Study of Cloudlet Deployed in Various Environment

Kiran R.Bidua[1] Geet Christian[2]
[1,2]Department of Computer Science
[1,2]RCC, Gujarat University, Ahmedabad, Gujarat, India

Abstract— Cloud computing is a newly emerging technology. Cloud computing is widely in use because it provides opportunity to dynamically scale the computing resources for applications. Also the resources are efficiently shared among customers using virtualization technology. Over the past two decades, the world economy has rapidly moved from manufacturing to more service-oriented. Since cloud computing environment has large setup and cost associated with it, testing applications and resource allocation policies are highly challenging. Also in today's generation, mobile devices are widely used like mobile phones, PDA, laptops, etc., which provides public the flexibility to access the cloud while on the move. So, it is difficult to manage the large amount of data on cloud server and allocate the resources efficiently to mobile users or mobile devices [3], [11]. Thus, a new architectural element called CLOUDLET which comes from the convergence of mobile computing technology and cloud computing technology came into existence [3]. It represents the middle tier of a 3-tier hierarchy: mobile device --- cloudlet --- cloud. A cloudlet can be viewed as a "data centre in a box" whose goal is to "bring the cloud closer".

Key words: Cloud Computing, Cloudlet, Latency, Mobile Computing, Virtualization

I. INTRODUCTION

A. General Virtualization

With the help of virtualization technology we are able to run multiple operating systems at the same time on a single physical machine which shares the same resources [4]. The reason behind usage of virtualization are a) sufficient capability of recent computers to run multiple operating systems, b) using multiple isolated operating systems, resource utilization can be maximized, c) ability to run different operating systems on single physical machine. The software layer which provides the virtualization of underlying hardware is called hypervisor. Hypervisor emulates the underlying hardware resources to the different operating systems (virtual machine). Normally operating system has the direct access to the underlying hardware of a physical machine. In case of virtualization, operating systems access the hardware through the hypervisor. Hypervisor executes the privileged instruction on behalf of virtual machine. Virtualization technology enables to allocate resources of single physical machine among multiple different users. Some of the popular virtualization technologies are XEN and KVM [4]. In cloud computing, virtualization technology is used to create/destroy virtual machine to dynamically allocate/reduce resources for an application. Also virtualization helps to co-locate virtual machines to a small number of physical machines, such that number of active physical machines can be reduced. This kind of efficient allocation of resources led to the evolution of cloud computing technology in which the resources are efficiently shared among multiple users by using general virtualization.

B. Cloud Computing

Computing is used as a utility in cloud computing. Cloud computing is a new technology for representing collection of resources which are shared and scaled dynamically , So as to fulfill the requirement of large number of users. It uses "pay as you use" model, means resources are used when they are required as per the necessity and released when the resources are no longer in use. Thus users have to pay only for the resources as per their usage which is measured by ceilometers for billing purpose in cloud computing. This servers act as application as a service and server as a service. To run multiple operating systems on a physical machine virtualization technology is used. Cloud computing is a collection of resources (servers in datacenters) which are interconnected and using virtualization can be scaled and adapted dynamically.

It provides customers to start their business without purchasing any physical hardware, whereas service providers can rent their resources to customers and make profit by earning money according to the service provided to customers. Customers also have the opportunity to scale up or down their services according to their requirements. For the large amount of storage of data of mobile users and to provide the service efficiently with negligible latency a new architectural element named CLOUDLET is used as a middleware between the mobile devices and cloud server [2].

C. Cloudlet

Cloudlet is a new architectural element which arises from the convergence of two technologies i.e., mobile computing and cloud computing. It is used to overcome the problem of large amount of data storage of mobile users which we were facing in mobile computing and cloud computing technology by acting as a middleware between these two technologies [4] .Thus, it reduces latency by acting as middle tier in three tier hierarchy in which mobile computing is the first tier, cloudlet is the middle one and cloud computing is the third tier [10]. This class represents the applications running on virtual machines. It encapsulates the number of instructions to be executed, amount disk transfer to complete the task.

Cloudlet class also provides workload generation model, identification of guest virtual machine on which it's running. Currently, parameter representing network and disk I/O has been added to Cloudlet class. Amount of network data to be sent or received, to or from another entity are encapsulated to Cloudlet class. To keep the workload generation mechanism same, disk read, disk write, amount of network data to be received or sent can also use existing workload generation models.

