Structured Cloud Federation for Carrier and ISP Infrastructure

Vamis Xhagjika*, Vladimir Vlassov†, Magnus Molin‡ and Simona Toma§
*UPC Technical University of Catalonia, Barcelona, Spain - xhagjika@ac.upc.edu
*KTH Royal Institute of Technology, Stockholm, Sweden - xhagjika@kth.se
†KTH Royal Institute of Technology Stockholm, Sweden - vladv@kth.se
‡Ericsson Sweden, Stockholm, Sweden - magnus.molin@ericsson.com
§Ericsson Sweden, Stockholm, Sweden - simona.toma@ericsson.com

***Abstract*—Cloud Computing in recent years has seen enhanced growth and extensive support by the research community and industry. The advent of cloud computing realized the concept of commodity computing, in which infrastructure (resources) can be allocated on demand giving the illusion of infinite resource availability. The state-of-art Carrier and ISP infrastructure technology is composed of tightly coupled software services with the underlying customized hardware architecture. The fast growth of cloud computing as a vastly consolidated and stabilized technology is appealing to Carrier Providers in order to reduce Carrier deployment costs and enable a future of Carrier Clouds with easily accessible virtual carriers. For such migration to happen software services need to be generalized, to decouple hardware and software, and prepared to move into the Cloud.**

The network backbone is centrally managed and only provides network connectivity. We believe this presents an opportunity. The edges of such networks and the core are interconnected with high performance links. If services could be deployed in these edges they would benefit from enhanced locality to the user. In this position paper we propose a distributed cloud architecture (precisely a structured multi-cloud federated infrastructure), with minimal impact on existing infrastructure, as a first step to incorporate the Cloud into the network infrastructure of such providers, enabling and enhancing novel and existing applications.

***Index Terms*—Cloud Computing, Cloud Architecture, Cloud Federation, Network Infrastructure, IAAS, ISP, Carrier.**

I. Introduction

Cloud computing is a general term referring to the successor of GRID computing and is generally known as commodity computing. The concept of the Cloud [1] enables infrastructure to be allocated on demand and to be managed by software services. Virtualization techniques are used to give the impression of possibly infinite resource allocation, with virtual resources sharing the same real resource by use of time sharing or computation alternation in multi-core architectures.

The Cloud is generally seen as a stack model composed of the IaaS layer (Infrastructure as a Service) providing to the user virtual machines allocatable on demand, PaaS layer(Platform as a Service) providing abstractions of components and a language to manage the infrastructure and SaaS layer (Software as a Service) providing software capable of providing services to multiple organizations at the same time. Recently a new term is being used to provide the entire real resource as a commodity resource, Machine as a Service. This permits real specialized hardware to be allocated on-demand when needed. The users of these services optimize costs by only paying for the resources they need. Some of the most well known open source cloud management software at the time of writing are Openstack [2], OpenNebula [3], CloudStack [4].

A multi-cloud federation system enables resource provisioning and life-cycle monitoring among different cloud providers. Users may need to access and create resources on multiple clouds for locality or economical advantages without penalizing their Quality of Services (QoS). A federated system handles cross-cloud interactions and integration in order to achieve higher levels of usability and locality.

Infrastructure Providers (IP) are the organizations providing the hardware and software that make possible the operation of a Carrier System and Internet Provider system. Compared to the IP we will classify into Service Providers(SP), ISPs and Carrier Providers as the organizations that use the infrastructure to provide network services. The architecture of the network backbone, maintained by these providers, is based on specialized hardware and follows a strictly centralized architecture, where the edges of the network provide access and the centralized Core Network provides services such as accounting, routing, etc. In this work we propose enriching these core and edge networks with our novel cloud architecture, enabling them to host cloud services. The proposed structured multi-cloud federation infrastructure is a first step in moving these providers into the cloud, and introduces a minimum technology impact on the existing infrastructure.

The rest of the paper is organized as follows. Section II provides the necessary background information for both Cloud Infrastructure and Network Topologies. In Section III, we describe a unified network model for Carrier and ISP networks. Section IV presents the system design for both cloud deployment and federation model and properties. In Section V, we describe different federation models and the model selected for our architecture. Section VII shows possible applications based on this infrastructure. We conclude in Section VIII.

