University of Würzburg
Institute of Computer Science
Research Report Series

A Survey of Cloud Services and Potential Applications of Social Awareness

Michael Seufert[1], Tobias Hoßfeld[1,†], Guilherme Sperb Machado[2], Thomas Bocek[2], Matteo Biancani[3], Paolo Cruschelli[3], Roman Łapacz[4], George Darzanos[5], Ioanna Papafili[5], Krzysztof Wajda[6]

Report No. 496   November 2015

[1] University of Würzburg, Institute of Computer Science, Würzburg, Germany
[†]Now at: University of Duisburg-Essen, Chair of Modeling of Adaptive Systems, Essen, Germany
seufert@informatik.uni-wuerzburg.de, tobias.hossfeld@uni-due.de
[2] University of Zurich, Department of Informatics, Zurich, Switzerland
machado@ifi.uzh.ch, bocek@ifi.uzh.ch
[3] Interoute S.p.A, Rome, Italy
Matteo.Biancani@interoute.com, Paolo.Cruschelli@interoute.com
[4] Poznań Supercomputing and Networking Center, Poznań, Poland
romradz@man.poznan.pl
[5] Athens University of Economics and Business, Department of Informatics, Athens, Greece
ntarzanos@aueb.gr, iopapafi@aueb.gr
[6] AGH University of Science and Technology, Department of Telecommunications, Kraków, Poland
wajda@kt.agh.edu.pl

A Survey of Cloud Services and Potential Applications of Social Awareness

M. Seufert, T.Hoßfeld[†]
University of Würzburg, Institute of Computer Science, Würzburg, Germany
[†]Now at: University of Duisburg-Essen, Chair of Modeling of Adaptive Systems, Essen, Germany
seufert@informatik. uni-wuerzburg.de, tobias.hossfeld@uni-due.de

G. Sperb Machado, T. Bocek
University of Zurich, Department of Informatics, Zurich, Switzerland
machado@ifi.uzh.ch, bocek@ifi.uzh.ch

M. Biancani, P. Cruschelli
Interoute S.p.A, Rome, Italy
Matteo.Biancani@interoute. com, Paolo.Cruschelli@ interoute.com

R. Łapacz
Poznań Supercomputing and Networking Center, Poznań, Poland
romradz@man.poznan.pl

G. Darzanos, I. Papafili
Athens University of Economics and Business, Department of Informatics, Athens, Greece
ntarzanos@aueb.gr, iopapafi@aueb.gr

K. Wajda
AGH University of Science and Technology, Department of Telecommunications, Kraków, Poland
wajda@kt.agh.edu.pl

Abstract

The variety of emerging cloud services and applications manifests in a multitude of network and hardware requirements, but also in different service and traffic characteristics. Due to the popularity of clouds, a set of issues emerges for the different stakeholders involved in providing and delivering cloud services to the end user. Not only pure network layer optimization, but especially the new field of socially-aware traffic management seems promising to overcome these issues. In this paper, the applicability of social awareness to different types of cloud services is discussed. For that purpose, cloud applications are classified according to relevant technical and non-technical characteristics. Based on this novel classification scheme, the benefits and challenges of social awareness are discussed and examples for the optimization of cloud services are given.

Keywords: cloud services, classification, survey, social awareness, traffic management, optimization potential

1

Contents

1 Introduction

Cloud computing, i.e., distributed computing over the Internet, has been gaining more and more popularity in recent years and is about to become the dominating computing paradigm. According to [1], 54% of US businesses already use cloud computing, and it is expected that nearly two-thirds of all workloads will be processed in the cloud by 2016 [2]. Major benefits for businesses using clouds are lower capital expenditure for IT and the ability to scale IT resources. Customers on the other hand can mainly benefit from convenience, flexibility, improved access to information and online content, automatic maintenance and updating, and potentially better security [3]. Despite its increasing popularity, cloud computing still suffers from several inefficiencies related to traffic from, to, within, or in between data centers, e.g., redundant traffic transmission, high latency connections, information asymmetry, or missing or limited abilities to react fast on changing network conditions. As these inefficiencies heavily influence efficient operation of cloud services and good user experience, all involved stakeholders, i.e., cloud and service operators, network operators, and service consumers, are interested in traffic management solutions which help to overcome these issues. However, there are no taxonomy and classification of cloud services in terms of service and traffic characteristics in literature. Understanding the characteristics of services considering its traffic and stakeholders is a fundamental knowledge for deriving and developing appropriate solutions. These solutions are not limited to pure network layer optimization, but recently the new field of socially-aware traffic management attracted attention [4]. Socially-aware traffic management aims at using social information, e.g., from online social networks (OSNs), to improve traffic management decisions in the network and is thus relevant for future network optimization.

This paper aims at providing a survey of currently available cloud systems, focusing on their technical and non-technical characteristics with the respect to the possible utilization of social information. The survey and its conclusions reflect the work and intervention ideas of authors also involved in the FP7 ICT SmartenIT research project[1]. The SmartenIT consortium is representative of a large community of researchers, and network and cloud operators who are deeply involved in advancing and deploying cloud systems by employing social awareness. Thus, the presented discussion of social awareness extends its impact and validity beyond the scope of the single research project and intercepts broader business and technical strategies. To highlight the benefits and challenges of applying social awareness to cloud services and their operation, this paper is organized as follows. In Section 2, we give an overview on current cloud solutions and categorize them. In Section 3, we present technical and non-technical characteristics of cloud services. Then, we introduce social awareness, discuss its applicability, and present examples in Section 4. Section 5 concludes our work.

---

[1]SmartenIT Project, URL: http://www.smartenit.eu/

2 Definition of Clouds, Existing Categories and Examples

Cloud computing is still a difficult term to define, even with vast discussions over the years in industry or academia [5, 6, 7]. However, in order to discuss and propose a new cloud classification based on optimization approaches, it is necessary to clearly present the authors' understanding of what is cloud computing, how clouds are formed, and how to classify cloud services based on the presented definition. Among several definition and classification proposals, e.g., [8, 6, 9], NIST definition [10] is the one that comprises the most complete and differentiated set of enablers, characteristics, service models, and deployment models. Therefore, the NIST's cloud computing definition should be considered throughout this paper: "Cloud computing is 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 is composed of five essential characteristics, three service models, and four deployment models."[10]

Complementing NIST's definition, the following five essential characteristics are also proposed [10]: (1) on-demand self-service, (2) broad network access, (3) resource pooling, (4) rapid-elasticity, and (5) measured service. In (1), any consumer should be able to consume services, anytime that is needed, in a fully automated fashion (no need of human interaction). In (2), services should present a standardized interface to interact with heterogeneous clients, through the network. In (3), cloud provider's resources are organized in pools, in a multi-tenant fashion in order to support multiple users. The resources are dynamically assigned according to customer demand. The geographic and/or logical distributions of these resource pools give the notion of location independence, which is also a key factor to understand the term cloud. In (4), resources should elastically be provisioned and released, scaling up or down according to consumer's demand. This specific characteristic, in combination with the others, provides to the consumer "the illusion of infinite computing resources always available" [8]. In (5), resource utilization should be metered, accounted, controlled, and reported (for consumers). The provider manages and optimizes the use of resources leveraged by the metering capability.

