Optimal Deployment Model of Cloudlets in Mobile Cloud Computing

Alakbarov R. G.
ANAS Institute of Information Technology
Baku, Azerbaijan
e-mail: rashid@iit.ab.az

Pashayev F. H.
ANAS Institute of Control Systems
Baku, Azerbaijan
e-mail: pasha.farhad@gmail.com

Alakbarov O. R.
ANAS Institute of Information Technology
Baku, Azerbaijan
e-mail: oqtayalakbarov@yahoo.com

***Abstract*—Article studies development dynamics of mobile cloud technologies. Advantages of this technology and problems occurring during its use are analyzed. At the same time, issues related to meeting the demand of computing and memory resources of mobile equipment using this technology are studied. Article studies conditions necessitating development of cloudlets on mobile computing clouds and solves issues of forecasting location time of cloudlets near certain base stations.**

Keywords-mobile computing clouds; mobile equipment; computing and memory resources; cloudlet; computing clouds; mobile trade; mobile healthcare; cloud services; forecasting

I. INTRODUCTION

Currently, intensive works are carried out world-wide to efficiently use computing and memory resources of data processing centers using Cloud Computing technology. Such systems with large computing and memory resources are developed based on computer networks with high speed connection channel. Cloud Computing technology allows using the computing and memory resources of data processing center's more efficiently. Using this technology, user information is stored, processed on computing clouds' servers and at the same time, results can be reviewed through browsers [1, 2]. Cloud computing services provide wide use of clustering and virtualization of computing and memory resources of processing centers. Hence, Cloud Computing system is an Internet service consisting of technical-software and allowing internet users to use remote computer resources (computing and memory resources, program and data etc.) through an appropriate web-interface.

Cloud computing technology allows users to acquire strong computing and memory resources and at the same time, user is not interested in location and settings of these resources. Using mentioned service, it is possible to provide efficient use of computing and memory resources at data centers, reduction of problem solution time and less loading of the network [3]. Cloud technologies optimally distribute the existing memory resources of the processing center among users, thus attracting more users to the system and obtaining the results faster. Recently, mobile users have started widely using the services of Cloud Computing technologies.

Rapid acceleration of world-wide use of mobile devices (laptop, tablet, smartphones etc.) and their connection to computing clouds via relevant telecommunication technologies (GPS, 3G, 4G, Wi-Fi etc.) lead to development of new technologies – Mobile Cloud Computing technologies. The article studies issues related to more efficient use of computing and memory resources of mobile computing clouds widely used recently. It is known that, capabilities of any mobile device (computing and memory resources) are limited. But users use these devices to solve issues that require large computing and memory resources. Computing cloud technologies are widely used for this purpose. Hence, computing and memory resource deficiencies existing in users' phones can be solved using cloud technologies.

II. ARCHITECTURE OF MOBILE COMPUTING CLOUD TECHNOLOGIES AND PROBLEMS

Mobile computing clouds are the new platform formed on equipment and computing clouds allowing mobile users to solve complex issues and store large-volume data [4].

Technical capabilities (computing and memory resources) of mobile devices are limited. Cloud computing capabilities are widely used in order to eliminate this limitation. Mobile users solve any issues use computing cloud services. Recent reduction in the cost of cloud servers enables mobile users to use these services more widely. Currently, many companies develop multiple applications (Google, Gmail, Maps and Navigation systems for Mobile, Voice Search, Mobil Me from Apple, Live Mesh from Microsoft) for mobile equipment users, which enables them to widely use mobile computing clouds [5]. According to the calculations of the Juniper Research centers' analysts, development speed on computing clouds increases at 88% due to software and applications.

In mobile computing clouds, processing and memorization of information in carried out outside of the mobile devices. Mobile devices connect to internet networks through base stations (GPS, 3G, 4G, and Wi-Fi etc.) and use

978-1-5090-4499-3/17/$31.00 ©2017 IEEE

necessary services in mobile computing clouds. Currently, users widely use three types of cloud services (IaaS, PaaS and SaaS). IaaS (Infrastructure as a Service) service allows users to use computing and memory resources (for example, Amazon Elastic Cloud Computing - EC2 and Amazon S3 - Simple Storage Service) of the cloud system. PaaS (Platform as a Service) service is a platform allowing users to use operation systems and special applications (Google App Engine, Microsoft Azure, Azache, My SQL etc.) located on virtual servers. SaaS (Software as a Service) service allows users to solve their issues using software located in the computing cloud servers (Google Apps, Google Docs, Autodesk) and applications. In the SaaS service, the user can obtain results by launching the software from computing cloud servers using internet service, without loading the resident portion of the necessary software to his/her own computer. Applications operate on servers provided by SaaS service provider and send the processed information to the user. Thus, user does not purchase the application and pays according to the use. Currently, millions of mobile users are widely using mobile applications (mobile commerce, mobile education, mobile health, mobile games etc. fields) using mobile cloud provider services [4, 6]. Developed mobile applications do not depend on the type of operation systems and device type of the mobile devices. For this reason, number of mobile device users that use cloud services increases rapidly every day. In this case, mobile devices with minimal computing and memory resources act as thin client terminal connected to Internet network.

