MINIMIZATION OF ACCESS DELAY USING CLOUDLET CACHE

S.Pushpalatha1, M.Reshma2,S.Jennykalaiarasi3, R.K.Kapila vani4

[1] Department of Computer Science and Engineering, Prince Shri VenkateshwaraPadmavathy Engineering College,Chennai,India
[2] Department of Computer Science and Engineering, Prince Shri VenkateshwaraPadmavathy Engineering College,Chennai,India
[3]Assistant professor,Department of Computer Science and Engineering, Prince Shri VenkateshwaraPadmavathy Engineering College,Chennai,India.
[4] Assistant professor,Department of Computer Science and Engineering, Prince Dr.KVasudevan College of Engineering and Technology,Chennai,India.

Abstract- *Mobile cloud computing mainly focus on performance and energy savings of mobile kind of devices. The main purpose of cloudlet is to reduce the latency between mobile Devices. Cloudlets bridge the gap between mobile devices and users. They are rich set of computing resources of server clusters connected in a LAN. In order to reduce the processing time, extend the cloudlet usages. The previous works consumes more time to give query response by the client. To avoid a delay we going to place a cache over every cloudlet and access point. The caches store the unique converted resource over by the client request. when the client make a second request of the same data by any other user, the cloudlet cache offer the cached copy of that searched data. So the converting time is reduces significantly.*

Key word-*Cloudlet, Access point, cache, Cloudlet cache*

1. INRTODUCTION

Nowadays, mobile devices have undergone a transformation from bulky gadgets with limited features to everyday accessories. Advance in mobile hardware technology have led to a high growth in mobile application markets. However mobile applications are advanced in computational-intensive, the computing capacity of mobile of mobile devices remains limited, because of the consideration of weight, size, battery life, etc. performance of the mobile applications is enhanced by enabling mobile devices to offload some of the workload of mobile devices to remote resource rich clouds. The communication delay between the clouds and mobile users are unpredictable because clouds are geographically far away from the mobile users. This make a problematic for mobile applications in which crisp response time is critical to the users, like speech recognition, navigation, language translation, etc. In order to bring the cloud, closer to the mobile users cloudlet is used. This cloudlet are resource-rich server clusters which is placed with Wireless Access Points(APs)in a local network, and mobile users can offload their task to local cloudlets for processing. The physical proximity between mobile users and cloudlets means that the cloudlet access delay on offloading task can be reduced there by improving mobile users experiences significantly. The cloudlet is mainly focus on offloading the tasks of mobile users to save the energy of mobile devices. The placement of cloudlet in the different access point is problematic, so we have to place little concentration on placement of cloudlet. In this paper we focus on the placement of cloudlet in a Wireless Metropolitan Area Network (WMAN) which provides wireless internet coverage for mobile users in a large- scale metropolitan area, where the WMAN is often owned and operated by local government as public infrastructure. By using this the mobile users can utilize a following benefits : (1)The cloudlet will be accessible by a large number of mobile users;(2)Cloudlet service is more affordable to the general public. However placing of cloudlet is challenging. The access delay and utilization of resource will be critical to the mobile users due to the location of the cloudlets, especially in a large-scale WMAN which consist of hundreds and thousands of Access Points (APs).Due to the large size of the WMAN, poor placement of cloudlet will result in long access delay and heavily unbalanced load among cloudlets, i.e, some of the cloudlet are overloaded while others are under-loaded even idle. In order to improve the performance of various mobile application such as the average cloudlet access delay we going to place a capacitated cloudlets. In this paper we going to see the novel cloudlet placement problem in a large-scale WMAN, where the service provider of cloudlet is planning to deploy K (≥1) cloudlets at some strategic

AP locations in a WMAN for mobile user access. The aim is to minimize the average cloudlet access delay between the mobile users and the cloudlet using the servers. The challenges faces with such placements are: which cloudlet should be placed to which location, and which user request should be assigned to which cloudlets so that the average cloudlet access delay among the mobile user is minimized. As the problem is NP-hard, is there any approximation algorithm with a guaranteed approximation for it? in this paper we will discuss this issues.