II. Background

A. General Cloud Infrastructure

Cloud management software are sometimes referred to as Virtual Infrastructure Manager (VIM) but this term is

978-1-4799-2730-2/14/$31.00 ©2014 IEEE

more adequate to describe local virtual management control components used for the local machine hypervisor. Yet another denomination found in the sites of cloud management open source software [2] refers to such components as a Cloud Operating System, which is easily confused with recent approaches to having the actual Operating System(OS) in the cloud. For the sake of this work we will refer to such management systems as the Cloud Infrastructure Manager (CIM).

The general architecture of a Cloud Infrastructure Manager is shown in Fig. 1. The main entities present in most of the CIM are the Cloud Controller, Virtual Infrastructure Manager (per node basis, or compute engine), Data Store Provider(DS) and Network Controller(NC). These components appear in different CIMs as a single component or as a family of components providing specialized behaviors, as an example Openstack has an Image Repository (Glance), a key-value store (Swift) and a block storage (Cinder) implementing different types of data stores. Cloud Controller (CC) is the main orchestration entity, governing cloud orchestration for an entire cloud. Apart from cloud orchestration the CC handles also user management, security policies, resource scheduling, resource deployment, monitoring and billing. The CC is the core of the cloud managing infrastructure from a centralized point of view. Virtual Infrastructure Manager (VIM) is the component that manages local resources to a physical node in a cloud, by handling tasks such as VM scheduling, creation and monitoring. The virtual networking infrastructure is managed by the Networking Controller (NC) and defines software defined networks between various VMs in a datacenter. The whole setup is managed through the CC interface and the user need only take care of specifying a logical architecture which is then actuated by the Cloud Controller. The Data Store (DS) generally provides various abstractions of distributed storage for the cloud. All these components together provide a full cloud experience ready for service deployment.

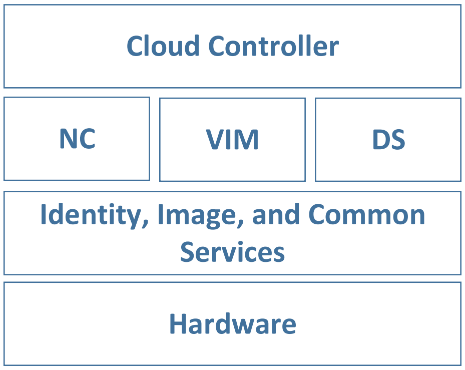

Fig. 1. Cloud Infrastructure Manager structure

B. Cloud Federation

Defining the term Cloud Federation is a difficult task, as in different cloud contexts it is used to represented different concepts. In general we have a federation when two or more administrative domains collaborate in order to achieve a common goal. In case of Cloud Federation, there are multiple types of federation possible depending on the type and the layer in the cloud stack in which federation is provided. The federation model we assume in this work is non transparent federation in which different sites have different Cloud Controllers and all of them know of each other, and collaborate through a Federation Middleware.

Extrapolating on the above definition of federation model, a cloud is a self-sustained cloud entity with a whole cloud software stack deployment. This means that every entity has its own cloud controller, VIM, network controller and data store. The Federation of such a conglomerate of clouds is conducted through a central federation software that has the ability to access the clouds APIs transparently and orchestrates the different clouds in order to provide resources.

C. Carrier and ISP to the Cloud

Carrier and ISP providers generally follow a close market with software and hardware tightly coupled in order to get the best performance from both sides of the environment. Providing a generalized enough software stack for all the services of a provider in order to be allocated on demand may prove to be not feasible without rewriting most of the subsystems. There are also other limitations on Cloud adoption for the core system of such providers and these limitations are generally related to QoS concerns that can be only met by specialized hardware and software.

In this work we propose a different approach to integrate cloud services inside the existing network environment in use by these providers, in order to achieve better and novel services and also permit SPs and third parties to easily deploy services in the IP networks. The Cloud Federation that will be described in the next sections will use the nature and topology of the existing IP based Carrier and ISP networks.

