, and its numerical points are provided in Table 6 with the variation of the different number of tasks. Response time is the total time taken from the request submission time in the cloud to the execution completion time. The response time linearly increases with the increase of the number of mobile application requests in the MC environment. The existing EERA approach consumes 16.66% of higher response time than the proposed EARO approach, even when the number of tasks is a minimum of 10. The existing approach lacks to consider the communication resource selection during the execution, which leads to ineffective RAO. Transfer time-based channel bandwidth selection and execution parallelization in the proposed EARO approach significantly reduce cloud execution time and transmission time. The proposed EARO approach consumes only 6.8 seconds of the response time when the number of tasks is 50.

**Table 6 Performance of Response Time**

| Number of Tasks | Response Time (s) |
|-----------------|-------------------|
|                 | EARO | AQ-MAC | EERA |
| 10              | 4.2  | 4.61   | 4.9  |
| 20              | 4.5  | 4.83   | 5.3  |
| 30              | 4.9  | 5.34   | 5.8  |
| 40              | 6.4  | 6.81   | 7.3  |
| 50              | 6.8  | 7.24   | 7.9  |
This research paper presented the computation and communication RAO model considering the energy consumption and response time constraints. Initially, it has sorted the mobile user requests concerning the estimated execution time. The EARO approach has allocated the cloud resources to the mobile application requests after prioritizing the requests based on the data size and the resource capabilities. It designed the ECET algorithm to optimize the resource allocation from the results of continuously monitoring the energy and utilization of the resources. Thus, the EARO approach has optimized the resource allocation with the reduced energy consumption and response time from the perspective of the execution time and migration time minimization with the assistance of the inter-VM bandwidth-based task migration. The experimental result has demonstrated the performance improvement of the EARO approach when compared to the existing resource allocation model in the MC environment.

