The Standardisation of Cloud Computing: Trends in the State-of-the-Art and Management Issues for the Next Generation of Cloud

Authors Name/s per 1st Affiliation *(Author)*
line 1 (of *Affiliation*): dept. name of organization
line 2: name of organization, acronyms acceptable
line 3: City, Country
line 4: e-mail address if desired

Authors Name/s per 2nd Affiliation *(Author)*
line 1 (of *Affiliation*): dept. name of organization
line 2: name of organization, acronyms acceptable
line 3: City, Country
line 4: e-mail address if desired

***Abstract*—Future cloud systems will be guided and restricted by regulation from the standardisation bodies, if rolled out across the community. Trends in cloud computing observed to date have not been guided by any regulatory standards, and resources have been deployed in an ad hoc manner as demanded according to the business objectives of service providers. This is the least costly and most quickly revenue-returning business model. It is not however, the most cost-effective approach on a long-term basis: As a consequence of this roll-out model to date, the interoperability of resources deployed across operators is restricted through inability to achieve their utilisation in a regulated and controllable manner. The absence of standardisation in cloud management is therefore now beginning to be accommodated such that the cost and performance advantages of interoperable operation may be exploited. In this paper, we review the state-of-the-art in the roll-out of standards across the field and trends in their development over time. We present a model which defines the drivers for cloud interoperability and the constraints which restrict the extent to which this may realistically occur in future scalable solutions. This is supplemented with discussion on future challenges foreseen with regard to cloud operation and the way in which standards require provision such that cloud interoperation may be accommodated.**

Keywords*—cloud management, regulatory standardisation, cloud interoperation, context awareness, SLA management.

I. INTRODUCTION

"*Cloud computing is an evolving paradigm*" [1], which has been free to evolve in any manner required by cloud operators. While this achieves the individual business objectives in the least costly and most quickly revenue-returning manner, there are consequences of doing so: By operating according to distinct sets of management procedures which are specific to an individual operator, the ability to be interoperable across resources is more limited because the policies recognised at one are unlikely to be recognised at another. When monitoring resource usage for billing objectives, for example, different operators may achieve this using a contrasting set of operation and performance data, and management of both using a single approach will therefore not be possible within the current set-up. This limits the cost-efficiencies which are achievable through the application of interoperability.

Interoperation across resources is however, becoming more desirable [2]; Interoperation describes the use of resources distributed across different clouds which are rolled-out and controlled by different service providers in a potentially inconsistent approach. Any implementation issues are hidden from users and they will not be aware of the use of hardware from different operators nor of any management functions executing behind the scenes while their SLA is being fulfilled – access control will be uniform across resources as a function of client-side requirements, billing will be integrated across service providers, and application SLA will be fulfilled. The industry therefore requires an over-haul in order that a standardised set of operational procedures may allow interoperability across technologies so that the associated performance and revenue benefits of doing so may be achieved.

Standardised cloud deployment management strategies applicable across cloud providers have not been defined to date, and mechanisms continue to be rolled out in an ad hoc and operator-specific basis. This may be due to reasons such as:

1. The ad hoc method of deployment to date has been sufficient given the way in which resources have been utilised and the typical demands placed on them. This refers to a level of demand which can comfortably be accommodated with the current cloud deployment and by which operators are not over-whelmed with demand
2. Change in application requirements and user behaviour has resulted in a constantly moving target for managing
3. The challenges associated with cloud standardisation due to the number of bodies now involved

Current approaches to cloud roll-out and operation do not guarantee that interoperation across resources is possible. There are unique challenges to managing clouds which support interoperable operation as opposed to rolling-out and managing stand-alone solutions. A unified billing approach is challenging, for example, where the costs from multiple operators should be presented to customers in a single document. Standardisation therefore has an important role to play so that this may occur in a consistent manner [3]. While there is some overlap in the

Table 1 Cloud Governance Techniques to Assist Interoperable Operation

Cloud Management Platform	Security						Monitoring						Auditing		
	McAfee MOVE	StillSecure	Check Point	SELinux	VMware vShield	McAfee ePO	ntop nProbe	Nagios	Zenoss	WireShark	Nova	CloudKick	HyTrust	VMware Compliance Checker	Trend Micro
VMware vCloud	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	✓	✓	✓
OpenStack	-	✓	-	✓	-	-	-	✓	✓	✓	✓	✓	-	-	-
CloudStack	-	-	-	✓	-	-	✓	✓	✓	-	-	-	-	-	✓
OpenNebula	-	-	-	✓	-	-	-	✓	-	-	-	-	-	-	-
Eucalyptus	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	✓
Microsoft App Controller	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-
Amazon Web Services	-	✓	✓	✓	-	-	-	✓	✓	✓	-	✓	-	-	✓
oVirt	-	-	-	✓	-	-	-	-	✓	-	-	-	-	-	-
Nimbus	-	-	-	✓	-	-	-	✓	-	-	-	-	-	-	-
RightScale	-	-	-	✓	-	-	-	✓	✓	-	-	-	-	-	✓
Scalr	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
enStratus	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	✓

way in which this occurs across platforms today (Table 1), full interoperability cannot be guaranteed. In Table 1, we capture the way in which cloud governance, as an exemplar effect of management, occurs across a range of platform types from the perspectives of security, monitoring and auditing capabilities. Solutions implemented are both licensed and open source. There is not, however, a single solution which is implemented across this selection of cloud platforms and their interoperability with each other is therefore uncertain from this perspective. In response, there are a number of bodies working to achieve standardised approaches to cloud management (e.g., the ITU-T, DMTF, SNIA) so that the operation and performance benefits may be achieved. From the perspective of operators, this includes ability to achieve services from resources deployed by other operators; from the perspective of customers, this includes improved resilience and ideally, increased QoS; and from the perspective of the environment, this includes reduced carbon cost in the provision of cloud architectures.