There are several placement problems in networks such as cache placements and server placements that has been referred [6][7].For example, the cache placements problem is to select K replicas or hosting services among N potential areas, such that the latency is minimized[7],which is reduced to the capacitated K median problem. Because of the NP-hardness of the latter, there are approximation algorithm for resource version of theproblem[5],[8],In 'Resource' user request can be served by only one data center[5],and in 'split' user request can be served by multiple centers[8].In spite of some similarities between the cache/server placement problem[6],[7]and cloudlet placement problem, they are essentially different. First existing studies examines that either there is no capacity of each cache/server is identical. second, existing studies simply examines that each user request has identical resource demands, while we assume that different user requests may have different amounts of resource demands. Thus the existing solutions to cache/server placement problem [5] may not be applicable to the cloudlet placement problem. there are several million mobile users in a metropolitan area, compared with only several hundred node(access point) in such a network. Therefore, new algorithm for cloudlet placement problem must be devised in order to deal with a large set of user request and different amounts of resource demands by different users.

The main purpose of this paper are, We study multiple cloudlet placements in a large-scale WMAN, by formulating a novel capacitated cloudlet placement problem with the aim of reducing the average cloudlet access delay. We first view that the problem is Hard, and propose an exact solution by formulating it as an Integer Linear Programming (ILP) Because of the worst scalability of the ILP, we then devise a fast, scalable heuristic. For special case of the problem where all cloudlet have similar computing capability, we device two approximation algorithms with assured approximation ratios, it is depend on whether all user requests have similar resource demands or not. we also propose an efficient online algorithm for dynamic user request assignment to the cloudlets, provided the K cloudlets have already been placed. further we evaluate the performance of the proposed algorithms through experimental simulations. The simulation results demonstrate that the proposed algorithm are very promising.

The rest of the paper is organized as follows, section 2Related work, section 3-Algorithm for capacitated cloudlet placement problem, section 4-Expremental result, section 5-Futher work, section 6-Reference

2. Related work

Most existing studies concentrates on reduces the task processing time for the users to remote clouds by using cloudlet and efficient task scheduling [9],[10],[11],[12],[13],[15],[16].However the average access is delayed from remote clouds to mobile users, we going to use a cloudlet which is used in order to reduce the average access delay for the mobile users. This will used as to save the energy of battery, processingtime,etc.[17],[18],[19],[20],[21],[22],[23], [24],[25],[26],[14],[27].In this the new term called fog is used. Fog is nothing but cloudlet based computing. The aim of the fog computing is to reduces the access delay between mobile users and remote clouds, by providing compute, storage, and networking service[1],[2].for example the system Odessa[23] is used to reduce the access proposed by Chun *et al[17]*.Hoang *et al.*[19]proposed a linear programming solution for task offloading by considering the Qos requirements of mobile users with an aim to maximize the revenue of service providers. Novel online algorithms for dynamically admitting user requests to a cloudlet is devised by Xia *et al*[26],[14].Cardellini*et al*[3].Game theory-based is used as a solution for offload their tasks to both remote clouds and local cloudlets. the multiplayer game in mobile is advanced by using the cloudlet assisted multi-player cloud gaming[4] by correlating the frames this will not consume more energy i.e the access time will be reduced. The increasing number of cloudlet will faces the problem where to place a cloudlet. The placement of cloudlet will be classified in two types they are trivial and non-trivial, In trivial the cloudlet will be placed with in a small area like college campus, building ect. Where in non-trivial the cloudlet will be placed in a wide range of area with the help of thousands and hundreds of Access Points. This placement of cloudlet in wide range of area is used to reduces the average access delay between the mobile users and cloudlet servers. There are two classical optimization problems closely related to the cloudlet placement problem: the *cache placement problem[7]and the server placement problem[6]*.Before we uses K replicas or hosting services among N potential sites to reduce the latency experienced by the user[7].After we placed a number of server replicas among some potential location such that the perceived delay of users is minimized under a given traffic pattern. Above discussed problem can be solved by a direct reduction to the capacitated K-median problem.

