D1.1.1

Draft Scenario and Requirements Report

Project number:	257243
Project acronym:	TClouds
Project title:	Trustworthy Clouds - Privacy and Resilience for Internet-scale Critical Infrastructure
Start date of the project:	October 1st, 2010
Duration:	36 months
Programme:	FP7 IP

Deliverable type:	Report
Deliverable reference number:	ICT-257243 / D1.1.1/ 1.0
Activity and Work package contributing to the deliverable:	Activity 1 / WP 1.1
Due date:	March 2011 – M06
Actual submission date:	01.04.2011

Responsible organisation:	UMM
Authors:	UMM (R. Glott); IBM (E. Husmann, M. Schunter); TU Darmstadt (A.-R. Sadeghi)
Dissemination level:	Public
Revision:	1.0

Abstract:	This document provides a literature-based discussion of cloud computing and possible cloud computing development trajectories. The aim of the paper is to define specifics and economic challenges and opportunities of cloud computing from the business point of view and to serve as a basis for the scenario building activities that form a core task of WP1.1.
Keywords:	Cloud computing, cloud computing definition, cloud computing ontologies, cloud computing ecosystems, cloud computing trends, cloud computing scenarios

Table of Contents

List of Figures

List of Tables

1 Introduction

TCLOUDS aims at a Future Internet where federations of standardized resilient and privacy-protecting global infrastructure clouds offer virtualized computing, communication, and storage resources that allow hosting of critical and non-critical ICT systems. While a large part of the debate how to realize this vision focuses on technological aspects (such as open standards), the implications of cloud computing for businesses should not be overlooked. Privacy and security concerns pose a significant risk towards cloud computing, but at the same time they offer tremendous opportunities for businesses to provide solutions and services in order to make cloud computing secure and protect the privacy of users and clients. This is well reflected in market analyses. For instance, Forrester Research (Penn 2010) expects the cloud security market to grow to 1.5 billion $ by 2015 and to approach 5 % of overall IT security spending. Whereas today identity management and encryption solutions represent the largest share of this market, particular growth can be expected in three directions:

- securing commercial clouds to meet the requirements of specific market segments
- bespoke highly secure private clouds
- a new range of providers offering cloud security services to add external security to public clouds

An example for the first category is the Google gov.app cloud launched in September 2009 that offers a completely segregated cloud targeted exclusively at US government customers. Similarly, IBM has launched a FISMA compliant Federal Community Cloud in 2010.

Cloudsourcing (Oclassen 2009) follows more or less the same economic rationale as traditional IT-outsourcing but provides more benefits, inter alia with regard to upgrades and patches, quick procurement services, avoidance of vendor lock-ins, and legacy modernization (Rajan 2010). For the security objectives when adopting clouds for hosting critical systems we believe that today's datacenters are the benchmark for new cloud deployments. Overall, the benefits need to outweigh the potential disadvantages and risks. While the cost and flexibility benefits of using clouds are easy to quantify, potential disadvantages and risks are harder to qualitatively assess or even quantitatively measure.

An important aspect for this equation is the perceived level of uncertainty: For instance, a low but contractually guaranteed availability (such as 98% availability) will allow enterprises to pick workloads that do not require higher guarantees. Today, uncertainty about the actual availability does not allow enterprises to make such risk-management decisions and thus will only allow hosting of uncritical workloads on the cloud. A related issues is that public cloud providers typically impose their standard terms of use on their customers and largely refrain from liabilities e.g. when it comes e.g. to business impacts of data loss or service downtime.

For security this argument leads to two requirements for cloud adoption by enterprises: The first is that with respect to security and trust, new solutions such as the cloud or cloud-of-clouds will be compared and benchmarked against existing solutions such as enterprise or outsourced datacenters. The second is that in order to allow migration of critical workloads to the cloud, cloud providers must enable enterprises to integrate cloud infrastructures into their overall risk management.

A fundamental problem for companies that are interested in cloud computing is the ambiguity of this term. While it is tedious to define cloud computing even when only its technological facets are considered, the problems of control, manageability and thus security and privacy multiply when enterprises, their business and service models and their organization of work and processes are considered in the light of cloud computing, as all these factors can contribute to a significant variation of the way how cloud computing can be designed and implemented, and what impacts it features. Thus, next to the above mentioned uncertainty, the complexity and unfathomableness the cloud shows towards businesses is another main reason for current observations such as: “many businesses are experimenting with Amazon’s EC2 to see how it works and what it can do for them, but not much of the enterprise computing workload has moved off businesses’ premises into the Amazon infrastructure cloud (Babcock 2010: 20).” In other words: While cloud computing has become widespread it is still far from unfolding its true potential.

This document contributes to the clarification of security and privacy risks and the meaning of cloud computing from a business perspective. After evaluating existent definitions of cloud computing from a business’ point of view (section 2), cloud ontologies are considered (section 3) in order to draw conclusions with regard to the composition and structure of cloud computing ecosystems and specifically new security and privacy risks deriving from cloud computing as a new technology (section 4). Section 5 discusses possible directions of cloud computing trends before the final section (6) provides first conclusions regarding the requirements that must be met in order to build possible and realistic cloud computing scenarios.

This paper provides the conceptual starting point for the identification and evaluation of cloud computing scenarios with regard to business opportunities and (privacy and security) challenges. The next steps of the scenario building process are based on detailed expert interviews and case studies in the field of cloud computing. The overall aim of Work Package 1.1 is to define relevant and realistic cloud computing scenarios that allow the identification and systematic assessment of typical security and privacy threats for cloud computing businesses under different context conditions (e.g. different sectors or ecosystems). In the final cloud computing scenarios emphasis will be laid on the TCLOUDS assets, i.e. principles, policies, security extensions at software component level, open cloud standards.

2 Cloud Computing Defined – the Business Perspective

Almost all articles, blogs, and other publications dealing with cloud computing start by pointing out that there is no recognized absolute definition of this term. The ambiguity of the term 'cloud' may be due to its longstanding usage as "…a euphemism for everything that was beyond the data center or out on the network. (…) the cloud was a mishmash of remotely connected parts and network protocols that didn't have much to do with the immediate problem (Babcock 2010: 1-2)." For businesses, such a 'cloudy' term is anything but helpful when identifying and evaluating new business opportunities.