II. KEY ATTRIBUTES

A. Only Soft State:

It does not have any hard state, but may contain cached state from the cloud. It may also buffer data originating from a

mobile device. The avoidance of hard state means that each cloudlet adds close to zero management burdens after installation: it is entirely self-managing [5].

B. Powerful, Well-Connected and Safe:

It possesses sufficient compute power (i.e., CPU, RAM, etc.) to offload resource-intensive computations from one or more mobile devices. It has excellent connectivity to the cloud (typically a wired Internet connection) and is not limited by finite battery life (i.e., it is plugged into a power outlet). Its integrity as a computing platform is assumed; in a production-quality implementation this will have to be enforced through some combination of tamper-resistance, surveillance, and runtime attestation [5].

C. Close At Hand:

It is logically proximate to the associated mobile devices. "Logical proximity" is defined as low end-to-end latency and high bandwidth (e.g., one-hop Wi-Fi)[5]. Often, logical proximity implies physical proximity. However, because of "last mile" affects, the inverse may not be true: physical proximity may not imply logical proximity.

D. Builds On Standard Cloud Technology:

It encapsulates offload code from mobile devices in virtual machines (VMs), and thus resembles classic cloud infrastructure such as Amazon EC2 and OpenStack[5]. In addition, each cloudlet has functionality that is specific to its cloudlet role.

III. Difference between cloudlet and cloud server [4].

Feature	Cloudlet	Cloud Server
Soft state	Yes	Yes
Hard state	No	Yes
Management	Self-managed	Professionally administered
Ownership	Decentralized	Centralized
Environment	Data centre in a box at business premises.	Machine room with power conditioning and cooling.
Network connectivity	LAN connection.	WAN or internet connection.
Sharing	Few users	More than 100 users at a time.

Table 1: Shows the difference between cloudlet and cloud server

IV. Related work

There are lots of applications developed using cloudlets. Cloud-Vision [8] is a face-recognition application which uses cloudlet to offload the compute-intensive and latency-sensitive task of face recognition. There are two stages in their application. Face Detection and Face Recognition. They have used cloudlet to provide initial pre-processing of image to reduce the data sent to the cloud servers. Cloudlet coordinates face detection and face recognition phases. They have used two algorithms: Fixed and Greedy for task partitioning.

In [12], authors have developed cloudlet based multilingual dictionaries which support 6 languages. This application uses Dynamic VM synthesis technology to rapidly instantiate multilingual dictionary to a nearby cloudlet. In this VM is delivered to the infrastructure dynamically. The VMs have to synthesized rapidly otherwise users relying on cloudlet will find it unacceptable if there are extended delays for service initiation in a new location occurs. The VMs are scheduled in FIFO order.

In [9], authors have presented a mobile-cloud-based compute infrastructure for battlefield application. In this infrastructure cloudlet is located in hostile environment housed in nearby military vehicle which enables access to the cloud servers through satellite links. Soldier's mobile device interact with the cloudlet in the tank or helicopter and the cloud via satellites. The mobile device in soldier's night vision goggles must use less battery power, so it only captures the image and sends it to nearby cloudlet. All the preprocessing on the image is done by cloudlet. This way soldier's mobile device can capture more images.

Cyber foraging [6] uses surrogates to offload heavy work from resource poor mobile devices. This term was coined by M. Satyanarayanan in 2001. This technique allows user to offload compute-intensive processing from a mobile device to more powerful servers, thereby saving more battery and allows more cloud-based computing.

V. Approach of building cloudlet:

There is an architecture called MOCHA (Mobile Cloud Hybrid Architecture) which introduced cloudlet in the cloud environment. It is also called Mobile-Cloudlet-Cloud architecture. It is developed by University of Rochester. It is useful for running applications requiring high computational power and low response time [14].

In this architecture instead of directly sending data to remote cloud, we can send data to cloudlet first. A cloudlet will store and update profile of the network latencies and their variation to reach different cloud servers.

