Telecom Grade Cloud Computing

Version 1.0, 2011-05-03

Questions pertaining to this document, or the terms or conditions of its provision, should be addressed to:

SCOPE Alliance,
c/o IEEE-ISTO
445 Hoes Lane
Piscataway, NJ 08854
Attn: Board Chairman

Or

For questions or feedback, use the web-based forms found under the Contacts tab on www.scope-alliance.org

1. PURPOSE

According to a recent analysis by Gartner [2], cloud computing is one of the top technology trends; at the same time, the market for cloud computing services is expected to grow to $112 billion by 2015. Consequently, most major IT companies and many telecommunication providers have started to roll out cloud computing offerings, at all service levels: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). From telecommunications' perspective however cloud computing has a number of issues that need to be addressed before NEPs and operators can fully leverage the benefits of the technology.

Cloud providers succeeded in pushing the cost of computation and storage down by concentration, virtualization and economies of scale; by doing so, they had to compromise on some fundamental issues: networking, security and real-time characteristics.

For data intensive applications the cost of networking can be one or two orders of magnitude higher than the cost of computation – essentially eating away the economical benefits of cloud computing. In fact, it has been noted that the cost of networking is the single most important cost factor for any cloud computing infrastructure, compared to which the cost of computation and storage becomes negligible. Hence, the trend of concentration taken to the extreme is no fit for high throughput applications: locality will be important and needs to be embraced in order to control the cost of networking.

Multi-tenancy and the addition of a third party – the cloud provider, beside the original service provider and consumer – will inherently weaken security, exposing the system to malware attacks, malicious employees and denial of service attacks. Resolving these issues – not just contractually, but through technology – is paramount for providing the level of security found in telecom networks.

Last but not least, telecom networks have multiple quality of service (QoS) requirements – in terms of networking, placement of computation and interaction between computational instances – that need to be supported and resolved in the context of a cloud computing infrastructure.

Therefore, the purpose of this paper is three-fold. First, it aims at defining and publicizing the differentiating factors that can make cloud computing usable for telecom and real-time services. In this context, we include the role and importance of inter-cloud architectures as well as the usage of private, public and hybrid solutions for real-time and telecom services. Second, it is intended to provide a telecom perspective for standardization efforts in the area of cloud computing. Finally, its goal is to create a common work agenda for the SCOPE Alliance in its relationship with various standardization bodies. The mission of Scope Alliance is to advance the objective of a vibrant and diverse ecosystem of COTS: carrier-grade platform components utilizing open standards. Cloud computing has the common goal of reducing the platform costs while continuing in the direction of increased openness of architecture.

2. AUDIENCE

This document is intended for the following audiences:

- Standardization bodies (such as DMTF, OASIS and other) that work on various aspects related to cloud computing

- Telecom vendors interested in cloud computing technologies
- Operators planning to use cloud computing infrastructure either for cost management or for generating new services
- Cloud computing infrastructure and service vendors who want to understand the requirements emerging from real-time and telecom service providers

3. REFERENCES

[1] T. Sridhar, "Cloud Computing: A Primer, Part 1: Models and Technologies," *The Internet Protocol Journal,* Volume 12, No. 3, September 2009.

[2] Gartner, “Gartner Identifies the Top 10 Strategic Technologies for 2011”, http://www.gartner.com/it/page.jsp?id=1454221, last accessed on 31.3.2011

[3] Gartner, “Gartner’s Top Predictions for IT Organizations and Users, 2011 and Beyond: IT’s Growing Transparency” Available on demand from Gartner

[4] A Border Gateway Protocol 3 (BGP-3), and related other RFCs, available at the address http://tools.ietf.org/html/rfc1267

[5] OSPF Version 2, and related RFCs, available at http://www.ietf.org/rfc/rfc1583.txt

[6] Guidelines for creation, selection, and registration of an Autonomous System (AS), and related other RFCs at http://tools.ietf.org/html/rfc1930

[7] Web Services Standards Overview Poster; Posted by Mark Little on Mar 06, 2007 at http://www.infoq.com/news/2007/03/innoq-ws-standards-poster

[8] *Virtualization Management (VMAN) Initiative,* Distributed Management Task Force, Inc. at http://www.dmtf.org/standards/mgmt/vman/

4. INTRODUCTION

The landscape of cloud computing is today exceptionally fragmented with a large number of companies offering services at different levels – IaaS, PaaS or SaaS – based on multiple technology platforms (such as hypervisors and operating systems) and using largely proprietary mechanisms and formats. This fragmentation – while it certainly encourages competition – makes it nearly impossible to design cloud applications that can be deployed across multiple technology platforms.