The current situation is characterized by a multitude of definitions of cloud computing. Already in 2009, Vaquero et al. found and compared 22 definitions of cloud computing, some of which emphasizing technical aspects, such as scalability and virtualization (Klems 2008) or automation (of data center management) (Gruman & Knorr 2008), while others highlighted aspects of business models, such as collaboration and pay-as-you-go (Kaplan 2008, Cohen 2008), or cloud computing as a realization of utility computing (Kaplan 2008, Cohen 2008; Buyya et al. 2008, McFedries 2008). Out of this huddle of confusing, sometimes coinciding and sometimes contradictory notions, Vaquero et al. (2009) tried to derive something that would deserve to be called 'the essence of cloud computing'. The comprehensive characterization of cloud computing that Vaquero et al. (2009: 51) finally concluded from this multitude of definitions was the following:

"Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and / or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs."

Interestingly, Vaquero et al. (ibid.) point out that if one seeks for a definition of cloud computing representing the minimum common denominator of all 22 cloud computing definitions analyzed by the authors, there would be no definition at all. The reason for this is that, according to Vaquero et al., no single feature is shared by all these definitions. The features that are most widely shared across the 22 definitions are "scalability, pay-per-use utility model and virtualization (Vaquero et al., 2009: 51)."

Essential parts of this early notion of the term cloud computing still hold true for definitions that have been developed after 2009, i.e. after cloud computing has advanced from a rather abstract option and few real-life experiments in some companies to a relatively widespread business reality. Especially the sharing of resources, virtualization, scalability, pay-per-use (or pay-as-you-go) models and service level agreements (SLAs) are widespread components in more recent definitions of cloud computing.

The National Institute of Standards and Technology (NIST) defines, as a draft, cloud computing as "…a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models. (Mell & Grance 2011: 2)."[1]

Though there are strong similarities to Vaquero et al.'s definition, e.g. the notion of a pool of computing resources, there are also fundamental differences between these two concepts. Most notably the NIST definition is more generic and does e.g. not specify a technology such as 'virtualization' as a necessary element of cloud computing. On the other hand it puts additional emphasis on business-level service characteristics such as ubiquity of access as well as minimized management effort and service provider interaction.

Another interesting approach to define cloud computing is provided by Armbrust et al. (2009), as they focus strictly on the technical side of cloud computing, although their definition starts from acknowledging the pervasive character of cloud computing: "Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The services themselves have long been referred to as Software as a Service (SaaS). The datacenter hardware and software is what we will call a Cloud. When a Cloud is made available in a pay-as-you-go manner to the general public, we call it a Public Cloud; the service being sold is Utility Computing. We use the term Private Cloud to refer to internal datacenters of a business or other organization, not made available to the general public. Thus, Cloud Computing is the sum of SaaS and Utility Computing, but does not include Private Clouds (Armbrust et al. 2009: 1)."

It is evident that the latter definition is more conservative than the previous ones, as it derives its definition of cloud computing primarily from the two older concepts of SaaS and Utility Computing. The cloud is here essentially the data center that allows to provide computing resources as a utility and software as a service. Furthermore, this definition explicitly eliminates private clouds – i.e. clouds built, used and managed within the boundaries of an entity, such as an enterprise – from cloud computing. In addition, this definition recognizes business and service models and their components (such as prizing models and revenue streams) as external features of cloud computing, which the authors deem, however, essential to make cloud computing available to the public.

This neglects that whereas cloud computing has clearly a technological heritage in previous "on demand", utility computing and SaaS approaches, a particular interesting aspect of cloud computing is that it has given rise to new types of cloud services and business models – in particular by allowing advanced user configuration and user-side control. Therefore cloud computing has also gone beyond the analogy of ICT as a "utility" – such as energy or water – as in cloud computing services are not passively consumed but cloud resources are actively

[1] The five characteristics Mell & Grance refer to are on-demand self-service (the user does not rely on interference of a service provider), broad network access, resource pooling (also called multi-tenancy), rapid elasticity ((automated) scalability with the opportunity for the consumer to purchase seemingly unlimited resources in any quantity at any time) and measured service (the possibility to measure exactly and optimize resource usage through means for measuring storage, processing, bandwidth etc.). The three service models Mell & Grance see as components of cloud computing are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). Finally, Mell & Grance consider private clouds, community clouds, public clouds and hybrid clouds as the four deployment models of cloud computing (Mell & Grance 2011: 2-3). We will discuss these concepts in the following sections of this document.

configured and controlled from the user-side even though they ultimately belong to the provider.

Finally, the authors define that private clouds do not belong to cloud computing as they are not offering their services and resources as utility to the general public. However, this definition bears a number of problems, most notably its inherent terminological ambiguity. If – as in the first part of the definition - the datacenter hardware, software and services define the cloud, it appears difficult to understand why private clouds – from the viewpoint of these characteristics very similar to public clouds – would not fall under the term cloud computing. So this exclusion seems rather to be imposed by the authors than a logical consequence of the previous definitions. Here essentially the provider decision about a service being specific to a customer or open to the general public defines if is finally called a cloud or not.

A large part of the problems that appear common to most (if not all) definitions mentioned above can be explained by the circumstance that they try to capture the phenomenon of cloud computing at a stage at which it has not fully developed and differentiated. Given the dynamics of the phenomenon, authors in this subject matter tend to focus on 'the really new features' that come along with cloud computing (as, for instance, explicitly pointed out for the hardware of cloud computing by Armbrust et al. 2011: 1) and to succumb to the temptation to 'freeze' the volatile picture that 'the cloud' shows towards them in an angle that highlights these features, instead of examining the driving forces and directions of key drivers of the dynamics they see. For instance, there is not much sense in cloud computing definitions that demand a certain billing model (e.g. pay-as-you-go) when in reality many cloud services exist that are sold in forms of a fixed rate where the user has to determine, for instance, how many servers he wants to use, and to pay for them regardless of whether or not he actually does make use of them. According to Yousseff et al. (2008: 7), there are three different forms of prizing models for cloud services: tiered pricing, per-unit pricing and subscription-based pricing, whereby the latter, which does not allow charging for actual usage rates (i.e. pay-as-you-go) is claimed to be the dominant pricing model for SaaS. This clarification illustrates the limitations of including a certain pricing model in the components of a definition of cloud computing. The problem for businesses, in this case, is threefold: Enterprises interested in using cloud services may get confused when something that looks very much like a cloud does not offer a pay-as-you-go model, potential cloud providers that do not offer a pay-as-you-go model would not be allowed to label their offering 'cloud computing' regardless of how many other elements of the various cloud definitions they might comply with, and the fact that pricing models seem rather to be determined by what services the user wants the cloud to provide than by the fact that he uses a cloud in general may contribute to users' confusion, too.