According to the calculations of the Gartner company analysts, number of Internet users will reach 2.16 billion in 2016. This number will rise to 2.56 billion in 2018.

Currently, despite centralized clouds have high computing and memory resources, processed information cannot be delivered to the users on a high speed. Rapid increase of number of users on computing clouds loads the network which causes significant delays in delivery of processed information to the user. Cloud computing resources have to be located closer to the users in order to eliminate this deficiency.

Computing cloud service users use two modes: offline and online. Let’s assume that virtual computing devices are acquired from the cloud for solution of issues that require vast computing power. He/she (One) sends the issues to the cloud and obtains the result after a certain period of time. In this mode, there is no direct connection between the user’s computer and the cloud during the period necessary for the solution of the problem. But there are such issues, where user and cloud are connected until the problem solution process is complete. This, leads to loading of the network during solution of the mentioned problem. Mobile computing cloud technologies are used to eliminate the network load. Sometimes it is also referred to as decentralized cloud technologies in literature. Architecture of mobile computing clouds is shown (figure 1).

As seen on mobile computing cloud architecture, system consists of several components: mobile users (mobile device, smartphone etc.), mobile connection operators, wireless connection devices (Wi-Fi access point), internet provider (ISP), and computing cloud providers (Amazon IBM, Google, Microsoft etc.). as seen on the diagram, mobile users connect to Cloud Computing over internet using base stations (Mobile connection, Wi-Fi Access Point) such connections cause high load of the networks and delays in provision of processed information. At the same, these services are expensive.

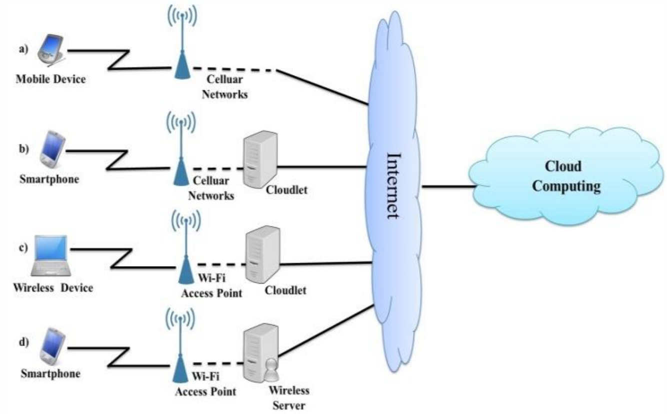

Figure 1. Architecture of Mobile computing clouds.

Cloudlets (small clouds) are developed near base stations (cellular networks, Wi-Fi Access Point) using Cloud Computing connection options b, c, and d by the users in order to eliminate indicated deficiencies [4, 7].

Thus, mobile servers (Cloudlets) beside base stations of mobile operators are developed for vaster use of cloud technologies by mobile equipment. Cloudlets (small computing clouds) are devices (server) located close to the users and provide faster delivery of necessary information to the users from the central servers. After the user completes his/her work, if necessary, information is re-loaded to the central servers. Cloud services necessary for the users can be implemented through Cloudlet servers [7, 8, 9]. Let’s assume that, if a user needs to work with a software (SaaS service), he calls that software to the closest mobile service, works on it for a while and then, sends that software to the central server. This, reduces the use cost of the clouds, reduces the issue solution period and loads the network less.