D. Mobile Cloud Computing

Mobile cloud computing uses the cloud in order to optimize mobile device experience, power consumption, resource availability. In this approach a mobile device could use computation offloading, by sending the computation to be executed into the cloud, in order to gain more available and powerful resources but also to optimize battery consumption [5].

Our approach enhances this discipline with a new cloud paradigm that exploits locality in order to have better support for Mobile Cloud Computing as the cloud is brought closer to the mobile devices providing higher access bandwidth and decreasing traffic on the backbone as all of the interactions can happen between the mobile device and the local cloud close to the points of presence.

III. Carrier and ISP System Topology

Carriers for Mobile and WiFi networks have a pretty close theoretical system topology to the ISP providers infrastructure. Some of the differences are in the protocols and the mediums of the last mile and the routing, else these two topologies are very affine to each other. In Fig. 2 we find a high level simplified star topology that unifies the different topologies and assumes IP as network transport protocol. The high level topology view simplifies dealing with different protocols at different stage of the carrier network and gives a much simpler framework to work with by decoupling the architecture from the actual communication protocols in the real topology. However there may be some protocol restrictions to the real topologies as some protocols may not include IP in all of the points of presence. This point will be a focus of further discussion when dealing with real deployments of such technology.

The subsystems of a normal Carrier and ISP provider are introduced in Fig. 2, where we have a main access to the internet protected by a firewall then we have the Core Network (CN), which compromises the main routing activities for data and voice traffic. The CN takes also care of the main activities like monitoring, provisioning of resources, accounting and also intra provider connectivity and handover. We assume we have one CN for each country which constitutes the main backbone of the infrastructure. In this generalized view of the architecture we have geo-distributed points of presence which provide network access to the clients. These points of presence are the edges of the access network and the one responsible to give connectivity to the clients. In case of the Carrier Network we have the Radio Base Stations providing local points of presence, while in the case of the ISP we have the Point of Presence (POP) which offer routing facilities and access and are connected with fiber to the CN.

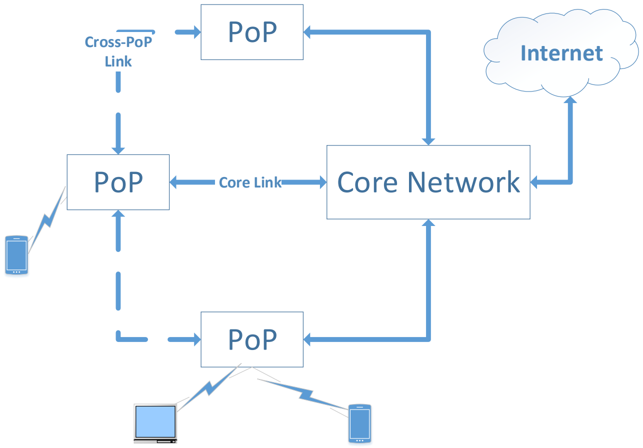

Fig. 2. High Level Provider Network Topology

The clients then connect to one of the POP with a different or same media but the bandwidth of the POP is subdivided between the clients that connect to the POP. In the case of Carrier networks, the transmission medium from POP to client is radio while in the case of ISP, it can either be radio (WiFi access points), cable or fiber. The total bandwidth toward the internet of the clients is less or equal to the bandwidth from the POP to the backbone. In general, different deployments are setup so that the total client bandwidth does not saturate the backbone link in order to permit also control signal bandwidth for inbound control. In the Carrier cases neighboring base stations have cross-connections that are used for traffic handover on user mobility. These links provide also a major way to optimize locality and present additional bandwidth that normally is used only in particular cases.

In general the topology (Fig. 2) is a star topology at the core, with core connections build up by high performance mediums (like fiber) and have lower bandwidth connections at the edges toward the client. In this work a multi-cloud distributed environment will be discussed in order to use the strong points of the topology and enable novel application and services to be deployed on the providers networks.

IV. Structured Multi-Cloud Architecture

Section III introduces the abstract network topology for normal deployments of Carrier and ISP provider networks. The Star topology described previously as the main topology has the benefits of high speed links at the core and separate local bandwidth at the edges of the network (the clients). Let us discuss a small user-case to show potential benefits and give a realistic view on benefits brought by the new multi-cloud architecture.