The remainder of this paper continues as follows: In Section II, we review the state-of-the-art in cloud management standardisation from both the standardisation bodies and independent research community. In Section III, drivers behind cloud interoperability are discussed, together with exemplar design issues which limit the extent to which this may occur in a scalable cloud solution. The challenges of cloud management standardisation are also discussed in relation to the requirements placed on cloud architectures in the future and developments from the standardisation bodies to date. Finally, the paper concludes and presents future work in Section IV.

II. STATE-OF-THE-ART IN CLOUD STANDARDISATION

State-of-the-art cloud management processes should be able to accommodate the range of cloud types available, including public, private, community and hybrid clouds. Similarly, management processes should also be able to accommodate the various functions for which the cloud has been provided, including Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). Interfaces which enable interoperability between the different cloud types which mask the differences between each require careful provision and management for optimisation of operation and overall system performance from the perspective of customer and, ultimately, revenue return to service operators. The most standardised cloud management solutions are based on contributions from the Distributed Management Task Force (DMTF) (which includes the Global Inter-Cloud Technology Forum (GICTF), the Object Management Group (OMG) and the Storage Network Industry Association (SNIA)), the International Telecommunications Union – Telecommunication Standardisation Sector (ITU-T), and the National Institute of Standards and Technology (NIST) (Figure 1). In this work, we consider the contributions from these bodies in terms of '*how*' they have recommended that cloud resources should be deployed, '*why*', in their opinion, standardised solutions are necessary, and '*what*' cloud resources will commonly be available and will require standardised solutions.

Each regulatory body has contributed different aspects to the cloud standardisation process to date: The National Institute of

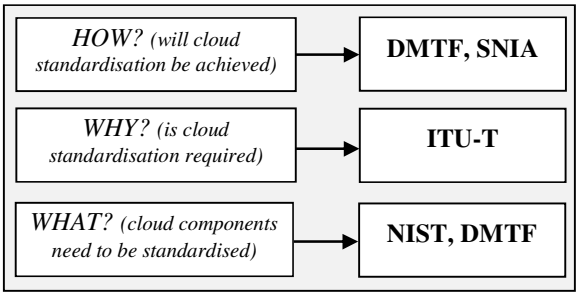

Figure 1 Cloud Management Standardisation and the Main Contributors across the Cloud Field

Standards and Technology, for example, has defined a cloud reference architecture [23] which highlights the components of a cloud which require management in future standards (the '*what*'). The ITU-T, on the other hand, outlines the reasons that cloud management is needed (the '*why*') [9]. The DMTF and SNIA describe a selection of ways in which cloud resources and data management should be provisioned (the '*how*'). A summary of their roll-out over time is presented in Figure 2 and a selection are described in more detail in the following sections.

A. *Contributions to Cloud Regulation from the Standardisation Bodies*

The DMTF, in association with a number of bodies involved in cloud roll-out, contributes to definitions for the ways in which clouds should be managed for optimised operation and performance. The DMTF has divided the overall work task by forming subsidiary working groups which respond to individual research challenges: The Cloud Management Working Group (CMWG) [10], for example, responds to challenges of interoperability across clouds by defining a common approach to dynamically provision, configure and administer cloud usage with an interface which masks the complexity of the management process. The Network Services Management Working Group within the Platform Management category of the DMTF, is working to provision integrated management profiles for the transport and routing protocols such that they can be used to manage both the Physical and Virtual network associated with clouds, as defined in their Charter [4]. The Platform Management Components Inter-communications (PMCI) Working Group deals with communication and functional interface features between the components of the platform management subsystem which are 'inside the box' [5]. The System Virtualisation, Partitioning and Clustering (SVPC) Working Group is providing a standard packaging format for virtual machines to support the interoperability of virtualisation management [6]. This includes extension of the Open Virtualisation Format (OVF) [7] and definition of a standardised data set to feed into the Common Information Model [8].

The Cloud Management Working Group of the DMTF is of particular interest for this work as they have developed the Cloud Infrastructure Management Interface (CIMI) specification [12] to improve cloud management and provide an interface to the infrastructure. The CIMI model specifies that interfaces within clouds use HTTP. Each entity in the system is identified by a unique ID. The interface with the cloud occurs via a Cloud Entry Point (the unique ID of which must be known) and other bodies within the cloud can then be accessed by following paths across the cloud using their unique IDs. Server activities are defined using the series of HTTP status codes from 100 (Continue) to 503 (Service unavailable) and resource operations which may be provided by a Cloud Provider include Create, Read, Update and Delete. Due to the range of technologies on which CIMI may be implemented, the model defines a range of formats using which it may be applied. Irrespective of provision for support across platforms however, it enforces that a selection of metadata should be available regardless of the technology used; this includes a URI, a name, a namespace, a type, whether or not it is required, and any constraints e.g., a maximum for the 'cpu' attribute. The model defines the range of identifiers using which entities may be identified; a common list of attributes by which all entities will be defined include the URI, the name, description, created and properties. Properties of the Cloud Entry Point are defined as part of the CIMI. This includes a catalogue of entities such as Systems, System Templates, Machines, and Machine Templates that can be queried by the consumer.