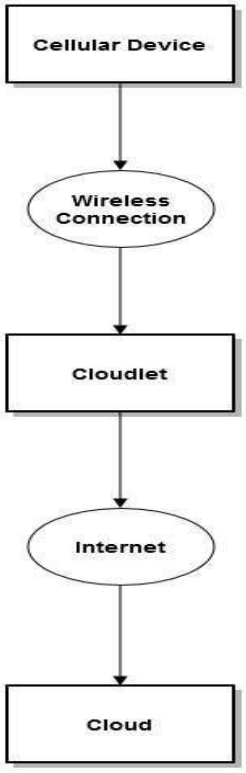

Fig 1: Mobile-Cloudlet-Cloud connection in MOCHA Architecture [1] [7]

A. Challenges in Cloudlet

There are different strategies for mobile to cloudlet data collecting and cloudlet to cloud data collecting. The primary information to measure is the two-way communication latency, also called response latency. Number of hops in the path between mobile-cloudlet and cloudlet-cloud connection is to be recorded. Latency can be found using traceroute command.

After measuring latencies between mobile and cloudlet and between cloudlet and cloud, data analysis is performed. In data analysis average latency as well as variance in latency is to be examined. Here latency is a function of data size and number of hops in the route.

1) Latency Measurements

Latency is used as a primary performance metric. Here latency is defined as the round-trip time delay. Latency is defined as the amount of time spent between sending the first packet by the client and receiving acknowledgement for the last packet from destination via same connection. It is nearly impossible to measure one-way latency because clocks at sender and receiver are not synchronized.

Maximum, mean and minimum latencies is to be calculated as the file size changes for one hop and multi-hop Wi-Fi. The results [14] show that latencies increases with the file size for both single and multi-hop as well as testbed nodes. Also the mean latency increases with data size in linear fashion. These measurements can be saved into latency models and can be stored and updated in the cloudlet to optimize the server selection and minimize the overall communication latency.

2) Latency Variance Behavior

The results [14] of latency variance calculation shows that for mobile to cloudlet, standard deviation increases with file size for both single and multi-hop connections. For cloudlet to cloud results [14] show that latency variance increases as the latency increases.

3) Latency Linear Model

If we have latency distribution model we can predict latency for certain server and certain data size with a certain probability. When original data splits into multiple chunks and sent to multiple servers, these models are used to estimate latency to a target server for an arbitrary data size. For example, if we send 1 MB of data to 3 servers, the data splits into 3 341 KB chunks approx. Then we use latency and variance linear models to predict the latency and standard deviation for transferring 341 KB data.

4) Server Selection Algorithm

If face recognition is to be performed on 1000 faces. We can send all 1000 faces to one server, or each face can be sent to 1000 servers. Client has an option to divide the data into little chunks of data and send these chunks to different servers for computation. Servers will perform the computation and return the result. The results gathered from all the servers is aggregated to obtain final result. There are three ways to divide data among different servers.

Random Algorithm	**Fixed Algorithm**	**Greedy Algorithm**
In this algorithm the task is assigned to the random server.	In this algorithm the task is evenly distributed among available cloud server	In this algorithm, each task is assigne greedily to the server.
Using redundancy provides reduced latency.	Cannot use redundancy.	Using redundancy provides fault tolerance in the case where selected server is unavailable.
Latency information is not required.	Latency information is required.	Latency information is required.
It is easy to implement.	It is complex than random algorithm.	It is complex than random and fixed algorithm.
The optimal latency is higher compared to fix and greedy algorithm.	The optimal latency is higher compared t greedy algorithm and lower compared to random algorithm.	The optimal latency is lower compare to random and fixed algorithm.
We don't divide task into small data chunks	We divide task into equal sized data chunks.	We divide data chunks into arbitrary sized data chunks.
Random algorithm is useful when there is no cloudlet and mobile device is not feasible to perform network profiling.	Fixed algorithm is useful to find optima number of servers.	Greedy algorithm is useful to find be server selection set.

Table 2: Comparison of server selection algorithms

5) Other Related Aspects

It is required that path between nodes do not change frequently because if it does than it is ineffective to model the latency of different servers. Research [13] shows that 68% of the routes in the Internet never changes for seven days. Only one pair out of 10 changes its path and number of hops several times, and all other never change their paths.