The problem that is inherent to 'freezing' definition approaches is their ignorance towards cloud computing's high pervasiveness. Cloud computing permeates and changes hardware, software, services, business processes, business models, commercial and private IT usage, and the interplay of actors in the digital economy. It is this high pervasiveness that makes cloud computing so iridescent, complex, and rich of opportunities, and therefore – from a business point of view - definitions should try to highlight these dynamics and opportunities instead of selected and static features of cloud computing technologies or services.

A way to approximate such a definition is to look at what cloud computing changes in the interplay of users and providers of services, computing resources and storage. Babcock (2010: 15) postulates to examine clouds "…less through a microscope and more through the lens of business and technology convergence." Instead of debating cloud computing with a focus on new technological features pertaining to data centers the debate needs to be stand on its head, according to Babcock: "It's not its most prominent feature, the huge Internet data center that is the cloud's defining element. Rather, that's just one building block. The cloud is actually a number of advances – the data centers, the Web's setting of conventions for loosely coupled systems (two systems that don't know very much about each other) and an ability to activate virtualized servers remotely via standard Web services – that converge to give the cloud its enticing power (Babcock 2010: 7).”

Such an approach towards defining cloud computing allows to conceptualize cloud computing across all actors and levels involved. Babcock (2010: 8-9) describes cloud computing as a new distribution model for computing coupled with a business model that provisions computing resources at very low cost, due to so far unachievable economies of scale, and a new relationship between final consumer and technology, giving the user (which may be a SME that is equipped by the cloud with the same computing power as a large company) "…"programmatic control" over a part of the data center, the ability to command a server in the data center to run the program she has selected and sent (Babcock 2010: 9)."[2]

Babcock (2010: 19) describes the new relationship between user and service provider / technology as a shift from the overcome master/slave relationship to a peer-to-peer relationship, whereby the peer-to-peer relationship is considered to "… [give] the cloud its defining characteristic and [affect] businesses the most (Babcock 2010: 23)." In this understanding, Apple's iTunes store, which would fall under many of the above mentioned cloud computing definitions, appears only as an advanced but nevertheless 'old school' e-commerce offering, as it operates on the basis of huge data centers accessible through the Internet but keeps control over the user at the side of the data center, i.e. preserving the old master/slave relationship between provider and user (Babcock 2010: 6).

Comparing the various attempts to define cloud computing, the interplay of technology, business models and 'programmatic control' of the end user, as suggested by Babcock, appears indeed as a conceptual model that captures best the opportunities cloud computing opens up for businesses, thereby allowing for a vast variety of technological architectures, business models and end-user involvement in cloud computing instead of narrowing cloud computing down to certain technology and business model elements.

[2] Interestingly, Mell & Grance (2011: 2, 3) explicitly deny any user control over the data center when it comes to SaaS or PaaS: "The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings." Regarding IaaS, Mell & Grance acknowledge that cloud computing usually demands from the user to create or determine virtual appliances to run on the cloud's virtual machines, which gives the user some control. Whether or not this holds true for all current SaaS and PaaS offerings and especially for future forms of cloud computing must remain an open question, at this stage.

3 Cloud Computing Ontologies

All attempts to define cloud computing are interwoven with explicit or implicit assumptions about the components and layers clouds consist of. A typical example of this is provided by the definition of the NIST (Mell & Grance 2011). As laid out above, besides the five characteristics that are described above, Mell & Grance distinguish between three service models and four deployment models[3] of cloud computing. The categorization of service models follows the distinction between Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) proposed by, for instance, Knorr & Gruman (2008)[4] and Foley (2008), which has been referred to in D1.1.3, too:

Cloud Software as a Service (SaaS) means to run a single application in a data centre, and deliver the functionality via the Internet to the users. Examples of enterprise SaaS vendors are Salesforce.com (for sales force applications), Oracle/Siebel (CRM applications), Workday (for ERP applications), Citrix (meeting applications). Examples of SaaS desktop applications for end users are Google Apps, Zoho Office, Microsoft WindowsLive, etc. Finally, Internet portal sites, Internet search engines, and Internet social networking sites are essentially SaaS vendors for end customers.

A more restricted sub-type of SaaS are Internet platforms that offer Web services - specific APIs that application developers can use in developing applications that integrate services of the platform. Examples are Google Maps, ADP payroll processing, the U.S. Postal Service, Bloomberg, credit-card processing services, etc.

A similar sub-type of SaaS are Managed Services, such as virus-scanning services for email, spam-filtering services (Google/Postini, etc.), security services (SecureWorks, IBM, Verizon, etc.), desktop management services (CentreBeam, Everdream, etc.).

Cloud Platform as a Service (PaaS) delivers an application development environment (platform) as a service, usually also equipped with computing resources for hosting the applications developed on the platform. Examples are Amazon, Salesforce.com (Force.com), Coghead, Google (Google App Engine), Yahoo (Pipes), and Dapper.net. The Amazon Web Services (AWS), which is an important platform for many SaaS startups, consist of Simple Storage Service (S3), Elastic Compute Cloud (EC2), Simple Queuing Service (which uses S3), and SimpleDB. Examples for startups operating on top of AWS are Desktop Two, Zimdesk, GOPC, and Sun Microsystems' Secure Global Desktop. Another example, a startup in this field, is provided by Skytap (http://www.skytap.com), which offers a platform for virtual software testing as an on-demand service. Users can choose from a variety of operating systems (Linux, Solaris, and Windows) and databases. Skytap's Virtual Lab supports testing software for function, performance, and quality assurance, and it can be used for preproduction staging.

Cloud Infrastructure as a Service (IaaS) means to offer computing resources in form of virtual servers and storage as utility computing service. Examples are Sun Microsystems, IBM, Amazon and AT&T; and new vendors such as Nirvanix, Hatsize, Joyent, Cloudworks.

PaaS and IaaS show in particular the aspect of programmatic control that has been mentioned in the previous chapter and are defining new services models in the context of cloud computing. Whereas SaaS had already existed before the advent of the term cloud computing and still represents the largest part of the overall cloud computing market, PaaS and IaaS represent its strongest growth segments.