The Global Inter-Cloud Technology Forum (GICTF) is a Japanese conglomerate involved with the work of the DMTF on the development of standardised cloud computing solutions (the collaboration was established in June 2012) [24]. Their 'Work Register' was released in June 2012, in which their contributions are specified as including a Cloud Resource Data Model and an Intercloud Protocol [25]. The 'Intercloud Interface Specification Draft' highlights the way in which they approach the challenge [26]: In this, the interface is defined in terms of interoperation between intercloud controls and operation systems across clouds. It uses protocols at lower stack layers, including an intercloud protocol and cloud resource data model. The intercloud protocol facilitates connection between cloud service providers, and necessary to accommodate a unique ID and connect clouds in an approach consistent across all systems operating under the common management control protocol. This can be achieved using an IP address or domain name, and the connection will be identified the first time the system is used.

Across the cloud environment, storage is a key component of services available from operators. A standardised cloud storage business stream is under development by the Storage Networking Industry Association (SNIA) in association with the DMTF, as part of their Cloud Storage Initiative (CSI) [27] and Storage Management Initiative Specification (SMI-S). The CSI delivers storage which is elastic and offers resources in an on-demand fashion in that it bills for only those resources which are consumed. Management of clouds which incorporate this storage feature is provisioned using the Cloud Data Management Interface (CDMI) [28]; this specification describes operations which may be applied to cloud resources, such as create, read, update and delete, and the ways in which these actions may be enforced as a function of the entity being a CDMI or non-CDMI type. The SNIA is composed of a number of Technology Communities, such as 'Analytics and Big Data', 'Cloud Storage', 'Green Storage', and 'Storage Security' to accommodate cloud management in a manner anticipated to be compatible with the current and future challenge of standardised cloud operation.

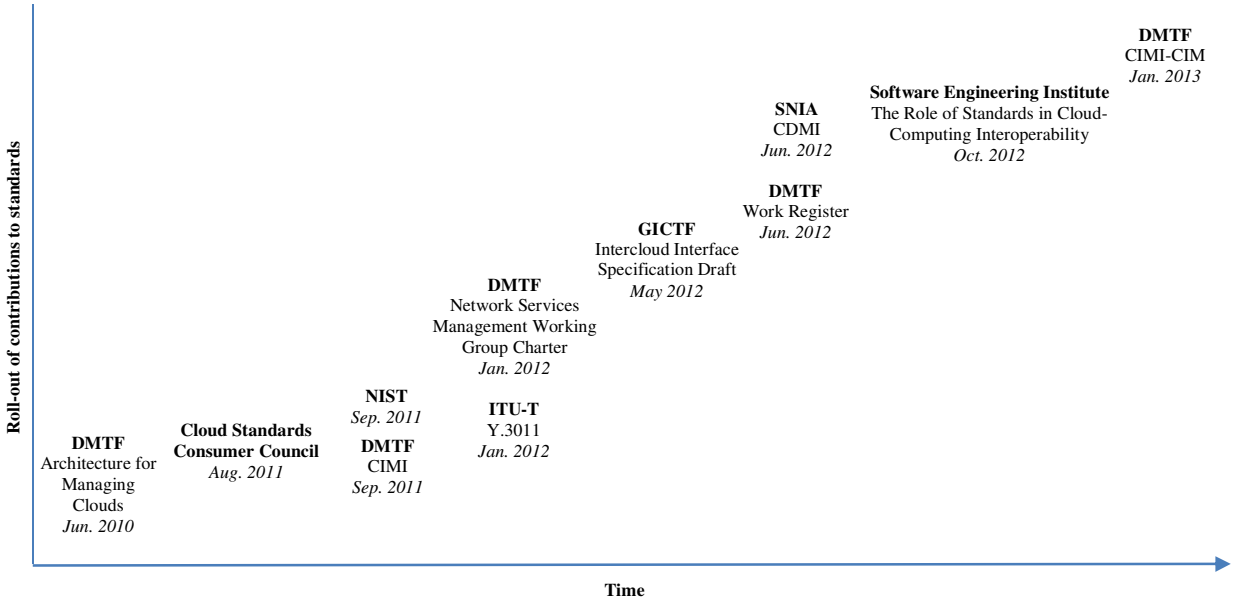

Figure 2 Roll-out of Cloud Management Standardisation and Efforts Towards Standardisation

The DMTF also provides a reference architecture for cloud management [22]. This model includes a provider, a provider interface, a service developer, a service consumer, data artifacts (e.g., request, SLA, contracts, agreements) and DMTF profiles which describe the associated behaviour for a management domain, such as server virtualisation. In contrast to other reference architectures (e.g., [23]), the DMTF model considers management on a geographical basis. This takes into account the geographical location in which cloud data is stored and compliance requirements in terms of geographical constraints when managing the interoperability process across clouds.

The DMTF therefore provides one of the more wide-spread definitions supporting a standardised approach to cloud management which take into account the range of ways in which cloud management requires standardisation. Nonetheless, there are also important contributions from other players in the field: Cloud management from the perspective of the ITU-T, for example, is identified as being required due to a need for more effective resource allocation [13]. Resource allocation describes an umbrella of requirements within the scope of managing the cloud, such as resource modelling and description, resource offering, resource discovery and monitoring, resource selection and resource allocation. Study Group 13 [11] within the ITU-T responds to research challenges associated with the challenges of future networks and NGN, mobility management and fixed-mobile convergence, and cloud computing. The cloud computing focus group is divided into two working groups: 'Cloud computing benefits & requirements' and 'Gap analysis & Action plan for development of relevant ITU-T Cloud Standard'. Working Group 1 is segmented into six core areas, which include: 'Cloud Definition, Ecosystem & Taxonomy', 'Uses Cases Requirements & Architecture', and 'Cloud Services & Resource Management, Platforms and Middleware'. Working Group 2 is divided into two core function areas, namely 'Overview of Cloud Computing SDOs activities', and 'Gap analysis & Action plan for development of relevant ITU-T Cloud Standards'.

