# Building Virtual Private Clouds with Network-aware Cloud

J. Soares[1,2], J. Carapinha[1], M. Melo[1,2], R. Monteiro[1,2]
[1]Department of Exploratory Innovation
Portugal Telecom Inovação
Aveiro, Portugal
{ joao-m-soares, jorgec, marcio-m-melo, romeu-r-monteiro}@ptinovacao.pt

Susana Sargento[2]
[2]Instituto de Telecomunicações
University of Aveiro
Aveiro, Portugal
susana@ua.pt

***Abstract*— Cloud computing presupposes on-demand network access to pool of computing resources. However, network access through the WAN is usually not compliant with any kind of service guarantees, including reliability, security and performance. In this work, two types of network services able to fulfill cloud requirements are presented. In addition, an extension of the Virtual Private Cloud concept is proposed by integrating these network services. Managing cloud and network resources in an integrated way is a need and an obvious challenge, thus resource management in such environment is a major focus. We identify the main inherent challenges in resource management and how they can be overcome. Further, an experimental platform is presented, along with a preliminary analysis of results.**

***Keywords-cloud computing; cloud networking; virtual private cloud; network-as-a-service; connectivity-as-a-service.***

## I. INTRODUCTION

Today, the proliferation of broadband access gives users the possibility to use services available directly through the Internet, which represents a change of the paradigm for using applications and communicating, thus popularizing the so-called Cloud Computing (CC).

CC brings with it requirements at two different levels: at the cloud level, i.e. data centers; and at the network level, where required levels of performance, reliability and security must be guaranteed.

So far, the cloud and the network have been seen as two separate entities in this picture, with the network playing a relatively minor role, mostly as provider of connectivity between the cloud resources and the user premises. We argue that, to provide assured levels of performance to cloud services, cloud and network resources need to be provisioned, managed, controlled and monitored in an integrated way.

There are several reasons for the integration of network and cloud. First, the establishment of Service Level Agreements (SLA) is essential to encourage customers, particularly enterprises, to adopt cloud services. Today, the lack of reliability and performance guarantees is one of the main obstacles against the widespread use of cloud services. It is clear that these SLAs can only be implemented through network integration. Just like the Wide Area Network (WAN) component of enterprise networks is usually based on reliable managed network services such as Virtual Private Networks (VPNs), rather than the public Internet, there is no reason to believe that future enterprise cloud services will require a lesser degree of reliability and performance guarantees from the network.

Secondly, it is essential that cloud properties such as elasticity and self-provisioning be also extended to network resources. Quite often, expanding or reducing cloud resource capacity, or provisioning new cloud resources, requires a corresponding reconfiguration of network resources, e.g. bandwidth admitted into the network. Today, by contrast, reconfiguration of network services is supposed to be relatively infrequent and usually involves a significant amount of manual effort.

Thirdly, the dynamism of the cloud will often require live migration of resources (e.g. from a local enterprise data center to the cloud, or between two different sites of the cloud service provider) without interrupting the operating system and any noticeable impact on the running application. This requires IP addressing to remain unchanged after migration and all relevant QoS, security and traffic policies applied on network equipment (e.g. routers, switches, firewalls) to be adapted appropriately in real time.

For the reasons stated above, it is clear that next generation of cloud services must handle network and cloud resources in an integrated way. This paper presents the concept of Cloud Networking (CN) to achieve this integration in the context of virtual private environments, and its resource management aspects to develop an integrated view and allocation of both network and cloud resources.

This paper is organized as follows. Section II summarizes relevant work in the area and how this work intends to progress, and section III presents how the concept of CN can be applied in the context of virtual private environments. Further, section IV provides an overview of the resource management challenges that arise in a CN environment. The experimental platform that embraces this approach is described in section V. Section VI describes how a Virtual Private Cloud (VPC) request is instantiated and also presents the prototyping results achieved so far to demonstrate the concept of CN. Finally, section VII provides general conclusions and indicates directions for future work.

## II. RELATED WORK

Based on recent trends and evolvements, it is clear that the network will play a key role in the provisioning of cloud

ISBN: 978-1-61208-172-4

computing services, by giving the necessary guarantees to access the cloud. The importance of this role will be increasingly evident. In this area, some research works have been presented in the literature, such as [2] where the critical impact of network performance on the applications is shown and an extension of [3] is presented based on a platform for provisioning of virtual infrastructures, to extend the traditional cloud paradigm to network provisioning. [4] presents a software-based network resource management system for VPCs, able to handle heterogeneous network equipment and proposes a virtual network point for multipoint network provisioning.