The same fragmentation is apparent in standardization efforts. Most major standardization bodies have now established one or several working groups that focus on some aspects of cloud computing, along with a growing number of new bodies that target specifically cloud computing related standard issues. In fact, this fragmentation is one of the driving forces behind this white paper, as it defines a coordinated view of the telecommunication industry with regard to the most important aspects that need to be addressed as part of the standardization efforts.

Despite the large number of companies providing cloud computing services, some of the key issues from a telecommunication industry perspective are not yet properly ad-

dressed. For example, the primary focus of infrastructure build-up efforts was the reduction of the cost of computation and storage, through concentration and virtualization, while networking has been until recently largely ignored, both from usage and virtualization point of view. This however exacerbated the impact of network latency and bandwidth on the performance of cloud based applications and led to an almost complete lack of mechanisms to enforce quality of service constraints. Similarly, security concerns are today addressed largely through contractual enforcement, with little support available that would mitigate the most stringent concerns by providing a secure-by-design infrastructure.

5. TERMS AND DEFINITIONS

IaaS	Infrastructure as a Service
NEP	Network Equipment Provider
OVF	Open Virtualization Format
PaaS	Platform as a Service
QoS	Quality of Service
SaaS	Software as a Service
SLA	Service Level Agreement
VM	Virtual Machine

6. CLOUD COMPUTING BENEFITS AND OPPORTUNITIES

6.1 Network Equipment Providers

For NEPs, cloud computing can represent the underlying technology for both internal and external efficiency as well as for new business opportunities.

For *internal efficiency*, cloud computing provides the same benefits as for most IT-intensive businesses: economies of scale, flexibility and pay-as-you-go model. In itself the usage of cloud computing technologies is not radically different; however, when coupled with the re-use of the cloud infrastructure across internal and external offerings, it will require additional functionality, outlined in this paper.

For *external efficiency*, cloud computing technologies enable NEPs to improve their responsiveness and delivery schedules by shifting to software as a service models, away from traditional, box based approaches. Such an approach can be applied to both traditional telecom infrastructure sales as well as to managed service deals; the net effect is the reduction of turn-around delivery times, flexibility and reduced operational costs that can benefit both NEPs and their customers.

Cloud computing can enable NEPs to enter *new business areas*, by leveraging on cloud based offering of services beyond their traditional customer base. While this is a model in its early phases, it clearly will require the build-up of cloud infrastructure solutions with strong telecom grade characteristics.

6.2 Operators

For operators cloud computing technologies provide the basis for both improving the efficiency and flexibility of traditional service delivery and to enter new business areas.

On the *efficiency side*, cloud computing provides multiple opportunities. It enables deployment of some of the services on standard, large-scale computing environments. It also enables larger scale outsourcing of business process and network management tasks.

Many operators have already started offering cloud computing services to their enterprise and private customers. The key asset operators can leverage on is control of the networking infrastructure, which enables them to provide a much better experience, making the need for telecom grade cloud infrastructure even more stringent.

7. TELECOM GRADE CLOUD INFRASTRUCTURE

7.1 Definition

In the context of this white paper we define a telecom grade cloud as a cloud computing infrastructure suitable for the deployment, in a cost efficient manner, of applications with stringent availability, reliability, quality of service and security requirements. Telecom grade and real-time applications clearly fall under this category; however this definition also defines those aspects of a cloud infrastructure offering where NEPs and operators can differentiate themselves and offer a better service to their customers.

It is important to emphasize that these requirements shall be met through technology rather than just contractual enforcement. In most cases – especially with regards to security – pure contractual enforcement is just not enough and will act as a hinderer for acceptance by customers. Hence a telecom grade cloud shall have built in mechanisms to enforce these requirements.

