On Networking and Computing Environments' Integration: A Novel Mobile Cloud Resources Provisioning Approach

Prodromos Makris, Dimitrios N. Skoutas and Charalabos Skianis
Department of Information & Communication Systems Engineering
University of the Aegean, Greece
Karlovassi, Samos Island, GR 83200
{pmakris, d.skoutas, cskianis}@aegean.gr

Abstract— **Mobile/wireless networking and cloud computing are nowadays revolutionizing the ways of communication and computation and the newly founded Mobile Cloud Computing (MCC) research area is inspired by the notion of complete networking and computing environments' integration. As a result, efficient MCC resource management frameworks should simultaneously take into consideration both: a) wireless/radio access resources pool aiming at always-best connectivity contexts and b) computing resources pool for data processing/storage aiming at flexible virtualized infrastructure sharing solutions. In this paper, we propose a Mobile Cloud Resources Provisioning (MCRP) scheme, which is flexible enough to adapt to the various general MCC reference use cases being described. The main novelty feature of the employed MCC Service Admission Control algorithm lies in the fact that it jointly handles radio and computing resources rather than confronting the problem as two independent resource management sub-problems. The performance of MCRP scheme is evaluated via simulation results showing that the assumed context-aware service admission control policies can lead to efficient mobile cloud resources provisioning outcomes.**

Keywords – mobile cloud computing, cloudlet, context aware, JRRM, resource allocation

I. INTRODUCTION

As long as the integration of computing and networking environments is continuously boosted with the assistance of innovative ICT architectures and technologies, implications of Cloud Computing (CC) paradigm, which are applicable in mobile and wireless networking area are increasingly gaining ground [1]. Indeed, Mobile Cloud Computing (MCC) is an emerging research area and is introduced as the integration of CC into the mobile environment [2]. The ultimate vision is to fulfill the dream of providing "information at everyone's fingertips anywhere at anytime" and as computation capabilities of mobile terminals (MTs) will always be a compromise, MCC aims at efficiently using CC techniques for data storage and processing on MTs, thereby reducing their limitations [3] [4]. Other major MTs-related technical restrictions are short battery lifetime, varying wireless channel conditions and high network latency, all of them hindering the remote display functionality of cloud applications on mobile devices [5].

Going back to the late '90s, when 3G networks were standardized, a hierarchical cell structure model was proposed in order to better cope with the next generation mobile networking continuum challenges. More specifically, pico, micro, macro and global cell concepts were introduced based on cells' geographical expansion and till now, this hierarchical cell structure is considered as the basis for architectural mobile networking innovations [6]. Nowadays, inspired by this widely accepted abstract paradigm, the future computing continuum is expected to embrace distant cloud infrastructures, proximate cloudlet infrastructures, communicating objects and smart devices.

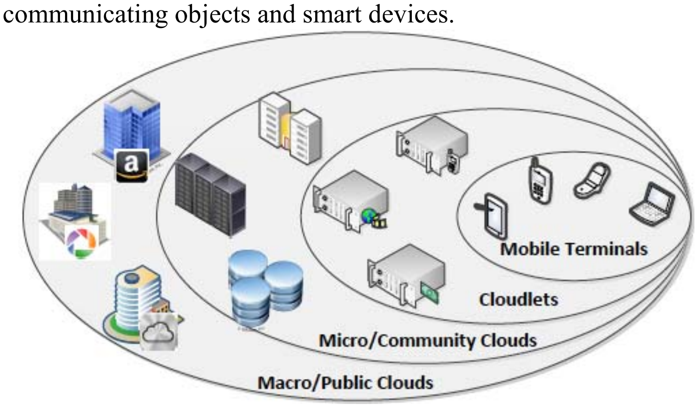

Figure 1. CC Hierarchical Structure.