The Internet Draft, 'Cloud/Data Centre SDO Activities Survey and Analysis', [21] summarises the activities from industrial Standards Development Organisation (SDO) partners, including the DMTF, GICTF, and ITU-T FG Cloud with regard to standardisation activities. In analysing the contributions made from these bodies, the objective is to identify areas where gaps exist and where overlaps in efforts are occurring so that a more streamlined and focused effort may result in the future.

A series of Technical Reports have also been published by the ITU-T Study Group (e.g., in [9] and [14]). Of specific interest to this research is that on 'Cloud Resource Management Gap Analysis' [19], which is currently under development. In this Technical Report, the ITU-T is concerned with interactivity across multi-cloud environments, and specifically an awareness of real-time availability of residual resources, reservation of resources based on client demand and take up or release of resources based on actual demand.

From the perspective of the National Institute of Standards and Technology (NIST) [23], the cloud is considered to be an architecture of five components, which include a cloud consumer, provider, broker, auditor and carrier. Within the NIST architecture, the provider is responsible for service orchestration, cloud service management, security and privacy. Cloud resource management is considered in the NIST reference architecture from the perspective of business support, provisioning and configuration, and portability and interoperability requirements. Business support includes accounting and billing, and customer/contract management. Of specific interest to our research, provisioning and configuration management within the scope of this model refers to the provisioning of resources, their change in availability over time, monitoring and reporting of resource allocation, metering to

record their usage and SLA management. This cloud reference model can be contrasted with the reference model developed by DMTF, which does not take into account characteristics associated with geography in decisions made.

The Internet Engineering Task Force (IETF) released a Cloud Reference Framework in December 2011, which expired in June 2012 [15]. This framework deals with both intra-cloud and inter-cloud operational issues at each protocol stack layer. Layers across cloud resources are considered in terms of: Data/Content layer, Application/Service layer, Resource control layer, Resource Abstract and Virtualisation layer, and Physical Resource layer. Capabilities within the resource control layer support configuration management, auditing, security management and Service Level Agreement (SLA) management. Another draft has also been released from the IETF which responds to mobility management issues for cloud-like architectures [16]. This responds to a need to separate the control and data plane by separating the Home Agent and Mobile Access Gateway functionalities into the control and data planes.

In addition to the DMTF, NIST, ITU-T and IETF, a number of other smaller bodies are also involved in the development of cloud standardisation efforts: The Cloud Standards Customer Council [20], for example, operates from the perspective of cloud users and advises on the way in which organisations may use the open standards available for customer benefit. This Council was founded by members such as IBM and Rackspace and is supported by leading organisations such as Citigroup and North Carolina State University. Current Working Groups supporting the overall work effort of the Council include Government, Healthcare, Security, XaaS and Big Data. The TM Forum [17], as another example, defines frameworks by which business are advised to use to operate more effectively using off-the-shelf products. Their frameworks include a Business Process framework, Information framework, Application framework and Integration framework. The Topology and Orchestration Specification for Cloud Applications (TOSCA) is also contributing to cloud standardisation efforts, and will soon be published by the Organisation for the Advancement of Structured Information Standards (OASIS) [18]. This standard is working to improve the portability of cloud applications and services, and to enable interoperable description of applications and infrastructure cloud services.

B. Contributions to Cloud Regulation from the Independent Research Community

Common across documents from the standardisation bodies is a description of '*what*', but not the associated '*how*' of the ways in which cloud management can be achieved in terms of, for example, policy development and implementation aspects. Due to the failure of regulatory bodies to define a set of best practices which are implementable across all cloud architectures and the subsequent absence of standardised solutions across the cloud environment to date, the independent research community is also contributing suggestions with regard to the way in which cloud management can be achieved to fulfil the competing operational objectives. We therefore look to the independent research community to complete this gap in knowledge. An exemplar management procedure is proposed in [29]: Management capability in this system is responsible for responding to queries for host and VM monitoring information, for estimating VM resource demands, for scheduling VMs in response to application requests, for sending VM management enforcement requests, for transmitting summary management information to local controllers, and for announcing VM presence. System operation is driven by periodic state monitoring and policies are applied which relate to scheduling, optimising and planning functions. VM placement is periodically optimised and the planning function facilitates any movement required.

In [31], cloud management is considered from a contrasting aspect in terms of scalability challenges across cloud computing solutions available to date. It is their opinion that cloud scalability is limited by the consideration of only a subset of performance characteristics in decisions which trigger resource scaling, such as CPU utilisation, for example. They advocate however, that a wider range of metrics, which accommodate the cloud's compute, storage and network resources, should be considered in any decision made for optimised performance overall and heightened long-term cloud operation.

In parallel with identification of a need to include cloud geography in decisions made as proposed by the DMTF, the authors in [32] present a cloud management solution which provides a solution to demonstrate the way in which such context may be used. In this, they consider the security implications associated with geographical boundaries, subsequent legal/political consequences and cloud operation as a consequence.