The European Commission, through the Seventh Framework Programme (FP7), has also been supporting research in this area, FP7 Projects SAIL (Scalable & Adaptive Internet soLutions) [5] and GEYSERS (Generalised Architecture for Dynamic Infrastructure Services) [6] are two relevant examples.

From the industry side, both standardization bodies and enterprise efforts have highlighted the need for cloud and network resources to be handled together. Verizon has been working on the extension of VPNs for Private Clouds and an Internet Engineering Task Force (IETF) Internet-Draft has been released on this matter [7]. Meanwhile, IBM already offers enterprises a cloud data backup supported by Verizon's VPN services.

Although there are works on management addressing VPCs and others addressing virtual networks, there are few addressing the management of both in an integrated way. This paper addresses this subject and proposes an extension of the VPC concept to also provide WAN guarantees. A platform able to provide this VPC model is also presented in the paper.

## III. Cloud Networking for Virtual Private Clouds

Virtualization has been the key enabler of agility in data centers, which in turn led to the emergence of CC. The fundamental breakthrough offered by virtualization is the separation of operating systems (OSs) and applications from the underlying physical infrastructure.

Applying the same concept to networks has been often advocated – by decoupling networks from infrastructure through virtualization, it should be possible to establish and reconfigure (virtual) networks with great flexibility, nearly on-demand. Network virtualization has been explored by different research initiatives in multiple contexts and application scenarios. The idea of on-demand provisioning of network services has been demonstrated in practice [9].

Providing the network infrastructure with the ability to match the dynamism of the cloud would be required to overcome the problems and limitations identified in the previous section. From this point of view, network virtualization would be the perfect companion for virtualization in the data center, in order to build seamless end-to-end elastic and agile offer of cloud services.

A virtual network (VN) is supposed to fully replicate the behavior of a physical network, from all points of view. While this replication may be useful in some cases (e.g. when the customer is itself a service provider), in most cases the effort of managing a VN is a burden that customers would prefer to avoid.

Thus, just like the CC service model defines three basic services - Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) - we propose a similar approach for networks. From this perspective, we define two types of network services (Figure 1): *Network as a Service* (NaaS) and *Connectivity as a Service* (CaaS). NaaS allows the user to request a network by specifying precisely the network topology, link bandwidths, routers' computing capacities, routing protocols, as well as possibly other features (e.g. physical location, security properties). As for CaaS, the user is provided with the ability to define, just like in VPNs, a set of customer edge equipments (CEs) (e.g. enterprise sites and possibility the cloud CE, however this latter does not necessarily needs to be defined) and certain characteristics such as bandwidth at ingress/egress points and routing protocols between the CE and the provider edge equipment (PE). Everything that runs inside the service provider network domain is not visible to the customer.

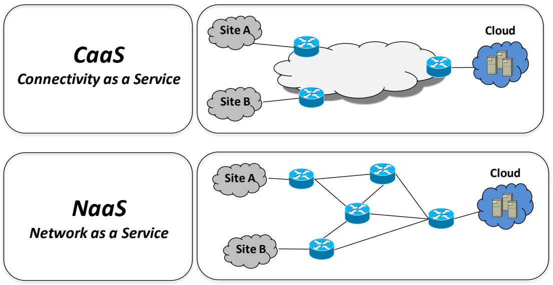


Figure 1. Network layering model.

Both network service options can be materialized in multiple ways. A fully virtualized network is the "natural" way to materialize NaaS, whereas managed VPNs (e.g. BGP/MPLS VPNs) are a typical example of CaaS.

In this paper, we propose a solution that is able to embrace both types of network service.

At this point it is useful to define the concept of VPC. In [8] a VPC is defined as "*a combination of cloud computing resources with a VPN infrastructure to give users the abstraction of a private set of cloud resources that are transparently and securely connected to their own infrastructure*". In the context of this paper, we propose to generalize this concept, in order to embrace any kind of private network service, either materialized as a VPN, or as more advanced service types, including those based on fully virtualized networks. Moreover, a network service should allow the handling of network resources (e.g. bandwidth, add/remove a costumer site) with a certain level of freedom in order for it to keep up with the cloud.