One of the key decision points when implementing cloud computing services is whether the infrastructure shall be private (owned by the provider) or leased (from an established cloud computing provider). In the telecom sector we will likely see a mixture of these two: some NEPs and operators will be large enough to benefit from a customized, internal cloud offering; others will see efficiency benefits from relying on large scale providers.

The relationship of telecom clouds with the rest of the ecosystem is shown in *Figure 1.*

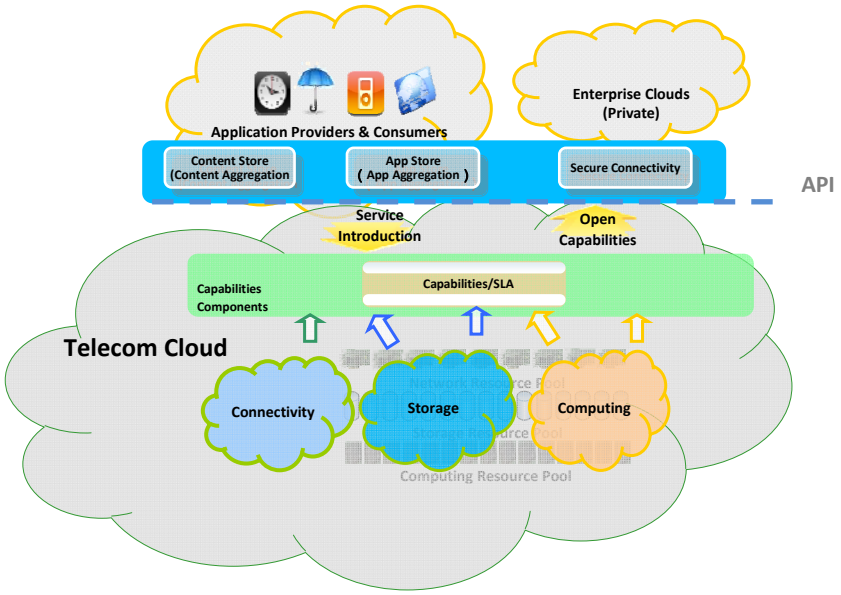

Figure 1 Telecom Cloud Computing within the broader Ecosystem

7.2 Differentiating factors for telecom grade clouds

There are four fundamental differentiating factors for telecom grade clouds that need to be considered when deploying such infrastructures: locality, SLA management, security and trust management and the usage of inter-cloud technologies.

7.2.1 Locality

Real world experiences from leading cloud computing providers, NEPs and operators have shown that there is a limit to how far concentration of computing and storage can be pushed. When networking represents the bulk of the cost and some applications have very strict quality of service requirements with respect to latency and throughput – which puts network performance at the top of concerns with respect to operating a cloud infrastructure - concentration becomes simply un-sustainable and a more distributed, locality-aware cloud infrastructure is required. Intuitively, this goes against the underlying assumption of unlimited, amorphous cloud, “floating” in an unspecified virtual space; however, in order to meet the needs of applications with telecom grade requirements where network performance is a top concern, a balance shall be found between concentration (that reduces the cost of computing and storage but increases the cost of networking) and distribution (which has the reverse affect). In a similar manner, availability requirements will also influence where the ideal locality balance lies: certain level of distribution is mandatory for satisfying end-to-end availability requirements of complex applications.

Obviously, exposing the network topology to the applications is unrealistic and would just shift the burden of managing locality back to the application. We believe it is a sound architectural choice to keep locality management within the cloud infrastructure; however

the cloud management framework has to cater for a more distributed environment where the application needs in terms of network performance will be a key factor in deciding where certain computations will be placed – a departure from today's load balancing geared primarily towards balancing computational load.

7.2.2 Service level agreements

Telecom grade applications are characterized by a number of availability and quality of service related service level agreements (SLAs) that need to be fulfilled, as the application will depend intimately on whether and how these are met. Therefore, SLA management – definition, analysis and enforcement – is a key feature required from telecom grade cloud infrastructures.

An SLA management framework can also support locality management without the need to expose the infrastructure architecture to the application. Information on locality requirements can be extracted from SLAs provided by the application and the optimization of network aware deployment can be managed as an SLA enforcement activity.

