Wireless and wired network convergence in support of cloud and mobile cloud services: The CONTENT Approach

A. Tzanakaki[1*], M. P. Anastasopoulos[1], G. Landi[2], G. Bernini[2], N. Ciulli[2], S. Peng[1], B. Rofoee[1], Y. Yan[1], D. Simeonidou[1], J.F. Riera[3], E. Escalona[3], J. A. Garcia-Espin[3], K. Katsalis[4], A. Korakis[4]
[1]University of Bristol, UK, *E-mail: anna.tzanakaki@bristol.ac.uk, [2]Nextworks, Pisa, Italy [3]i2CAT, Barcelona, Spain, [4]University of Thessaly, Greece

***Abstract*—The project CONTENT proposes a next generation ubiquitous converged infrastructure to support Cloud and mobile Cloud computing services. The proposed infrastructure facilitates interconnection of fixed and mobile end users with computational resources through a heterogeneous network integrating optical metro and wireless access networks. In this paper, we present the CONTENT converged network infrastructure and layered architecture which deploys cross-domain virtualization as a key technology. In accordance to the CONTENT proposal this paper also presents a novel virtual infrastructure planning scheme that takes a holistic approach considering jointly the network and computational resources.**

***Index Terms*— Mobile Cloud Computing, energy efficiency, queuing theory, Virtual Infrastructure Planning, Converged Infrastructures.**

I. INTRODUCTION

During the last decade, large-scale computer networks supporting both communication and computation were extensively employed to run distributed applications that deal with customer support, internet control processes, web content presentation, media services, file sharing etc. The current technology trend is cloud computing offering on-demand delivery of infrastructures, applications, and business processes in a commonly used, secure, scalable, and computer based environment over the Internet for a fee [1]. Cloud computing allows users to gain access to remote computing resources that they do not have to own. This introduces new business models and facilitates new opportunities for a variety of business sectors. At the same time it increases sustainability and efficiency as it reduces the associated capital and operational expenditures as well as the overall energy consumption.

It is predicted that cloud computing services are emerging as one of the fastest growing business opportunities for Internet service providers and telecom operators [2]. In addition, mobile internet users are expected to exceed in number the desktop internet users after year 2013, introducing a huge increase in mobile data, a big part of which will come from Cloud computing applications [2]. To address these emerging requirements cloud computing services are also becoming available to mobile users, introducing the concept of Mobile Cloud Computing (MCC), where computing power and data storage are moving away from mobile devices to remote computing resources [3].

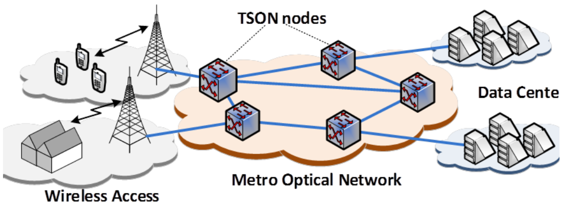

Fig. 1 Physical Infrastructure

Existing mobile cloud computing solutions allow mobile devices to access the required resources by accessing a nearby resource-rich cloudlet, rather than relying on a distant “cloud,” [4]. In order to satisfy the low latency requirements of several content-rich mobile cloud computing services such as high definition video streaming, online gaming and real time language translation [5], one-hop, high-bandwidth wireless access to the cloudlet is required. In the case where a cloudlet deploying small data centres (DCs) is not available nearby, traffic is offloaded to a distant cloud such as Amazon’s Private Cloud, GoGrid [6] or Flexigrid [7]. However, the lack of service differentiation mechanisms for mobile and fixed cloud traffic across the various network segments involved, the varying degrees of latency at each technology domain and the lack of global optimization tools in the infrastructure management and service provisioning phases make the current solutions inefficient.

To address these issues, a next generation ubiquitous converged infrastructure suitable to support cloud and mobile cloud computing services has been proposed in the context of the European project CONTENT [8] (Fig. 1). This infrastructure facilitates the interconnection of data centres (DCs) with fixed and mobile end users through a heterogeneous network integrating optical metro and wireless access network technologies. The proposed architecture integrates an advanced optical network solution offering fine (sub-wavelength) switching granularity with wireless Long Term Evolution (LTE) access network technology supporting end user mobility through wireless backhauling. To support the Infrastructure as a Service (IaaS) paradigm as well as the diverse QoS needs of future Cloud and mobile Cloud services, the concept of virtualization across all technology domains is adopted.