TABLE I. CC HIERARCHICAL STRUCTURE ATTRIBUTES

CC Hierarchical Structure	Expansion	Applicability	Pros & Cons
MacroCloud	Public Clouds	Amazon/Apple/Google/ Microsoft large datacenters/datafarms	(+) high availability, (-) high WAN latency
MicroCloud	Community Clouds	e-gov/health/learning/ banking/commerce/ logistics datacenters	(+) targeted industrial/institutional sectors applicability (-) high WAN latency
Cloudlet	Local/Private Clouds	cafes, shopping malls, airports, stadiums, museums, campuses, train stations, etc.	(+) low latency, real-time services (-) context-aware applications range restrictions
Mobile Terminal	Personal MT Computing	smartphones, tablets, laptops, etc.	(+) security, privacy (-) battery power restrictions

As shown in figure 1 and table I, the CC hierarchical structure can be composed of four different computation layers. Today's MTs (i.e. laptops, smartphones, tablets, PDAs) can provide advanced computing capabilities (compared to the past decade), but the inherent problems of battery power restrictions are inevitable. Moreover, with the explosion of mobile applications market, the average mobile user demands for computing/storage power is much higher than the one that can be supported by an average MT and this gap is continuously growing [4]. As a result, mobile users need to access servers located in virtualized CC infrastructures in order to meet their increasing functionality demands. A cloudlet infrastructure consists of a cluster of servers well-connected to the Internet and available for use by nearby MTs. A cloudlet can be contained in a MCC hotspot together with a wireless access point comprising thus a datacenter-in-a-box concept [7]. These MCC hotspots can be placed in cafes, shopping malls, airports, stadiums, museums, city squares, campuses, train stations, etc. In these environments, mobile users may meet the demand for real-time interactive responses for specific-purpose context-aware mobile applications by low-latency, one-hop and high-bandwidth wireless access to the cloudlet. The problem with cloudlets is their restricted computing/storage capabilities, taking into account that they have to (by priority) meet the QoS demands of numerous users for specific-purpose context-aware mobile applications in a restricted geographical space. Consequently, for general e-* related MCC services (i.e. e-government, e-health, e-banking, e-commerce, e-learning, etc), the access to a community cloud infrastructure would be more appropriate for mobile users requesting corresponding specific e-* mobile applications. A community cloud or micro-cloud (cf. microcells in mobile networking continuum) is controlled and used by a group of industrial/institutional organizations, which have common or shared financial, security and legal objectives [8]. Despite of their high availability and targeted applicability to specific mobile users' demands, high WAN latency is an inevitable trade-off. The same problem is valid for public clouds or macro-clouds (cf. macrocells in mobile networking continuum). The main difference is that Amazon/Apple/Google/Microsoft etc large datacenters can provide virtually infinite computing/storage capabilities for any type of MCC service.

The main contribution of this paper lies in the assertion that in the integrated networking and computing continuum, state-of-the-art resource management frameworks and techniques have to be enhanced in order to confront the related research challenges from both networking and computing perspectives simultaneously. The paper is organized as follows: in section 2, we formulate the MCC problem from a resource management perspective by classifying related MCC variants in general use cases and by mapping state-of-the-art MCC challenges found in the international literature to the assumed use cases. We then describe the assumed system model and specify the applicability of the proposed Mobile Cloud Resources Provisioning (MCRP) scheme. In section 3, the assumed context-aware service admission control policies of MCRP scheme are presented and its joint resource management rationale is explicitly explained. In section 4, the MCRP performance is evaluated through event-driven simulation demonstrating its capability in meeting all general MCC use cases requirements. Finally, concluding remarks and future research guidelines are provided at the end of the paper.

II. Problem Formulation & System Model

A. Problem Formulation

In order to formulate the problem we are investigating, we first need to have a clear MCC paradigm realization. In figure 2, we present five MCC reference use cases. In a nutshell, different kind of MTs (i.e. cell phones, smartphones, tablets, PDAs, laptops, etc) have first to be connected to one of the available heterogeneous radio access technologies (e.g. LTE, WiFi, WiMAX, UMTS, HSPA, femtocell, WPAN, etc). Once a MT is connected to a wireless access technology, a plethora of different alternatives arise, regarding the ways that efficient partitioning of the computation workload of a mobile application can be achieved between the MTs and the cloud infrastructure. Dynamic partitioning of mobile applications can offer fine-grained flexibility on the "what to process/store where" problem in order to cope with the MCC heterogeneity challenges (e.g. workload, network, device, etc) [9].