In an ISP provider, if some services could be moved at the POP, the clients would have a full 100Mbs connection to such service, under the assumption of having cable as the last mile medium, but only a 2-10Mbs connection to the internet from the backbone. Assuming the required service from the user could be elastic enough to be moved at the POP the user could interact with the service at a far higher bandwidth than that of a cloud service positioned somewhere in a centralized datacenter reachable via the backbone. Thus we effectively augment the available bandwidth toward services from the user and also generate new bandwidth by using previously not usable bandwidth.

The simple case described in the previous paragraph introduces hints to a more optimized cloud environment in which the whole Carrier and ISP provider network could be transformed into a service enabled multi-cloud federation. Let us discuss in details the architecture design of such cloud and also various aspects of performance, control, usability, applications and stability of such architecture.

A. General Architecture

The proposed cloud architecture is constructed in order to take advantage of the network topology in order to achieve better services but also the possibility for both Carriers and ISP to enable in-house and third-party application and protocols to be deployed in a most secure and isolated manner. The deployment of such applications and protocols would not impact the existing architecture and their deployment would be as easy as requesting the cloud manager to deploy some service images.

In the current state of the art such deployment would require provisioning of new hardware to the POPs or even software modifications to the POPs operating system which may lead to down-times if not done correctly and may require downtime in order to do the necessary system reconfiguration. The structured multi-cloud federated deployment architecture is shown in Fig. 3. The main aspects of this architecture are Tier-1 Cloud (T1, a datacenter grade cloud) and POP Tier2 (T2, POP local cloud). The combination of this two clouds permits a hierarchical cloud architecture that exploits locality and also the network topology for optimized bandwidth usage and sometimes also augmented bandwidth usage. This novel cloud architecture, could use local cross-POP links if present, that at the moment are used only for particular cases, in order to have full link usage and much more bandwidth than the normal POP bandwidth. This links in Carrier networks at the moment are used only for handover procedures.

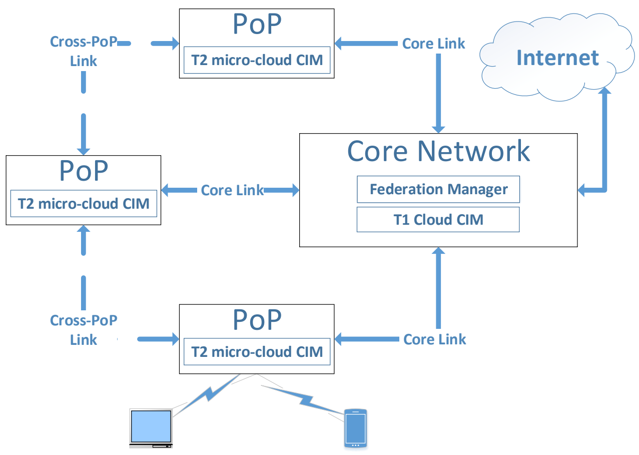

Fig. 3. Extended Infrastructure and Federation Model

The T1 cloud is a datacenter cloud deployment located, inside or highly connected to, the core network. For all intent and purposes, connection wise, it is an integral part of the core network and has high performance links to the internet. This core datacenter provides the necessary strong backbone in order to have a hybrid architecture. Data and application if deemed necessary can be moved around from the T2 cloud to the T1 cloud in order to have stronger Service Level Agreement (SLA) guarantees.

The T2 clouds are geographically distributed close to the POP or part of the POP OS itself, in which a VM acts as a routing facility, giving the possibility to allocate resources on demand very close to the users of the service. In some cases bringing the user as close as one-hop distance from the desired service. Each POP has its own T2 micro-cloud and the conglomeration of all the micro-clouds builds the T2 distributed cloud layer. T2 micro clouds are self-managed full cloud deployments in order to have full support for application deployment.

The micro-clouds fabric is federated or managed by a centralized federation controller (CIM) installed in the T1 cloud. The federation type that is used is non-transparent federation which will be explained in describing possible control mechanisms of such a cloud concept.