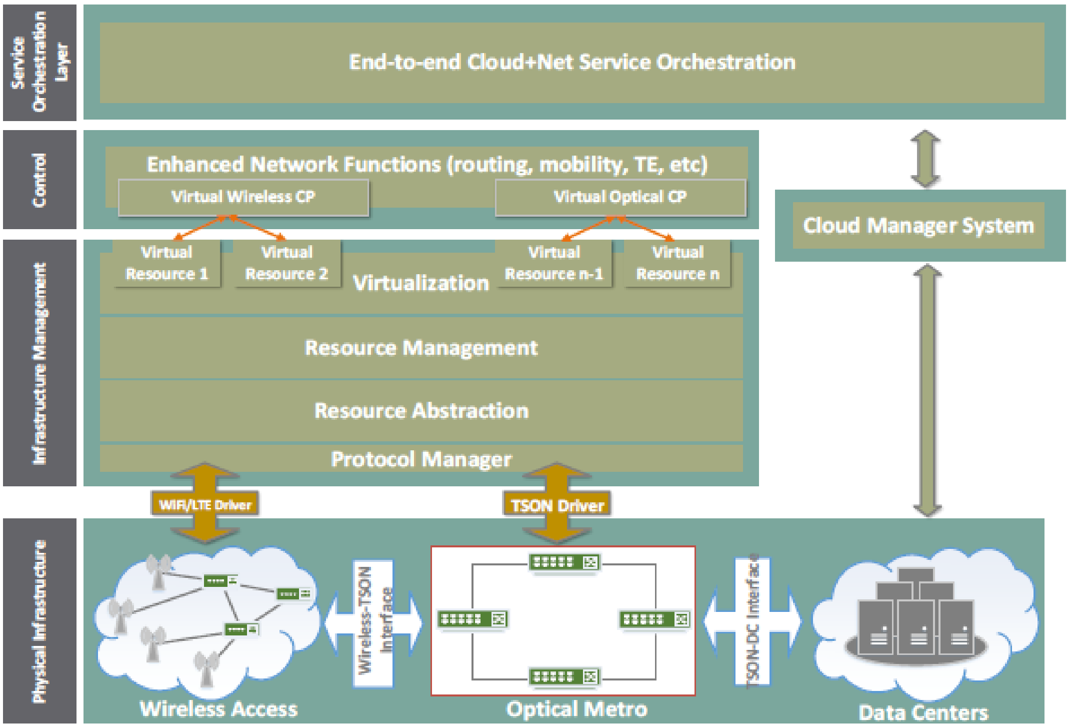

Fig. 2 The CONTENT architecture

In this paper, we present in detail the CONTENT converged network infrastructure as well as the details of layered architecture proposed to support efficiently and effectively cloud and mobile cloud services. As in the context of the CONTENT architecture cross-domain virtualization is a key enabling technology this paper also presents a novel virtual infrastructure (VI) planning scheme that takes a holistic approach considering jointly the network and IT technology segments. The adoption of this approach ensures allocation of the required resources across all domains extending the work presented in [8]. In addition, it enables the support of service requests and their specific characteristics such as low latency, QoS differentiation and mobility of end users and facilitates globally optimized solutions in terms of objectives such as energy consumption and resource allocation. Our modeling results identify trends and trade-offs relating to resource requirements and energy consumption levels across the various technology domains involved that are directly associated with the services characteristics.

The remaining of this document is structured as follows: Section II provides a functional description together with a detailed structural presentation of the proposed architecture. This includes the details of the individual layers involved. Section III, includes a discussion on the modelling framework developed with the aim to evaluate the CONTENT architecture and the associated results. Numerical results are provided in Section IV. Finally, section IV summarizes the conclusions.

II. Vision and Architectural Approach

The infrastructure model proposed by CONTENT is in accordance to the Infrastructure as a Service (IaaS) paradigm and is based on a layered architecture (Fig. 2). To support the IaaS paradigm, physical resource virtualization, generating virtual infrastructure slices, is enabled by a cross-domain infrastructure management layer. Connectivity services are provided over the virtual infrastructure slices, created by the infrastructure management layer, through the virtual infrastructure control layer. Integrated end-to-end network, cloud and mobile cloud services are orchestrated and provisioned through the service orchestration layer. More details on the individual architecture layers are provided below.

A. Physical Infrastructure Layer

The heterogeneous physical infrastructure (PI) comprises a hybrid wireless access network (LTE/WiFi) domain, and an optical metro network domain interconnecting geographically distributed DCs (Fig. 1). The optical metro network is based on the Time Shared Optical Network (TSON) technology supporting frame-based sub-wavelength switching granularity [10]. TSON will offer connectivity to the wireless access and DC domains by providing flexible rates and a virtualization friendly transport technology. The wireless access part comprises a converged 802.11 and 4G (LTE) access technology network, used to support cloud computing services through the NITOS wireless testbed [11]. The backhaul network comprises the packet core network that is used to transport traffic to the Gateway that will interact with the TSON Gateway.