7.2.3 Security

Identity and location confidentiality, as well as communication and data security are some of the most sensitive areas with respect to security in a cloud environment. While telecom companies have a strong track record in being a trusted partner, cloud computing can open up for new security threats that need to be mitigated: multi-tenancy may lead to new attack opportunities; involvement of a third party (the cloud provider, in a public cloud scenario) means a new potential source of leak or unauthorized access; in general, remote access to data and computation introduces a degree of uncertainty that needs to be mitigated.

All these concerns are not specific just to the telecommunications industry. However, maintaining the high level of trust existing today requires a particular focus, well beyond green field approaches. In addition, there are few specific aspects related to telecommunications that require special consideration:

- *Handling of data about the customer, which can be subject of legal requirements on storage of data, accessibility and protection:* laws in different countries stipulate which data must be stored (or, conversely, may not be stored), where the storage shall be done etc. These requirements generate similar location-aware requirements as mentioned under the networking chapter

- *Need to maintain strict identity and trust management.* Contractual commitments are not sufficient – large organizations rely on the security of their communication infrastructure that shall not be exposed to e.g. rogue employees of the cloud infrastructure provider.

- *Need to fulfill traceability requirements,* imposed by regulatory bodies. The security of features that fulfill these requirements is of utmost importance and should ideally undergo additional analysis compared to what is required for many other features to reduce risks of abuse. The telecom industry is experienced in this area and should be well equipped to handle the issues.

Therefore, cloud computing infrastructures shall provide mechanisms to manage legal constraints on data accessibility, readability and localization, embedded into the technical fabric of the system, rather than enforced solely through the contractual mechanism. Identity and communication data shall be as safe as managed in-house and as isolated as in separate houses – i.e., with technology enforced isolation from the other users of the cloud infrastructure. This is generally considered a core aspect of virtualization, but management solutions and implementation must be carefully engineered to preserve the isolation.

7.2.4 Inter-cloud

Besides service providers constructing new, planet-scale virtualized datacenters, as software and expertise becomes more available, enterprises and smaller service providers are also building cloud computing implementations. Interoperability amongst the varied implementations of these clouds, from the lower level challenges around network addressing, to multicast enablement, to virtual machine mechanics, to the higher level interoperability desires of services, is an area deserving of much progress and will require the cooperation of several large industry players.

Identifying a profile of protocols and formats is one part of the interoperability puzzle; a set of common mechanisms must also be present, both inside the clouds, and in-between the clouds.

7.2.4.1 Presence and conversation transport

For presence and conversation transport there is a need for protocols. For the telecom characteristics the implementation must support requirements in terms of high availability, low latency and session retainability. The requirements are in the area of presence, structured conversation, lightweight middleware and content syndication.

7.3 Principles for a telecom grade cloud infrastructure

As highlighted through the requirements from the previous chapter, there are four main areas a cloud infrastructure targeting telecom services shall address: *locality and exposed internal topology* of the cloud; an *SLA management and enforcement framework* for managing and delivering on QoS constraints specific to telecom and real-time services; *security, trust and identity management,* both from a legal and 'prior practice' perspective; support for cloud interoperability. In this chapter we present the fundamental principles the design of a cloud infrastructure meeting these requirements shall adhere to.

7.3.1 Data-centric computing

The most fundamental change required from a cloud computing infrastructure is the need to expose internal topology and provide mechanisms for managing locality and data placement. Consequently,

- *Data shall be placed where it is used.* This principle requires an intelligent data replication, caching and distribution mechanism that can adjust availability of data based on application provided constraints and observed network behavior

- *Computation shall be placed where the data is.* This principle places the data at the center of how computations are organized.

These principles are based on the assumption that, compared to the cost of shifting data around, the cost of providing computation resources at the place where the data is available is negligible. The focus is not storage, but rather the prioritization of networking versus computation, making computational power available in the right nodes in order to minimize network transfer.

7.3.2 Networking as a first class resource

Traditionally, cloud computing focused on managing computations and storage, with just limited (usually local to the data center) consideration for networking. However, as locality is increasingly important, we believe there is a need to include networking as a first class component in the management of cloud resources.