In order to address the coupling of both cloud and network services, the concept of CN was put forward. Similarly to CC, CN has no standard definition, but we can say that it goes beyond classical networks, encompassing on-demand provisioning i.e. scalability, guaranteed performance, self-healing and extensible management. So far, no real attempt to merge networking and cloud resources in a common framework has actually taken place.

We pursue the concept of CN by envisioning a unified management framework for computing and communication, where the network operator can provide simultaneously the network and cloud resources (IaaS), in an integrated approach, optimizing overall resource allocations by considering network and computing resources as a unified whole. In this work, network services are materialized in VNs, however we do not exclude the possibility of other network approaches (e.g. VPN, OpenFlow).

In the following section we identify what we consider to be the most important challenges to CN that resources management raises.

## IV. Resource Management in Cloud Networking

Bringing together network and CC resources so that users can access services in the cloud with guaranteed performance and reliability raises several challenges. The discovery, allocation, adaptation and re-optimization of resources, addressing simultaneously both network and cloud resources, are the main inherent challenges of resource management in CN. The management of these resources lays upon concepts of virtual resource mapping in the physical infrastructure with self-organized reconfiguration of resources, devices and associated network, according to the services and user requirements, policies (with respect to e.g. location) and changes in the infrastructure.

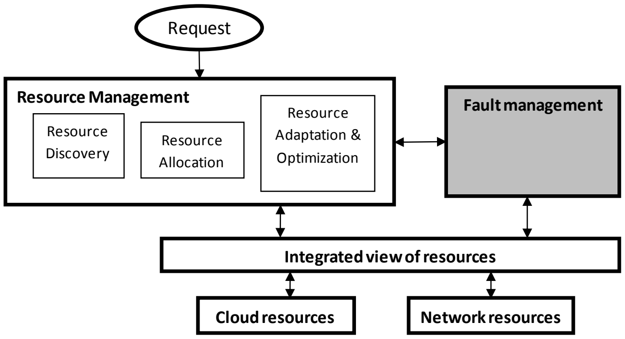


Figure 2. Cloud Networking Management diagram.

In Figure 2 we present a management block diagram composed by three main blocks: the *Resource Management* block (RM); the *Fault Management* block (FM); and an underlying block entitled *Integrated view of resources* (IVR). This work is focused on the RM and IVR. The former is composed by three sub-blocks: the *Resource Discovery* block (RD); *Resource Allocation* block (RA); and the *Resource Adaptation & Optimization* block (RAO). These sub-blocks will be detailed ahead. As for the latter, the IVR, it has the purpose of providing the upper blocks with the domain agnostic ability to view and interact with resources, whether they are cloud or network resources. Regarding the FM, it is illustrated in the picture to facilitate the interpretation of the management system, mainly regarding its interaction with the RAO sub-block.

### A. *Resource Discovery (and Monitoring)*

A fundamental requirement in virtualized environments is the integrated view of the existing physical and virtual topologies, the resources' characteristics, as well as the status of all network elements and links. This knowledge can be provided either by a centralized or by a distributed approach [10].

Today the cloud, i.e. data centers, and the operator's network are two distinct domains which CN aims at integrating. However, there are boundaries that cannot be crossed as these domains will not be willing to share full information about their domain. In this approach, we assume to have access to network information such as topology and physical resources, as well as the ability to retrieve information on the virtual resources that the physical resources may host. On the data center side, we do not expect to have such detailed information, we rather expect to see a data center as a single node in the network with unlimited capacity and a set of associated information elements (similar to today's cloud services, e.g., instance types, available OSs, pricing, plus location).

### B. *Resource Allocation*

Virtual resources should be provisioned and placed in an optimal location according to the available resources at the time of the request, based on a number of possible criteria from both cloud and network, e.g.: type of VMs and possible restriction on location of these VMs; latency, bandwidth topology, geographical places where users will access the service, and other possible restrictions.

In order to map resources, a combined mechanism, able to perform balanced decisions taking into account the abovementioned requirements of both network and cloud resources, is needed. This mechanism must be able to determine a possible solution, i.e., physical hosts able to allocate the cloud resources which, at the same time, can have an associated network service able to fulfill the requirements in the access to the cloud.