Communication between the two technology domains of optical and wireless networks will be established though Ethernet interfaces. The wireless domain will provide 1GE links, which are then aggregated in to 10GE and fed into the TSON system. The two domains will use VLAN technology for end-to-end data path integration for converged network slices and to define isolation between different coexisting virtual networks. The VLAN tagged traffic will then be directly sent into the TSON network. The DC nodes as TSON clients, will also communicate with the TSON nodes through Ethernet interfaces. Different Virtual Machines (VMs) or VM groups can be identified and tagged using VLAN tagging, and TSON will be able to transfer them between the DC and wireless domains.

B. Infrastructure management

The Infrastructure Management Layer (IML) is the architectural layer responsible to provide management of the physical resources. The IML functionality is twofold. On the one hand, it is the element of the architecture devoted to the converged management (e.g. monitoring, abstraction, discovery, or lifecycle management) of physical resources populating different technology domains. On the other hand, it is responsible for the creation of isolated virtual infrastructures composed of resources belonging to different technology domains. Additionally, the management layer, which lies directly over the physical infrastructure, deploys the Cloud Management System (CMS). CMS is used to facilitate management of computational resources.

The lower part of the IML contains components (i.e. Drivers) that are responsible for retrieving information and communicating with the various technology domains. Once the information has been acquired, the resources are abstracted and virtualized.

Virtualization is the key functionality of the infrastructure management layer and given that CONTENT adopts a cross-domain and cross-technology virtualization solution is also a key innovation point of the proposed approach. Cross-technology virtualization allows the creation and operation of infrastructure slices including subsets of the network resources connected to the different computational resources in support of fixed and mobile services.

The virtualization strategy adopted per-domain, as well as the cross-domain virtualization strategy [9] is based on two different approaches, namely resource-based and service-based virtualization. The functional architecture depicted in Fig. 2 supports any of the two virtualization implementation. The internal virtualization component is the element responsible for the creation and management of the different virtual resources used to compose the required isolated slices. The virtual resources created will hold a set of capabilities (e.g. protocols supported, or actions that can be performed over the resource), which will be accessible from the corresponding network control layer. These capabilities will depend on the virtualization approach selected per domain.

C. Virtual Infrastructure Control Layer

The converged Virtual Infrastructures, delivered through the Infrastructure Management Layer described in the previous section, are jointly operated through a unified control layer based on the Software Defined Networking (SDN) paradigm. This layer, called *Virtual Infrastructure Network Control Layer*, implements converged control and management procedures for dynamic and automated provisioning of end-to-end connectivity in support of QoS-guaranteed cloud services for mobile users. The network services span across the wireless and metro networks, and are coordinated to provide efficient utilization of the overall virtual network resources, while exploiting the specific benefits offered by the different technologies deployed in each domain.

The Virtual Infrastructure Network Control Layer is internally organized in a hierarchical structure of network functions, where the converged control of the Virtual Infrastructure relies on “enhanced” network applications built over elementary services of the optical and wireless domain. Following this approach, at the bottom level directly on top of the virtualization layer, basic control plane functions provide simple services like resource configuration and monitoring, intra-domain connection setup, etc. These functions are provided through an SDN controller, potentially distributed in several interacting controllers, which manages the entire infrastructure, but operates with separated per-domain scopes through specialized modules able to deal with the specific constraints of the underlying technologies. It should be noted that in the particular case of CONTENT this architectural concept is applied to virtual infrastructure domains characterized by specific technologies (e.g. TSON for the metro network). However the overall approach can be easily extended to multi-layer infrastructures based on different types of technologies (e.g. packet switching), through the usage SDN controllers able to manage the given domains.

The services at the bottom level expose interfaces that allow customizing and adapting their behaviour, so that they can be easily used as basic elements to compose and develop more complex network functions on top. Moreover, the low level functions can be also disabled and fully replaced by corresponding, more powerful, high-level functions. The CONTENT architecture can accommodate different types of enhanced network applications, which may cooperate together for a more efficient management of the heterogeneous Virtual Infrastructure. A first category is dedicated to the provisioning of end-to-end, multi-layer and cloud-oriented connectivity services. These functions may operate on a high level vision of the Virtual Infrastructure topology, abstracting the details of each technology domain, and they usually expose rich APIs towards the Service Orchestration Layer to enable the joint composition of network and cloud services. Another category of network applications is dedicated to the internal management and automated re-optimization of the virtual network infrastructure. They do not usually expose any operational interface towards the upper layers and may be specialized to work either over single network segments or with a cross-domain scope.