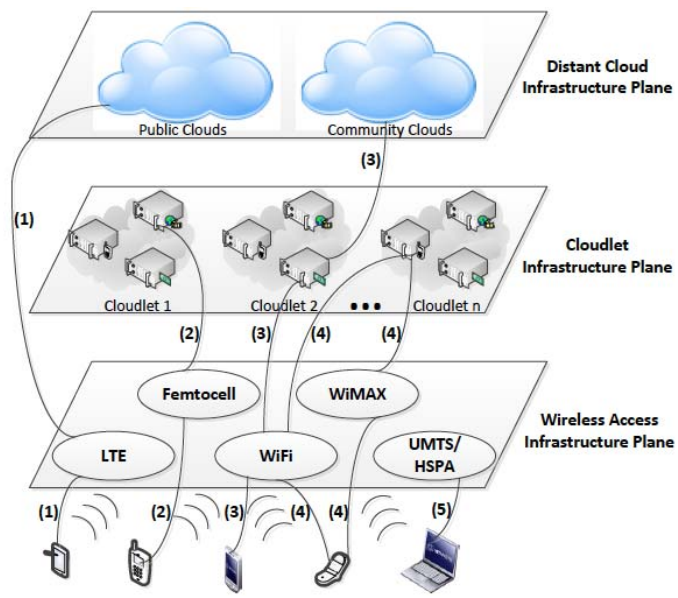

Figure 2. MCC Reference Use Cases.

More specifically, in use case (1), a MT is connected to a LTE network and uses computing resources from a public cloud infrastructure (e.g. gmail, dropbox, icloud, etc) in order a mobile application to be executed. In MCC context awareness use case (i.e. use case 2), a MT is connected to a femtocell and uses computing resources from a cloudlet infrastructure. For example, a user enters a shopping mall area and wants to know about all the events, discounts and social networking information related with his/her presence in this geographical area. If he/she is a regular visitor at the shopping mall, a clone of his/her smartphone can exist in one of the servers of the cloudlet infrastructure and thus real-time context-aware mobile applications could run both on the MT and the cloudlet. Furthermore, there is the case in which the dynamic partitioning of a context-aware mobile application can include computation offloading to a distant cloud infrastructure, too (i.e. use case 3). For example, for a e-gov mobile application, some “heavy” non real-time operations can be executed in a distant e-gov community cloud, some other real-time operations in the cloudlet before all data are displayed at the MT’s side. In the MCC mobility management use case, a MT is initially connected to a WiFi hotspot and cloudlet computing resources are used. While the mobile application is running, the user moves out of the coverage of the WiFi and thus seamless connectivity with another RAT (e.g. WiMAX) has to be achieved without violating QoS constraints. Finally, the fifth variant considers the trivial MCC use case, which assumes that complete execution of a mobile application on a MT is preferable compared to previous computation offloading scenarios [10].

TABLE II. MAPPING OF STATE-OF-THE-ART MCC CHALLENGES TO MCC REFERENCE USE CASES

MCC Challenges / MCC Use Cases	Limited bandwidth/ high latency	Mobility/ Connectivity	Context-aware MCC services/ Business logic	Security/ Privacy/ Trust	Energy efficiency
(1) Distant Cloud Connection	√	L	-	√	L
(2) MCC Context Awareness	L	L	√	P	-
(3) Optimal MCC resources distribution	L	L	L	√	√
(4) MCC mobility management	P	√	L	P	L
(5) Autonomous MT	L	L	-	-	√

Grade of impact √: high **P**: partial **L**: limited

Summarizing the MCC reference use cases depicted in figure 2, it is obvious that there are many more combinations between the three infrastructure planes. The classification of all possible combinations into five reference use cases was done in order an appropriate mapping with state-of-the-art MCC challenges to be realized (see table II). According to recent international literature [1] [2] [3] [4] [10] [11], the major MCC challenges are: a) limited bandwidth and high WAN latency, b) wireless access availability and mobility management, c) efficient context-aware MCC services provisioning and business logic issues, d) security, privacy and trust issues and e) energy efficiency issues. In table II, the grade of impact that MCC challenges have on the assumed MCC use cases is qualitatively evaluated. Hence, in “distant cloud connection” use case, the major problem is the high WAN latency incurred not allowing QoS constraints of real-time services to be met. Security, privacy and trust-related issues are also a major concern, because users often do not have an entire view of the ways that their data are processed in distant cloud infrastructures. In MCC context awareness use case, the major concern is on inventing ways that mobile applications can be useful to users exploiting location-based information and spatial augmented reality concepts boosting thus the MCC market towards introducing opportunities for new business players and models. In “optimal MCC resources distribution” use case, the major challenge is to dynamically partition computational operations of mobile applications between MTs, cloudlets and distant cloud infrastructures in order to optimize specific key performance indicators such as energy, security, QoS, etc. The fourth use case deals with mobility management in MCC and as such, seamless connectivity and handover frameworks have to be carefully designed [11]. Finally, in “autonomous MT” use case, energy efficiency issues are of major concern in order to extend the MT’s battery power autonomicity.