Service models are high-level abstract terms to describe how services are delivered to cloud customers. Within the cloud model, there are three basic service models [10]: Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS). SaaS is the highest level of cloud computing, meaning that users do not need to be concerned on PaaS or IaaS in order to use software that runs on the cloud. SaaS features a complete application offered as a service on demand. An example of this service model is online alternatives to local office applications, like text processors. PaaS is a service that provides the software platform where systems run on. The management of hardware resources demanded by the execution of services is transparent. One of the most well-known examples of PaaS is the Google Apps Engine, since it provides a well-defined API in a specific programming language to build scalable applications. Finally, IaaS is the lowest level of cloud computing and the one that represents the most immediate impact for enterprises [11]. IaaS is centered on computation capabili-

ties, which makes it possible to customize infrastructures from the lowest to the highest level at a low cost. Additionally, other service models were already discussed in literature. Storage-as-a-Service [12], Metal-as-a-Service [2], Network-as-a-Service [13], Policy Management-as-a-Service [14], Testing-as-a-Service [15], and Wireless Sensor Networks-as-a-Service [16] are some of the proposed models making use of resources as utility.

Cloud resources can be consumed or provisioned following four deployment models [10]: private cloud, community cloud, public cloud, or hybrid cloud. In the private cloud model, the cloud infrastructure is dedicated to one single organization. In the community cloud, the cloud infrastructure is dedicated to a set of customers with related matters (e.g., mission, security requirements, policy, and compliance considerations). In the public cloud, the cloud infrastructure is opened to the public in the sense that it can be operated by any organization and also provisioned to any customer. In the hybrid cloud, the cloud infrastructure is composed by a combination of at least two deployment models. For example, an organization can contract a service in a public cloud for non-critical tasks (e.g., testing), and have a private cloud (within its own premises) for the on-production environment. The categorization of private and public cloud is mainly based on open source vs. closed source. While many private cloud solutions are open source applications and can be freely downloaded and installed, public cloud solutions are typically closed source and services are provided by the cloud provider only. There are several exceptions such as WordPress which offers public cloud services while also providing open source software. The categorization of cloud models is based on the abstraction level, however, none of these categorizations consider performance. Moreover, performance characteristics of cloud service models depend on each other. If a SaaS or PaaS runs on top of an IaaS and the performance of IaaS is poor, the performance of SaaS/PaaS will be poor as well. In contrast if IaaS or PaaS perform well, the application running on top will determine the characteristics of the whole system. Moreover, applications have more heterogeneous requirements compared to IaaS and PaaS, which share common service characteristics (such as computation or storage).

Therefore, in order to generalize SaaS applications to find cloud services with common requirements, this paper introduces SaaS application categories. For example, SaaS applications within the "Video/music on demand" require content delivery with lower latency and lower delay, when compared to SaaS applications in the "File storage/sharing" category. Table 1 shows the proposed categories and gives some examples of SaaS applications.

---

[2]https://maas.ubuntu.com

Table 1: SaaS application categories and examples

SaaS Application Categories	SaaS Applications
Video/music on demand	YouTube, Vimeo, Youku, Netflix, Hulu, SoundCloud
Live video/HD TV/music streaming	Justin.tv, Ustream, Telekom Entertain, Spotify, Pandora
Video conferencing/VoIP	Skype, Hangouts, FaceTime, Asterisk, Jitsi, Yuilop
Remote desktop	TeamViewer, PocketCloud, XenDesktop, Windows Remote Desktop Services
Collaboration	WebEx, Adobe Connect, Zimbra, Gmail, Google Docs, RoundCube, WordPress
File storage/sharing	Dropbox, OwnCloud, iCloud, Skydrive, Picasa, Gallery, Rapidshare, Mega
Instant messaging	WhatsApp, WeChat, Viber, ICQ, Threema
Gaming	OnLive, Gaikai, Kalydo, GamingAnywhere
Online social networks	Facebook, Twitter, LinkedIn, Xing, Google+, Tencent Weibo, Diaspora

3 Major Characteristics of Cloud Services

In this section, we identify similarities as well as differences of current and emerging cloud services and group the services into categories. In contrast to the NIST definition [10] of cloud computing, we do not focus on deployment models. We rather consider characteristics that are important for the optimization of cloud services to assign the services to the categories. Therefore, our categories are complementary to the NIST definition and their aim is to facilitate an easy decision which optimization approaches are suitable for which category of cloud services. The methodology we adopted to identify and assess the different cloud characteristics consisted of two different steps. At first we classified the technical and non-technical aspects in order to identify two nearly orthogonal directions. Then we proceeded with a survey among the SmartenIT partners to rank relevance, intervention potentials, and expectations on the different characteristics. Our assessment was based on the research and commercial background of the SmartenIT partners, who represent a consistent European community of renowned network researchers, a global network equipment vendor, and big network and cloud operators. The primary aim of this assessment process was to select the most relevant optimization areas in which new research interventions are needed.

3.1 Technical Characteristics

Within the technical area we identified hardware and network requirements, traffic characteristics, service complexity, delay tolerance, and cacheability. These characteristics will be presented in the following sections in detail.

3.1.1 Hardware Requirements

Cloud services differ very much regarding the utilized hardware resources. The most important components of a cloud service are computing, storage, and networking equipment. These components can be deployed either in a single location or distributed over multiple locations. With potentially up to hundreds of thousands of servers at each location, the total number of servers of a cloud service can range from one to millions. Consequently, cloud service providers must have the ability to move the workload among different servers or different data centers. To account for increased workload (burst service requests, peak times, growing popularity), data backup, redundancy, and failure recovery, either hardware can be over-provisioned (i.e., more hardware is operated) or the service can be deployed in a virtualized environment. This means, virtual resources are allocated to the service when they are needed, and released (for other services) or terminated when they are idle. In this case, the workload has to be moved among the different virtual machines. One of the main issues related to cloud operator hardware is energy consumption, especially within data centers. The operated hardware not only consumes power but also needs to be cooled. Therefore, it is important to maximize the utilization of resources and decrease the energy consumption. In practice, this is achieved, e.g., by smart allocation of workload, consolidation of virtual machines among physical servers, or shutdown of currently unused resources.