REFERENCES

[1] Fernando, Niroshinie Seng W. Loke, and Wenny Rahayu, “Mobile cloud computing: A survey”,Elsevier transaction on Future Generation Computer Systems, pp.84-106, Vol.29, No.1, 2013
[2] Dinh Hoang T, Chonh Ho Lee, Dusit Niyato, and Ping Wang, “A survey of Mobile Cloud computing: architecture, applications, and approaches”, Wireless communications and mobile computing, Vol.13, No.18, pp.1587-1611, 2013
[3] Hussain, Hameed, Saif Ur Rehan Malik, Abdul Hameed, Samee Ullah Khan, Gage Bickler, Nasro Min-Allah, and Muhammad Bilal Qureshi et al., “A survey on resource allocation in high performance distributed computing systems”, Elsevier transaction on Parallel Computing, Vol.39, No.11, pp.709-736, 2013
[4] Arfeen, Muhammad Asad, Krzysztof Pawlikowski, and Andreas Willig, “A framework for resource allocation strategies in cloud computing environment”, IEEE transaction on Computer Software and Applications Conference Workshops (COMPSACW), pp.261-266, 2011
[5] Madni, Syed Hamid Hussain, Muhammad Shafie Abd Latiff, and Yahaya Coulibaly, “Recent advancements in resource allocation techniques for cloud computing environment: a systematic review”, Cluster Computing, Vol.20, No.3, pp.2489-2533, 2017
[6] Zhou, Bowen, and Rajkumar Buyya, “Augmentation techniques for mobile cloud computing: A taxonomy, survey, and future directions”, ACM Computing Surveys (CSUR), Vol.51, No.1, pp.3-38, 2018
[7] HamaAli, Kurdistan Wns, and Subhi RM Zeebaree, “Resources allocation for distributed systems: A review”, International Journal of Science and Business, Vol.5, No.2, pp.76-88, 2021
[8] Pallavi, L., B. Thirumala Rao, and A. Jagann, “Mobility Management Challenges and Solutions in Mobile Cloud Computing System for Next Generation Networks”, International Journal of Advanced Computer Science and Applications, Vol.11, No.3, pp.177-192, 2020
[9] Shu, Peng, Fangming Liu, Hai Jin, Min Chen, Feng Wen, Yupeng Qu, and Bo Li, “eTime:energy-efficient transmission between cloud and mobile devices”, Proceedings IEEE in INFOCOM, pp.195-199, 2013
[10] Rahimi, M. Reza, Nalini Venkatasubramanian, Sharad Mehrotra, and Athanasios V. Vasilakos, “On Optimal and Fair Service Allocation in Mobile Cloud Computing”, Distributed, Parallel, and Cluster computing, arXiv preprint arXiv:1308.4391, 21 pages, 2013
[11] Sokol Kosta, Andrius Aucinas, Pan Hui, Richard Mortier, and Xinwen Zhang, “ThinkAir: Dynamic resource allocation and parallel execution in cloud for mobile code offloading”, IEEE proceedings INFOCOM, pp.945-953, 2012
[12] Song, Jian, Yong Cui, Minming Li, Jiezhiqin Qu, and Rajkumar Buyya, “Energy-Tradeoff Cooperative Offloading for Mobile Cloud Computing”, IEEE 22nd International symposium of Quality of service, pp.284-289, 2014
[13] Ghasemi-Falavarjani, Saeed, Mohammadali Nematabakhsh, and Behrouz Shaghahi Ghahfarokhi, “Context-aware multi-objective resource allocation in mobile cloud”, Computers & Electrical Engineering, Vol.44, pp.218-240, 2015
[14] Sun, Huaqing, Huqin Yu, Guisheng Fan, and Liqiong Chen, “Energy and time efficient task offloading and resource allocation on the generic IoT-fog-cloud architecture”, Peer-to-Peer Networking and Applications, Vol.13, No.2, pp.548-563, 2020
[15] Wang, Kezhi, Kun Yang, and Chathura Sarathchandra Magurawalage, “Joint energy minimization and resource allocation in C-RAN with MOBILE CLOUD”, IEEE Transactions on Cloud Computing, Vol.6, No.3, pp.760-770, 2016
[16] Jin, A-Long, Wei Song, and Weihua Zhuang, “Auction-based resource allocation for sharing cloudlets in MOBILE CLOUD computing”, IEEE Transactions on Emerging Topics in Computing, Vol.6, No.1, pp.45-57, 2015
[17] Liu, Yanchen, Myung J. Lee, and Yanvan Zheng, “Adaptive multi-resource allocation for cloudlet-based MOBILE CLOUD computing system”, IEEE Transactions on Mobile Computing, Vol.15, No.10, pp.2398-2410, 2015
[18] Chen, Meng-Hsi, Ben Liang, and Min Dong, “Joint offloading and resource allocation for computation and communication in mobile cloud with computing access point”, In IEEE INFOCOM 2017-IEEE Conference on Computer Communications, pp.1-9, 2017
[19] Neha Jayant, Gagandeep, “Energy Efficient Dynamic Resource Allocation Technique in Mobile Cloud Computing”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCT) Vol.6, No.8, 2017
[20] Di Lorenzo, Paolo, Sergio Barbarossa, and Stefania Sardellitti, “Joint Optimization of Radio Resources and Code Partitioning in Mobile Cloud computing”, arXiv preprint arXiv:1307.3835, pp.1-14, 2013
[21] Angin Pelin, and Bharat Bhargava, “An agent-based optimization framework for mobile-cloud computing”, Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, Vol.4, No.2, pp.1-17, 2013
[22] Barbarossa, Sergio, Stefania Sardellitti, and Paolo Di Lorenzo, “Joint allocation of computation and communication resources in multi-user mobile cloud computing”, IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC), pp.26-30, 2013
[23] Kuribayashi, and Shin-ichi, “Optimal joint multiple resource allocation method for cloud computing environments”, International Journal of Research and Reviews in Computer Science (IJRCS), arXiv preprint arXiv:1110.1730, Vol.2, No.1, 2011
[24] Chaisiri, Sivadon, Bu-Sung Lee, and Dusit Niyato, “Optimization of resource provisioning cost in cloud computing”, IEEE Transactions on Services Computing, Vol.5, No.2, pp.164-177, 2012
[25] Zhang, Weiwen, Yonggang Wen, Kyle Guan, Dan Kilper, Haiyun Luo, and D. Wu, “Energy-Optimal Mobile Cloud Computing under Stochastic Wireless Channel”, IEEE Transactions On Wireless Communications, Vol.12, No.9, pp.4569-4581, 2013
[26] Hao, Liang, Gang Cui, Mingcheng Qu, and Xunde Ke, “Resource Scheduling Optimization Algorithm of Energy Consumption for Cloud Computing Based on Task Tolerance”, Journal of Software, Vol.9,
RESEARCH ARTICLE