B. *System Model*

According to the problem formulation descriptions, there are numerous different alternatives in the MCC resources provisioning problem. In any case, there is a need to allocate both networking and computing resources simultaneously following joint resource management principles [12]. That is, wireless heterogeneous networking resources allocation problem has to be stressed in conjunction with the cloud computing resources allocation problem and not as two independent sub-problems.

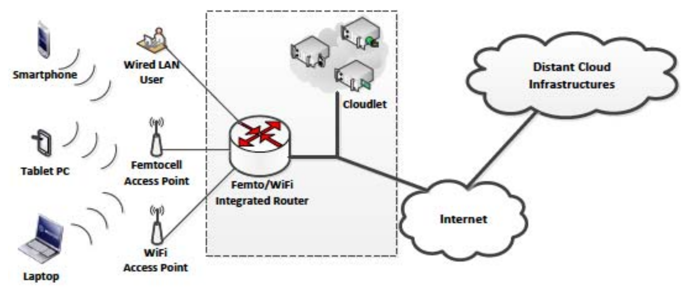

Figure 3. System Model.

In figure 3, the assumed system model for this paper is presented. The geographical area, in which our system model can be applied, is restricted and refers to scenarios of shopping malls, city squares, stadiums, museums, airports, train stations, campuses etc. In this kind of places, cloudlet infrastructures are soon going to be deployed in a wide range. These cloudlet infrastructures can be deployed much like WiFi and femtocell access points today and it would be relatively straightforward to integrate cloudlet and WiFi/femtocell APs hardware into a single easily deployable entity [4]. In [7], this integrated infrastructure (i.e. wireless access and cloudlet infrastructure) is introduced as “MCC hotspot” notion. In our system model, the two pre-referred infrastructures can also be physically separated (i.e. in different physical spaces) but in any case are virtually and functionally integrated.

As shown in figure 3, mobile users request for MCC services using their MTs such as smartphones, tablet PCs and laptops. These MTs can be connected to more than one heterogeneous radio access technologies (RATs) such as LTE femtocell and WiFi access points. Traffic can also be generated by wired LAN users. The resource management module, which resides at the femto/wifi integrated router has to dynamically partition the available backhaul traffic [13]. The available networking resources allocation procedure follows context-aware service admission control policies [1]. The resource management module, which resides at the cloudlet infrastructure, has to determine an integrated hybrid CC resources pool at any time instance and efficiently allocate computing resources using virtual partitioning methods and user-oriented, customizable infrastructure sharing approaches such as the ones described in our recent work in [8].

III. MOBILE CLOUD RESOURCES PROVISIONING SCHEME

The aim of the proposed MCRP scheme is to jointly handle both the radio and computing resources of the integrated system model as it is depicted in figure 3.

A. *MCC Service Classes and Initial Resource Partitioning*

Assuming four basic resources (i.e. Bandwidth, CPU, Memory and HDD Capacity partitions) to be managed by MCRP, each new Service Request (*SR*) can be expressed as a four-dimensional request vector of the form [$A_{(1,x)}$, $A_{(2,y)}$, $A_{(3,z)}$, $A_{(4,k)}$], x, y, z, k ∈ ℕ where $A_{(1,x)}$, $A_{(2,y)}$, $A_{(3,z)}$, $A_{(4,k)}$ are the requirements of the service for each of the four resources. Services that have similar requirements from the same resource are mapped to be served by the same partition.

An example of mapping two different MCC services, namely S_1 and S_2, to multiple resource partitions, is shown in figure 4. Both services require similar bandwidth and storage capacity and thus they are mapped to the same partitions $P_{(1,1)}$ and $P_{(4,1)}$. However, the two services have different processing (CPU and Memory) requirements and thus S_1 is mapped to $P_{(2,1)}$ and $P_{(3,1)}$, while S_2 is mapped to $P_{(2,2)}$ and $P_{(3,2)}$ respectively.