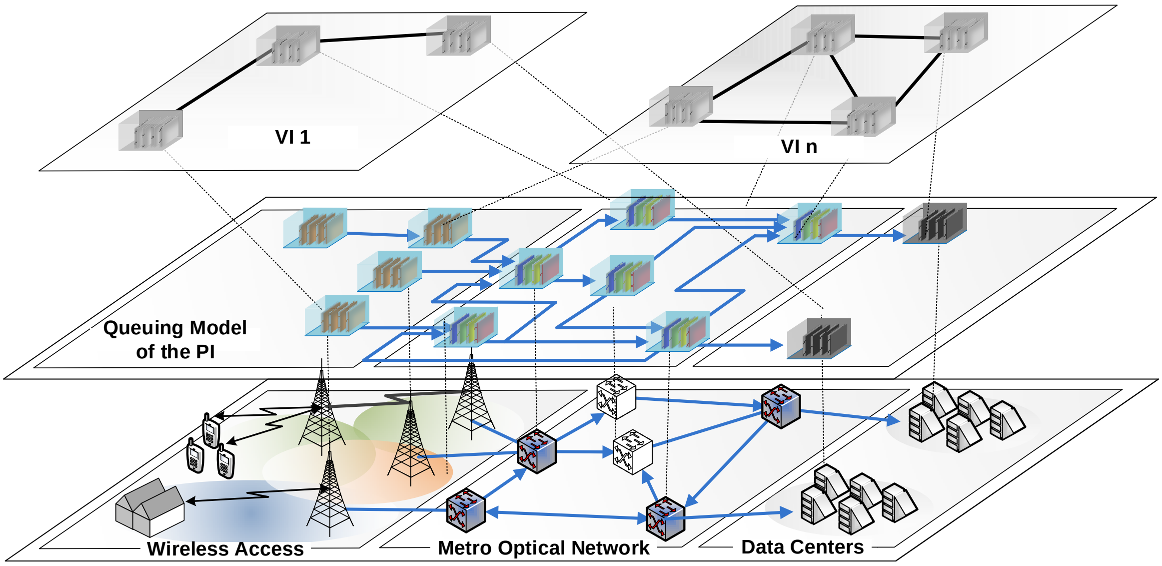

Fig. 3 Mapping of the VI requests onto the multi-queuing model of the converged wireless, optical and DC infrastructures

D. Service Orchestration Layer

On top of the Virtual Infrastructure Network Control Layer, a Service Orchestration Layer is in charge of composing and delivering cloud services to the mobile end-users, properly integrated with dedicated wireless connectivity services. The Service Orchestration Layer combines network and cloud resources available in the Virtual Infrastructure, and provides a complete and converged cloud service that matches the user's requirements. Cooperation between the control and the orchestration layers is the key factor for consistent and converged management of the entire Virtual Infrastructure, from the network to the cloud domains, while continuously fulfilling the requirements of the cloud services.

More specifically, Cloud services are dynamic entities that evolve during their runtime, with changes in the characteristics of the reserved computational resources and in the traffic load generated by the applications running on the distributed VMs. In addition, the mobile nature of the CONTENT cloud users may lead to further changes in the characteristics of the network access available for the end-users and in the profile of the user-to-DC traffic. To address these characteristics, the CONTENT solution operates the virtual infrastructure with flexibility, so that the whole composition of resources allocated for a given service instance can be automatically re-adapted in short time intervals.

III. ARCHITECTURE EVALUATION

A. Problem description

In this section, a stochastic non-linear programming model is proposed suitable from the planning of VIs over an integrated platform comprising a cellular LTE system for the wireless access domain and an optical metro network that interconnects the mobile computing devices with the computing resources. To achieve this, a multi-layer approach is adopted in which the lower layer contains the physical resources while the upper layer is comprised by a set $\mathcal{I}$ of I VIs $\mathcal{I} = (1, \dots, I)$. The PI is represented as a weighted graph $\mathcal{G}^p = (\mathcal{N}^p, \mathcal{E}^p)$ where $\mathcal{N}^p$ represents the set of PI nodes and $\mathcal{E}^p$ the set of PI links. Each VI_i $(i \in \mathcal{I})$ is modeled as an undirected graph $\mathcal{G}_i^v = (\mathcal{N}_i^v, \mathcal{E}_i^v, \mathcal{D}_i)$ where $\mathcal{N}_i^v$, $\mathcal{E}_i^v$ are used to denote the set of nodes and virtual links, respectively, and $\mathcal{D}_i$ is used to describe the set of demands. These demands may arise either from the optical or the wireless domain and need to be served by a set $\mathcal{S}$ of S geographically distributed DCs $\mathcal{S} = (1,2, \dots, S)$. Hence, each demand is associated with certain users groups, i.e., fixed or mobile, that require differentiated quality of service. For example, traffic demands corresponding to fixed cloud applications originate at the TSON edge nodes in the wired domain and need to be served by specific computing resources. A common characteristic of fixed cloud services is that due to the large amount of data that they generate they require very high level of network and computing capacities. Mobile traffic on the other hand is generated by the mobile devices and, compared to the fixed cloud services, requires lower levels of network and computing resources to operate. However, the main challenge it introduces is that in some cases it needs to traverse several hops before it reaches the IT resources through the optical metro network leading to increased end-to-end delays. The latter plays a key role in the VI planning process as it affects the decision of a mobile device to offload or not its traffic requests in the cloud. For example, if end-to-end delays are high, demands will be not offloaded to the cloud as there is increased probability that their corresponding QoS constraints to be violated. Apart from network delays, uncertainties related to the prediction of the locations of the mobile devices (in addition to the traffic volume demand uncertainties that exist in both fixed and mobile services) are also considered in the analysis.