Advantages of mobile computing clouds are listed below [7 ÷ 9]:

- Increasing battery life of mobile devices;
- Increasing computing and memory resources of mobile devices;
- Storage of user data on several back-up computers increases reliability, which in its turn reduces information loss risk on mobile devices;
- Dynamic distribution of resources. This provides timely provision of necessary resources to the user without prior ordering;
- Increases network bandwidth;
- Service charge costs less;
- Organizing the storage of user data on clouds enables to use this information from any place worldwide;

- Increased scaling. System meets the computing and storage demands of the user for a short-term due to flexible resource capability;
- Security capabilities of the system. When mobile users use computer services they encounter problems related to storage and protection of information on clouds. Despite service providers protect the user information through different security software, at the same time they can observe that information. For this reason, confidentiality problem is very relevant. Security applications developed for mobile devices are used in order to solve these problems and prevent illegal interferences. These pieces of security software use a certain part of computing and memory resources of mobile devices. For this reason, it would be reasonable to transfer threat detection software to cloud servers from mobile devices. Cloud AV platform has been developed for this purpose; it provides multiple services detecting malware on clouds and mobile devices alike.
- Easy integration. Meets user demands by integrating services of different provides through clouds and Internet.

Scope of technical capabilities (number of virtual machines, storage space, network bandwidth etc.) of servers used for development of small clouds near base stations is among important issues. At the same time, provision of connection among developed small clouds is a relevant issue. At the same time, provision of connection among developed small clouds is a relevant issue.

On one hand users require high computing resources, and on the other, they wish to use small mobile devices. Currently, cloud technologies are used in order to eliminate the mentioned deficiency of the mobile devices (computing and memory resource deficiency).

On the other hand, use of the traditionally centralized cloud services can meet the demand for high computing and memory resources. But results or data are obtained with delay in this type of cloud services. For example, Google has developed an online translation service (Google Translate API), which enables users that speak different languages to speak to each other. Mobile devices used for this service (smartphones) do not translate independently and send the words and sentences to Google Cloud platform servers. Translation is carried out on the servers and results are sent to people talking to each other. We feel that the service is operating slowly in such type of services. While translation is carried out within seconds, its delivery to necessary destination over the network is delayed. Thus, it would be reasonable if the translation software was located on local servers (cloudlet) close to its users and rapid implementation of translation in real-time mode was enabled. At the same time, location of the software on closest cloud servers provides cheap, fast and quality service to the users using the SaaS service.

Users are guided by following criteria while using cloud technologies:

- Expenses required for problem solution;
- Problem solution duration;
- Provision of security of user information;
- Fast and reliable delivery of data and the results to the users.

Indicated criteria depend on the geographical distance of the user from the cloud server, computing power of the virtual machine and network load level. For this reason, it is attempted to place mobile cloud technology servers near the users. It would be reasonable to create Cloudlets near each base station.

Article studies conditions necessitating development of mobile computing clouds (cloudlets) and solves issues of forecasting location time of Cloudlets near certain base stations.

III. Problem Statement

It is known that establishment of Cloudlets near base stations (cellular networks, Wi-Fi access point) in the architecture of the mobile computing clouds depicted in Picture 1 is related to increasing demand of the mobile users for computing resources [10]. In order to determine Cloudlet establishment necessity, traffic and separate parameters in base stations must be monitored. Following are among monitoring parameters used for analysis:

- Network delays. Delays of network request responses, cloud software responses etc. must be measured. Passage of a certain mark by the delay renders the use of the system impossible. This is the main parameter for assessment of the used network in systems created using global clouds and implemented operations. For this reason, this parameter becomes an object of study as it changes time-wise and based on number of users. For example, studies conducted at the Telecooperation lab of Technische Universität Darmstadt demonstrate that the marks of this parameter are continuously growing (figure 2). Here, dependence of network delays on traffic volume on Cloud Asia, Cloud US West and Cloud EU is demonstrated. İn this diagram, traffic volume is presented as Kb/sec on X axis. On Y axis, we can see that traffic delay time increases in multi-fold depending of traffic volume marks [11]. Traffic volume increases as the number of users increase and as the users use more complex software pieces;
- Using computing resources. Computing force of processors used during development of cloudlets must be higher than the computing force required by the users;
- Using memory resources. Memory resource requirements can limit Cloudlet development. Due to this requirement, memory size of the server located in the Cloudlet must be higher than the computing power required by the users.
- Completion time of considered necessary operation. This time is usually related to network delays between the base station and global cloud. Growth

of network delays increases completion time of the operations;

- Operation completion energy. This can be defined as energy used by all equipment in unit time considering all operations of the users connected to the base station. Amount of energy spent in unit time can reach the limit threshold at a certain number and condition of mobile users. After reaching this threshold, energy spent on operations does not increase despite network delays.
- Operation price. Here, price for each operation and all operations conducted by the users connected to the base stations can be monitored. This parameter increases with network delays.