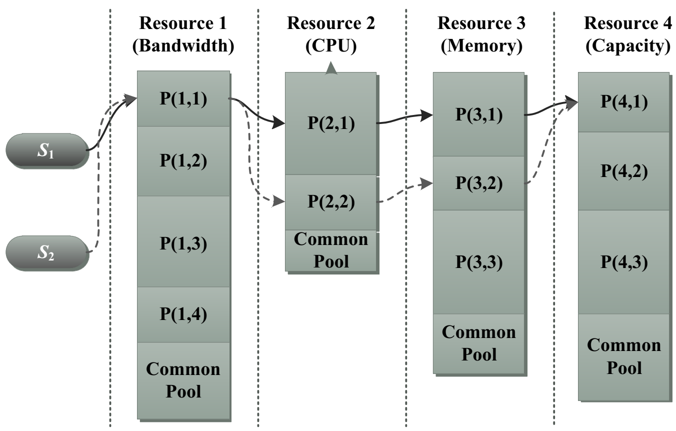

Figure 4. Example of mapping services to resource partitions.

Depending on the maximum value of the total offered traffic load intensity for each partition and for a given value of blocking probability, we calculate the size of each partition based on the *Erlang B* formula or, if required by the type of services, by following a more precise approximation [14]. In a nutshell, the request vector of a service can be corresponded to a partition vector of the form [$P_{(1,x)}$, $P_{(2,y)}$, $P_{(3,z)}$, $P_{(4,k)}$] x, y, z, k ∈ ℕ. MCC services with similar characteristics have the same partition vector and belong to the same Service Class (SC).

B. *Partitioning Limitations*

Although the size of the partitions is calculated on the basis of preserving a specific QoS in terms of blocking probability, in the general case we cannot always have a perfect alignment of the partitions among different kind of resources. This is due to various limitations, which are mainly imposed by the fact that we have to manage distributed resources, which are virtually integrated. For instance, the available bandwidth of the radio interfaces can be constrained by the available bandwidth of the backhaul line that provides access to a distant cloud. Furthermore, the excessive latency of accessing a distant cloud, may also exclude delay sensitive *SC*s by utilizing its computing resources. Consequently, the topology of such a distributed system as well as the technology used in its various components may crucially affect the degree of integration that can be achieved. In other words, the formed partitions may not always have the required capacity, even if the respective integrated resource is adequate to handle the total incoming traffic load.

C. *Partitions Adjustment*

Based on the above considerations, partitions that belong to the same partition vector [$P_{(1,x)}$, $P_{(2,y)}$, $P_{(3,z)}$, $P_{(4,k)}$] x, y, z, k ∈ ℕ of the *j*-th service class SC_j may offer different blocking probabilities, if they are independently calculated. In this case, the actual blocking probability B_j of the SC_j service class is lower bounded by the highest blocking probability that each of the individual partitions is able to provide:

$$B_j = \max\left\{B_{(1,x)}, B_{(2,y)}, B_{(3,z)}, B_{(4,k)}\right\},\ x,y,z,k \in N \quad (1)$$

Therefore, for example, there is no point to have low blocking probability at the radio access level, if a MCC service request is going to be blocked due to lack of computing resources. Having this in mind and aiming to preserve a high utilization of system resources, we have to re-examine all the partitions and determine which should be reduced in size.

Thus, subsequent to the initial calculation of the partitions, an adjustment phase follows. During this phase, the size of each partition that is allocated to *n* service classes SC_j, *j*=1, 2,.., *n* is downsized so as the blocking probability it offers is not lower than the minimum *L* of the blocking probabilities of these classes:

$$L = \min\left\{B_j\right\},\ j = 1,2,..,n \quad (2)$$

The resources that remain unallocated after this phase, if any, are forming a commonly shared partition (common pool) per resource (Fig. 4), which can be accessed by all *SC*s, as long as the limitations of III.B are met. Otherwise, if a part of these resources can be used only under specific limitations, then this part forms a separate partition. Therefore, the commonly shared partitions may be further divided, if necessary, in order to form smaller homogeneous partitions. These partitions are utilised by MCRP in order to increase system's utilization by absorbing small traffic load distribution variations.