A monitoring solution for private clouds is presented in [33]. In this work, the authors consider key components of the cloud monitoring process to include a Node Information Gatherer, Cluster Data Integrator and Monitoring Data Integrator, as three examples of system components. Unique to this approach is the fact that the cloud monitoring procedure is driven by the life cycle of VMs distributed across the cloud and the application of intelligent cloud management activities in response.

Larger research groups are also contributing cloud management solutions within the independent community: 'The Role of Standards in Cloud-Computing Interoperability' [3], for example, has been provided from the Software Engineering Institute of Carnegie Mellon in their attempts to improve software solutions, in general. The core areas of cloud operation where standardisation is considered to be required in this scheme include during the processes of user authentication, workload migration, data migration and management, and workload management. In their work, they consider the cloud efforts as occurring in first, second and third generation phases: They consider that near-term standards require focus on user authentication and workload management, while standards in the future will become more concerned with pricing and intelligent billing aspects.

An Open Cloud Manifesto [30] within the independent research community advocates use of an open set of standards, and is supported by a number of cloud platform providers, such as VMware, enStratus, Zenoss and Trend Micro. It is also supported by key players in hardware and software solutions, including IBM, Cisco and Sun Microsystems. Such an approach will be favoured by these players as it will ensure that their

overall business objectives will be supported – roll-out of their hardware will be more prevalent and current customers will be more likely to remain customers, with a probable growth in resource usage. From some perspectives, an open approach to standardisation may be thought necessary to increase the appeal of using such resources through improving the ease with which they are deployed. On the other hand a guaranteed regulated approach is considered essential for many to support the regulation and security of resources retained on the cloud.

III. Characteristics and Constraints of Interoperable Cloud Scenarios

There are a number of characteristics and constraints which affect the design of management processes for interoperability objectives across the range of operational requirements of future clouds, such as integrated billing, access control, energy efficiency, and resilience and security. These are defined in the following sections to reinforce the operational characteristics of interoperable clouds which may be exploited for performance optimisation and to highlight the ways in which standardised solutions should be provisioned.

A. Intra- and Inter-Cloud Environment: A Position Statement

A data centre cloud consists of one or more hosts $C = (D, L)$ where $D:=D(C)$ is the set of clouds associated with an operator and $L:=L(C)$ represents the paths which connect the clouds. A host describes a repository within which all virtual resources associated with a cloud exist. A host $N = (V, W, L)$ comprises the set of all servers V, switches W and links L which inter-connect them across the host. The host resides on a physical network, connected to the data centre with which it is affiliated. The overall cloud associated with an individual organisation (or operator) may therefore consist of one or more clouds positioned over internationally distributed sites whose resources may interoperably be used as a function of demand and supply. The volume of virtual resources provisioned across all clouds associated with the operator will be limited by the residual resources across the data centre(s) on which the cloud(s) reside.

A path p to a cloud $p=(v,i,j,..,k,h)$ is composed of sub-paths (v,i), (i,j),…,(k,h) between the source of the client request and the destination server. Network links across the inter-cloud environment may connect clouds from the same service provider and from different service providers. Virtual resources may move across the inter-cloud environment from one cloud to another in order to fulfil resource requirements as dictated by the management system in response to requests placed on cloud resources by consumers. Bandwidth availability across sub-paths is dependent on the number of nodes which are wakened in the intra-cloud environment and the number of client requests being serviced by the cloud.

With regard to management processes applied across the cloud environment in general, a mandatory set of management data $C=(c_i,\ldots,c_z)$ is collected for all z devices across the network to reflect operation and performance. Latency L associated with context collection across the intra- and inter-cloud environment is dependent on the number of hosts, the number of devices across hosts, the propagation distance between hosts and the proportion of devices which are sleeping and wakened:

$$C = n(d \times bc) \tag{1}$$

$$L = \frac{C}{bw} + \left(s_a \frac{\lambda}{\mu} \right) \tag{2}$$

where n is the number of clouds which are included in the interoperability mix, d is the number of devices across a cloud for which context is collected, bc is the bandwidth requirements of the baseline context data set per device, bw is the average bandwidth availability across the end-to-end path between the client requests and the context data MIB repository and sa is the number of intermediary paths across the end-to-end path between client and server. The number of sub-paths are included in the decision-making process due to the additional latency incurred at this point when pushing data through the device, calculated based on the relationship between the data arrival rate λ and the data service rate μ through the device queue. In order for clouds to be interoperable, it is also essential that there is a base set of operations which use this mandatory context data set which may be applied for all resources. Actions additional to this may be specific to the service provider.

B. Cloud Interoperability Drivers and Influences on Design of the Management Process

The probability that a managed cloud will be interoperable with services from multiple physical data centre clouds is a function of the probability that the volume of requests arriving and predicted to arrive into the cloud will exceed the volume of virtual resources provisioned and that, for some reason or other, provisioning additional virtual resources on the physical devices available is not possible or it is cost-inefficient to do so. This may be due to the volume of residual resources available or the inefficiencies of waking sleeping devices, for example, which may be dependent on the duration of time which the additional resources are expected to be required.