[3] We will examine these deployment models in section 5.

[4] Knorr & Gruman do not use the term Infrastructure as a Service, instead they use ‚utility computing', which seems to correspond largely with IaaS.

Wang & von Laszewski (2008) as well as Youssef et al. (2008) add to these cloud service models Hardware as a Service (HaaS) and Data as a Service (DaaS). Furthermore, Youssef et al. include Communications as a Service (CaaS) in their cloud ontology. Finally, Intel uses a cloud computing taxonomy that adds Service as a Service (e.g. billing as a service or security as a service) to the cloud service models, separates client software from SaaS and includes software, hardware and services provided or managed by the cloud client in order to develop a broad basis for cloud ecosystem analyses (see Figure 1).

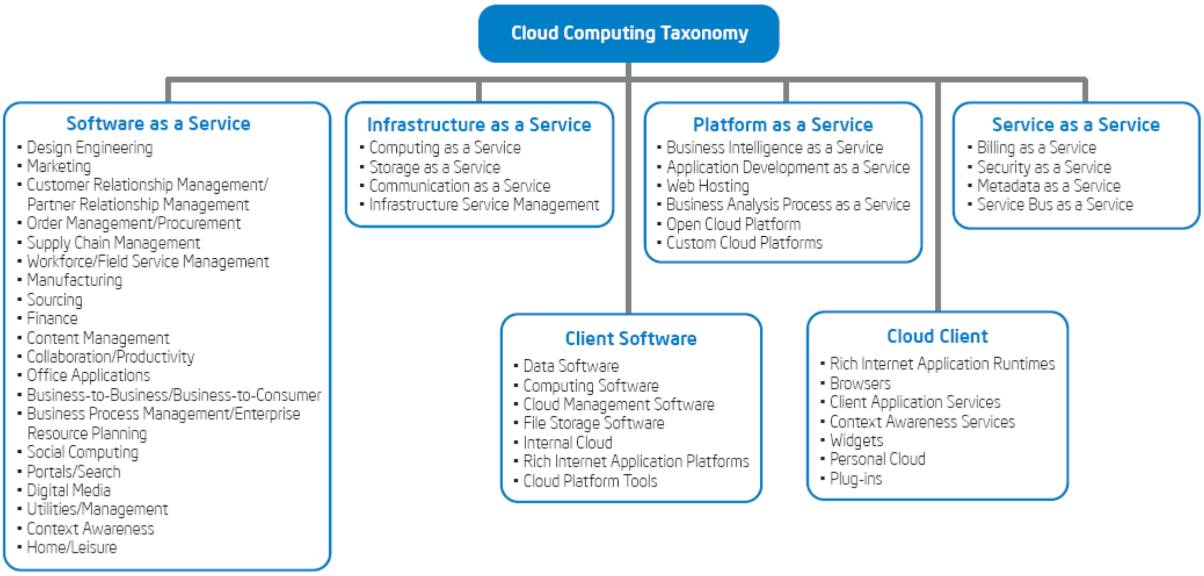
Cloud Computing Taxonomy
Software as a Service
• Design Engineering
• Marketing
• Customer Relationship Management/ Partner Relationship Management
• Order Management/Procurement
• Supply Chain Management
• Workforce/Field Service Management
• Manufacturing
• Sourcing
• Finance
• Content Management
• Collaboration/Productivity
• Office Applications
• Business-to-Business/Business-to-Consumer
• Business Process Management/Enterprise Resource Planning
• Social Computing
• Portals/Search
• Digital Media
• Utilities/Management
• Context Awareness
• Home/Leisure
Infrastructure as a Service
• Computing as a Service
• Storage as a Service
• Communication as a Service
• Infrastructure Service Management
Platform as a Service
• Business Intelligence as a Service
• Application Development as a Service
• Web Hosting
• Business Analysis Process as a Service
• Open Cloud Platform
• Custom Cloud Platforms
Service as a Service
• Billing as a Service
• Security as a Service
• Metadata as a Service
• Service Bus as a Service
Client Software
• Data Software
• Computing Software
• Cloud Management Software
• File Storage Software
• Internal Cloud
• Rich Internet Application Platforms
• Cloud Platform Tools
Cloud Client
• Rich Internet Application Runtimes
• Browsers
• Client Application Services
• Context Awareness Services
• Widgets
• Personal Cloud
• Plug-ins

Figure 1: Cloud computing taxonomy by Intel

Source: Intel 2010

Most descriptions of cloud computing remain at the level of phenomenological descriptions, distinguishing services and types of clouds (e.g. public vs. private clouds), but they fail to provide insights in the composition, structure and functioning or interplay of cloud components. A useful approach to capture these aspects in a generic ontology of cloud computing has been proposed by Youssef et al. (2008). The authors aim, inter alia, at determining the different layers and components of computing clouds and insights in the inter-relations between different cloud components. What makes this approach interesting from a business perspective is that it allows to allocate cloud services to distinct cloud layers and to examine opportunities and challenges of cloud computing for users and providers separately for each layer. The former bears potential for the analysis of the configuration and reconfigurability of cloud business models – which will be part of the next steps in WP1.1 -, the latter will be dealt with in the following section.

Overall, Youssef et al. (2008: 4, 2008a) propose to distinguish five layers of cloud computing, which are (see Figure 2):

- the firmware / hardware level, corresponding to HaaS
- the software kernel level
- the cloud software infrastructure level, composed of computational resources (corresponding to IaaS), storage (corresponding to DaaS) and communications (corresponding to CaaS)
- the cloud software environment level, corresponding to PaaS
- the cloud application level, corresponding to SaaS