D. *MCC Service Admission Control Algorithm*

Upon the arrival of an MCC service request that belongs to the *j*-th service class SC_j, MCRP has to check its partition vector. If the MCC service request is not mapped to a partition vector (i.e. excluded from service), then it is rejected. Otherwise, MRCP sequentially checks if the service request can have the required resources from the partitions included in the partition vector, or if this not possible, it checks if these resources can be allocated from the commonly shared partitions. If the resource allocation process is completed successfully, then the service request is accepted or else it is rejected.

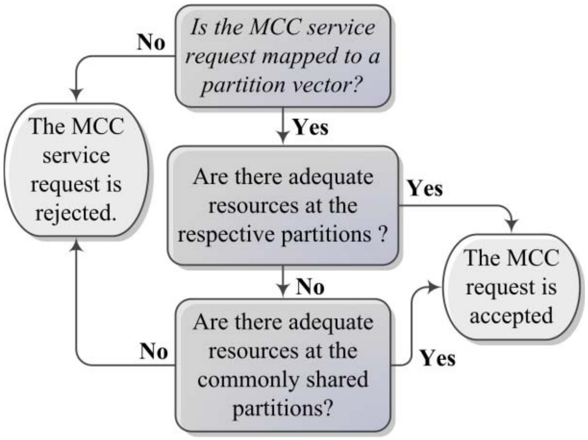

Figure 5. The proposed Mobile Cloud Resources Provisioning Scheme

IV. Performance Evaluation Results

In this section, we evaluate the proposed MCRP scheme through event driven simulation. MCC service requests are assumed to arrive according to a Poisson process, while their duration is exponentially distributed. For comparison purposes, we also evaluate a Complete Sharing Scheme (CSS), as well as a Complete Partitioning Scheme (CPS). CSS is a usual resource sharing approach for Cloud Computing environments [15], where the reserved resources are utilized as commonly shared pools, equally available for every incoming MCC request. On other hand the CPS scheme, performs typical complete partitioning of each resource independently, without addressing the need for joint resource management.

We assume, as shown at Fig. 4, that four main resources are shared among the end users, namely bandwidth, processing power, memory and storage capacity. For generality, we will refer to them as Resource 1, 2, 3 and 4, assuming that each MCC service requires a number of Basic Units (BU_m, m=1,2,3,4) from each of them. Consequently, the four-dimensional request vector of each new Service Request i (SR_i) can be expressed as (R_1 BU_1, R_2 BU_2, R_3 BU_3, R_4 BU_4) R_1, R_2, R_3, $R_4 \in \mathbb{R}$. For simplicity, we assume three kind of services: a High Demanding Service (HDS), a Low Demanding Service (LDS) and a Best Effort Service, which require resources that are multiples of a basic request vector of R_V=(1 BU_1, 1 BU_2, 1 BU_3, 1 BU_4). Mores specifically, the requirement of LDS in resources is R_{LDS}=R_V, HDS has a requirement of R_{HDS} =16·R_V and BES has a request vector of R_{BES} =5·R_V.

Our goal is to show that the management of mobile cloud resources should move from CSS schemes to Partitioning Schemes (PS) in order to be able to provide QoS provisioning. However, the employed PS schemes should take into account the peculiarities of a distributed system that is virtually integrated.

A. QoS Provisioning

In this scenario, we assume that we have to share the resources of a single cloudlet without any access to a distant cloud. The traffic load distribution is set to: HDS 20%, LDS 60% and BES 20%. The blocking probability of both HDS and LDS services should be less than 5%, while the BES service is served as long as there are available resources.

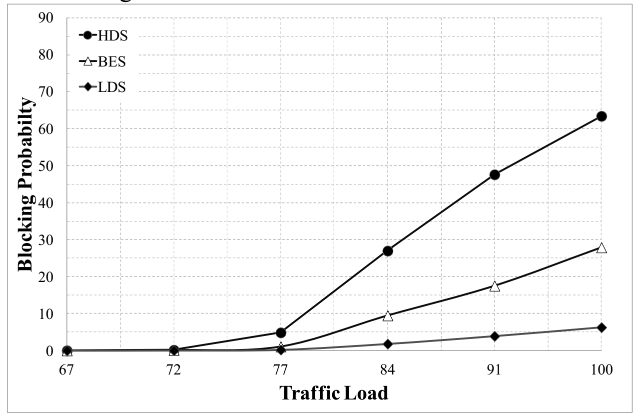

Figure 6. Blocking probabilities for the CSS scheme.