The overall system architecture is illustrated in Fig. 3 where in the physical layer the TSON solution [10] has been adopted to interconnect the DCs with the fixed and the mobile users. The proposed VI planning scheme aims at identifying the topology and determine the virtual resources required to implement a dynamically reconfigurable VI based on wireless optical network and IT resources. The objective of the proposed approach is to minimize the total energy that is consumed by the power dissipating elements of the resulting converged infrastructures keeping, at the same time, the end-to-end delays below a specific threshold so as to allow the offloading of computational intensive mobile applications to the cloud.

B. Mathematical Modeling

A set of randomly selected nodes both in the wired and in the wireless domain are considered to generate demands. These demands apart from computing requirements have to also support the associated network requirements. In this formulation it is assumed that the granularity of optical network demands is a portion of wavelength (e.g. λ/100), while in the wireless domain the granularity is assumed to be 1 Mbps. As already mentioned, the identification of the suitable DC resources is part of the optimization output. To formulate this requirement the binary variable a_{dsi} is introduced taking value equal to 1 if demand d of VI_i is assigned to server s; 0 otherwise. It should be also mentioned that once a demand has been assigned for processing to a specific DC, it cannot migrate to another DC. This assumption is mathematically formulated using the following equation:

$$\sum_{s} a_{dsi} = 1, \qquad d \in \mathcal{D}_i, i \in \mathcal{I} \qquad (1)$$

Now, let h_{di} be the volume of demand d of VI_i, $\mathcal{P}_{dsi}$ be a set containing all the possible paths that can be used in order to transfer the traffic volume h_{di} to the server s and x_{dpi} be the flow realizing demand d on path $p \in \mathcal{P}_{dsi}$. Since in MCC the requested infrastructure capacity can scale up and down on demand, while the location of the mobile computing devices is uncertain, information regarding h_{di} is not precisely available. However, it can be described by a probability distribution function (pdf) that can be estimated based on history observations. Now let ξ be a random vector that contains all the uncertain parameters that are involved in the planning process. It is assumed that ξ can be described through a finite number of possible realizations, called scenarios, say $\xi = (\xi_1, \xi_2, \dots, \xi_K)$, with probabilities $p(\xi) \in [0,1]$. Therefore, h_{di} are scenarios in ξ, namely $h_{di}(\xi)$, with known pdfs.

Then, the following *demand constraints* should be satisfied in the VI:

$$\sum_{s}\sum_{p} a_{dsi}\, x_{dpi}(\xi) = h_{di}(\xi), \quad d \in \mathcal{D}_i, i \in \mathcal{I} \quad (2)$$

Summing up the paths through each link e ($e \in \mathcal{E}_i^v$) of the VI_i, the necessary virtual capacity $y_{ei}(\xi)$ of link e that can support all demands for the scenario ξ is given by the following expression:

$$\sum_{d}\sum_{p} \delta_{edpi}\, x_{dpi}(\xi) \le y_{ei}(\xi), \;\; e \in \mathcal{E}_i^v, i \in \mathcal{I} \qquad (3)$$

where δ_{edpi} is a binary coefficient taking value equal to 1 if link e of VI_i belongs to path p realizing demand d at server s; 0 otherwise.

The next step is to identify the necessary resources in the physical layer. Initially, the virtual capacities $y_{ei}(\xi)$ are treated as demands that need to be supported by specific PI resources. Then, following the same rationale with the previous analysis the PI's candidate paths q realizing virtual link capacity $y_{ei}(\xi)$ are determined, which in their turn, are used to evaluate the capacity $u_{gi}(\xi)$ that is required by each PI link g to support the demands of VI_i for the scenario ξ. An additional constraint to be taken into account is that apart from the network capacity constraints, the requested processing power (measured in Million Instructions per Second-MIPS) at each server s should not exceed its capacity ϕ_s.