3.1.2 Network requirements

Although cloud services are depending on network access by definition, different requirements and characteristics can be seen. With increased adoption of mobile and fixed bandwidth-intensive applications, end user download speed is an important characteristic. This indicator will continue to be critical for the quality of service delivered to both businesses and consumers, especially when large amounts of data have to be transmitted, e.g., in media streaming, content retrieval, software updates, and collaboration services. With the increased adoption of cloud services, upload speeds are especially critical for delivery of content to the cloud over both fixed and mobile networks. The importance of upload speeds will remain increasing over time, promoted by the dominance of cloud computing and data center virtualization, the distribution of large files in virtual file systems, and the demand for consumer cloud game services, collaboration, remote desktop, and backup storage. Delays experienced with voice over IP (VoIP), viewing and uploading videos, online banking on mobile broadband, or viewing hospital records in a healthcare setting, are due to high network latencies (usually reported in milliseconds). Reducing delay in delivering packets to and from the cloud is crucial to delivering today's advanced services (and ensuring a high-quality end-user experience). [2] suggests to distinguish between basic, intermediate, and advanced cloud applications based on their network requirements.

3.1.3 Traffic characteristics

Traffic of cloud services is ever increasing [2] and can be characterized by its volume and its pattern. Volumes are high, for example when processing big data, replicating databases, or moving virtual machines and workload through datacenters. User generated content or video streaming typically account for medium traffic volumes, and lower volumes are seen for example with web applications or collaboration tools. Moreover, the traffic can be produced in regular patterns such as general periodic (e.g., database replication), diurnal (mainly user driven data), or nocturnal (mainly data driven by service provider operations). Due to the rapid growth of cloud services and consequently traffic between data centers (geographical replication, big data transfer), WAN optimization is a key success factor. Thus, a distributed architectural framework should have as main aim to optimize traffic over the underlying transport network.

3.1.4 Service Complexity and Degree of Interactivity

Service complexity aggregates the complexity of a service inside the cloud in terms of resources, geographical distribution, service management (provisioning, failure recovery, customer care), and SLA offered to customers (availability, security). A high complexity is necessary for services such as cloud gaming or video conferencing for which users both upload and download real-time information. Collaboration tools, OSNs, and live video streaming have a medium complexity, as they require a concerted server and network infrastructure. Services such as video/music on demand, file storage/sharing, and instant messaging are easy to deploy and thus have only a low complexity. On the contrary, degree of interactivity covers interactions between the end user and the cloud (human–computer interaction), e.g., to send user commands, and interactions between different end users. Services with high degree of interactivity are for example voice/video conferencing, remote desktop, and cloud gaming. Services such as cloud office products have medium, and, for example, live or on demand video streaming have only a low degree of interactivity.

3.1.5 Delay Tolerance of Data Transmissions

Delay tolerance of data transmission is a category that can be used to distinguish between traffic that must be delivered instantly and traffic for which its transmission can be arbitrarily delayed (the typical maximum delay is about one day). According to [17], applications must be designed with tailored storage facilities in order to implement delay tolerance. Typical cloud services with a high delay tolerance are cloud backup solutions, the delivery of large software updates, or the migration of data between data centers. It can be noted that delay tolerance is directly related to the degree of interactivity, as the more interactive a service, the less delay-tolerant it is. Thus, interactive services such as video conferencing or gaming are especially delay-sensitive and even very small delays of some hundred milliseconds may turn such services unusable. Due to their inherent elasticity to delay, delay-tolerant transmissions can be shifted to off-peak hours when interactive traffc is low. Thereby, increasing the transit costs paid at charged links under

95-percentile pricing is avoided, and negative impacts on the QoS of interactive traffc can be prevented.

3.1.6 Cacheability of Data Transmissions

The last characteristic that we consider here is the cacheability of the data transmissions resulting from cloud services. In general, caching data close to multiple receivers can reduce the network load considerably [18], i.e., it addresses the problem of redundant data transmissions. To this end, so called caches keep a copy of a specific file or data and serve this data to requesting users on behalf of the original source. Therefore, the data has not to traverse all links from the source to the requesting users, but only the ones from the cache to the user, thereby reducing network load and improving QoE of end users. The remaining challenge is to proper tune the caching algorithm in order to optimize the caching strategy keeping the network traffic as low as possible and reducing the number of stored copies of the content. This implies that caching is only effective if the content is static, i.e., it does not change over time, and if the same content is requested by a number of users located in the same area or network at about the same time. This means, a certain spatial and temporal locality needs to exist in order to permit that caching can reduce redundant data transmissions. A high cache-ability can therefore be attributed to video-on-demand services such as YouTube. The service itself and in particular a number of videos enjoy a huge popularity and the content of these videos does not change over time. Similarly, content-distribution networks such as Akamai and file storage/sharing services deliver a large amount of static content, and are very cacheable. In contrast, services such as IPTV, live streaming, or cloud gaming exhibit a very low degree of cacheability since the content is dynamic and changes very fast.

3.2 Non-technical characteristics

In the non-technical area we assessed more social aspects like popularity and target customers. Moreover, we investigated possible directions of optimization like mobility requirements or energy efficiency, and the intervention potential. These characteristics are described in the following sections.

3.2.1 Popularity of Services

Popularity of service is a non-technical dimension that is difficult to define since there is a great number of different cloud services (PaaS, Saas, Iaas) intended for different users (business user, consumers) over different deployment models (private/community/public/hybrid). Additionally, geographical aspects have to be taken into account as a service can be globally popular, or its popularity can be limited to a specific region. The two main key factors to be considered are, first, the number of users of certain service and, second, the amount of cloud traffic related to that service. This gives us the idea that popularity can be defined as a two dimensional category, and a service must be given a high popularity score according to the following method:

- If a service must be ranked from cloud service provider perspective, services with high number of users should receive a high score, giving the possibility to achieve a QoE enhancement for more end-users.

- If a service must be ranked from Internet service provider perspective, services which produce a high amount of cloud traffic should receive a high score, giving the possibility to reduce the network load.

Obviously services who score highest in both direction, i.e., both have a high number of users and produce a high amount of traffic, should be considered first when developing or applying an optimization solution. Finally, we have to note that an overall popularity index is time-varying. Cloud services in general are increasing their popularity, but over the years some services are growing in both aspects of traffic and users (e.g., Facebook, Google Apps), others see a decline or they shift towards private deployment models with service customized for big IT companies (which limits the potential for intervention).

3.2.2 Target Customers

Cloud services can be categorized on the basis of the target customer. Two main categories of customers can be outlined: business customers and consumers. Business customers rely on cloud service mainly to reduce their IT infrastructure cost in terms of computing capacity and storage, as well as in terms of software licences. They count on rapidity, on elasticity, and on demand provisioning and de-provisioning features offered by cloud operators. Cloud services (mainly IaaS and PaaS services) are offered with enhanced SLAs, and can even be offered to big companies with proper customization such as a deployment over a private model to account for security and traffic isolation concerns. Consumers, on the other hand, purchase cloud services for entertainment purposes (video/music on demand/streaming, gaming), for communication purposes (collaboration, instant messaging/conferencing) and for content sharing purposes (file storage/sharing, OSNs). Cloud services (mainly SaaS) are always offered to consumers as standard products with no customization and are deployed mainly over a public cloud model.