Operational objectives of cloud interoperability therefore include:

1. Minimising latency to respond to application requests

From the cloud operator perspective, objectives of interoperability also include:

2. Minimising the total carbon footprint at the data centre on which the virtual cloud is deployed
3. Minimising the financial cost to fulfil to application QoS from the perspective of the cloud service provider

Enforcement of objectives by the management procedure will vary on a transmission-specific basis in response to its SLA, in response to the cloud operator, and the general opportunities exploitable for interoperability across the cloud environment. Achieving these objectives may be based, for example, on application requirements for latency, resilience, or reliability. In the case of prioritising cloud operation for reliability, for example, the latency overhead of the decision-making process may not be prioritised while searching for paths with sufficient bandwidth to support application requirements. As with many optimisation challenges, achieving interoperability across clouds is one with competing objectives: In the quest to optimise

the financial or carbon cost associated with cloud operation, it may be expected that the latency cost, which is important from the customer perspective, will be compromised. Financial cost of operation, as an operator objective, will be optimised through interoperability from the service provider perspective. Operators will not incur additional expenses to create or power on additional virtual resources in the event that they exist in another cloud, which may be operated by a different service provider.

Based on these interoperability objectives, there may become a point at which further interoperability across clouds should no longer be supported, and that only those resources currently included in the interoperability mix should be used:

- The latency to respond to an application request in an interoperable cloud scenario will be dependent on the need to refresh management context when the request is received, the availability of resources such as bandwidth and memory across the cloud and the number of cloud(s) with which it is interoperable. Overhead grows as interoperability is increasingly exploited, as captured in Eq. (2).
- The carbon footprint at a cloud is also dependent in part on the context monitoring process due to the associated overhead and the management decisions which are invoked in response. This will therefore also increase as the interoperability mix grows.

The extent to which each of the three factors affect operational efficiency is dependent to a large part on the monitoring process which drives management in terms of the size of the context data set used to monitor clouds, the need to wake sleeping nodes to support application requests or the need to deploy additional resources across the cloud in response to application demand. This decision can be made based on overhead incurred when considering further interoperability, in terms of the subsequent latency increases which result from activities performed to achieve interoperability (i.e., context collection) and resources consumed in the decision-making process.

It is most likely to expect that decisions regarding interoperability across clouds will be made dependent on firstly latency, to optimise the customer experience and the competitive advantage which a cloud operator may hold within the field in terms of its service to customers. This takes into account an assumption of network resilience, security of data and services, and effective billing and access control mechanisms. The second objective most likely to be prioritised is the financial cost associated with cloud operation, a fundamental requirement from the perspective of all service providers. In prioritising these two objectives, revenue achieved should be maximised through maintaining service to customers in the most cost-effective way, with lastly, ability to achieve carbon footprint constraints.

C. Cloud Interoperability Constraints

As the number of clouds associated with a cloud operator increases, the latency associated with the management process will also increase in parallel (Eq. (2)). This should be restricted to occur within the limits of the residual resources across the end-to-end cloud path. In order for resources to be interoperably used across clouds, it is essential that a base line minimum set of context is collected. This base line context set should be defined so that the basic operations required of all clouds, irrespective of the operator or the hardware within, may be enforced. This should also allow security and resilience of the monitored resources to be maintained. This base line context data set can be supplemented with additional context specific to the service provider, which is not essential to support interoperability with other service providers. Interoperability across clouds, and the specific data set collected, is also dependent on the fact that the same opportunities for configuration are available across clouds: both the baseline context data set and the additional data set may therefore influence operations within the cloud. Context collected which is additional to the baseline set should be controlled as a function of the overall data centre on which the virtual resources are deployed. This takes into account the residual resources available across the network and hardware-specific aspects across the data centre which require monitoring.

Context monitoring should therefore occur to a degree which is proportional to the volume of resources being monitored, to allow context awareness to be gained in a manner which is efficient, and occur to a degree which allows resilience to be maintained. This requires that real-time change in the state of the network is captured in a timely manner to allow any potential attacks on security to be protected against. In parallel, it also requires that residual resources do not become fully consumed by the overhead of this management aspect. When monitoring for competing efficiency objectives, the operational cost incurred should also be taken into account in decisions made. Monitoring of this base line set of context should also occur at a rate which does not negatively contribute significantly to the carbon footprint of the data centre on which the virtual resources have been rolled out.

Interoperability constraints may also exist as a result of the use of incompatible types of hardware across clouds. The occurrence of ‘vendor lock-in’ is often observed in cloud roll-out across organisations and restricts the extent to which interoperability may occur. This refers to the fact that once an organisation has selected a specific vendor on which their business function will be achieved, they are more likely to continue to grow using their services than to use a composite solution or switch to adopt entirely the devices of competitors. This fact results in continued use of the product over time and potential growth in the volume of hardware used and number of software capabilities incorporated into the overall cloud solution. Achieving interoperability across clouds which may exhibit occurrences of vendor lock-in will require that services supported at the hardware resources of a cloud which uses one vendor can be used interchangeably with the resources of other vendors. This requires that a sufficient range of context is collected to support the awareness process across vendor types, and relevant evaluation and decision-making processes such that all devices across the clouds may be suitably managed.

The extent to which cloud interoperability occurs will also be affected by breeches of security on any individual organisation or cloud, with an overall objective of management processes to minimise the negative impact on resource resilience. Attacks on resilience can quickly spread across all resources associated with a cloud; consequences between clouds may therefore subsequently be disabled to minimise the extent

to which the negative effects of such an occurrence are felt. Being an organisation or cloud-specific event, breeches in security may result in the interoperability across organisations being restricted until the attack is over and the revised state of the network has been communicated to all relevant parties.

Across geographical domains, political and legal implications have ability to affect the exploitation of interoperability. This may enforce, for example, a restriction on the extent to which data may interoperability be passed across political domains in accordance with a country's regulations. Future interoperability across clouds is also likely to become dependent on the enforcement of legalities, such as acceptable data protection and data exchange across boundaries. This will require close monitoring of the sites from which client requests are originating to ensure that the legal breeches do not occur, a management decision which continues to be challenging due to the prevalence of mobile devices.