2.1 System model

We consider a WMAN G=(V∪S,E)consisting of many APs and a set of potential locations for cloudlets, here V is the set of APs and Sis the set of potential location of cloudlets, and E is the set of links between two APs or between an AP and links in v and E, respectively, i.e, n=| | and m=|E|. For each AP vjin V, let *w(vj)* represent the e*xpected number of user requests* using the AP to access cloudlets in the network, which is a positive integer. Aps are usually placed at strategic locations such as shopping malls, train stations, schools, libraries, etc., the number of user request w(v) at each AP v per unit time can be estimated by the population density in that area, or the historic AP access information through a linear regression technique. Let Rj may be set of user requests at AP vjwith w(vj)=|Rj|,and different user requests in Rjmay have different amount of computing resource demands. Denote by rma user request in Rjwith the computing resource demand γm in G. In addition, for each link(vj,vl) in E, denote by djlthe data transmission latency between its two endpoints(APs)vjand vl.

Consider that there are K cloudlets to be placed to K different locations in S. Assume that the cloudlets will be co-located with some APs, i.e, S⊆V.we also consider that K<<|S| and each cloudlet Cihas limited computing resource to process user request. Let Cibe the computing resource capacity of Ci with 1≤i≤K.Given the K placed cloudlets, mobile users can offload their tasks to the cloudlets through their local APs. If the cloudlet is placed in all the access points the access delay will be minimized. Otherwise, the user request at that AP must be relayed to near by cloudlets for processing, resulting in a cloudlet access delay due to the accumulative delay of multiple hop relays. *fig.1 illustrates a* WMAN network.

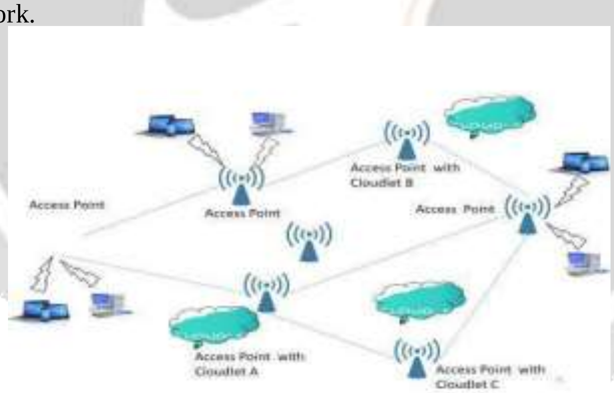

Fig.1.A WMAN G=(V,E) with K=3 cloudlet.

3. Algorithm

We now consider the cloudlet placement issues can be resolved by the following algorithm. We propose the algorithm which requires different user demands for the resources.

In this we use a scaling technique to fast up the user demands from the cloudlets. Assign the user demand of each request to be dm∈Djof Access point Aj is dm.It can be classified into two different cases let us assume the demand to be ∝maxand∝min.we can relate the demand as

=∝max/∝min for each user request for each user d`m1,d`m2,…..d`mπm.with same requests, request from AP can be divided to individual request and assume the execution to be ∞.The user's demand will be delivered not only by the single cloudlet, it is done by the different cloudlets enclosed with the access point. It can be resolved by following cases.

Case 1: If the request demands d_i is given to the cloudlet associated Access Point, the process will be efficiently performed without the delay in the response.

Case 2: If the request demand is with the cloudlet without access point, the process is not performed efficiently until it requires the information from various cloudlet.

The cases can be resolved further with in the rang of $d \leq n \leq N$. Assume the cloudlet faces maximum access delay between the user requests D_i.

3.1 Algorithm analysis:

To analyze the algorithm we mainly focus on complexity of algorithm. Let us consider the Graph G`= (V,E) be the requests and Demands over the wireless network. Let the two APs be Ai &Ajin A, Each request is assigned to different cloudlet.

Procedure:

Input:

G(x,t),a set of user demands at each AP aj∈ the number of user request in Dj, the delay d(r) or each request,N cloudlets with volume vs.