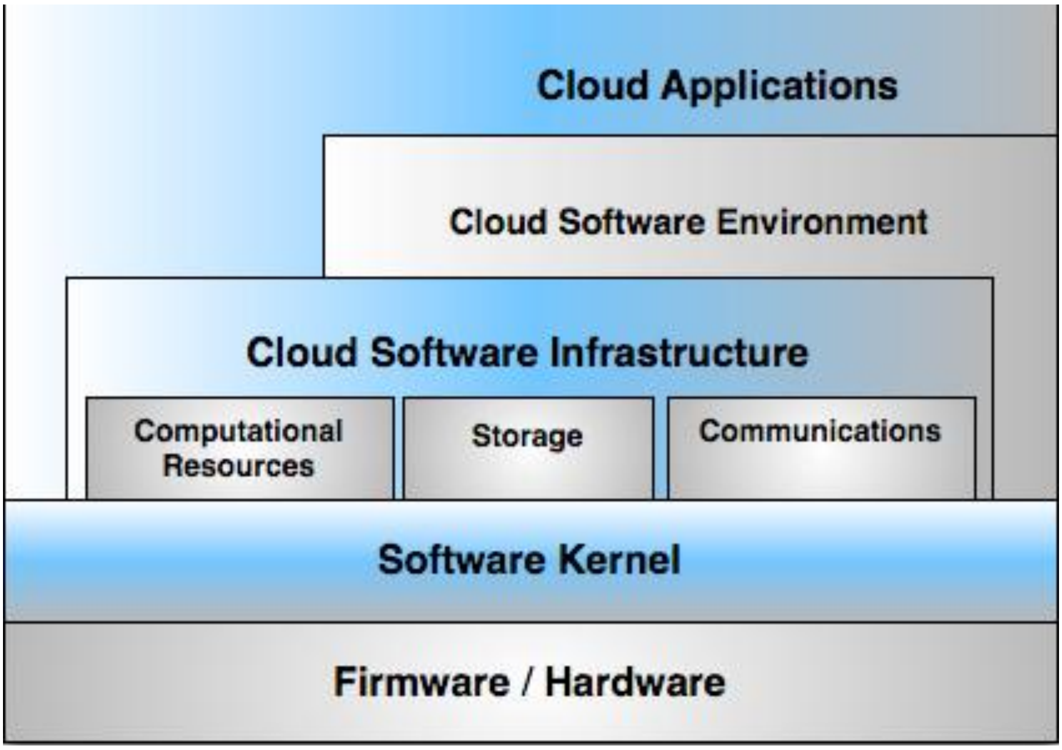

Figure 2: Cloud Computing Ontology proposed by Yousseff et al. (2008)

Source: Yousseff et al., 2008a

4 Actors, Opportunities and Challenges

4.1 New Business Opportunities

From the above derives that cloud computing offers a multitude of business opportunities at all levels of the cloud computing ontology. Kim (2009) sees "ample room for hundreds or even thousands of players in the market" that will give opportunities to many "large clouds" and a lot more "small clouds". Many of the small clouds, so Kim, will benefit from the infrastructure offered by the large cloud providers. Kim (2009) sees a need for companies that provide technological advice or integration and services to manage clouds effectively. Kim predicted the emergence of cloud technology integrators as a new form of cloud-based services. Firms like Kaavo, RightScale or CohesiveFT offer already such services (Foley 2008).

The market dynamics can thus result in the emergence of a cloud computing ecosystem composed of different types of players. The first group would be provided by the vendors that offer cloud services to the users. A second type offers solutions that help enterprises to provision and manage virtual data centres from commodity servers and storage, (e.g. 3Tera (AppLogic), Cohesive FT (Elastic Server on Demand)). Another type offers solutions for deploying and managing applications in a data centre (within an enterprise). Examples for such service providers are Elastra and Maavo. "Other types of players will include cloud computing platform vendors, the usual application software and middleware vendors, system integrators, and consultancies" (Kim 2009).

Armbrust et al. (2009: 7-8) consider the following kinds of applications as significant opportunities and drivers for cloud computing:

- Mobile interactive applications, which require high availability and rely on large data sets that can best be hosted in large datacenters. Armbrust et al. see particularly opportunities for services combining two or more data sources or other services, e.g. mashups.
- Parallel batch processing because cloud computing is tailor-made for batch-processing and analytics jobs of extreme huge amounts of data.
- The growing importance of business analytics and decision support, for the same reasons as laid out in the point above.
- The growth of compute-intensive desktop applications, such as the mathematics software packages Matlab and Mathematica, which are capable of using Cloud Computing to perform expensive evaluations.[5]

[5] Armbrust et al. also mention "earthbound" applications that require the capacity to handle (migrate into and out of the cloud) very huge datasets, but since the authors themselves relativize their statement in this regard so that finally they consider these services at current not really cloud-usable we have decided to ignore this subject here.

4.2 The Cloud Ontology Revisited

The crucial task for businesses is to develop an awareness of their position within cloud ecosystems in order to become able to thoroughly analyze its dependencies and impact and to identify opportunities and challenges aligned to business in the cloud. The ontology provided by Yousseff et al. (2008) appears meaningful in this regard, as it allows distinguishing opportunities and challenges for users and providers of services, software or hardware at every layer of their model.

- **Cloud application layer**

The cloud application layer provides users with access to the cloud through, e.g., Web portals. Advantages for users are the shifting of software maintenance tasks and computation workloads from their own computers to the cloud, decreasing operation and support costs, lower hardware requirements, and increased computing power and performance without huge capital investments in their IT equipment (Yousseff et al. 2008: 3).

For cloud application providers, advantages are simplified and accelerated software code upgrading and testing, better protection of their intellectual property. If it is possible to build cloud applications on the two underlying layers (cloud software environment or cloud software infrastructure) additional advantages may be tapped, such as simplification of cloud apps development, shorter time to market of the apps, and increased reliability through approved APIs. At the downside of this opportunity are limited flexibility of the applications and limited capacities of developers to optimize the performance of the applications (Yousseff et al. 2008: 3).

The challenges of cloud computing that occur at this layer are security and availability issues of the cloud applications, outages, difficulties to integrate legacy applications and the migration of end-users' data to the cloud. Finally, due to the security and availability issues users are often offered service level agreements (SLAs) that reflect rather the interest of providers than those of the users (Yousseff et al. 2008: 4).

- **Cloud software environment layer**

For users (= cloud application developers) the advantages are that they are provided with a adaptive (scalable) platform and APIs to develop and run their applications, automatic scaling and load balancing, integration with other services like authentication and communication services, and accelerated deployment time. (Yousseff et al. 2008: 4).

Cloud software environment providers achieve a large (potentially unlimited) number of users and can thus accelerate and increase the number of applications (i.e. the usability) and innovativeness of the cloud.

- **Cloud software infrastructure layer**

This layer provides computational resources, such as virtual machines, data storage, and communication to the higher level layers of the cloud and allows users to develop new cloud software environments and cloud applications. As common challenges, Yousseff et al. (2008: 6) describe availability and security issues as well as the lack of standard user interfaces.

a) Computational resources

Users gain control and flexibility through ‘super-user access to virtual machines, which results in increased customization, performance and efficiency. Providers’ advantages are protection of their physical infrastructure of their data center (through virtualization) and better occupancy rates through multi-tenancy (Yousseff et al. 2008: 5; Babcock 2010).