Therefore we argue for an *coordinated management of networking, computation and storage* – these shall be allocated as one single end-to-end virtual resource and mapped to the underlying physical infrastructure, including the last mile access– hence extending the concept of what belongs to the cloud computing infrastructure all the way to the end consumer of cloud computing resources. We believe that such an approach is the only one that can achieve the fulfillment of the requirements of telecom services. We also believe that there is a need for efficient network routing techniques in order to improve network performance with good load balancing, support for multi-tenancy and reliability.

7.3.3 QoS management

While from a cloud infrastructure point of view the internal structure of the cloud becomes relevant in the context of telecom services, we believe that this structure shall not be fully exposed to the user of the cloud infrastructure – for complexity as well as security reasons (in order to avoid facilitating attacks based on e.g. involuntarily exposed potential shortcomings).

Hence, we believe *QoS management shall be based on a constraint specification and enforcement mechanism.* Cloud users shall express their needs in terms of constraints that the cloud infrastructure shall resolve based on the available underlying infrastructure and other constraints expressed by other users. Constraints may cover a wide range of issues, such as:

- Networking: inter-VM bandwidth, latency, redundancy etc
- Co-location / disjoint placement of VMs: VMs may need to be grouped together or may be required to be located on different HW (e.g. in case of application-implemented redundancy schemes or for disaster recovery); proximity to specific external locations (e.g. IP address ranges) etc
- Placement constraints on data: data may need to be placed within geographical certain area or, in case of e.g. a content delivery network, content may need to be placed close to the consumers

The key characteristic of this principle is that the application is not made aware of the actual structure of the cloud infrastructure, but rather is given the possibility to drive the allocation of resources in a way that can meet its quality of service constraints.

7.3.4 Resource management

Cloud management today is largely concerned with usage of storage, CPU and memory, which is a fairly simple model to handle, allocate and charge for. However, as mentioned, networking is also an important type of cloud resource. It is of interest to support resource monitoring and accounting at a detailed level and encompassing all aspects of computing – CPU, storage, memory and networking - using e.g. introspection technologies.

7.3.5 Security

Security shall be an integral part of all three major components of any cloud infrastructure: computation, storage and networking.

On *computation level*, telecom services shall be tamper-resistant, e.g. the same level of security shall be achieved as in a closed system ("*as safe as in separate houses"*). While this is not specific to telecoms, it is an important principle that will impact on how VMs shall be isolated and monitored (e.g. in terms of where the execution is placed).

On *storage level*, the infrastructure shall guarantee that no customer data can be decoded without the presence (virtual or physical) of the customer. This can be achieved .e.g through *partitioning of data between the cloud and the user* in such a way that none of it is usable individually. In this way, the cloud can actually provide an enhanced level of security. Further, the customer resources may be encrypted while at rest, in particular VM images shall not be readable by either cloud provider or other entities without customer being present.

On the *networking level*, all communication shall be secured and tamper-resistant. Somewhat surprisingly – given the emphasis of telecom services on the networking aspect, this part is usually considered to be under control. Virtualized firewalls and antivirus along with bandwidth monitoring and encrypted channels may all be required to provide a secure cloud environment.

7.3.6 Inter-cloud requirements

One of the most basic resources which cloud computing delivers is the Virtual Machine. One way or another, a subscriber requests the provisioning of a particularly configured virtual machine with certain quantities of resources such as memory processor speeds and quantities. The format of this request varies by cloud computing platform and also is somewhat specific to the type of hypervisor.

Similar or even worse differentiation exists with respect to naming, discovery and conversation setup for storage interoperability.

8. STANDARDIZATION AGENDA

8.1 SLA Management

As was discussed in the previous sections applications need not know the specific internal architecture of the cloud, but they should be able to use abstract methods of defining their requirements. Based on this principle, applications must be able to express their SLA requirements to the underlying infrastructure and measure the quality of service provided by a telecom grade cloud. In order to minimize fragmentation of the application space and enable a vibrant ecosystem of application providers that can take advantage of the advanced capabilities of the infrastructure, there is a clear need for standardizing the methods that applications can use both for expressing their SLA constraints as well verifying and measuring performance. Application providers must not need to rewrite their applications for different cloud infrastructures and users and enterprises must not feel locked in specific architectures. This will farther enable the ability to easily migrate application between cloud providers and technology choices.