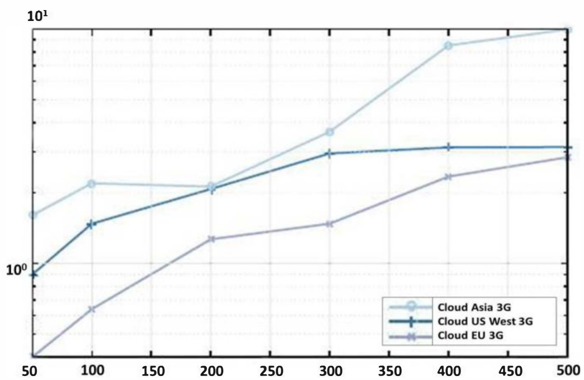

Figure 2. Comparative diagram of dependence of network delays of traffic volume on Amazon clouds.

In order to provide monitoring and analysis results' adequacy, mentioned parameters must be calculated and analyzed with equal time intervals. Article aims to solve issues such as creation of a final integral indicator based on parameters calculated at equal time intervals, smoothing the values on this indicator with wavelets transforms and forecasting future value of the integral indicator applying Neuron networks. Cloudlet development can be decided by analyzing approaching limit time.

IV. Forecasting Cloudlet Development Time

Let's assume that, during a certain T period M number of values are obtained by monitoring the values of J number of parameters. If we mark parameters as $P_1, P_2,, P_3, \dots,, P_J$ we will get this matrix.

$$\begin{pmatrix} P_{11} & \dots & P_{1M} \\ . & . & . \\ P_{J1} & \dots & P_{JM} \end{pmatrix} \quad (1)$$

In some cases, normalizing parameter values and writing their values as untitled values with maximum level of 100 can be useful. Thus, value of each parameter is written as the percentage of its own maximum value. If we denote the corresponding maximum value of each jth parameter as P_{jmax}, the following can be written:

$$P_{jm} = P_{jm} * \frac{100}{P_{jmax}}$$

Thus, the values of (1) matrix will be normalized to 100 as maximum value.

The importance or weight of each parameter in Cloudlet development can be determined by experiments. Those weights can be written as $(w_1, w_2, \dots, w_J)$ vector, which satisfies the following condition:

$$\sum_{j=1}^{J} w_j = 1 \quad (2)$$

If we multiply (2) weight vector by (1) matrix by using the rule of multiplication of matrices, the values of integral parameters can be obtained [12]:

$$(I_1, I_2, \dots, I_M) = (w_1, w_2, \dots, w_J) * \begin{pmatrix} P_{11} & \dots & P_{1M} \\ . & . & . \\ P_{J1} & \dots & P_{JM} \end{pmatrix} \quad (3)$$

here, I_m is calculated as $I_m = \sum_{1=1}^{J} w_i * P_{mi}$.

The values of $(I_1, I_2, \dots, I_M)$ vector obtained for integral parameter can be desctibed as $I_1, I_2, \dots, I_M$ time series and used for predicting the time of Cloudlet development. Time serious acquired in necessary cases can be smoothed with applying various methods [13, 14].

The analysis of the change of integral parameter generated as a result of monitoring of base stations with time enables to predict the approaching time of Cloudlet development. For this purpose, neural networks with one hidden layer can be employed (figure 3) [15, 16]. In this case, N1 number of inputs of neural network are smoothed and specified from repaired $I_1, I_2, \dots, I_M$ values. To study neural network,
$\{I_{M-(N1+k-1)}, I_{M-(N1+k-1)+1}, \dots, I_{M-(N1+k-1)+(N1-1)}\}$
can be defined as k number of input values and $I_{M-(k-1)}$ as output values corresponding to each input set.

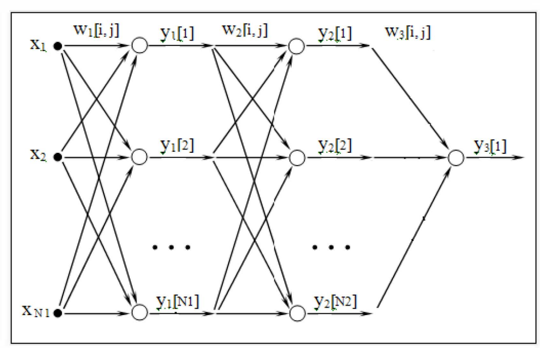

Figure 3. Neural network for predicting time series.

The process continues until output values are sufficiently close to designated values. Further value (M+1) and (I_{M+1}) of integral parameter after learning Neuron can be calculated by neuron network.