Challenges at this layer are again unsatisfactory (for the users) SLAs, which may also affect the quality of SLAs at the higher cloud layers (Yousseff et al. 2008: 5).

b) Data storage

Users can store and access from anytime anywhere scalable and potentially unlimited amounts of data on the cloud (Yousseff et al. 2008: 5), provided that data transfer bottlenecks (Armbrust et al. 2009: 16-17) are solved.

Challenges are provided by the numerous and sometimes conflicting requirements data storage systems have to meet. These ambiguities result in a lack of storage systems that are capable to meet all requirements and a preference by providers of some fe atures over others, which again results in unsatisfactory SLAs (Yousseff et al. 2008: 5).

c) Communication

Yousseff et al. (2008: 5-6) describe communication as a key precondition to meet requirements regarding a guaranteed quality of services but claim at the same time that this model is neither fully developed nor widely implemented in clouds.

- **Software kernel layer**

This layer provides the software management for the cloud’s servers, in form of an OS kernel, hypervisor, virtual machine monitor or clustering middleware. According to Yousseff et al. (2008: 6), there is some need for developing usable ports and interfaces for the cloud at different software layers and for a national infrastructure for utility computing.

- **Hardware and firmware layer**

This layer consists of the actual physical hardware and switches of the cloud. Users at this layer are usually big companies.

A future task will be to allocate all relevant challenges and possible solutions, as illustrated in the following, to the layers of the cloud computing ontology of Yousseff et al.

4.3 New Security and Privacy Risks and Emerging Security Controls

Armbrust et al. (2009: 14-19) have identified ten substantial challenges of cloud computing and developed suggestions to solve these

	Obstacle	Opportunity
1	Availability of Service	Use Multiple Cloud Providers to provide Business Continuity; Use Elasticity to Defend Against DDOS attacks
2	Data Lock-In	Standardize APIs; Make compatible software available to enable Surge Computing
3	Data Confidentiality and Auditability	Deploy Encryption, VLANs, and Firewalls; Accommodate National Laws via Geographical Data Storage
4	Data Transfer Bottlenecks	FedExing Disks; Data Backup/Archival; Lower WAN Router Costs; Higher Bandwidth LAN Switches
5	Performance Unpredictability	Improved Virtual Machine Support; Flash Memory; Gang Scheduling VMs for HPC apps
6	Scalable Storage Invent	Scalable Store
7	Bugs in Large-Scale Distributed Systems	Invent Debugger that relies on Distributed VMs
8	Scaling Quickly Invent	Auto-Scaler that relies on Machine Learning; Snapshots to encourage Cloud Computing Conservationism
9	Reputation Fate Sharing	Offer reputation-guarding services like those for email
10	Software Licensing	Pay-for-use licenses; Bulk use sales

Table 1: Top 10 Obstacles to and Opportunities for Adoption and Growth of Cloud Computing

Source: Armbrust et al. (2009: 14)

Many of the obstacles in Table 1 relate to data processing in general. However, like all new technologies, cloud computing introduces new security risks (CSA 2010) that need to be mitigated:

a) Isolation Breach between Multiple Customers

Cloud environments aim at efficiencies of scale by increased sharing resources between multiple customers. As a consequence, data leakage and service disruptions gain importance and may propagate through such shared resources. An important requirement is that data cannot leak between customers and that malfunction or misbehaviour by one customer must not lead to violations of the service-level agreement of other customers.

Traditional enterprise outsourcing ensures the so-called “multi-tenant isolation” through dedicated infrastructure for each individual customer and data wiping before re-use, whereby resources-sharing and multi-tenant isolation can be implemented on different levels of abstraction. In order to mitigate this risk in a cloud computing environment, multi-tenant isolation ensures customer isolation. One way to implement such isolation is labelling and flow control (Cabuk et al. 2010; Basak et al. 2010, Brassil 2010; Ristenpart et al. 2009).

b) *Insider Attacks by Cloud Administrators*

Accidental or malicious misbehaviour of insiders has increased due to global operations and a focus on low cost. Examples may include a network administrator impacting database operations or administrators stealing and disclosing data. This risk is hard to mitigate since security controls need to strike a balance between the power needed to administrate and the security of the administrated systems.

A practical approach to minimize this risk is to adhere to a least-privilege approach for designing cloud management systems. This means that cloud management systems should provide a fine-grained role hierarchy with clearly defined separation of duty constraints. The goal is to ensure that each administrator only holds minimized privileges to perform the job at hand. While today, operators often have god-like privileges, by implementing a least privilege approach, the following objectives can be met:

- Infrastructure administrators can modify their infrastructure (network, disks, and machines) but can no longer access the stored or transported data.
- Security administrators can design and define policies but cannot play any other roles.
- Customer employees can access their respective data and systems (or parts thereof) but cannot access infrastructure or data owned by different customers.

This so-called privileged identity management system is starting to be implemented today and should be mandated for cloud deployments. n the long run, practical approaches such as privileged identity management may be complemented with stronger protection by, e.g., trusted computing (Santos et al. 2009) or computations on outsourced data (Sadeghi et al. 2010).

c) *Failures of the Cloud Management Systems*

Due to the highly automated nature of the cloud management systems and the high complexity of the managed systems, software quality plays an important role in avoiding disruptions and service outages: Clouds gain efficiency by industrializing the production of IT services through complete end-to-end automation. This means that once errors occur in such complex and automated systems, manual intervention for detecting and fixing faults may lead to even more errors. It is furthermore likely that due to the global scale, errors will be replicated globally and thus can only be fixed through automation.

Another source of failure stems from the fact that large-scale computing clouds are often built using low-cost commodity hardware that fails (relatively) often. This leads to frequent failures of machines that may also include a subset of the management infrastructure.

The consequence of these facts is that automated fault tolerance, problem-determination, and (self-) repair mechanisms will be commonly needed in the cloud environment or recover from software and hardware failures.