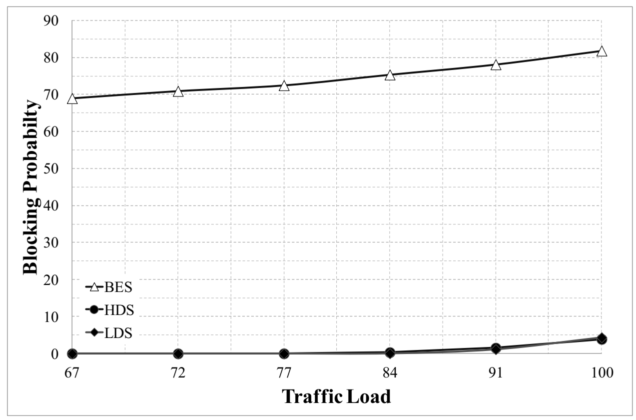

Figure 7. Blocking probabilities for the MCRP scheme.

As shown in figure 6, CSS is not able to provide QoS differentiation and thus the blocking probability of a service is based only on its requirements. The more resources a service requires, the more difficult is to be accepted. As a result, the blocking probability of HDS services exceeds 5% well before the system is fully loaded. On the other hand, MCRP is able to keep the blocking probability of both HDS and LDS services below the required blocking probability, as shown at figure 7, even when the system is fully loaded.

B. Adjustment of the partitions

Extending the previous simulation scenario, we assume that 15% of the system's storage capacity is derived from a distant cloud. We further assume that the HDS and LDS services are delay sensitive and they cannot tolerate the high latency of a connection to the distant cloud. By following a typical CPS approach, each of the LDS and HDS partitions will include approximately a 10% of resources, which will never be used due to the limitations at the storage level. On the contrary, MCRP reduces the size of the LDS and HDS partitions, at all resource levels, in order to be adjusted to the available storage capacity of the cloudlet. Subsequently, MCRP adds the released resources to the existing common pool partitions.

As a result, both CPS and MCRP schemes are able to provide the same, increased, blocking probability for the LDS and HDS services, while the extended common pool partitions formed by MCRP provide significantly lower blocking probability for the BES service requests. Thus, MCRP is able to provide better utilization of the available resources and this can be verified at figures 8 and 9, where the performance of CPS and MCRP schemes is shown respectively.

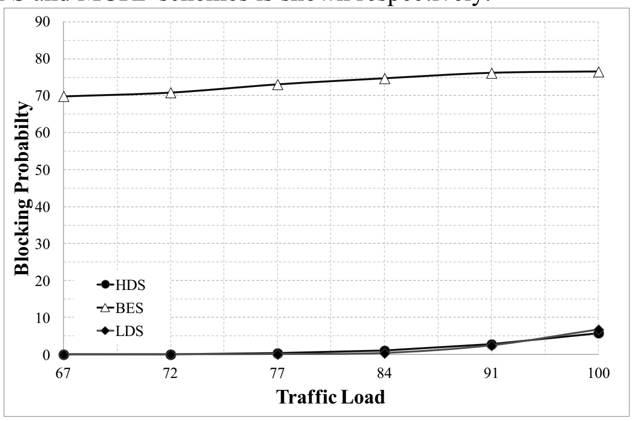

Figure 8. Blocking probabilities for the CPS scheme.

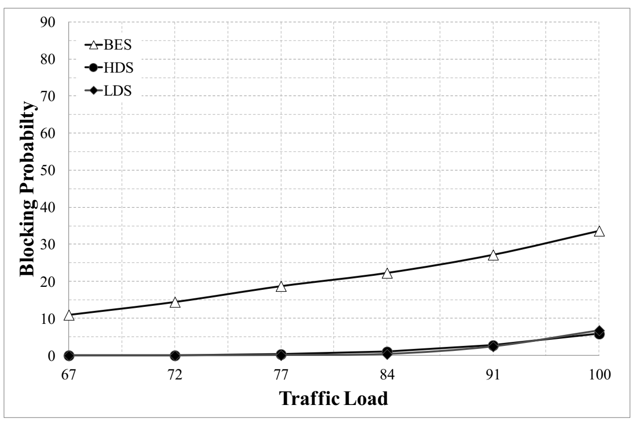

Figure 9. Blocking probabilities for the MCRP scheme.