Clients connecting to the WAN topology have generally higher bandwidth to the POP than the total client bandwidth toward the internet backbone which is limited by the backbone link from the POP to the CN. By having the T2 micro-clouds to the edges of the network the clients can interact with services with higher bandwidth and the total available bandwidth is increased as compared to the case in which the services are deployed on an internet cloud.

Since resources in the T2 cloud are limited we have the T1 cloud acting as a cloud helper in which load could be redirected in order to handle high traffic services which would saturate the resources in the T2 micro-clouds. In this case performance of the service in terms of locality and connectivity are limited as compared to the local micro-clouds, but still better than an internet cloud, since we are still inside the same WAN.

This architecture can further increase apparent total bandwidth if cross-POP links exist. This links can be used to shortcut application data between services running on neighbor POPs so that the backbone is left free for user internet access. In this case we have new data paths that normally are not used for user access becoming available to service data and thus increasing the efficiency of the network usage.

B. Bandwidth Control

The proposed structured multi-cloud architecture optimizes locality and efficiency of network links and also enables novel applications to be deployed on the described network topologies. The efficiency of the network is increased by rendering available to service and user traffic unused network bandwidth, that is normally not usable as limited by the network backbone and also new data paths that normally are used only for control or exceptional operations.

The efficiency comes with some restrictions as now data paths that were used for system specific tasks handle also user and cloud service traffic. The two types of traffic would be contending the same network resource and without some kind of management this may compromise system integrity.

We propose a solution based on previous work done in [6] a network bandwidth manager for cloud services. The principle following this network manager is in having bandwidth managed so that user-centric workloads and system-centric workloads could be managed and allocated at the endpoints independent of the network topology. This model permits to have SLA guarantees independent of the topology of the underlying physical network. In case this bandwidth management is still not enough to have system stability, services could be migrated from the T2 micro-clouds to the T1 micro-cloud. The system in that case would revert to a Carrier Cloud infrastructure where the providers enabled connectivity and services are on the CN cloud.

C. Service Migration

In case of availability of cross-POP links, as previously mentioned we have new data paths on neighbor POPs that can be used to deliver data and build new out-of-band services. This links can be used as newly available data paths but also for a more advanced usage of cross-POP service migration. In case a micro-cloud being saturated the federation manager could migrate some of the services to a neighbor POP and the clients can exclusively use the cross-POP links.

In this scenario we have multiple possibilities for service migration, we can either migrate the service from the micro-cloud to a neighbor micro-cloud or to the centralized T1 datacenter cloud. This migration process could be implemented on a policy based approach so that it can be possible to be modified on a per deployment basis depending on the system administrator priorities. Further study of such scheduling policies is delegated to future work.

Service migration can serve also as a mean to deal with mobility of devices in terms of geographical sparsity. Frequent movement of clients between different POPs can be accounted by moving the data slowly between POPs, but such model works only if the speed with which the client is moving between POPs is lower than the cost in terms of speed of moving the data between POPs. When clients have POP switching speed of a highly sporadic nature or of high frequency the data can be moved higher in the hierarchy to the T1 cloud so the services migrate but the data is static in T1.

V. Federation Manager

The proposed cloud infrastructure leads to multiple control possibilities for the cloud federation, the Federation management infrastructure, the entity coordinating multiple clouds together to produce a usable service. The control system for the federation leads also to design choices concerning the cloud infrastructure. In the Star topology we have assumed for this work, clouds are distributed in two variants; a micro-cloud fabric composed of multiple T2 local clouds on the edges of the network and T1 cloud a datacenter level cloud part of the WAN. By considering the Star topology and the placement of the clouds, control can either be centralized thus having transparent federation or decentralized by having a non transparent federation. Both of the federation models are viable alternatives for the proposed federation architecture. The chosen federation model for this architecture will be centralized and non transparent.

A. Federation Layer

In a multi-cloud federation, federation could happen on any layer of the cloud system. Some system may implement federation on the IaaS layer by rendering invisible to the users of the PaaS or IaaS that the federation exists. The IaaS layers of all the clouds would handle resource provisioning between multiple IaaS providers and hide to the upper layers the fact that a federation exists at all. Or by exporting federation specific functions though the IaaS interface, but normally the upper layers PaaS, SaaS and client don't need to necessarily know that the IaaS is actually a multi-cloud federation [7].