For building such resilient systems, important tools are data replication, atomic updates of replicated management data, and integrity checking of all data received (see, e.g., Vukolić 2010). In the longer run, usage of multiple clouds may further improve resiliency (e.g., as pursued by the TClouds project www.tclouds-project.eu or proposed in Guerraoui & Yabandeh 2010).

d) Lack of Transparency and Guarantees

Security incidents are largely invisible to a customer. Data corruption may not be detected for a long time, data leakage by skilled insiders is unlikely to be detected. Furthermore, the operational state and potential problems are usually not communicated to the customer except after an outage has occurred.

An important requirement in a cloud setting is to move away from today's "black-box" approach to cloud computing where customers cannot obtain insight on or evidence of correct cloud operations. A related challenge is how to best foster trust of customers into correct operation of the cloud infrastructure. There are partial solutions to this, such as the so-called best effort approach where operators promise "to do their best" but do not give any guarantees, but a well-accepted best practice is still lacking. An improvement to this approach is third-party audits, i.e. certification of (cloud) service providers by an independent organization. This approach is common best practice today but still only ensures compliance at a point of time and due to its spot-check approach may miss areas of non-compliance that by accident were not checked.

In the mid-term, it is important that cloud providers provide automated interfaces for observation and incident handling (Grobauer & Schreck 2010). This will allow customers to automatically identify incidents and to analyze and react to such incidents. In the long run, the ideal transparency mechanisms would guarantee that processes are implemented such that the agreed upon procedures are followed, the functional and non-functional requirements are met, and no data is corrupted or leaked. In practice, these problems are largely unsolved,, but a practical solution may be provided by using Trusted Computing to verify correct policy enforcement (Chow et al. 2009).

e) Privacy Risks

To enable trusted cloud computing, privacy protection is an essential requirement (Weichert 2009). In simple terms, data privacy aims at protecting personally identifiable data (PID). Since cloud computing often means outsourcing data processing, the user as well as the data subject might face risks of data loss, corruption or wiretapping due to the transfer to an external cloud provider. There are three particular challenges that need to be addressed by all cloud solutions: Transparency, technical and organizational security safeguards and contractual commitments (e.g., Service Level Agreements, Binding Corporate Rules).

Transparency might be technically realized by, e.g., installing informative event and access logs that enable the user to retrace in detail what happens to his data, where they are stored and who accesses them. Legally, the compliance of the cloud service providers with the European law may be ensured by a commitment to Binding Corporate Rules (BCR). Another method is the implementation of Service Level Agreements (SLAs) into the contracts, which guarantee the adherence to the spelled out privacy requirements.

This applies all the more in cases of cross-border cloud computing with various subcontracting cloud service providers. Unlike local data centers residing in a single country, cloud infrastructures often extend over multiple legislation and countries. Therefore, the question of applicable law and safeguarding the users' responsibilities regarding data privacy in cross-border cloud scenarios is a matter of consequences for the use of these cloud services. So to avoid unwanted disclosure of data, sufficient protection mechanisms need to be established. These may also extend to the level of technical solutions, such as encryption, data minimization or enforcement.

5 Cloud Evolution

According to Babcock (2010: 19, 23) , due to the lack of skilled end users and standard interfaces that allow integrating services between different service providers on the same cloud and even between different clouds, the full potential of cloud computing is not at all tapped at current. To him, "…eBay, Gmail, MySpace, GoogleApps, Facebook and Office Live are all just crude early signposts of where cloud computing can take us (Babcock 2010: 23)." Babcock (2010) agrees with Kim (2009) that there is an evolution of cloud computing, largely driven by the demand for availability and security that will result in a growing adoption of hybrid clouds. However, they disagree on the expected trends of prices that users will have to pay when more mature clouds and services exist. While Kim assumes increased prices for advanced clouds and services, Babcock sees a trend towards decreasing prices. He explains this with increased competition and a broader deployment of commodity parts in line with growing numbers of private clouds, which he foresees to integrate with public clouds, thus forming hybrid clouds (Babcock 2010: 69-101).

However, there are a number of technical and organizational obstacles to overcome in order to unfold this potential. Preconditions for this evolution are more skilled users that are capable of creating virtual applications, a wide adoption of open standards and open source software in order to secure interoperability on large scale, experience of firms how to build and manage computing clusters, and the overcoming of current security and privacy threats (Babcock, ibid.).

It appears therefore questionable to follow Mell & Grance (2011) in their claim that at already at current four different deployment clouds can be differentiated. Babcock's insights in the market suggest that private clouds are just starting to emerge – but since they provide the cornerstone for hybrid clouds – and possibly also for the adoption of community clouds – it appears necessary to distinguish rather between clouds that are already business practice and clouds that are rather theoretical concepts than business reality.

6 Cloud Computing Scenarios – Requirements and Constraints

This first overview of conceptualizations and realizations of cloud computing provided meaningful insights and conceptual foundations for the building of significant and realistic cloud computing scenarios. Conclusions to be drawn from the sections above are that:

- A clear distinction should be made between cloud computing as it is already practiced in businesses and cloud computing that is rather at a conceptual state than implemented in commercial environments.
- From the business perspective it makes sense to abandon concepts of cloud computing that highlight either technical or business aspects and to focus, instead, on the interplay between technology, business models and empowered (with programmatic control) cloud users.
- The obstacles for a fast and wide adoption of cloud computing within the economy are identified.
- There are useful ontologies of cloud computing that allow, in principle, to allocate challenges of cloud computing and possible solutions to specific layers and actors within the cloud and within the cloud ecosystem.
- Cloud computing has specific security and privacy challenges and requirements (going beyond those of traditional Internet services and data centers). These depend on the specific class of services being provided. Federated or hybrid cloud models further increase the complexity of these challenges. A first categorization of these challenges has been started in this paper.
- It appears there is a development trajectory for future cloud computing that leads from the currently dominating dichotomy of public clouds (with limited privacy and security) versus private clouds to integrated models like community clouds or public clouds with differentiated security and privacy levels as well as in the longer perspective to complex hybrids and federations of those.

The next steps of the scenario building in WP1.1 “Requirements and Roadmap” will be to secure and maybe differentiate these conclusions through interviews with cloud experts representing all layers and actors in current cloud ecosystems, to determine the interplay of business-related conditions and trends of cloud computing with regulatory (legal, political) factors, and to derive from this additional input first substantial cloud computing scenarios that can be discussed with experts at a scenario building workshop.

Elements of this work – in particular those related to categorizing and defining cloud security and privacy challenges and requirements - have been submitted as chapter to the Future Internet Book 2011.[6]

[6] To appear in FIA Book 2011, Springer Publishing: Glott R., Husmann E., Sadeghi A.-R., Schunter M., “Trustworthy Clouds Underpinning the Future Internet”.