Output:

Let Xmin and Xmax be resource demand of a user request at various AP in
A. For each AP aj∈ 4 do
For each of user request dm∈Djdo
Construct ∝m ∝/min user request with demand divide by
∝min.
end for end
for
Let d`m1,d`m2,…..d`mR,dmL, in which each request is assigned to various cloudlet.
If k user is assumed to different cloudlets then find the cloudlet with maximum access in delay in access.
Finally the request from one cloudlet to another will be sent through AP.
end if end
for

4 .Experimental Result:

Mobile cloud computing mainly focus on energy savings of mobile devices by offloading computingintensive jobs from mobile devices to remote clouds, the access delays between mobile users and remote clouds usually are long and sometimes unbearable. In order to over come this delay we using the technology called cloudlet which has a capability to bridge this gap, and it can enhance the performance of mobile devices significantly while meeting the crisp response time requirements of mobile users. The existing studies focused on offloading tasks of mobile users to cloudlets for energy savings of mobile devices. However we have to pay little attention on placing of cloudlet. Due to the large size of the WMAN, poor cloudlet placements will result in long access delays and heavily unbalanced load. There is no capacity constraint on caches/servers or the capacity of each cache/server is identical. In order to overcome this drawback we extend the cloudlet usage to reduce the processing time. This process of work consumes more time to access the request, to avoid that delay we going to place a cache over on every cloudlet. In this process the user will request for resource, If the requested resource is present in access point it will directly response to the client, if the requested resource is not present in the access point means, that access point will sent the request to the access point which as cloudlet then the response from the cloudlet associated access point will be given to the access point and that access point will sent response to the client. The intermediate access point was placed with cache, this cache store the unique converted resource over by the client request. While a client make request second the same data by any user, the cloudlet cache offer the cached copy. so the converting time is reduces significantly. This process significantly save the access time, energy consumption, etc.

5 Conclusion and Further works

Cloudlets have been emerged as an important technology that can extend the computing capabilities significantly of resource-constrained mobile devices. In this paper we first studied the capacitated cloudlet placement problem in a large-scale Wireless Metropolitan Area Network with the objective to reduce the average cloudlet access delay between mobile users and the cloudlets serving their requests. We then provide an solution for the problem when the problem size is small, otherwise, we proposed a fast yet scalable heuristic for it. For a special case of the problem when all cloudlets have identical computing capacities, we devised two novel approximation algorithms, depending on whether identical resource demands by all user requests. we finally evaluated the performance of the proposed algorithms byexprements. This exprement gives the very promising solution for the users. In the further we will study this problem by investigating the delay impact between users and their APs.

6 References

[1] F.Bonomi,R.Milito,J.Zhu,andS.Addepalli.fog computing and its role in the internet of things.Procs of the MCC workshop on Mobile cloud omputing,ACM,2012.

[2] E.A.Lee et al.The swarm at the edge of the cloud.Journal of Design Test,Vol.31,pp.820,IEEE,2014. [3] V.cardellini,V.D.N.Persone,V.D.Valerio,F.Facchinei ,V.Grassi,F.L.Presti,andN.Piccial.A game-Theretic approach to computation offloading in mobile cloud computing.Technicalreport,available online at http://www.optimizationonline.org/DB_FILE/2013/08/3 981.pdf,2013.

[4] W.Cai,V.C.M.Lrung,andL.Hu.A cloudlet-assisted multiplayer cloud gaming system.Journal of Mobile Networks and Applications,Vol.19,pp.144-152,2014.

[5] M.Charikar, S.Guha, E.Tardos,andD.B.Shmoys,A constant factor approximation algorithm for the kmedianproblem.proc.of STOC,ACM,1999.

[6] L. Qiu, V. N. Padmanabhan, and G.M.Voelker.On the placement of web server replicas.Proc. ofINFOCOM,IEEE,2001.

[7] H.Yin,X.Zhang,T.Zhang,Y.Zhang,G.Min, and D.Wu.NetClust:a framework for scalable and paretooptimal media server placement.IEEETrans.multimidia,vol.15,pp.2114-2124,2013.