Another way a multi-cloud federation could be implemented is by federating at the PaaS layer in which the IaaS-es of different clouds don't have any idea of each others existence. The federation is executed on the upper layer, the PaaS. In such federation scenario it is the platform or the client, in case a platform is not present, who is responsible for the federation mechanisms. It needs to implement metadata and multi-cloud resource provisioning to interact with each of the clouds in the federation. When resources are requested from a client application, the PaaS or the client contacts and interacts with each cloud in order to satisfy this requests. This approach has a low footprint on existing IaaS because of no need of actual modification, but elevates the complexity level on the PaaS side. For each cloud a driver would need to be implemented for accessing the clouds homogeneously and problems may arise as different IaaS providers may use different cloud technologies and may be unable to hide the heterogeneous nature of the multi-cloud.

B. Transparent Federation Model

With transparent federation we describe physically separated clouds on the same or different WANs (T2 micro-clouds) in which only one CIM (T1 Cloud CIM) exists for the whole cloud federation. In this approach no separate federation controller is needed as the CIM manages resources in the federation as if it was one big cloud.

The multiple micro-clouds (T2 micro-clouds), composed of at least a server grade machine, are distributed in different LANs. Of these clouds there exists one, which provides also the CIM (T1 Cloud CIM). The other clouds are connected to the CIM enabled cloud, using virtual LAN or tunneling technologies to build a unique LAN overlay and give the impression that all of these clouds exist on the same physical network.

This approach provides an easy implementation of a federation system as no additional changes to the existing CIMs would be needed. However problems arise from such a configuration as there is no actual distinctions between VMs on different clouds, or this distinction needs to be added to the CIM, maybe by separating different clouds in different IP ranges and use latency as metric.

Another drawback of this technique is that the added complexity of having a network overlay deteriorates performance of the network between different clouds but also between the VMs on the same machine as traffic would always need to pass through the VLAN or LAN tunnels. Performance deterioration is due to the overhead of the tunneling technologies as packets need to be encapsulated/decapsulated in order to reach the machines and VMs on different LANs.

Apart from performance concerns this model assumes that the connection between the clouds and the CIM is always persistent, if such connection breaks then the distributed partitioned micro-clouds would be left without any control system and thus rendered for all intent and purposes unusable.

C. Non-Transparent Federation Model

Non-transparent federation model is a model in which the distributed multi-clouds (T2 micro-clouds) are full cloud deployments with each having its own CIM (T2 micro-clouds) management system, in this case a third party software is needed to perform the federation. The Federation Controller (FC) would run in a centralized fashion on one of the datacenters (T1 Cloud) or a distributed software running on each one of the federated clouds.

The FC would be the entry point to all the resource provisioning system and also would need to care about scheduling, error recovery and multi-cloud monitoring and authentication mechanisms. Also the federation controller would need to have a universal interface to access heterogeneous clouds uniformly.

The network topology of the VMs from such federation would be by using public IPs thus flat networking or a reserved private network in case of private multi-cloud. In this scenario the networking is not penalized in performance as no amount of tunneling or virtual networking is involved. All the machines reach each other through the WANs or LANs by using the flat network IP addresses.

This model also accounts for topology partitioning as each cloud is a self-sustained entity. Each cloud has its own CIM manager so even in case that the main federation controller is offline operation of the cloud can still continue in an unsupervised fashion. When the main federation controller is returned to full functionality then the only data who needs to be updated would be the federation metadata. By having such behavior this model provides network partition stability for the cloud management and the cloud can still be operated locally, even in the absence of the federation controller.

This approach also minimizes traffic needed between the federation manager and the distributed clouds as monitoring and normal operation commands are delivered locally by the local CIM (T2 micro-cloud CIM), only resource provisioning and resource scheduling are done non-locally by iterating with the federation controller. Thus the majority of the external traffic would be functional traffic inherent to the application running on the machines and few federation management and monitoring traffic.

D. Storage and Identity

Identity management and distributed storage are two aspects crucial to the implementation of any such cloud federation. Depending on the infrastructure in place for such topologies there are different approaches to implementing these aspects.