7 References

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M., 2009. Above the Clouds: A Berkeley View of Cloud Computing. Berkeley: UC Berkeley Reliable Adaptive Distributed Systems Laboratory. Available online at: http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf

Babcock, C., 2010. “Management Strategies for the Cloud Revolution. McGraw Hill: New York et al.

Basak, D., Toshniwal, R., Maskalik, S., Sequeira, A., 2010. Virtualizing networking and security in the cloud. SIGOPS Oper. Syst. Rev. 44, 86–94

Buyya, R., Chee, S. Y., Venugopal, S., 2008. Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities. Available online at: http://arxiv.org/ftp/arxiv/papers/0808/0808.3558.pdf

Cabuk, S., Dalton, C.I., Eriksson, K., Kuhlmann, D., Ramasamy, H.V., Ramunno, G., Sadeghi, A.-R., Schunter, M., Stüble, C., 2010. Towards automated security policy enforcement in multi-tenant virtual data centers. J. Comput. Secur. 18, 89–121

Chow, R., Golle, P., Jakobsson, M., Shi, E., Staddon, J., Masuoka, R., Molina, J., 2009. Controlling data in the cloud: outsourcing computation without outsourcing control. In: ACM Workshop on Cloud Computing Security (CCSW’09), pp. 85–90. ACM Press, New York

Cohen, R. 2008. Definition of Cloud Computing. In: Geelan, J. (ed.), Twenty one Experts Define Cloud Computing. Available online at: http://cloudcomputing.sys-con.com/node/-612375/print

Cloud Security Alliance (CSA), 2010. Top threats to cloud computing, version 1.0.Available online at http://www.cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf

Foley, J., 2008. 20 Cloud Computing Startups You Should Know. Available online at http://www.informationweek.com/news/cloud-computing/software/showArticle.jhtml?articleID-=210602537

Grobauer, B., Schreck, T., 2010. Towards incident handling in the cloud: challenges and approaches. In: Proceedings of the 2010 ACM workshop on Cloud computing security workshop, Chicago, Illinois, USA. CCSW ’10, pp. 77–86. ACM Press, New York

Guerraoui, R., Yabandeh, M., 2010. Independent faults in the cloud. In: Proceedings of the 4th International Workshop on Large Scale Distributed Systems and Middleware, Zürich, Switzerland. LADIS ’10, pp. 12–17. ACM Press, New York

Intel, 2010. Intel Cloud Computing Taxonomy and Ecosystem Analysis. IT@Intel Brief, Intel Information Technology. Available online at: http://www.software.intel.com/file/30570

Kaplan, J., 2008. Definition of Cloud Computing. In: Geelan, J. (ed.), Twenty one Experts Define Cloud Computing. Available online at: http://cloudcomputing.sys-con.com/node/-612375/print

Klems, M. 2008. Definition of Cloud Computing. In: Geelan, J. (ed.), Twenty one Experts Define Cloud Computing. Available online at: http://cloudcomputing.sys-con.com/node/-612375/print

Kim, W., 2009. Cloud computing – today and tomorrow. Journal of Object Technology, Volume 8, No. 1, 2009; pp. 65-72. Available online at: http://www.jot.fm/issues/issue_-2009_01/column4.pdf

Knorr, E. & Gruman, G., 2008: What cloud computing really means. Available online at: http://www.infoworld.com/d/cloud-computing/what-cloud-computing-really-means-031?page=0,0

McFedries, P., 2008. The Cloud is the Computer. IEEE Spectrum. Available online at: http://spectrum.ieee.org/computing/hardware/the-cloud-is-the-computer

Mell, P. & Grance, T., 2011. The NIST Definition of Cloud Computing (Draft). U.S. Department of Commerce. Available online at: http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf

Oclassen, G., 2009. Why not cloudsourcing for enterprise app user adoption/ training? Available online at: http://velocitymg.com/explorations/why-not-cloudsourcingfor-enterprise-app-user-adoptiontraining/

Penn, J., 2010. Security and the cloud: Looking at the opportunity beyond the obstacle. Forrester Research (October 2010)

Rajan, S.S, 2010. Cloudsourcing vs outsourcing. Available online at: http://cloud-computing.sys-com.com/node/1611752

Ristenpart, T., Tromer, E., Shacham, H., Savage, S., 2009. Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In: Proceedings of the 16th ACM conference on Computer and communications security, Chicago, Illinois, USA. CCS ’09, pp. 199–212. ACM Press, New York

Sadeghi, A.-R., Schneider, T., Winandy, M., 2010. Token-Based Cloud Computing Secure Outsourcing of Data and Arbitrary Computations with Lower Latency. In: Acquisti, A., Smith, S., Sadeghi, A.-R. (eds.) Proceedings of the 3rd international conference on Trust and trustworthy computing, Berlin, Germany, June 21-23, 2010. LNCS, vol. 6101, pp. 417–429. Springer, Heidelberg

Santos, N., Gummadi, K. P., Rodrigues, R., 2009. Towards Trusted Cloud Computing. Available online at: http://www.mpi-sws.org/~gummadi/papers/trusted_cloud.pdf

Vaquero, L. M., Rodero-Morino, L., Caceres, J., Lindner, M., 2008. A Break in the Clouds: Towards a Cloud Definition. In: ACM SIGGCOMM Computer Communication Review, Vol. 39, No. 1., pp. 50-55.

Vukolíc, M., 2010. The byzantine empire in the intercloud. SIGACT News 41, 105–111

Wang, L. & von Laszewski, G, 2008. Scientific Cloud Computing: Early Definition and Experience. Rochester: Service Oriented Cyberinfrastruture Lab, Rochester Institute of Technology. Available online at: http://cyberaide.googlecode.com/svn/trunk/papers/08-cloud/vonLaszewski-08-cloud.pdf

Yousseff, L., Butrico, M., Da Silva, D., 2008. Toward a Unified Ontology of Cloud Computing. Available online at: http://www.cs.ucsb.edu/~lyouseff/CCOntology/CloudOntology.pdf

Yousseff, L., Butrico, M., Da Silva, D., 2008a (Presentation). Toward a Unified Ontology of Cloud Computing. Available online at: http://www.collab-ogce.org/gce08/images/7/76/Lamia-Youseff.pdf