No.4, pp.895-901, 2014

[27] Chrysa Papagianni, Aris Leivadeas, Symeon Papavassiliou, Vasilis Maglaris, Cristina Cervello Pastor, and Alvaro Monje, “On the optimal allocation of virtual resources in cloud computing networks”, IEEE Transactions on Computers, Vol.62, No.6, pp.1060-1071, 2013

[28] Daren Fang, Xiaodong Liu, Lin Liu, and Hongji Yang, “OCSO: Off-the-cloud service optimization for green efficient service resource utilization”, Springer Open access Journal of Cloud Computing: Advances, Systems and Applications, Vol.3, No.9, pp.1-17, 2014

[29] Younis, Ayman, Tuyen X. Tran, and Dario Pompli, “Bandwidth and energy-aware resource allocation for cloud radio access networks”, IEEE Transactions on Wireless Communications, Vol.17, No.10, pp.6487-6500, 2018

[30] Liu, Li, and Qi Fan, “RAO based on mixed integer linear programming in the multi-cloudlet environment”, IEEE Access, Vol.6, pp.24433-24452, 2018

[31] Meng, Sachula, Ying Wang, Zhongyu Miao, and Kai Sun, “Joint optimization of wireless bandwidth and computing resource in cloudlet-based mobile cloud computing environment”, Peer-to-Peer Networking and Applications, Vol.11, No.3, pp.462-472, 2018

[32] Zhao, Junhui, Quiping Li, Yi Gong, and Ke Zhang, “Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks”, IEEE Transactions on Vehicular Technology, Vol. 68, No.8, pp.7944-7956, 2019

[33] Zhang, Jing, Weimei Xia, Feng Yan, and Lianfeng Shen, “Joint computation offloading and RAO in heterogeneous networks with mobile edge computing”, IEEE Access, Vol.6 pp.19324-19337, 2018

[34] Ramasubbareddy, Somula, G. Vedavasu, K. B. Gopi Krishna, and Alekhyaa Savithri, “PIOCM: Properly Identifying Optimized Cloudlet in Mobile Cloud Computing”, Journal of Computational and Theoretical Nanoscience, Vol.16, No.5-6, pp.1967-1971, 2019

[35] Pallavi, L., A. Jagan, and B. Thirumala Rao, “ERMO2 algorithm: an energy efficient mobility management in mobile cloud computing system for 5G heterogeneous networks”, International Journal of Electrical and Computer Engineering, Vol.9, No.3, pp.1957-1967, 2019

[36] Kiran, K. Tara Phani Surya, K. V. V. Satyanarayana, and P. Yellamma, “Advanced Q-MAC: optimal Resource allocating for dynamic application in mobile cloud computing using Qos with cache memory”, International Journal of Engineering & Technology Vol.7, No.3.1, pp.143-146, 2018

[37] Kosta, Sokol, Andrius Aucinas, Pan Hui, Richard Mortier, and Xinwen Zhang, “Unleashing the power of MOBILE CLOUD computing using thinkair”, arXiv preprint arXiv:1105.3232, 2011.

Authors

Dr. R. Shankar, received B.E in CSE from University of Madras, Chennai, M.Tech in CSE from Anna University, Chennai and Ph.D in CSE from Manonmaniam Sundaranar University, Tirunelveli. He is now working as Associate Professor in Department of Computer Science and Engineering at K L Deemed to be University, Vaddeswaram, Vijayawada, Andhra Pradesh. He has 21 years of Academic experience. He published his research work in several reputed journals and conferences in the field of Mobile Computing, Cloud Computing and Artificial Intelligence.

Dr. Tharani Vimal completed B.E in EEE from Bharathidasan University, M.E in Computer Science from Anna University and Ph.D in Computer Science and Engineering from Manonmaniam Sundaranar University. She is now working as Professor and Head of Department of Information Technology in St.Peter's College of Engineering & Technology, Avadi, Chennai, Tamilnadu. Published various journal in conferences and journals in the field of Wireless Adhoc Network, Networking, Cloud computing etc She has more than 18 years of experience in teaching and over 5 years of experience in software industry.

How to cite this article:

R. Shankar, Tharani Vimal, “Optimization of Computation and Communication Driven Resource Allocation in Mobile Cloud”, International Journal of Computer Networks and Applications (IJCNA), 9(2), PP: 189-201, 2022, DOI: 10.22247/ijcna/2022/212335.