The results [14] generated by histograms of latencies of one hop and multi hop connections for 10 MB and 1 MB data transfers shows that the peak is close to the average and then gradually decreasing as the latency increases. This feature can be used to build a latency model for each data size. This latency model is stored and updated in the cloudlet for optimizing the server selection in order to minimize overall communication latency.

VI. CONCLUSION

Introduction of cloudlet in cloud computing offers increase in performance of applications. To build cloudlet, we need proper infrastructure and it is costly. But overall it is scalable. It can help lower the latency which is useful for latency intensive applications like real-time face recognition. It also saves client device processing power and battery consumption which is useful for mobile device. Based on cloudlet characteristics we have created cloudlet parameters in different parameters.

- Yes= Mandatory
- No=Not Mandatory

	Public Network	Private Network	Hostile Environment	Disaster Recovery

Security	No	Yes	Yes	No
Reliability	No	No	Yes	Yes
Scalability	Yes	No	No	Yes
Interoperability	Yes	No	Yes	Yes
Manageability	No	No	Yes	Yes
Performance	No	Yes	Yes	Yes
Affordability	No	No	Yes	No
Timeliness	No	No	Yes	Yes
Integrity	No	No	Yes	Yes

REFERENCES

[1] Bilogrevic, I., Satellite Communications: Internet challenges and strategies for low-latency applications, MS Project: Royal Institute of Technology, Sweden (2008), Proceedings of INET '97, June 1997
[2] Debessay Fesehaye, Impact of Cloudlets on Interactive Mobile Cloud Applications, Enterprise Distributed object computing conference, IEEE(16th international) , USA 2012,pp.3
[3] Kiryong Ha, Padmanabhan Pillai, The Impact of Mobile Multimedia Applications on Data Center Consolidation IEEE International Conference on Cloud Engineering, IEEE,2013,pp.1
[4] Krishnadhan Das, CLOUD COMPUTING SIMULATION, Thesis, Department of computer science,IIT-Mumbai,21st june,2010,pp.15
[5] Mahadev Satyanarayan, The Case for VM-Based Cloudlets in Mobile Computing, Pervasive computing,IEEE (volume-8,issue-4), 2009,pp.6
[6] M. Satyanarayanan (2001), Pervasive Computing: Vision and Challenges, IEEE Personal Communications (IEEE)
[7] M. Satyanarayanan, G. Lewis, E. Morris, S. Simanta, J. Boleng, and K. Ha, The role of cloudlets in hostile environments, IEEE Pervasive Computing, 12(4), Oct-Dec 2013T.
[8] Soyata, R. Muraleedharan, C. Funai, M. Kwon, and W. Heinzelman, Cloud-Vision: Real-time face recognition using a mobile-cloudletcloud acceleration architecture, in Proc. IEEE ISCC '12, Cappadocia, Turkey, Jul. 2012, pp. 59–66.
[9] T. Soyata, R. Muraleedharan, J. Langdon, C. Funai, S. Ames, M. Kwon, and W. Heinzelman, COMBAT: mobile-Cloud-based cOmpute/coMmunications infrastructure for BATtlefield applications, Proc. of SPIE, vol. 8403, 84030K-1, 2012.
[10] Tsu-Shiuan Lin, Feeney, L. M., & Nilsson, File sharing in Mobile Cloudlet networks, Twentieth Annual Joint Conference of the IEEE Computer and Communications Society, 2007, pp.6.
[11] Tim Verbelen, Cloudlets: Bringing the cloud to the mobile user, Proc. of the 3rd ACM Workshop on Mobile Cloud Computing & Services, MCS '12, 2012.
[12] V. S. Achanta, N. T. Sureshbabu, V. Thomas, M. L. Sahitya, and S. Rao, Cloudlet-based multi-lingual dictionaries, Services in Emerging Markets (ICSEM), 2012 Third International Conference on. IEEE, 2012, pp. 30–36.
[13] V. Paxson, End-to-end routing behavior in the internet, Proc.of ACM SIGCOMM, pp. 41–56, Aug. 1996.
[14] Zuochao Dou, "Benefits of Utilizing an Edge Server (Cloudlet) in the MOCHA Architecture", Master's thesis, University of Rochester. Department of Electrical and Computer Engineering, Rochester, NY, 2013.