So far, the proposed scheme ensures that there are sufficient network and processing capacities to support the requested services. Apart from network bandwidth requirements, end-to-end delay guarantees should be also provided. However, given that in highly loaded networks queuing delay is the dominant part of the end-to-end delay, the Virtual Infrastructure Control Layer described in Sec. II needs to be considered by applying relevant delay constraints in the service provisioning process across all the technology domains involved. These constraints should allow the VIs to reserve a specific portion of the receivers'/transmitters' queues at a TSON edge node or at an eNodeB, with the objective to maintain the end-to-end delay below a predefined threshold.

In order to mathematically formulate this issue, the PI is modeled as an open queuing network, in which its node $n \in \mathcal{N}^p$ has m_n service modules (in the wireless access domain, m_n corresponds to the number of input queues at an eNodeB, while in the optical domain it corresponds to the number of receiver/transmitter queues in the TSON edge node) with service rate μ_n. At this point is should be mentioned that in order to keep the analysis tractable m_n has been taken equal to 1. However, the analysis can be easily extended to the multiple service modules case. Due to the uncertainties introduced in MCC environments, we consider the general case where the inter-arrival times of the demands are not necessarily exponentially distributed. Assuming that:

- the external arrival process of the demands of VI_i is any renewal process with mean inter-arrival time $1/\lambda_i$ and coefficient of variation σ_{Ai}.
- the service times at the nth node of PI can follow any distribution with mean service time $1/\mu_n$ and coefficient of variation σ_{Bn} and,
- the demands are served on a First In First Out policy,

a closed form approximation for the end-to-end delay for the services that are provided by each VI can be extracted after applying the *diffusion approximation method* [12]. According to this method, the approximated steady-state probability of the

PI resources that are assigned to VI_i can be calculated as the product of the state probabilities of the individual nodes [13], that is,

$$\hat{\pi}_i(\kappa_{1i}, \kappa_{2i}, \dots, \kappa_{ni}) = \prod_n \hat{\pi}_{ni}(\kappa_{ni}) \qquad (4)$$

where κ_{ni} is the number of VI_i demands at the nth node of the PI. The marginal probabilities $\hat{\pi}_{ni}(\kappa_{ni})$ can be approximated by

$$\pi_{ni}(\kappa_{ni}) = \begin{cases} 1-\rho_{ni}, & \kappa_{ni} = 0 \\ \rho_{ni}(1-\rho_{ni})\hat{\rho}_{ni}{}^{\kappa_{ni}-1}, & \kappa_{ni} \geq 1 \end{cases} \qquad (5)$$

In (5) ρ_{ni} is the *utilization* at the nth node of the PI with respect to VI_i ($\rho_{ni} = \lambda_i/\mu_n$) and $\hat{\rho}_{ni}$ is a correction factor given by

$$\hat{\rho}_{ni} = exp\left(-\frac{2(1-\rho_{ni})}{\sigma_{Ani}{}^2\rho_{ni} + \sigma_{Bn}}\right) \qquad (6)$$

σ_{Ani} is the coefficient of variation of inter-arrival times at the nth node of the PI and is approximated by

$$\sigma_{Ani} = \frac{\left(1 + \sum_{j\in\mathcal{N}^P}(\sigma_{Bn}{}^2 - 1)\right)Pr_{jni}^2 e_{ji}}{e_{ni}} \qquad (7)$$

where Pr_{jni} is the routing probability of VI_i demands from node n to node j, while e_{ni} is the *relative arrival rate* (also known as *visit ratio*) of a VI_i demand at the nth node. Once the approximated steady probabilities of the network have been determined, the mean response time, T_{in}, of a VI_i demand at the nth node can be evaluated. Finally, in order to bound the end-to-end cloud delay of the services offered by VI_i below a specific threshold, $\mathcal{L}_i$, the following constraint should be satisfied:

$$\sum_n \vartheta_{in} T_{in} \leq \mathcal{L}_i, \quad i \in \mathcal{I} \qquad (8)$$

where ϑ_{in} is a binary variable taking value equal to 1 if node n in the PI is used by the VI_i.