3.2.3 Inter-Cloud Communication

A general definition of inter-cloud communication can be summarized by addressing all cross-domain traffic due to the interaction between different cloud operators. For that purpose interconnected clouds (cloud connecting together but maintaining separate administrative domains) and federated clouds (cloud connecting together in a super-cloud entity for resource sharing purposes) are distinguished. Traffic related to inter-cloud communication is a critical key issue since it is highly intensive with respect to the usage of Internet service provider infrastructure, thus, imposing high costs on cloud operators. Moreover, cloud services which generate inter-cloud traffic are subject to degradations of QoE because of possibly increased distance from the content location to its destination. The typical inter-cloud services are storage and computation resources

services offered via the IaaS paradigm over a federated/interconnected cloud model. A great amount of inter-cloud traffic is generated by typical operations of geographically distributed data centers in terms of virtual machine migration, and database replica and synchronization. Thus, this traffic is characterized by high volumes and some periodic patterns. Additional inter-cloud traffic is created by the transmission of so called "big data" which is generated by business custumers, as well as science and research datacenters. Cloud applications and services which are ranked high with respect to inter-cloud communication are highly interesting for optimization because introducing a traffic management solution could reduce network loads and enhance the QoE perceived by end users.

3.2.4 Global Service Mobility

Cloud services are consumed on a large variety of devices ranging from workstations and desktop computers to laptops, tablet PCs, and smart phones. Each of these devices has particular restrictions, e.g., screen resolution, and requirements such as power consumption. As a consequence, the special characteristics of the user devices play an important role when optimizing the corresponding cloud services. For example, mobile devices as lightweight nodes should only visualize the results processed in the cloud, monitor and control the resources. However, some cloud services, e.g., the replication of data between data centers, do not even directly involve any user device. Therefore, to keep things simple, we distinguish only between services used in mobile scenarios (this also includes that users want to access a service from different locations), services not used in mobile scenarios and a third intermediate group of services which are partially used in mobile scenarios. A large number of cloud services are very frequently accessed from mobile devices or different locations. Concrete examples range from video/music on demand, live video/HD TV/music streaming, OSNs, and instant messaging solutions. Additionally, file storage/sharing as provided by Dropbox or iCloud becomes more and more popular on both fixed and mobile devices, and is often used to synchronize between different devices. In contrast, business applications such as scientific computing, simulations and numerical calculations, or collaboration services, are mainly accessed from the same location and only rarely involve mobile devices.

3.2.5 Exploitability of Social Networks Information

Social networks contain huge amounts of information about their users, for example their geographic location, their relations, and the popularity of content. Recent approaches of network and traffic optimization like SocialTube [19] and TailGate [20], which will be detailed in Section 4, try to leverage it. In this section, we group the services according to their suitability to benefit from those approaches, i.e., whether information from social networks can be used for that purpose. However, open issues which still remain are: which information is publically available, and which information would have to be revealed by an operator or by the users of a social network. OSNs like Facebook and Google+ contain personal data of millions of users. Many OSNs view this data as a

valuable asset that is at the core of their business model. The idea of social awareness exploitation is that higher similarity in the interests/preferences of online social group members favors collaborative, and even altruistic, behavior in content replication and content dissemination scenarios, thus optimizing the underlying transport network. The use of social network information has also some limitations:

- The size of social network that can bring valuable information may be limited. Dunbar's number [21] is a suggested cognitive limit to the number of people with whom one can maintain stable social relationships. This limits the usability of information on social awareness. Typical values are 150-300.

- Both OSN users and OSNs have strong incentives to restrict large scale crawls of this data. OSN users want to protect their privacy and OSNs their business interest.

Both aforementioned approaches aim at the optimization of video on demand services. Therefore, we conclude that the exploitability of social networks information is very high for this type of services. A similar example is sharing of files and folders via services like Dropbox as friendship relations could be a good indicator of who is likely to share data with whom. In addition, information about the popularity of content can be used to select appropriate objects for caching. In contrast, business specific applications like remote desktop or collaboration services are likely to be optimized for specific purposes and scenarios anyway, and the available information inside a company might already be more detailed for the concrete purpose than the one in social networks. Therefore, business applications are expected to benefit less from the exploitation of social network information.

3.2.6 Energy Efficiency

Cloud computing offers efficient mechanisms to manage energy consumption. Section 3.1.1 already mentioned about maximized utilization of resources that can be achieved by smart allocation of workload, consolidation of virtual machines among physical servers, or shutdown of currently unused resources. The dynamic nature of clouds allows that the operations are executed in selected time and location, with suitable resources and without human interactions. Decisions about the time and location may include the energy efficiency aspect. An example of a powerful solution for optimization inside the data center is live migration of virtual machines. Cloud resources and their parameters are monitored such that the management entity of the data center infrastructure knows where and when the operations of cloud services should be running to optimize energy consumption. Any time, if needed, a virtual machine can be moved to another server within the cloud infrastructure. Another example from ISP point of view related to energy efficiency is an intelligent caching or pre-fetching of content. If a content is close to the user, less resources (e.g., network links) are utilized. However, a crucial requirement is the correct prediction of users' interest in the type of content. Because mobile devices are used ubiquitously nowadays, the limitations of energy power

on the client side have to be taken into account. An economic energy consumption on the client side will improve the QoE of end users. As energy consumption and its costs are growing fast, optimization in this field is important and should be considered for the classification of cloud services. To simplify the classification we define three groups: the first contains services with a high potential for energy efficiency both in the cloud, in the network, and on the end user device. The second group is composed of services which can only partially reduce their energy consuption, and the last group contains services with low potential for energy efficiency. We consider almost all applications as having a high potential, e.g., by switching from client-server to a peer-to-peer model like video conferencing or instant messaging, or by utilizing caching and delay tolerance. However, some services like media on demand/streaming, gaming, and OSNs have a much lower potential as they are not delay tolerant nor cacheable, or have high hardware and infrastructure requirements.