D. Management Decisions for Interoperability Objectives

Defining the minimum mandatory set of context data on which standardised cloud management systems may be built is therefore a first step in their development. This data should be sufficient to identify that the interoperable use of resources at another cloud may lead to a higher level of performance. A cloud management system may therefore interoperably use the resources of another cloud in instances that:

- Bandwidth availability on paths between sites making requests and the cloud will allow application QoS requirements to be fulfilled, when QoS is measured according to latency, reliability and accuracy
- Server loading within the cloud allows QoS requirements of the application to be fulfilled
- The cloud is located within a proximity to the site of the client request which allows latency QoS to be achieved
- The context monitoring of cloud resources occurs within a latency in which the real-time network state will not be expected to have changed significantly
- There are inefficiencies associated with waking sleeping devices in the cloud belonging to the operator and that the resources at another cloud can be used more efficiently and effectively
- There are not any legal implications with regard to utilising cloud resources in another country

At a minimum, context data should therefore be collected which is able to identify any of these events as being true. However, while there are widespread advantages of exploiting interoperability, management systems will also be empowered with coping abilities when this feature may not be exploitable. A cloud management system may, for example, decide to create additional resources within the cloud in instances that:

- There are insufficient resources at other clouds with which the system is interoperable, which may be due to total capacity within the data centre being unable to support further additional virtual resource creation
- Additional resources may be created at the cloud within a latency which will allow application latency QoS requirements to be fulfilled
- The residual bandwidth available on paths between clients and servers is insufficient to support the volume of resources requested
- Clouds may be interoperable but the specific application being requested is not supported at the other cloud

Enabling management processes such that each set of capabilities are achievable requires context data and management processes representative of these properties.

As an additional consideration in the design of the management procedure for interoperability objectives, one or more clouds may access the context data Management Information Base (MIB) used to guide operation of the cloud management system. A cloud system which uses the resources of a number of service providers should have access to a repository of context data which is centralised across the inter-cloud environment to optimise the performance hit incurred through its access. Given the dynamic nature of clouds, this position will vary over time. The cost overhead of re-locating the centralised repository should therefore be proportional to the performance benefits to operation, measured in terms of the duration of time which this will be an optimum re-location decision. Evaluations need therefore to be accommodated within management processes to capture such events.

As with a number of aspects associated with networking today, cloud computing is a constantly moving target. A number of studies review the challenges associated with the future of cloud computing from the perspectives of security [34], resource management [35] and SLA provisions [36]. Achieving a standardised approach to operation will therefore prove challenging for this reason: Standardised solutions are required for each aspect of cloud operation which is currently challenging before the full utility of cloud computing may be exploited. Integrating these solutions then introduces an additional overhead into the management process.

IV. CONCLUSION & FUTURE WORK

"*The hype around cloud has created a flurry of standards and open source activity leading to market confusion*" [20]. Relationships across clouds from different providers are more likely to be seen in the future as complementary as opposed to offering competitive services. Resources from one cloud to another may be used interoperably, and the performance benefits achievable from the consumer perspective and financial benefits from the cloud operator perspective exploited to allow a more positive experience for all. Due to the evolving nature of clouds and their management, the policies behind their operation will be forced to evolve in suite. There are a set of core capabilities required, however, which should remain constant regardless of any more minor issues, such as a minimum mandatory set of context which requires collection regardless of the cloud operator or management approach used, context monitoring at a rate which achieves efficiency and ability to reconfigure quickly while fulfilling a minimum set of resilience and security objectives. The challenge is achieving this in a scalable way.

In spite of the Open Cloud Manifesto which promotes the development of open standards and is supported by a number of key players in the cloud field today, it is the opinion of the authors that standardised operation is essential to perform cost-effective management of utilisation across resources and to benefit operations from the client perspective to allow future solutions which are customer-driven. The focus of Open Standards is not the needs of customers but rather the groups which benefit financially from cloud services such as the cloud operators, software producers and hardware manufacturers.

In future work, we will review the ways in which cloud platforms across service providers achieve their business function and ways in which interoperability may be achieved across competing systems. This will look into, for example, the extent to which resources may migrate across contrasting cloud platforms. We will highlight the ways in which technologies can support the anticipated requirements of next generation clouds and future standards in operation and interoperability.

ACKNOWLEDGMENT

This work is supported by the India-UK Advanced Technology Centre of Excellence (IU-ATC) in Next Generation Networks, funded by the UK Engineering and Physical Sciences Research Council Digital Economy Programme and Government of India Department of Science and Technology.

REFERENCES

[1] National Institute of Standards and Technology, "The NIST Definition of Cloud Computing," Special Publication 800-145, Sep. 2011.

[2] Distributed Management Task Force, "Interoperable Clouds, A White Paper from the Open Cloud Standards Incubator," Version 1.0.0, Nov. 2009.

[3] Software Engineering Institute, "The Role of Standards in Cloud-Computing Interoperability," Technical Note CMU/SEI-2012-TN-012, Oct. 2012, pp. 1-28.

[4] Distributed Management Task Force, "Network Services Management Working Group Charter," Version 2.1 Draft, Jan. 2012.

[5] Distributed Management Task Force, "Platform Management Components Intercommunications Working Group," Aug. 2009; Available at: www.dmtf.org/sites/default/files/PMCIWGCharter.pdf.