### C. *Resource Adaptation and Optimization*

With the dynamism of the cloud, reconfigurations and re-optimizations become common operations, whether to cater for possible side effects of new virtual resources being instantiated and existing virtual resources being resized, released or migrated, business policies, or triggered by unexpected events (e.g. node or link failure). These unexpected events are triggered by the FM which is responsible for monitoring the resources, detecting faults and collecting performance metrics.

Depending on the specific environment, actions can be taken at different levels: in the cloud, in the network, or in both. Thus, mechanisms are required for extending or moving cloud resources to other data centers, creating new network paths and reconfiguring existing ones (need for more bandwidth, less latency, failure, load balancing network resources). Those algorithms must decide on (1) when to reconfigure and (2) how to reconfigure. These decisions must be done based on information provided by the FM, or by an explicit request from the user.

Towards this aim, the Network Virtualization System Suite (NVSS) presented in [9] was extended with the control and management of the Suite enabling now cloud (IaaS) and network services (NaaS and CaaS) to be provisioned together to meet the user's requirements, apart from its original feature, the deployment of VNs. The implementation work presented in this paper specifically targets the challenge of resource discovery (and monitoring) of cloud resources and resource allocation of both cloud and network. The next section gives an overview on the evolved NVSS.

## V. Network-aware Cloud System Suite

The Network-aware Cloud System Suite (NCSS) is a platform that provides integrated deployment and management of cloud and network resources in a single tool. This platform is an evolution of NVSS [7], an experimental platform that provides VN design, embedding, creation, discovery, monitoring, and management. NCSS extends NVSS in two fundamental ways: it handles cloud resources (rather than just network resources) and CaaS (rather than just NaaS).

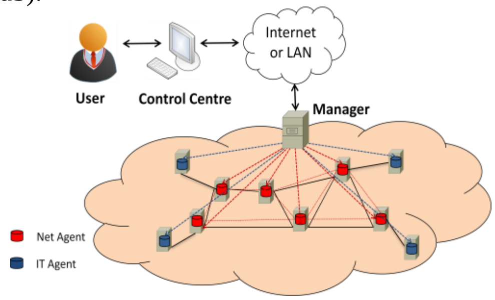


Figure 3. NCSS Architecture

NCSS is composed of 3 software modules: the Agent module, the Manager module and the Control Centre module. Their hierarchical decomposition can be analyzed on Figure 3. The Agent module is designed to run on network nodes ('Net Agent'), as well as on those acting as computing nodes ('IT Agent'), in order to act upon them and periodically gather data from them. The two types of Agents, besides interacting with each other, receive and send requests to the Manager, which is a centralized entity in charge of aggregating all Agents' knowledge and sending them commands. Additionally, the Manager also communicates with the Control Centre, which is the user's front-end, and provides him with graphical and simple to use VN creation, management, and monitoring functionalities.

### A. *Functionalities*

The NCSS platform provides a set of main functionalities: distributed network and cloud resource discovery, network and cloud mapping and creation, network and computing monitoring, and network and cloud resource management. These functionalities are described below.

*1) Distributed Network and Cloud Discovery.*

Network and cloud resource discovery is not only an administrator's utility that provides a fast and easy way of viewing how the cloud resources and the network resources are been used and where they are been consumed, but it is also fundamental when embedding new cloud and network resources, since the embedding process requires an accurate and up-to-date view of the substrate and currently running cloud and network resources.

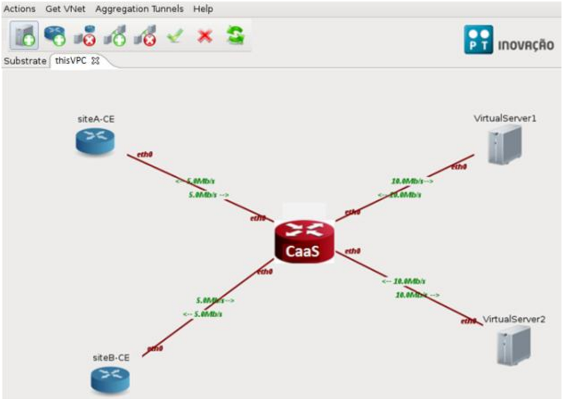


*Figure 4. Control Centre– VPC Requirement Example*

*2) Network and Cloud Mapping and Creation*

The Control Centre module provides the user with means to create and embed new cloud and network resources in runtime. By selecting and placing either cloud or network resources, i.e. servers or routers, on the platform GUI and by connecting them with links, as depicted in Figure 4. The user can specify both cloud and network resource capabilities, CPU, RAM amount, location, number of interfaces and also perform network addressing configurations.