3.2.7 Intervention Potential

Cloud infrastructure and cloud services may offer the interfaces that could be used by external entities to fetch a set of information, execute some operations, or even manage the resources. Such interfaces are useful to optimize default operations or add new functionalities. Due to security and business models this kind of openness is usually present only in public clouds or clouds established by some communities, e.g., for research purposes. Moreover, it is limited to fetching a set of information or other similar basic functionalities. Popular cloud services (like Dropbox, Gmail, etc.) do not provide information about underlying infrastructures and topologies. Only some statistics or basic APIs are offered, which are little useful for optimization efforts. In case access to the network information and management functionalities is needed, the intervention potential is also lower because ISPs are not willing to disclose the information about network topology and measurements. Additionally, if a cloud covers more than one network domain, access to network related information is even more difficult to obtain because ISPs usually do not share internal details. We consider services to have a high intervention potential if the service is open source and can be influenced by the optimization approaches. Moreover, there are services with medium intervention potential to which non intrusive optimization can be applied, i.e., optimization without changing the service, and services with low intervention potential.

3.3 Classification of Cloud Services and Applications

Two types of characteristics, technical and non-technical, were proposed to be used for classification of cloud services. The main goal of classification described in this document is to collect a set of information required to select the most promising cloud services for optimization and suitable solutions. The list of characteristics helps to understand how a service works and how much potential for improvements there is. The proposed classification is focused on SaaS – not on IaaS or PaaS – since characteristics are application-specific. With IaaS and PaaS the cloud customer has the the freedom

Table 2: SaaS characteristics

	Hardware requirements	Network requirements	Traffic volume	Degree of interactivity	Service complexity	Delay-tolerance	Cache-ability
Video/music on demand	High	High	High	Low	Low	Medium	High
Live video/HD TV/music streaming	High	High	High	Low	Medium	Low	Low
Video conferencing/VoIP	Medium	High	High	High	High	Low	Low
Remote desktop	Low	Medium	Medium	High	Medium	Low	Low
Collaboration	Medium	Low	Low	Medium	Medium	Medium	Low
File storage/sharing	Medium	Low	Medium	Low	Low	High	High
Instant messaging	Low	Low	Low	Medium	Low	Medium	Low
Gaming	High	High	High	High	High	Low	Low
Online social networks	Medium	Low	Medium	Medium	Medium	High	Low

	Popularity	Target customers	Degree of inter-cloud communication	Degree of global service mobility	Potential for social awareness	Potential for energy efficiency	Intervention potential
Video/music on demand	High	Consumer	High	High	High	Medium	High
Live video/HD TV/music streaming	High	Consumer	Medium	High	High	Medium	High
Video conferencing/VoIP	Medium	Business/ con- sumer	Low	Low	Low	High	Medium
Remote desktop	Low	Business	Low	Medium	Low	High	Medium
Collaboration	Low	Business	Low	Low	Low	High	Low
File storage/sharing	High	Consumer	High	High	High	High	High
Instant messaging	Medium	Consumer	Low	High	Low	High	Low
Gaming	Medium	Consumer	Medium	Medium	Medium	Low	Medium
Online social networks	High	Consumer	Medium	High	High	Medium	Medium

to build up an own infrastructure solution with an own architectural design and own optimization strategies. However, the design and the resulting performance of the whole system is depending on the software running on top. Therefore we will now only focus on the classification of SaaS applications whose characteristics are shown in Table 2.

We qualitatively evaluated 9 major categories of overlay applications (cf. Table 1), considering both their technical and non-technical characteristics presented in Section 3.1

and Section 3.2, respectively. Concerning the technical characteristics of the SaaS services examined, we have found that applications employing high quality video, either video on demand, live streaming, or even gaming, have high requirements in terms of hardware. Applications requiring inter-connection of geographically distributed nodes, e.g., for real-time collaboration or to perform quick synchronization of files, are considered to have medium hardware requirements. Low quality video conferencing applications, as well as VoIP and instant messaging applications have the lowest requirements in terms of hardware of all overlay applications. Additionally, applications generating high traffic volumes, e.g., video on demand and real-time applications, or applications that combine both characteristics, have strict requirements in terms of network conditions. Regarding generated traffic volumes, video-related applications generate high volumes of traffic. Online storage, remote desktop, and OSNs are assumed to generate medium traffic volumes, while the remaining applications, e.g., VoIP or instant messaging, produce low traffic volumes. Note that especially for online storage and OSNs traffic is expected to significantly increase in the near future [2].

Applications requiring real-time interaction of geographically distributed nodes have the highest system complexity, while live video streaming or OSNs seem to have medium complexity. Applications such as video on demand, online storage and instant messaging are rather simple to deploy. Furthermore, real-time and video applications such as live streaming, video conferencing, or gaming, are clearly not delay-tolerant. Non-real-time applications such as file storage/sharing and OSNs are considered to be highly delay-tolerant. Finally, video on demand and file storage are considered to be highly cacheable, e.g., their data can be handled by a caching approach, because of their static content, while traffic generated by the remaining ones cannot be easily addressed by a caching solution.

Concerning non-technical characteristics, all video-related applications, file storage/sharing, and OSNs are found to be highly popular, followed by medium popular video conferencing/VoIP and gaming, and the less popular collaboration and remote desktop applications. The popularity of these applications is directly related to their target group, i.e., applications of high and medium popularity address mainly consumers, while applications of lower popularity address business customers. With inter-cloud communication, global service mobility, social awareness, and energy efficiency, we investigated different scenarios from which optimization could emerge. Combined with the potential for intervention we found that video/music on demand and file storage/sharing seem to be the most promising services. Within all services we recognize not only popular but also some emerging applications. The reason for this is the fact that they can constitute nice examples for evaluating traffic management mechanisms where intervention potential is higher than other well established proprietary applications where intervention may not be applicable.

Summarizing the investigation conducted in this chapter, we have provided a description of the aforementioned applications from multiple points of view, in terms of their hardware, network and system requirements, as well as popularity, relevance to significant terms and approaches, such as energy efficiency, and their intervention potential. The identified characteristics will play fundamental role and are provided as input to

the next section of this document in order to identify and discuss potential applications of social awareness for the improvement of cloud services.

4 Social Awareness

Users voluntarily publish lots of information about themselves, their interests, their friends and their actions in online social networks (OSNs), especially about current situations or exceptional events. Such social signals, however, can not only be collected from OSNs (e.g., friendships, interests, trust-relevant metadata), but also from cloud services (e.g., interactions and usage data) and sensors (e.g., location). Social awareness harvests these ubiquitous signals, extracts useful and re-usable information (e.g., users' social relationships, activity patterns, and interests), and utilizes them in order to improve the quality of a cloud service.

Socially-aware traffic management is a new research field which aims to exploit available social information about Internet users in order to enhance existing traffic management strategies. Solutions like prefetching and caching can yield a win-win situation for both, the end user and the Internet service provider. On the one hand, the ISP can better utilize its resources when it is known from social information where, when, and what traffic volumes will be generated. On the other hand, the user benefits from shorter delivery times which generally improves quality of experience (QoE) of Internet services.