Although there are several efforts across the IT industry today of standardizing the interfaces to the data center infrastructure itself, it is clear that these efforts ignore the importance of interface of compute and network resources and lack the ability to provide end-to-end visibility and control.

We believe that there is a need for an SLA specification language that can be generic and flexible enough to apply to a wider range of cloud infrastructures. The language can define several types of requirements, including but not limited to the following:

- Physical and logical affinity attributes that will define how virtual machines must be placed in the cloud. Physical attributes can instruct co-location within a physical machine, rack, availability zone, data center or even region. Logical attributes can define location relationships across the logical topology of the cloud.

- Compute and storage relationships that define the dependencies between applications (compute) and data that will enable the correct relative placement of compute to storage resources.

- Performance metrics that will define requirements in terms of network or storage bandwidth resources.

- Quality-of-service metrics that will define requirements in terms of end-to-end latency, jitter both for network and storage connectivity.

- Availability metrics that will define requirements in terms of availability of connectivity between compute resources and end-users or within individual data centers.

- High availability installations that will automatically protect applications with redundant compute and storage resources.

Obviously SLAs are not of much use unless applications can verify them and more importantly can react to fault or quality of service issues. One of the main benefits of the cloud is that it allows the creation of elastic applications that can dynamically allocate resources based on load. For example an application will use metrics of quality derived from the network infrastructure to add/remove processing or network resources. There-

fore applications will also need a standard mechanism for requesting performance metrics from the infrastructure and/or receive notifications of anomalies. Examples of metrics that applications will be interested in include, but are not limited to:

- End-to-end bandwidth allocation.
- Latency and jitter encountered by a particular application across different elements in the path spanning both compute and network resources, including hypervisors, LAN switches, appliances, WAN interfaces etc.
- Computational load as seen by hypervisors.
- Storage load as seen by storage devices and/or dedicated storage networks.
- Faults in compute, storage or network infrastructure components.

Obviously any effort towards standardization of expressing and verifying SLAs must be generic enough to allow NEPs and Operators to differentiate in the type of services that are offered. At the same time it should be designed around the basic principles of extensibility and open interfaces.

8.2 Networking

As indicated in section 7, performance and cost considerations drive the need for locality of processing and storage to meet the typical criteria of telecoms services in a cloud. Achieving the control plane and data plane aspects of directing data to local resources will require some integration of the IT and networking platforms. Such integration would also achieve economies in the infrastructure if owned by one entity.

A typical use case requiring locality may be the support of thin client solutions which are delay sensitive. The need then is to provide processing near to the user. While the requirement comes from the SLA and abstracted management, the underlying network must have mechanisms to respond. This may be through optimized routing and switching mechanisms. A further use case for locality is content delivery network solutions. For in-network storage, the IETF is trying to address this in the DECADE project.

With the high number of addressable resources in a typical data centre compounded by the virtualization used for cloud services there is consideration needed for performance and flexibility between using layer 2 switches or Layer 3 routers. Layer 2 is the simpler mode, where the Ethernet MAC address and *Virtual LAN* (VLAN) information are used for forwarding. The disadvantage of Layer 2 networks is scalability. When we use Layer 2 addressing and connectivity in the manner typical for IaaS clouds, we end up with a flat topology, which is not economic when there are a large number of nodes. An option is to use routing and subnets—to provide segmentation for the appropriate functions but at the cost of forwarding performance and network complexity [1].

A further approach is to implement programmable switches. These are a new design of network equipment where, unlike conventional routers, simple packet switching mechanisms and control functions are separated. Users can freely develop and operate control middleware independently of the switching mechanism. This equipment enables the realization of advanced new technologies that link networks and cloud computing. Stan-

dards for this are developed by the Open Networking Foundation (http://www.opennetworkingfoundation.org/).