The area of federated identity manager is still a hot research topic on cloud federation technologies, approaches to achieve such identity federation include modification to single cloud proprietary protocols to include multi-cloud identity management and third party authentication authorities. A work conducted on OpenStack [8] uses both approaches by modifying the local cloud identity management protocols to include third party authentication servers.

As for the distributed storage the model that best fits the topology in our opinion would be local storage for each T2 micro-cloud and Peer-to-Peer deployment of images in local repositories. This solution provides also a reliable system as in presence of partitions micro clouds can continue operations without central supervision.

E. Final remarks on federation dynamics

The federation model chosen for the multi-clouds infrastructure described in this work is the non-transparent federation with centralized controller. In our opinion this model provides the best stability of operations and also minimizes clouds management overhead on the links that in the Carrier and ISP case is a primary resource.

The Federation Controller is placed in the Core Network T1 datacenter while each one of the micro-dataclouds if treated as a full cloud deployment on at least a server grade machine. The FC provides the main activities for resource scheduling and resource provisioning, and delegates to the local CIMs the management of the lifecycle of the provisioned resources, monitoring and local resource scheduling.

VI. Real Implementation on Carrier Networks

In this section, a possible deployment for the cloud infrastructure is described in the context of the Carrier Network infrastructure. Fig. 4 shows a generalized view of the architecture and the placement of the T1 and T2 cloud enabled hardware. This deployment is one possible way to deploy the clouds, each provider could customize it to fit production and deployment needs. As shown in Fig. 4 the T1 datacenter cloud is an integral part of the Core Network. We can suppose for the sake of 3GPP standard that this data center, connection wise, is placed between the Core Network and the firewall connecting it to the internet. This way Core Network functionality is not compromised and routing can be done easily while maintaining good connectivity.

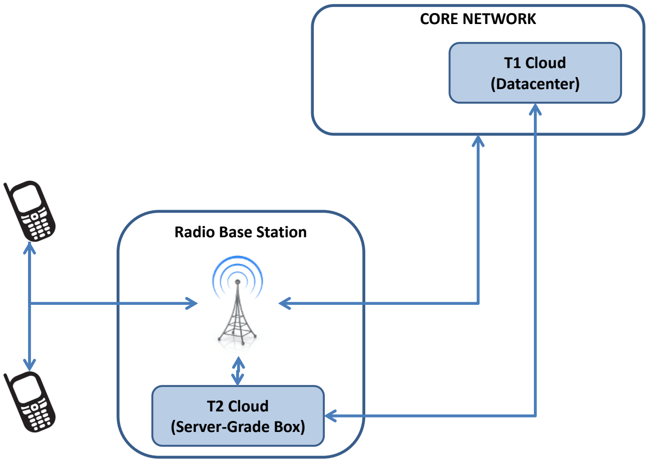

Fig. 4. Carrier Network Implementation

As for the T2 micro-cloud fabric, it can be either an external cloud enabled hardware plugged into the base station or dedicated cloud enabled hardware inside the base station hardware. We have a separate control network connecting the

T2 clouds with the T1 cloud. The base station is considered to have routing capabilities, forwarding data either to the T2 cloud or to the Core Network.

In the current 3GPP standard it is not possible for the base station to have such routing facilities, but such functionality could be provided through different techniques outside of such standard. The VMs started on both of the T1 and T2 clouds could be part of a private network IP allocation range or be provided with public IPs and provide network isolation through VLAN an Network Overlays.

VII. Enabled and Enhanced Applications

In this section we discuss some potential novel applications, as use-cases in this context, that would benefit from this cloud infrastructure and cloud provide hints on how to develop this technology further.

A. Internet of things

Recent developments have seen a lot of attention shifting to the so called ”Internet of Things”, or the idea of having all of the electronics present in our environment to be connected to the internet [9]. By having a cloud infrastructure close to the clients and thus close to the devices a client uses, these devices could use the micro-clouds in order to achieve better performance or even as a helper for their tasks.