In this section, social awareness is defined and its applicability for cloud services is discussed. Therefore, benefits and challenges that arise from social awareness for different cloud services are presented as well as monitoring related issues are covered. Finally, examples are given how socially-aware traffic management can improve online storage and video streming services.

4.1 Definition and Terminology

Social signals are any signals which are emitted in the Internet by interactions of an end user of an Internet application. A signal itself does not contain any information, but information can be generated out of them when interpreted in the right context. Examples of social signals range from simple logins to a cloud service to complex service requests which might include interactions with other users or the environment. In the context of online social networks, these signals are, e.g., friendship requests and confirmations, indications of interest or liking, or postings about activities. Another example are location data which are created by sensors of mobile devices and are communicated when using an Internet service.

Social information are insights about certain users or relationships between users which can be derived when bringing social signals into an appropriate context. The generated information depends on the particular evaluation of the social signals, and might require additional (external) information in order to create new social information. Usually, such partial information that requires external information to produce new social information is called meta information.

Social information providers gather social information and make them available. Thus, they are the sources of social information and their goal is to benefit from the provision, e.g., in terms of money, additional information, or improved service quality.

Many social information is available from online social networks about their users, but access might be limited. Moreover, Internet service providers or cloud service providers can collect and provide information about their customers. Finally, end users themselves can publish information about themselves to other stakeholders, and thus can also be considered social information providers.

In general, the term **social awareness** comprises the utilization of social information for a specific purpose. In the context of cloud services, we will consider social awareness to be the utilization of social information to improve a cloud service. Social awareness can include the monitoring of social signals and production of social information but also a collaboration with a social information provider. Taking provided or generated social information as an input, social awareness will utilize this information in order to improve QoE of end users and/or to deliver the cloud service more efficiently.

4.2 Benefits and Challenges

The exploitation of social awareness information can be performed in many different areas, from marketing to traffic engineering. In this paper we will concentrate on the latter. Social information can potentially provide interesting answers to the following question: "Who requests a content?", "Where is the content located and where and when should it be transferred?". Such answers are very relevant from the point of view of network providers since every time content is moved across the network the generated traffic has a well-defined cost. More in detail, the exploitation of social awareness can generate information that can be used to predict where content will be requested. Such predictions can be used by ISPs (or CDNs) for caching and prefetching purposes. Moreover, those social information can be used for traffic engineering purposes, e.g., to choose the optimal destination for the content placement.

In such a framework, the obvious benefits of exploiting social awareness are two-fold and can be classified in benefits from the user point of view and benefits from the operator (cloud service providers, content providers, ISPs) point of view.

From the user point of view, the obvious benefit is the potential enhancement of the Quality of Experience related to the requested content. Imagine a scenario in which a user requests content and the cloud service provider serving its request has predicted the request by leveraging social information related to the social groups to which the user belongs, and has already placed the content in an optimal location. As a result, the user will enjoy a higher QoE due to the immediate content availability.

On the other hand, operator benefits can be identified in a better control of content placement across the network (potentially resulting in lower inter-domain traffic with accordingly lower cost) and a better market positioning (the users are more satisfied with the provider). Furthermore, the exploitation of social information can enable a more dynamic response in case of flash crowds, i.e., events that are correlated with sudden increase in resources utilization. If a cloud service provider is capable of predicting such events, he is able to withstand the peak requests by dynamically provisioning the popular services. Moreover, social awareness can be employed in order to achieve energy efficiency for cloud operators and ISPs. For instance, energy consumption in the

network takes a significant proportion of the total power consumption for cloud storage services. However, if files that are stored online are accessed, the energy consumption is higher than accessing the file from the local disk due to network power consumption. Thus, social information could be employed to predict the future access to specific files, i.e., both high-popular ones and low-popular, such as user-generated content. Then, prefetching and caching of most likely accessed files to the storage of a user's device could significantly reduce energy consumption compared to frequent online requests.

The two approaches (which represent exactly the two scenarios relevant for SmartenIT) have similar goals, but differ in the granularity of the social information to be analyzed. This means a finer granularity tailored on user activities and groups is needed for the end-user scenario, and a coarser granularity for the operator scenario.

The exploitation of social information has a certain number of barriers and challenges from technical, legal, and business perspective:

1. *Availability of social information.* Social information providers could not willingly disclose social information related to users and groups to protect the privacy of their affiliates. As an alternative, proper 'social crawler' software must be deployed by operators, taking into account that such solution could lead to incomplete datasets and to non-optimal prediction, which could result in the deleterious effect of increasing the inter-domain traffic and lowering the QoE for the end user.

2. *Analysis of huge amount of information.* We might assume that social information is perfectly known without any security or privacy concerns. As detailed in Section 4.1, the raw information extracted by social information providers must be interpreted and correlated with different information sets coming from different providers and from different sources (network information, geographical information, etc.). This yields a huge amount of data that should be processed in real-time and results should be made available to different points of presence of cloud service providers.

3. *Moving towards new cloud-based applications.* The so-called "big data", i.e., an umbrella term for the explosion and diversity of high frequency digital data generated by cloud applications, such as mobile banking transactions, tweets, or online storage, will have unpredictable patterns and cross many different domains (from time to time) due to the mobility of cloud users or the distribution of the cloud resources. Thus, the synchronization of data between multiple data centers providing the same application causes large amounts of traffic on intra and inter-domain links of ISPs.

4. *Achievement of global service mobility.* Due to the rapidly increasing number of mobile devices, it is foreseen that a service should "follow" its receiver in the future. For example, a user uses a service on a notebook or tablet and wants to receive high QoE regardless of his location, e.g., at home, in the train, at the airport, or in a hotel. To enable access to the service with the same level of QoE anywhere in the world, a cloud service provider may employ social information to predict not only when but also where the service will be requested (within its footage).

The exploitation of social information and its benefits may enable a cloud or data center operator, a cloud service provider, or a CDN provider to make decisions aiming to optimize their own operation. Some approaches already exist which show that this novel concept works. For instance, TailGate [20] collects data on social relationships and behaviors, in order to distribute long-tail content among geographically separated participants in off-peak hours. This approach demonstrates significant benefits for the data center operators, e.g., in terms of low energy cost due to operation in off-peak hours, and the cloud service providers, e.g., resulting in improved performance due to content placement/prefetching techniques. More examples are discussed in Section 4.3.

4.3 Examples

In this section, we focus on two examples of cloud-based applications: i) online storage, and ii) content/video delivery over an OSN, which, according to our SaaS classification in Table 2, expose the following characteristics:

First, both of them contribute significantly to inter-cloud communication due to distribution purposes and replication/content placement activities, they both express a similar requirement for global service mobility, e.g., service provided seamlessly regardless the space or time that it is requested, and both are anticipated to provide a high potential for intervention. Additionally, video streaming, e.g., YouTube, is currently the most popular application; it generates huge traffic volumes and it requires high capacity (i.e., bandwidth) to achieve adequate QoS, while it is forecast to continue generating high volumes in the future. On the other hand, online storage applications, such as Dropbox or Google Drive, are becoming increasingly popular, and are expected to be among the highest traffic contributors in near future. Moreover, online storage, falling under the file storage category, generates mainly delay-tolerant traffic, which is therefore cacheable. Note that video streaming can also be delay-tolerant and cacheable, especially in the case of on-demand services.