V. CONCLUDING REMARKS

In this paper, we dealt with the problem of future networking and computing continuums convergence via the realization of the MCC paradigm. The main contribution of the paper lies in the assertion that in the integrated networking and computing continuum, state-of-the-art resource management frameworks and techniques have to be enhanced in order to confront the related research challenges from both networking and computing perspectives simultaneously and thus basic design guidelines for the development of corresponding resource management frameworks are introduced. The proposed MCRP scheme jointly handles both radio and computing resources of a customized MCC infrastructure and can be flexibly adapted to various MCC reference use cases such as the ones described in this work. Performance evaluation results show that MCRP outperforms typical complete sharing and partitioning schemes by exploiting context-aware service admission control policies. As future work, we aim to further elaborate on assumed MCC use cases concepts and deal with mobility management and seamless mobile services QoS provisioning challenges.

REFERENCES

[1] P. Makris, D. N. Skoutas and C. Skianis, “A Survey on Context–Aware Mobile and Wireless Networking: On networking and computing environments’ integration”, IEEE Communication Surveys & Tutorials, 2012, doi: 10.1109/SURV.2012.040912.00180.

[2] H. T. Dinh, C. Lee, D. Niyato and P. Wang, "A Survey of Mobile Cloud Computing: Architecture, Applications, and Approaches", *Wireless Communications and Mobile Computing (WCMC)*, doi: 10.1002/wcm.1203.

[3] L. Guan, X. Ke, M. Song and J. Song, "A Survey of Research on Mobile Cloud Computing", in Proc. IEEE/ACIS 10th International Conference on Computer and Information Science (ICIS), pp. 387-392, May 2011.

[4] M. Satyanarayanan, P. Bahl, R. Caceres and N. Davies, “The Case for VM-based Cloudlets in Mobile Computing”, IEEE Pervasive Computing Journal, vol. 8, no. 4, pp. 14-23, 2009.

[5] P. Simoens, F. De Turck, B. Dhoedt and P. Demeester, "Remote Display Solutions for Mobile Cloud Computing", IEEE Computer, vol. 44, no. 8, pp. 46-53, 2011.

[6] H. H. Chen and M. Guizani, “Next Generation Wireless Systems and Networks”, John Wiley & Sons Ltd., 2006.

[7] D. T. Hoang, D. Niyato and P. Wang, "Optimal admission control policy for mobile cloud computing hotspot with cloudlet", in Proc. *IEEE Wireless Communications and Networking Conference (WCNC)*, Paris, France, 1-4 April 2012.

[8] P. Makris, D. N. Skoutas, P. Rizomiliotis and C. Skianis, "A User-Oriented, Customizable Infrastructure Sharing Approach for Hybrid Cloud Computing Environments," in Proc. *IEEE 3rd International Conference on Cloud Computing Technology and Science* (CloudCom), pp. 432-439, 2011.

[9] B. G. Chun and P. Maniatis, “Dynamically Partitioning Applications Between Weak Devices and Clouds”, in Proc. *1st ACM Workshop on Mobile Cloud Computing and Services (MCS)*, San Francisco, USA, 2010.

[10] K. Kumar and Y. Lu, "Cloud Computing for Mobile Users: Can Offloading Computation Save Energy?", IEEE Computer, vol. 43, no. 4, pp. 51-56, 2010.

[11] A. Klein, C. Mannweiler, J. Schneider and H. D. Schotten, “Access Schemes for Mobile Cloud Computing”, in Proc. 11th International Conference on Mobile Data Management (MDM), pp. 387-392, May 2010.

[12] O. E. Falowo and H.A Chan, “Joint Call Admission Control Algorithms: Requirements, Approaches and Design Considerations”, Elsevier Computer Communications, vol. 31, no. 6, pp. 1200-1217, 2008.

[13] D. N. Skoutas, P. Makris and C. Skianis, “Optimized Admission Control Scheme for Coexisting Femtocell, Wireless and Wireline Networks” Springer Telecommunication Systems, Special Issue on Mobility Management in Future Internet, in press, 2012.

[14] Matthieu Jonckheere and Jean Mairesse “Towards an Erlang formula for multiclass networks”, *Queueing Syst. Theory Appl.* vol. 66, no. 1, pp. 53-78, 2010.

[15] B. Rochwerger et al., "Reservoir - When One Cloud Is Not Enough," Computer, vol. 44, no. 3, pp. 44-51, 2011.