The VI planning scheme described so far aims at minimizing the total power consumption of the converged wireless, optical DC network. However, for traffic demands that are generated in the wireless domain, computation offloading is beneficial for the mobile device, if the total energy consumed in the mobile terminal for transmitting and receiving data to the DC, is at least equal to the total energy that is consumed for data processing in the mobile device itself [16]. Let, P_m^p (watt) be the power that is consumed in a mobile device for data processing, P_m^i (watt) while being in idle mode, and P_m^t (watt) during the phase of data transmission/reception, with $P_m^t > P_m^p > P_m^i$. If a traffic demand with volume $h_{di}(\xi)$ is processed locally by mobile processor with speed S_M, the energy that is consumed is $h_{di}P_m^p/S_M$. However, if the same traffic demand is offloaded to the VI_i, the energy consumption in the mobile device is: $P_m^t T_{iW} + P_m^i\left(T_{iO} + T_{iDC}\right)$, where T_{iW}, T_{iO} and T_{iDC} are the delays that are introduced in the wireless, optical and DC segments of the VI_i, respectively. From the above it is deduced that a necessary condition for a mobile device to offload its demands to the cloud is:

$$P_m^t T_{iW} + P_m^i\left(T_{iO} + T_{iDC}\right) < h_{di}(\xi)P_m^p/S_M$$

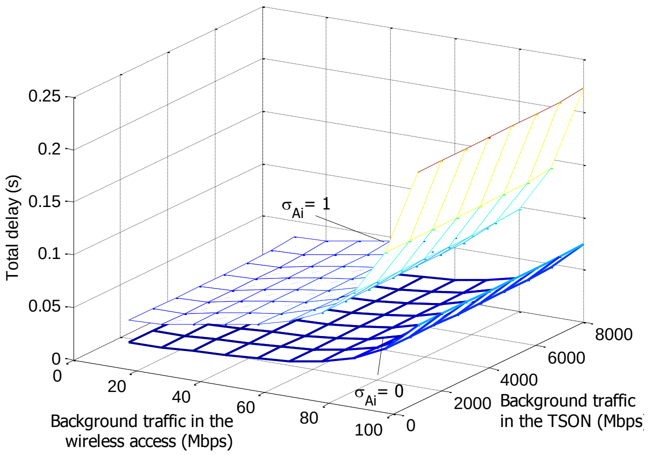

Fig. 4.End-to-end delay for a mobile cloud demand with traffic volume 1 MB under various background traffic profiles and traffic demand variations

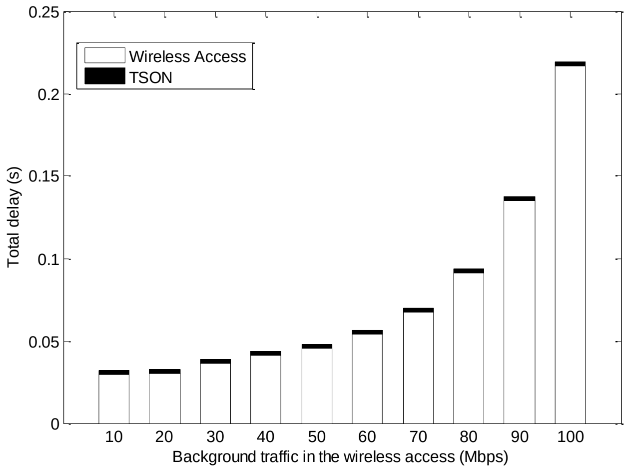

Fig. 5 Delays introduced in the various domains of the converged infrastructure as a function of the traffic load in the wireless access network (Load in the TSON=3Gbps)

The stochastically planned VIs are obtained by minimizing the expected cost:

$$\min \mathbb{E}[Q(x, y; \xi)] \qquad (9)$$

where

$$Q(x,y;\xi) = min\sum_{dpi} k_{dpi}x_{dpi}(\xi) + \sum_{s} P_s(\upsilon_s(\xi)) \qquad (10)$$

IV. NUMERICAL RESULTS

The performance of the proposed VI planning scheme across the multiple domains involved is studied based on the infrastructure illustrated in Fig. 1. For the PI, a macro-cellular network with regular hexagonal cell layout has been considered similar to that presented in [17], consisting of 12 sites, each with 3 sectors and 10MHz bandwidth, operating at 2.1 GHz. The inter-site distance has been set to 500m to capture to scenario of a dense urban network deployment. Furthermore, a 2×2 MIMO transmission has been considered, while the users are uniformly distributed over the serviced area. Each site can process up to 115 Mbps and its power consumption ranges from 885 to 1087W, under idle and full load, respectively [17]. For the computing resources, three "*Sun Oracle Database Machine Basic Systems*" [15] have been considered where each server can process up to 36Gbps of compressed flash data. The physical TSON topology assumed is illustrated in the lower part of Fig. 3 where the dimensions of the optical rings are below 5 km and the supported data rate is 8.68Gbps. The power consumption of the TSON equipment is measured to be 50W for the EDFAs and100mW for the PLZT chip is. The mobile devices are equipped with an Intel XScale processor with $P_m^t = 1.3$W, $P_m^p = 0.9W$ and $P_m^i = 0.3W$ (see [16] for a similar power consumption model).