[6] Distributed Managmeent Task Force, "System Virtualisation, Partitioning and Clustering Working Group Charter," Jul. 2010; Available at: http://dmtf.org/sites/default/files/SVPCWGCharter_0.pdf.

[7] Distributed Management Task Force, "Open Virtualisation Format Specification," DSP0243, Version: 2.0.0, Dec. 2012.

[8] Distributed Management Task Force, "Common Information Model Infrastructure," DSP0004, Version: 2.7.0, Apr. 2012.

[9] ITU-Telecommunication Standardisation Sector, "Part 2: Functional requirements and reference architecture," FG Cloud Technical Report Part 2, Feb. 2012.

[10] Distributed Management Task Force, Cloud Management Working Group; Available at: http://dmtf.org/standards/cmwg.

[11] ITU-Telecommunication Standardisation Sector, ITU-T SG13: Future networks including cloud computing, mobile and next-generation netwroks; Available at: http://www.itu.int/en/ITU-T/studygroups/2013-2016/13/Pages/default.aspx.

[12] Distributed Management Task Force, "Cloud Infrastructure Management Interface (CIMI) Model and REST Interface over HTTP," Version 0.0.35, Sep. 2011, pp. 1-105.

[13] ITU-Telecommunication Standardisation Sector, Cloud Computing Resource Management and Virtualisation; Available at: http://www.itu.int/ITU-T/studygroups/com13/sg13-q28.html.

[14] ITU-Telecommunication Standardisation Sector, "Part 4: Cloud Resource Management," FG Cloud Technical Report Part 4, Feb. 2012.

[15] B. Khasnabish, J. Chu, S. Ma, Y. Meng, N. So, P. Unbehagen, M. Morrow, and M. Hasan, "Cloud Reference Framework," 'work in progress' Internet Draft, Dec. 2011.

[16] B. Sarikaya, "Mobility Management Protocols for Cloud-Like Architectures," 'work in progress' Internet Draft, Oct. 2012.

[17] TM Forum; Available at: http://www.tmforum.org/.

[18] Advancing Open Standards for the Information Society, "OASIS Topology and Orchestration Specfiation for Cloud Applications (TOSCA)"; Available at: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=tosca.

[19] ITU-Telecommunication Standardisation Sector, FG Cloud; Available at: www.itu-int/en/ITU-T/focusgroups/cloud/Pages/default.aspx.

[20] Cloud Standards Customer Council; Available at: http://www.cloud-council.org/.

[21] B. Khasnabish and C. JunSheng, "Cloud/DataCenter SDO Activities Survey and Analysis," 'work in progress' as Internet Draft, Dec. 2011.

[22] Distributed Management Task Force, "Architecture for Managing Clouds, A White Paper from the Open Cloud Standards Incubator," DSP-ISO102, Jun. 2010.

[23] National Institute of Standards and Technology, "NIST Cloud Computing Reference Architecture," Special Publication 500-292, Sep. 2011.

[24] Distributed Management Task Force, "DMTF Announces Partnership with Global Inter-Cloud Technology Forum (GICTF) for Cloud Management"; Available at: www.dmtf.org/content/dmtf-announces-partnership-global-inter-cloud-technology-forum-gictf-cloud-management.

[25] GICTF/DMTF, "Work Register," Version 1.0a, Jun 2012.

[26] Global Inter-Cloud Technology Forum, "Intercloud Interface Specification Draft (Cloud Resource Data Model)," White Paper, May 2012, pp. 1-39.

[27] Storage Networking Industry Association, Cloud Storage Initiative; Available at: snia.org/forums/csi.

[28] Storage Networking Industry Association, "Cloud Data Management Interface," Version 1.0.2, Jun. 2012, pp. 1-224; www.snia.org/cdmi.

[29] E. Feller, L. Rilling and C, Morin, "Snooze: A Scalable and Autonomic Virtual Machine Management Framework for Private Clouds," in Proc. of IEEE/ACM Int. Symp. on Cluster, Cloud and Grid Computing, 2012, pp. 482-489.

[30] Open Cloud Manifesto; Available at: www.opencloudmanifesto.org.

[31] M. Hasan, E. Magana, A. Clemm, L. Tucker, S. Gudreddi, "Integrated and Autonomic Cloud Resource Scaling," in Proc. of IEEE Network Operations and Management Symposium, Apr. 2012, pp. 1327-1334.

[32] S. Yan, B. Sung Lee, G. Zhao, D. Ma and P. Mohamed, "Infrastructure Management of Hybrid Cloud for Enterprise Users," in Proc. of Int. DMTF Academic Alliance Workshop on Systems and Virtualisation Management, Oct. 2011, pp. 1-6.

[33] S. Aparecida De Chaves, R. Brundo Uriarte, C. Becker Westphall, "Toward an Architecture for Monitoring Private Clouds," IEEE Communications Magazine, Dec. 2011, pp. 130-137.

[34] A. Behl, "Emerging Security Challenges in Cloud Computing, An Insight to Cloud Security Challenges and their Mitigation," in Proc. World Congress on Information and Communication Technologies, Dec. 2011, pp. 217-222.

[35] K. Al Nuaimi, N. Mohamed, M. Al Nuaimi and J. Al-Jaroodi, "A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms," in Proc. 2nd IEEE Symp. On Network Cloud Computing and Applications, Dec. 2012, pp. 137-142.

[36] R. Buyya, S. K. Garg and R. N. Calheiros, "SLA-Oriented Resrouce Provisioning for Cloud Computing: Challenges, Architectued and Solutions," in Proc. Int. Conf. on Cloud and Service Computing, 2011, pp. 1-10.