The final step in creating a new set of cloud and network elements is to commit it to the Manager, which will then map it in the physical infrastructure.

The embedding problem of cloud and network elements is a complex one, which requires a trade-off between computation time and embedding optimization. In order to lower the computational requirements, a heuristic mapping algorithm was developed, which aims to embed both types of elements taking into consideration both the load of physical links and load of network and computing nodes.

*3) Cloud and Network Monitoring*

Dynamic resource monitoring is fundamental to provide an accurate view of the state of both types of resources and to quickly react to failures or configuration problems. The implemented monitoring functions periodically update the information on resources; therefore it is possible to quickly identify diverse situations, such as failures and high resource usage.

*4) Cloud and Network Management.*

The management feature provides functionalities like the change of the resource state (i.e., reboot, shutdown, suspend or power up), the change of the assigned RAM memory in runtime and the deletion of either a single resource or a complete set of resources, which greatly simplifies the administrator work.

The following section describes the process of establishing a VPC using the NCSS.

## VI. Virtual Private Cloud Establishment & Evaluation

The establishment of a VPC using the NCSS can be supported by a NaaS or CaaS. The process of establishing a VPC, from the moment a user requests a VPC until the moment it is ready to be enforced in the physical infrastructure, will be detailed in this section. Two VPC requests will be considered, one for NaaS and another for CaaS. In addition, results on the request's processing time are presented. Note that these results do not intend to perform any comparison between a VPC with NaaS and CaaS since they are two different services.

The process of establishing a VPC supported by a NaaS service is divided in 6 main phases, as Figure 5 shows: *request formulation*; *request conversion; resource discovery; node mapping*; *link mapping*; *node* and *link embedding*. Phase 1 encompasses the formulation of a request, in which the user defines all resources: virtual servers (CPU, RAM and HDD); network topology and the characteristics of all nodes and links. The request is then sent to the *Manager* which first performs the conversion of the request from XML to a structured topology (phase 2). Moreover it is performed the resource discovery (phase 3), and the nodes and links are mapped (phase 4 and 5) using a mapping algorithm which considers both the occupation of physical nodes and links. Finally, the VPC is enforced (phase 6).

Based on the tools already available in the NCSS, which already supported NaaS, we developed the necessary mechanisms to support CaaS. The request of a CaaS service was done taking into consideration some aspects of the widespread VPN concept as example - PE, CE, fully mesh topology.

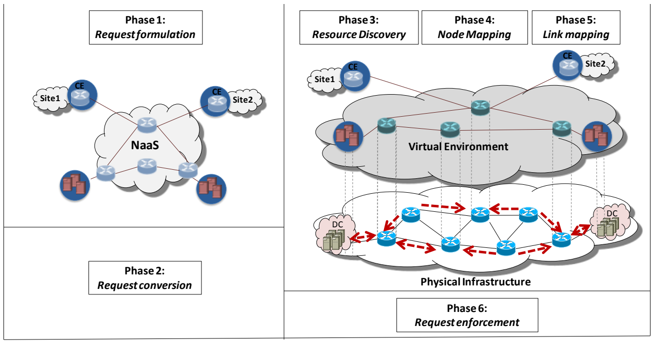


Figure 5. Virtual Private Cloud mapping process with NaaS.

The user is able to define CEs as if he was requesting a VPN, specify ingress and egress bandwidth requirements for each network endpoint (hose model), rather than specifying the requirements between all pairs of endpoints, as in NaaS. The user can drag and drop graphical depictions of routers and servers, which represent the sites and cloud resources of the network, and then connect them to a central graphical element representing an abstraction of the network (Figure 4). These connections contain information about the bandwidth from each element to and from the network. In the end the user just has to press the commit button and the CaaS information is processed and the request enforced. In the end, the user gets a VN which connects the customer sites and cloud resources, according to the user's configuration parameters.