Initially, the impact of the background network traffic on the total end-to-end delay is analyzed for a scenario where 1 MB of data needs to be exchanged between the mobile device and an IT server. In Fig. 4 it is observed that, due to the scarcity of resources in the wireless access network, the increase of the background load in the wireless domain leads to an exponential increase of the end-to-end delay. On the other hand, with the increase of the background traffic in the optical domain, the end-to-end delay remains almost unaltered (the total delay is increased by less than 2%). It is also observed that with the increase of the demands' uncertainty the end-to-end delays are also increased. This is explained by the fact that for larger variances, the demand request experience longer waiting times in the queues leading to an overall increase of the total delay. Similar results are presented in Fig. 5 where the total end-to-end delay when applying the proposed approach is depicted as a function of the background traffic load in the wireless access domain. It is observed again that with the increase of the traffic load in the wireless domain from 10 to 100Mbps, the end-to-end delay is increased by a factor of 6. At the same time, the optical network is responsible for less than 1.5% of the overall network delay.

V. CONCLUSIONS

This paper studied the problem of virtual infrastructure planning over converged wireless, optical network and computing resources. A novel problem formulation based multi-objective non-linear programming has been presented in order to a) minimize the energy consumption of the converged infrastructures and, b) maximize the lifetime of the mobile systems. Numerical results indicate hat there are a number of trade-offs relating to end-to-end service delay, resource requirements and energy consumption levels of the infrastructure across the various technology domains closely associated with the service characteristics.

ACKNOWLEDGMENT

This work was carried out with the support of the CONTENT (FP7-ICT- 318514) project funded by the EC through the 7th ICT Framework Program.

REFERENCES

[1] M. Rappa, The utility business model and the future of computing systems, *IBM Systems Journal*, 43 (1), 32-42, 2004.

[2] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012–2017", White Paper, 2013

[3] H. Dinh, C. Lee, D. Niyato, P. Wang, "A survey of mobile cloud computing: architecture, applications, and approaches," *Wirel. Commun. Mob. Comput*. Oct. 2011

[4] M. Satyanarayanan et al.,"The Case for VM-Based Cloudlets in Mobile Computing." *IEEE Perv. Comp*. 8 (4), pp.14-23, 2009

[5] K. Mun, "Mobile Cloud Computing Challenges", TechZine Magazine, http://www2.alcatel-lucent.com/techzine/mobile-cloud-computing-challenges/

[6] http://www.gogrid.com/

[7] http://www.flexiscale.com/

[8] A. Tzanakaki *et.al*., "Virtualization of heterogeneous wireless-optical network and IT infrastructures in support of cloud and mobile cloud services," *IEEE Communications Magazine.* vol.51, no.8, pp.155-161, Aug. 2013

[9] CONTENT Deliverable D4.1, Definition of the virtualization strategy and design of domain specific resources/services virtualization in CONTENT, 2013

[10] MAINS project website, http://www.ist-mains.eu/

[11] D. Giatsios, A. Apostolaras, T. Korakis, L. Tassiulas, "Methodology and tools for measurements on wireless testbeds: The nitos approach," in Measurement Methodology and Tools, LNCS, Springer Berlin Heidelberg, 2013, vol. 7586, pp. 61–80.

[12] R. Bruno, M. Conti, A. Pinizzotto. "A Queuing Modeling Approach for Load-Aware Route Selection in Heterogeneous Mesh Networks," in proc. of IEEE WoWMoM 2009.

[13] F. Baskett, K. M. Chandy, R. R. Muntz, F. G. Palacios, "Open, Closed, and Mixed Networks of Queues with Different Classes of Customers," *J. ACM* , Vol. 22, No. 2 pp., 248-260, 1975,

[14] Z. Davis, "Power consumption and cooling in the data center: A survey" [Online]. Available: http://www.greenbiz.com/.

[15] http://www.oracle.com/us/industries/healthcare/058454.pdf?ssSourceSiteId=ocomjp

[16] K. Kumar, Yung-Hsiang Lu, "Cloud Computing for Mobile Users," Computer, vol.PP, no.99, pp.1,1,

[17] G. Auer, V. Giannini, "Cellular Energy Efficiency Evaluation Framework", in Proc. of IEEE VTC 2011

[18] 3GPP TS 23.203, "Technical Specification Group Services and System Aspects"