If estimated value is greater than specified boundary value (I_{lim}) of parameter, Cloudlet development time can be predicted:

$$I_{M+1} \geq I_{lim}$$

V. CONCLUSION

The article explores specific parameters necessary for Cloudlet development in Base stations, which establishes connections between mobile users and Mobile calculation clouds. Integral parameters were generated to exhibit the importance of Cloudlet development but using weight vector; the latter shows the importance of investigated parameters. The values of integral parameter are considered as time series and smoothed by applying wavelet conversions, and some missing values were repaired. To predict the future value of this time series, three-layered neural network with hidden layer was implemented. The future value of parameter was calculated in neural network output. If the value exceeds the specified maximum value of parameter, it can predict the time of Cloudlet development.

It is to be noted that, any important parameter among the parameters monitored in base station can be used instead of integral parameter.

The tools and methods presented in the article can also be implemented for the solution of similar problems.

ACKNOWLEDGMENT

This work was implemented through financial aid of Scientific Development Fund under the President of the Republic of Azerbaijan – Grant N EİF-2014-9(24)-KETPL-14/02/1.

REFERENCES

[1] Marios D. Dikaiakos, George Pallis, Dimitrios Katsaros, PankajMehra, Athena Vakali.Cloud Computing -Distributed Internet Computing for IT and Scientific Research// IEEE internet computing, 2009. № 9, pp.10-13.

[2] Alguliyev R.M., Alekperov R.K. Cloud Computing: Modern State, Problems and Prospects// Telecommunications and Radio Engineering, 2013, vol.72, no.3, pp. 255-266.

[3] Rashid Alakbarov, Fahrad Pashayev, Mammad Hashimov.A Model of Computational Resources Distribution Among Data Center Users . International Journalof Advancementsin Computing Technology. Volume 7, Number 2, Mar ch 31, 2015,pp. 1-7

[4] Hoang T. Dinh, Chonho Lee, Dusit Niyato, and Ping Wang. A survey of mobile cloud computing: Architecture, applications, and approaches. Wireless Communications and Mobile Computing. Volume 13, Issue 18. 25 December 2013 .Pages 1587–1611

[5] Han Qi , Abdullah Gani. Research on Mobile Cloud Computing: Review, Trend and Perspectives. https://arxiv.org/ftp/arxiv/papers/1206/1206.1118.pdf.

[6] Mukesh Goyal, Sukhwinder Singh. Mobile Cloud Computing. International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463 Vol. 3 Issue 4, April-2014, pp: (517-521),

[7] Fangming Liu, Peng Shu, Hai Jin, Linjie Ding, Jie Yu, Di Niu, Bo Li. Gearing resource-poor mobile devices with powerful clouds: architectures, challenges, and applications. IEEE Wireless Communications (Volume: 20, Issue: 3, June 2013) Page(s): 14 - 22

[8] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for vm-based cloudlets in mobile computing," Pervasive Computing, IEEE,vol. 8, no. 4, pp. 14–23, 2009.

[9] Yujin Li and Wenye Wang. The Unheralded Power of Cloudlet Computing in the Vicinity of Mobile Devices. Globecom 2013 - Wireless Networking Symposium Pages 4959-4964

[10] Usman Shaukat, Ejaz Ahmed, Zahid Anwar, Feng Xia. Cloudlet deployment in local wireless networks: Motivation, architectures, applications, and open challenges Journal of Network and Computer Applications 62 (2016) pp. 18–40

[11] Christian Meurisch, Alexander Seeliger, Benedikt Schmidt, Immanuel Schweizer, Fabian Kaup, and Max Munlhauser. Upgrading Wireless Home Routers for Enabling Large-Scale Deployment of Cloudlets. https://www.tk.informatik.tu-darmstadt.de/fileadmin/user_upload/Group_TK/Slides/CM/2015_MobiCASE_Presentation.pdf

[12] David Cherney, Tom Denton, Rohit Thomas and Andrew Waldron. Linear Algebra. First Edition. Davis California, 2013. 436 p.

[13] Daubechies I. Ten lectures on wavelets. Izhevsk: Scientific-Resarch Center "Regular and chaotic dynamics", 2001, 464 pp.

[14] Daubechies, I. Orthonormal Bases of Compactly Supported Wavelets, Comm. Pure Applied Mathematics, vol. 41, 1988, pp. 909-996.

[15] R. Rojas. (1996) Neural networks. Springer-Verlag, Berlin, pp. 479.

[16] Stanislav Osovsky. (2002) Neuron networks for information processing. Publisher: «Finance and statistics» 343 pp.