B. Mobile cloud Computing

As previously stated, Mobile Cloud Computing is to be understood as taking computation and data away from mobile devices into the cloud [5], enabling reduced power consumption and availability of additional resources. Drawbacks of this approach are introduced by the sporadic nature of network latency and that of network partitioning. Both these drawbacks can be addressed by having the T2 micro-cloud close to the user, so that the user can offload computation to the local T2 cloud. In case of user mobility as mentioned, the data of the service could be moved either to the next local T2 cloud or higher up in the hierarchy to the T1 cloud.

C. Third party applications

Third parties will be able to deploy their own services on top of the Carrier or ISP networks, and the Carrier and ISP would be able to charge for such services as they provide the infrastructure. These applications could be deployed easily through a appstore approach and different business model may be applicable. At the moment the network providers are unable to charge service providers for the network utilization. This approach would be acceptable by both parties as the service provider is assured to have better connectivity to the clients and stability of execution environment, while the network provider is able to charge the service for network and execution environment costs. This model may lead to new streams of income for all the involved providers.

As a real life example, Akamai a well known CDN provider at the time of writing of this paper, provides its own boxes to network providers in order to deploy it’s cache-ing services. If the proposed architecture on this paper would be in place then the network providers may provide such infrastructure and charge for it, while Akamai would have better scalability for their distributed caches and better locality to the users.

VIII. Conclusion

To summarize the contributions of this positioning paper, this work devises a cloud enabled architecture for Carrier and ISP Networks in which the topology is augmented with cloud infrastructure in order to provide cloud services.

The proposed cloud architecture is based on a structured multi-cloud federation, in which micro-clouds are distributed in the PoPs of the network topology (namely T2 micro-clouds) and a central datacenter cloud, namely the T1 cloud. As a control mechanism for such distributed architecture we chose non transparent federation, managed by a centralized cloud federation manager running on the T1 cloud. The T2 clouds provide locality augmentation for services and the T1 cloud provides augmented performance for more performance oriented services.

In order to give a complete view of the architecture various aspects were treated as mobility, cloud properties and service migration policies, etc. To conclude the discussion of the cloud architecture, possible applications benefiting such technology are presented. Future work will focus on further study of such cloud federation middleware and on possible services running on such a distributed architecture.

Acknowledgement

This work was supported in part by the ErasmusMundus Joint Doctorate in Distributed Computing (EMJD-DC) funded by the Education, Audiovisual and Culture Executive Agency (EACEA) of the European Commission under the FPA 2012-0030, in part by the End-to-End Clouds project funded by the Swedish Foundation for Strategic Research (SSF) under the contract RIT10-0043, and in part by Ericsson AB.

References

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “A view of cloud computing,” *Commun. ACM*, vol. 53, no. 4, pp. 50–58, Apr. 2010. [Online]. Available: http://doi.acm.org/10.1145/1721654.1721672

[2] Openstack. (2014, Feb.) Openstack cloud software. [Online]. Available: http://www.openstack.org/software/

[3] OpenNebula. (2014, Feb.) Opennebula flexible enterprise cloud made simple. [Online]. Available: http://www.opennebula.org/

[4] A. CloudStack. (2014, Feb.) Apache cloudstack open source cloud computing. [Online]. Available: http://cloudstack.apache.org/

[5] D. Huang *et al.*, “Mobile cloud computing,” *IEEE COMSOC Multimedia Communications Technical Committee (MMTC) E-Letter*, vol. 6, no. 10, pp. 27–31, 2011.

[6] Y. Liu, V. Xhagjika, V. Vlassov, and A. Al-Shishtawyz, “Bwman: Bandwidth manager for elastic services in the cloud.”

[7] B. Rochwerger, D. Breitgand, E. Levy, A. Galis, K. Nagin, I. M. Llorente, R. Montero, Y. Wolfsthal, E. Elmroth, J. Caceres *et al.*, “The reservoir model and architecture for open federated cloud computing,” *IBM Journal of Research and Development*, vol. 53, no. 4, pp. 4–1, 2009.

[8] D. W. Chadwick, K. Siu, C. Lee, Y. Fouillat, and D. Germonville, “Adding federated identity management to openstack,” *Journal of Grid Computing*, pp. 1–25, 2013.

[9] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey,” *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.