Thus, information about social relationships and social activity patterns can be utilized to predict spatial and temporal (i.e., *where* and *when*) requests for some files (in the case of online storage) or video items (in the case of video streaming). This prediction capability can be considered when employing optimization techniques such as content/file placement (push) or prefetching (pull). In the following, we discuss how social awareness can be employed to improve the efficiency of the aforementioned applications.

4.3.1 Online Storage

The online storage use-case corresponds to the "File storage/sharing" SaaS application category (cf. Table 1). Based on the characteristics we defined in Section 3 for this application category, file storage/sharing applications show high cache-ability and delay-tolerance. Also, the applications of this type are highly popular and have high potential for social awareness and energy efficiency. Finally, online storage is characterized by high inter-cloud communication and global service mobility. In this section, we focus on online storage and the exploitation of social awareness to improve its efficiency. In particular, we

consider the utilization of social information derived from OSNs to improve the internal decision making algorithms in advanced distributed hierarchical storage management systems.

Hierarchical Storage Management (HSM) is an approach that aims to efficiently handle large volumes of data, i.e., the categorization of data and the decision-making on where to move data between storage types to reduce the energy consumption and operating cost of data storage management. Specifically, in the context of HSM, a hierarchy level is assigned to a storage media. Usually, three levels of storage hierarchy are defined. The first hierarchy level is represented by high-speed, high-cost devices, such as hard disk drive arrays, destined for data sets that are frequently accessed (e.g., by applications or users), whereas oher data, e.g., older and thus less popular, can be automatically moved to a slower, low-cost storage media. The second one is slower, such as optical storage, and the last one, i.e., the slowest, may be implemented as magnetic tape drives. As the technology of the first level is the most expensive, the size of it is smaller than the storage sizes of other levels; thus, there is a trade-off between the speed and the size of the storage type (level). Figure 1 depicts the three storage levels including performance and cost trends in each level.
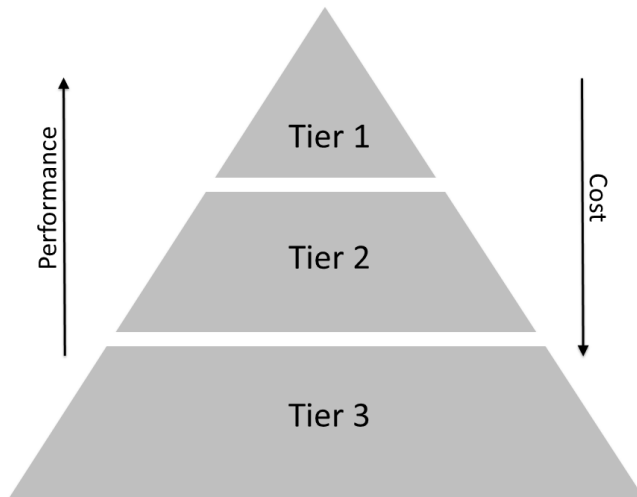

Figure 1: Storage tiers in HSM along with performance and cost trends. (Source: inspired by [22])

Nowadays, when the amount of data is rapidly growing, HSM offers a substantial benefit from managing storage devices efficiently, especially in large-scale networks, storage and computational environments, such as clouds. In particular, a common deployment scenario involves resources of a cloud residing in remote geographical locations, while end users perceive its resources as a consistent pool available for allocation (e.g., IaaS model [10]). Moreover, a cloud operator may utilize storage resources or assign specific works to another cloud operator, e.g., in the context of a cloud federation, to achieve load balancing, reduction of his individual energy consumption, etc. In order to optimize

these operations of data migration, social signals can be exploited by the cloud operators to predict not only the amount but also the where and when of future demand. As a result, end users will experience better QoE, i.e., faster access to data, while the cloud operators will achieve more accurate utilization of their storage hierarchies (tiers), and thus, in consequence lower energy consumption and operating cost.

4.3.2 Video Streaming

Video streaming, is related to the SaaS application categories of "Video/music on demand" and "Live video/HD TV/music streaming". Video streaming applications are very popular and usually require high level of inter-cloud communication and global service mobility. Also, this type of applications show potential for social information exploitation and intervention. Finally, the volume of traffic created from videos streaming applications is high, thus requirements in hardware and network components are high. In this section, we consider a use-case inspired by the evaluation scenario described in [20]. Specifically, we consider an OSN having users around the globe. The OSN users share videos via the OSN, which are stored in a third-party-owned online video streaming platform (e.g., YouTube). This content can be viewed by their online friends, their friends' friends, etc.

In order to not only meet the content demand by users of the video streaming platform but also provide them with high QoE, the video streaming platform comprises multiple Points-of-Presence (PoPs) distributed globally. These PoPs are connected to each other by links, which can be either owned by the video platform, or leased from network providers. Then, each user is assigned to and served out of his geographically nearest PoP for all of his requests as depicted in Figure 2.
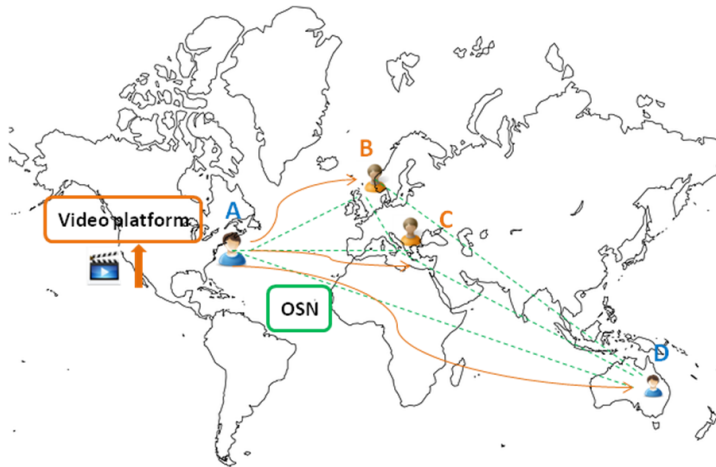

Figure 2: Content delivery in geographically distributed PoPs.