Protocols for data center networking are addressed mainly by the IEEE, such as IEEE 802.1aq defining shortest-path bridging. There is also similar work in the IETF TRILL (*Transparent Interconnect of Lots of Links*) working group. The key motivation behind this work is the relatively flat nature of the data-center topology and the requirement to forward packets across the shortest path between the endpoints (servers) to reduce latency, rather than a root bridge or priority mechanism normally used in the *Spanning Tree Protocol* (STP). The shortest-path bridging initiative in IEEE 802.1aq is an incremental advance to the *Multiple Spanning Tree Protocol* (MSTP), which uses the *Intermediate System-to-Intermediate System* (IS-IS) link-state protocol to share learned topologies between switches and to determine the shortest path between endpoints.

FCoE is important for converged storage network environments. The IEEE is working to enable FCoE guarantees through an Ethernet link in what is known as "Lossless Ethernet." FCoE is enabled through a *Priority Flow Control* (PFC) mechanism in the 802.1Qbb activities in the IEEE. In addition, draft IEEE 802.1Qau provides end-to-end congestion notification through a signaling mechanism propagating up to the ingress port, that is, the port connected to the server *Network Interface Card* (NIC) [1].

8.3 Inter-cloud

To address the interoperability issues between clouds, certain commonalities amongst clouds must be adopted. With the Internet, interoperability foundations were set with the basics of IP addressing, DNS, exchange and routing protocols such as BGP [4], OSPF [5], and peering conventions using AS [6] numbering. Clearly, analogous areas in cloud computing need to investigated and similar technologies, but for computing, need to be invented.

We call the protocols and formats, collectively, *Inter-cloud Protocols*. We call the common mechanisms, collectively, an *Inter-cloud Root*.

A well choreographed conversation using several inter-cloud protocols and formats need to occur between the two clouds in question. Depending on the exact interoperability scenario, many different parts of the Inter-cloud protocols would be utilized. We envision some parts of the inter-cloud protocols to specify aspects around communications, some to reference the physical payouts of the clouds, and so on. Just as in Web Services, as the actual interoperability scenarios get more complete, the number of protocol areas expands. In the Web Services area [7], there are almost 100 interoperability conventions, protocols, and formats specified.

8.3.1 Virtual Machines in the Inter-Cloud Context

Most cloud computing implementations have a capability to deliver a Virtual Machine “on demand” to a subscriber, who requests the provisioning of a particularly configured virtual machine with certain quantities of resources. At that point the Virtual Machine is “booted” with an image (or via instructions) to result in a running system.

The metadata which specifies the image or the system is a crucial abstraction which is at the centre of VM interoperability, a key feature for Inter-cloud. Our goal is to promote an

open, secure, portable, efficient, and flexible format for the packaging and distribution of one or more virtual machines.

VM Mobility is that feature in a particular hypervisor which allows a running system to be moved from one HW to another HW. As far as the running system is concerned it does not need to be reconfigured, all of the elements such as MAC and IP address and DNS name stay the same; any of the ways storage may be referenced (such as a World Wide Name in a SAN) stay the same. Whatever needs to happen to make this work is not the concern of the running system.

Typically, VM Mobility is restricted to a Layer 3 subnet and a Layer 2 domain (for VLANs) because the underlying network will support the VM operating outside of the local scope of those addresses. Needless to say, the network addressing scheme in a cloud operated by an entirely different service provider is not only a different subnet but a different class B or class A network altogether. Routers and switches simply would not know how to cope with the “rogue” running system.

Another aspect is that the instantiation instructions of the VM for the running system are specific to that cloud computing platform and the hypervisor which it uses. If the new cloud takes an entirely different set of instructions, this is another barrier to VM Mobility. Open Virtualization Format (OVF) [8] is a platform independent, efficient, extensible, and open packaging and distribution format for virtual machines. OVF is virtualization platform neutral, while also enabling platform-specific enhancements to be captured. We are encouraged by the possibility of convergence of this space on OVF by some recent open source conversion utilities which are a proof point that VM meta-data for instantiation and for mobility can be solved eventually.

8.4 Security

As we outlined in chapter 7, contractual enforcement of security has proven to be insufficient from the perspective of many potential clients of cloud computing. According to a recent Gartner analysis [3], by 2015 80% of enterprises using external cloud services will demand independent certification that providers can restore operations and data.