A VPC supported by CaaS releases the user from the NaaS complexity, but adds an intermediate process step. Nevertheless this does not imply an increase of time from the request to the enforcement moment.

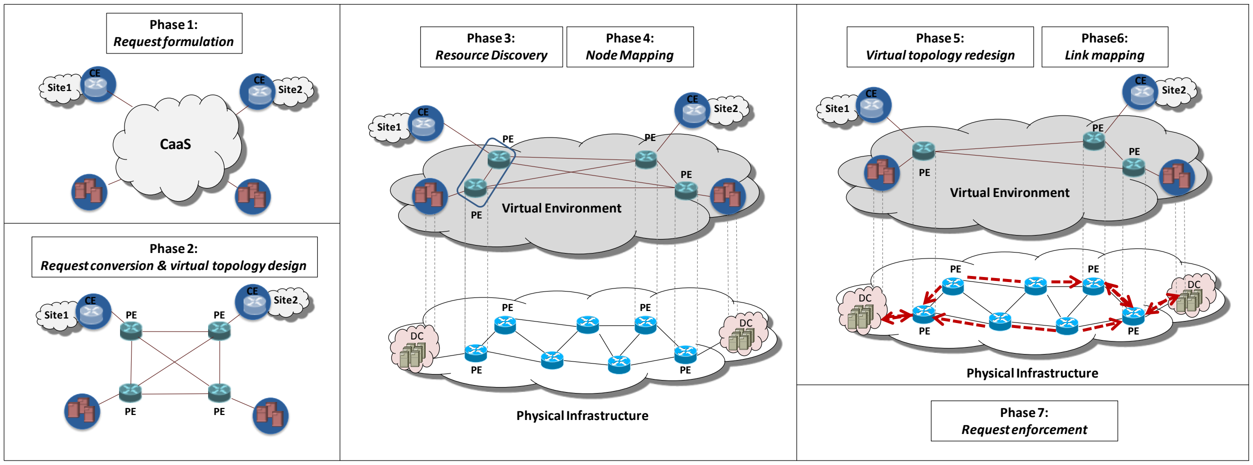
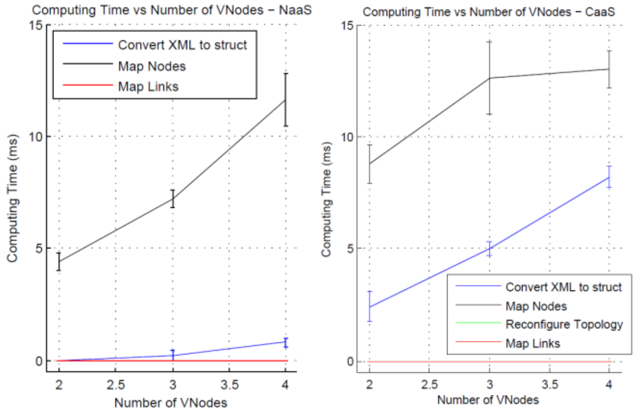


Figure 6. Virtual Private Cloud mapping process with CaaS.

As depicted in Figure 6, the process has 7 phases: *request formulation*; *request conversion* (includes the creation of the virtual topology which is not entirely defined); *resource discovery*; *node mapping*; *virtual topology reconfiguration*; *link mapping*; *node* and *link embedding*. First, the user configures the CaaS requirements on the GUI (phase 1). Once the *Manager* has received the request, it converts it to a topology structure, where a virtual PE router is connected to each site and cloud resource, with the PE routers connected in full mesh (phase 2). The bandwidth of the links is set to the minimum necessary to fulfill the worst case hose-model requirements. Then resources are discovered (phase3), and virtual PE routers and cloud resources are mapped to physical PEs and datacenters according to the temporary topology (phase 4). Note that the set of candidate PE routers for a CE encompasses those located near that CE, which eases the mapping process. Phase 5 comprises the virtual topology reconfiguration, so that virtual PEs mapped to the same physical PEs are joined together in one virtual PE. The resulting links are mapped onto the physical substrate in phase 6. In the end, all elements are enforced, phase 7.

To finalize, Figure **7** presents the time that the *Manager* takes to process a VPC request. The presented values are an average of 5 requests.