[8] S. Li. An improvement approximation algorithm for the hard uniform capacitated k-median problem.http://arxiv.org/abs/1406.4454,20146.

[9] L.Gu,D.Zeng,A.Barnawi,S.Guo, and I.Stojmenovic.optimal task placement with Qos constraints in geo-distributed data centers using DVFS.IEEE Transactions on computers,Vol.64,pp.2049-2059,IEEE,2015.

[10] H.Jin, X.Wang,S.Wu,S.Di,andX.Shi.Towards optimized fine-grained pricing of IaaS cloud platform.IEEETrans. On Cloud computing,Vol.PP,IEEE,2014.

[11] R.Kemp,N.Palmer,T.Kielmann,andH.Bal. Cuckoo:acomptation offloading framework for smartphone.Mobilecomputing,Application, and Service,Springer,pp.59-79,2012.

[12] J.Niu,w.Song,andM.Atiquzzaman.Bandwidtha daptive partitioning for distributed execution optimization of mobile applications.journal of network and computer application,Vol 37,pp.334347,Elsevier,2014.

[13] S.Ren and M.Van der Schaar.Dynamic scheduling and pricing in wireless cloud computing.IEEETrans. On Mobile Computing,I vol.13,pp.22832292,IEEE,2014.

[14] Q.Xia W. Liang, Z.Xu,andB.Zhou.Online algorithms for location-aware task offloading in twotiered mobile cloud environments Proc. Of 7th international Conference on Utility and Cloud Computing,IEEE/ACM,2014

[15] Z.Xu and W.Liang.Operational cost minimization for distributed data centers through exploring electricity price diversity.Computer Networks,vol.83,pp.5975,Elservier,2015.

[16] Y.Zhang,H.Liu,L.Jiao,andX.Fu.To offload or not to offload:an efficient code partition algorithm for mobile cloud computing.proc.of CLOUDNET,IEEE,2012.

[17] B.G Chun,S.Ihm,P.Mainatius,M.Naik and A.Patti.Clonecloud: elastic execution between mobile device and cloud.Proc. of EuroSys,ACM,2011.

[18] E.Cuervo,A.Balasubramanian,D.Cho,A.Wolm an,S. Saroiu,R.Chandra,andP.Bahl.MAUI:making smartphones last longer with code offload.Proc. of MobiSys,ACM,2010.

[19] D.T.Hoang,D.Niyato,andP.Wang.Optimal admission control policy for mobile cloud computing hotspot with cloudlet.Proc.of WCNC,IEEE,2012.

[20] M.Jia,J.Cao,andW.Liang.Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks.Toappear in IEEE trans.on cloud computing,2015

[21] K.Kumar and Y.Lu. Cloud computing for mobile users:canoffloading computation save energy.Journal of Compute,Vol:43,pp.51-56,2010.

[22] S.Kosta,A.Aucinas,P.Hui,R.Mortier,andX.Zha ng.ThinkAir:Dynamic resource allocation and parallel execution in the cloud for mobile code offloading.Proc. of INFOCOM,IEEE,2012.

[23] M.R.Ra,A.Sheth,L.Mummert,P.Pillai,D.Wethe rall,andR.Govinden.Odessa:enabling interactive perception applications on mobile devices.Proc. ofmobisys,ACM,2011.

[24] M.R. Rahimi,N.Venkatasubramanian,andA.V.Vasilakos.MU SIC:mobility aware optimal sevice allocation in mobile cloud computing. Proc. Of 5^{th} international Conference on cloud computing.IEEE,2013

[25] M.Shiraz,S.Abolfazli,Z.Sanaei,andA.Gani.A study on virtual machine deployment for application outsourcing in mobile cloud computing.The journal of dsupercomputing,Spring,Vol.63,pp.942-964,2013.

[26] Q.Xia,W.Liang,andW.Xu.Throughput maximization for online request admissions in mobile cloudlets.proc. of local computer Networks,IEEE,2013.

[27] Z.Xu,W.Liang,W.Xu,M.Jia,andS.Guo.Capacit ated cloudlet placements in wireless metropolitan area networks.Proc. of local computer Networks,IEEE,2012.