There exists a plethora of generic standards for authentication, authorization, secure communication. In recent years, there have also been standardization efforts to develop standards which directly or indirectly target the cloud (such as trusted computing standards that enable e.g. remote attestation). In many cases, entire frameworks have been standardized that allow flexible configuration and profiling to target specific use cases. A first priority from telecom cloud computing perspective is to bundle and profile these standards to address cloud computing specific issues, rather than developing new standards from the ground up.

In addition, we need a overall security schema that defines different security needs/requirements at different layers and if possible the interactions between these layers to provide a coherent security framework. These layers include networking, hardware, hypervisor, virtual machines, operating systems as well as middleware solutions for PaaS and SaaS.

Standards or cloud specific profiles of existing standards shall be defined for secure data management (e.g. data confidentiality and integrity) and data life cycle (e.g. secure data deletion), including addressing issues related to privacy. Features shall include standard

enforcement and tracking of data placement, data partitioning within the cloud infrastructure and outside of it, data access, data backup, including mechanisms that can guarantee – in a verifiable manner – various levels of redundancy and access isolation for customer data. Such a framework shall take into account telecom specific requirements on data placement and compliance with legal requirements. We shall consider how security SLAs can be integrated and managed together with networking SLAs to visualize in a clear manner the trade-off that exists between these.

Standards shall be defined to provide cloud providers with secure management of the cloud. There is also need for auditing standards in the cloud; these can be the extensions to ISO 27001, but given the cloud specifications there is a need to have cloud specific standards. Security certification standards should be defined for handling sensitive data in addition to computations and networking inside the cloud. We believe a security attestation framework similar to other security critical industries that can be independently verified and certified has the potential to generate the trust level needed for an increased uptake of cloud computing. The experience accumulated in this area by telecom operators forms a good basis for working out such a framework.

9. CONCLUSION

Cloud computing undoubtedly represents one of the most important technology and business model shifts happening right now in ICT. Even though cloud providers succeeded in pushing the cost of computation and storage down by concentration, virtualization and economies of scale, some issues have been left unsolved: networking, security and soft real-time characteristics are the most important ones from the telecom industry's point of view.

In addition cloud computing suffers from fragmentation both in concrete solutions and in standardization efforts. Most major standardization bodies are working on some aspects of cloud computing, along with a growing number of new bodies that target specifically cloud computing related standard issues.

These are the main underlying issues that led SCOPE Alliance to consolidate the view of leading network equipment vendors in a white paper. The main findings of this white paper are as follows:

- There are fundamental issues that are yet to be addressed properly both in commercial solutions and through widely accepted standards. These include security, cloud networking, enforcement of service level agreements and inter-operability of various cloud solutions

- There is a need to create profiles of existing, widely recognized standards that target specific cloud computing. This is applicable primarily for security, where we believe there is a good foundation which however requires specialization for the specific needs of cloud computing infrastructures

Perhaps the most important aspect of telecom and soft real-time applications is the enforcement of specific *service level agreements*. We believe there is a need for a set of standard mechanisms to express the requirements of applications executing on a cloud infrastructure, without exposing the full details of the system. In order to verify how SLAs are realized, we also envision a set of standard metrics that can be used across multiple types of cloud infrastructures.

For *cloud networking,* the most important aspect is management of locality in order to meet the requirements of delay sensitive and soft real-time applications. We need mechanisms in the networking infrastructure that enables flexible, efficient, coordinated management of not only computing and storage, but also networking resources.

For supporting *cloud interoperability* which can have a major significance e.g. for operator-provided cloud infrastructures, there is a need for a set of inter-cloud protocols that can provide standard, widely used mechanisms for migrating computations between clouds, similarly to how networking interoperability is addressed for the Internet.

Security considerations are perhaps the most stringent ones and have been identified as the biggest hinderer for the adoption of cloud computing. Besides profiling existing, well established standards for use in cloud computing, we advocate the need for standards that can provide the foundation for third party attestation of cloud security as well as the need for a standard schema that defines different security requirements at different layers and if possible the interactions between these layers to provide a coherent security framework.

In conclusion, we believe that standardization bodies have their work cut out both in terms of adapting existing standards to the new realities of computing as a third party provided utility as well as in terms of covering the specific needs of various domains such telecommunications, the prime focus of this white paper.