(a) Computing times for mapping NaaS&Cloud

(b) Computing times for mapping CaaS&Cloud

Figure 7. Virtual Private Cloud mapping process with NaaS.

Figure 7 (a) shows the results for NaaS. The time to convert from the XML message to a structure increases with the number of nodes (0-1ms), but is still a small value when compared to the time to map the nodes (4.5-11.5ms).

As for the mapping with CaaS, Figure 7 (b), we can see an extra time stage, topology reconfiguration, which is not visible because it is at a 0ms value in both cases, and thus is below the red line. The conversion process takes a little longer with CaaS, 3-14ms, since there is the need to create the full topology. Node mapping values for CaaS are higher, 9-13ms. This might be explained by the fact that the node mapping process in CaaS is made using the temporary topology, which includes 1 PE router per site or server + 1 server element per virtual server. Times for topology reconfiguration in the CaaS are very small, which might be explained by the simple nature of this action, which mainly consist in removing some virtual nodes and links. Link mapping on both cases is always close to 0 (not perceptible since the out time unit while measuring was ms). This might be due to the small number of nodes in both VNs and substrate.

## VII. Conclusion and Future Work

A major limitation of CC is the lack of coordination between CC resource control and network resource control. To overcome this limitation, elasticity and agility of the cloud must be extended to the network infrastructure. In this sense we have associated the concept of VPC with network virtualization, allowing cloud and network resources to be handled as a single set in a dynamic and flexible way.

Two network service models are proposed, CaaS and NaaS. The former roughly corresponds to the traditional managed network-based VPN paradigm. The latter provides a service which is functionally identical to a network. Both models should have a role to play in future networks. Moreover, we present a platform able to handle NaaS and CaaS services along with cloud resources.

One of the tasks that is still open for future work is the integration of BGP/MPLS VPNs in the platform as a CaaS, since VPNs are today in the market and represent the strongest short-time deployment possibility. The integration with the cloud using standardized application programming interfaces (e.g. Open Grid Forum Open Cloud Computing Interface) is also in the evolution roadmap.

## Acknowledgment

The authors are thankful to the members of the SAIL project, particularly those involved in Work Package D, for the collaboration and fruitful discussions.

## References

[1] P.Mell and T. Grance, “The NIST Definition of Cloud Computing (Draft)”, National Institute of Standards and Technology, January 2011.

[2] T. T. Huu, G. Koslovski, F. Anhalt, P. Vicat-Blanc Primet, and J. Montagnat. “Joint elastic cloud and virtual network framework for application performance optimization and cost reduction”. Journal of Grid Computing (JoGC), pp. 27-47, 2010.

[3] F. Anhalt, G. Koslovski, and P. Vicat-Blanc Primet. “Specifying and provisioning virtual infrastructures with HIPerNET”. Int. J. Netw. Manag., pp. 129-148 May 2010.

[4] T. Miyamoto, M. Hayashi, and K. Nishimura, , "Sustainable Network Resource Management System for Virtual Private Clouds," Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on , vol., no., pp.512-520, Nov. 30 2010-Dec. 3 2010.

[5] FP7 Project SAIL, http://www.sail-project.eu/.

[6] FP7 Project GEYSERS, http://www.geysers.eu/.

[7] So et al, “VPN Extensions for Private Clouds” IETF Internet Draft, February 2011.

[8] T. Wood, A. Gerber, K. Ramakrishnan, P. Shenoy, J. Van der Merwe, “The Case for Enterprise-Ready Virtual Private Clouds”, HotCloud’ 09, 2009.

[9] J. Nogueira, M. Melo, J. Carapinha, and S. Sargento, “Network Virtualization System Suite: Experimental Network Virtualization Platform”, in TridentCom 2011, 7th International ICST Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, April 2011

[10] J. Nogueira, M. Melo, J. Carapinha, and S. Sargento, "A distributed approach for virtual network discovery," GLOBECOM Workshops (GC Wkshps), 2010 IEEE , vol., no., pp.277-282, 6-10 Dec. 2010

[11] J. Nogueira, M. Melo, J. Carapinha, S. Sargento, “Virtual Mapping into Heterogeneous Susbtrate Networks”, Computers and Communications (ISCC), IEEE Symposium, pp.438-444, 2011.