Placing data close to the users is an approach followed by most content delivery networks (CDN). Therefore, all content uploaded by a user A is first uploaded to the nearest PoP, i.e., PoP_A. When content is requested by another user B, the nearest

PoP to B, i.e., PoP_B, is contacted and if the content is available there, the request is served. The content can be already present at PoP_B, if it was initially uploaded there or was brought there by an earlier request. If the content is not available in PoP_B, then a request is made to PoP_A and the content is brought to PoP_B. Though, the latter operation increases the service time, and thus, may deteriorate the users' QoE which is an undesirable effect.

In such as setup, social information, such as the social graph or the type of social relationships between the users of the OSN, can be employed to predict the space and time of the next request of a piece of content, i.e., in which PoP. For instance, it can be anticipated due to their social relationships that an OSN friend of user A will request to view the piece of content that A uploaded with higher probability than users who have no social relationship with A. Such predictions can be utilized by socially-aware mechanisms such as TailGate proposed in [20], which will employ pre-fetching of the uploaded video to the relevant PoPs.

5 Conclusion

Cloud services and applications are currently receiving increased attention for several reasons. Industry and business companies are attracted by the ease of use in terms of deployment, administration and maintenance, and high scalability and flexibility to create new services. From the end user's point of view, the usability, the reliability, and the quality of service compile the success story of cloud services. In particular, the end user is able to consume the service typically anytime and anywhere on any device which only needs a broadband Internet connection. However, there are several challenges to overcome in order to operate and deliver such a service with the quality that is expected by end users. The service quality will become an important differentiator between providers, and QoE as perceived by the end users has the potential to become the guiding paradigm for managing quality in the cloud. Nevertheless, the multitude of similar cloud services and applications migrated to the cloud makes prices decrease, and competition between providers increase. These trends create conflicting, challenging demands on the network operators and service providers involved: on the one hand, they need to develop and offer sophisticated high-performance infrastructures and services that enable high quality experiences that lead to customer satisfaction and loyalty. On the other hand, they have to operate on a profitable basis in order to remain successful in the long run. Thus, the optimization of cloud services with respect to QoE, traffic reduction, and energy savings is of major interest.

Socially-aware traffic management is a promising research field which aims at improving cloud services, their operation, and delivery. The major contribution of this article is a survey of cloud services and an introduction to social awareness in the context of different types of cloud services. For that purpose, cloud applications are classified according to relevant technical and non-technical characteristics. Based on this novel classification scheme, the applicability of social awareness is discussed. We introduced social awareness and its terminology and discussed expected benefits and challenges for cloud services. Moreover, we elaborated on the required monitoring of social information and presented examples for existing socially-aware traffic management solutions and cloud services. It could be seen that promising results are already yielded, especially for service categories like video/music on demand and file storage/sharing, which showed to have a high optimization potential and can benefit from employing social awareness.

Acknowledgements

References

[1] Neovise, Use of Public, Private and Hybrid Cloud Computing, Tech. rep., Neovise (2013).

[2] Cisco, Cisco Global Cloud Index: Forecast and Methodology, 2011-2016, Tech. rep., Cisco (2012).

[3] Civic Consulting, Cloud Computing, Tech. rep., European Parliament's Committee on Internal Market and Consumer Protection (2012).

[4] M. Seufert, G. Darzanos, I. Papafili, R. Łapacz, V. Burger, T. Hoßfeld, Socially-Aware Traffic Management, in: Socioinformatics - The Social Impact of Interactions between Humans and IT, Springer International Publishing, 2014, pp. 25–43.

[5] Sun Microsystems Inc., Introduction to Cloud Computing Architecture, Tech. rep., Sun Microsystems Inc. (2013). URL https://java.net/jira/secure/attachment/29265/CloudComputing.pdf

[6] D. Milojicic, Cloud Computing: Interview with Russ Daniels and Franco Travostino, Internet Computing, IEEE 12 (5) (2008) 7–9.

[7] I. Foster, Y. Zhao, I. Raicu, S. Lu, Cloud Computing and Grid Computing 360-Degree Compared, in: Grid Computing Environments Workshop, 2008.

[8] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, M. Zaharia, Above the Clouds: A Berkeley View of Cloud Computing, Tech. Rep. UCB/EECS-2009-28, EECS Department, University of California, Berkeley (2009). URL http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html

[9] A. Lenk, M. Klems, J. Nimis, S. Tai, T. Sandholm, What's inside the Cloud? An architectural map of the Cloud landscape, in: ICSE Workshop on Software Engineering Challenges of Cloud Computing, 2009.

[10] P. Mell, T. Grance, The NIST Definition of Cloud Computing, Tech. rep., Recommendations of the National Institute of Standards and Technology (2011).

[11] F. Gillet, Conventional wisdom is wrong about IaaS, Tech. rep., Forrester (2009).

[12] C. Wang, Q. Wang, K. Ren, N. Cao, W. Lou, Toward Secure and Dependable Storage Services in Cloud Computing, IEEE Transactions on Services Computing 5 (2) (2012), pp. 220–232.

[13] J. Gabrielsson, O. Hubertsson, I. Más, R. Skog, Cloud computing in telecommunications, Tech. rep., Ericsson (2010).

[14] H. Takabi, J. Joshi, Policy Management as a Service: An Approach to Manage Policy Heterogeneity in Cloud Computing Environment, in: Hawaii International Conference on System Science, 2012.

[15] L. Yu, W.-T. Tsai, X. Chen, L. Liu, Y. Zhao, L. Tang, W. Zhao, Testing as a Service over Cloud, in: IEEE International Symposium on Service Oriented System Engineering, 2010.

[16] F. Delicato, P. Pires, L. Pirmez, T. Batista, Wireless Sensor Networks as a Service, in: Engineering of Computer Based Systems, 2010.

[17] N. Laoutaris, M. Sirivianos, X. Yang, P. Rodriguez, Inter-datacenter Bulk Transfers with NetStitcher, SIGCOMM Computer Communication Review 41 (4) (2011), pp. 74–85.

[18] L. Fan, P. Cao, J. Almeida, A. Z. Broder, Summary cache: a scalable wide-area web cache sharing protocol, IEEE/ACM Transactions on Networking 8 (3) (2000), pp. 281–293.

[19] Z. Li, H. Shen, H. Wang, G. Liu, J. Li, SocialTube: P2P-assisted Video Sharing in Online Social Networks, in: IEEE INFOCOM, 2012.

[20] S. Traverso, K. Huguenin, I. Triestan, V. Erramilli, N. Laoutaris, K. Papagiannaki, TailGate: Handling Long-Tail Content with a Little Help from Friends, in: International Conference on World Wide Web, 2012.

[21] R. I. M. Dunbar, Neocortex size as a constraint on group size in primates, Journal of Human Evolution 22 (6) (1992), pp. 469–493.

[22] B. Ganley, Optimize the Virtual Desktop Experience Through Strong Back-end Design, Tech. rep., Dell Power Solutions (2013). URL http://i.dell.com/sites/doccontent/business/solutions/power/en\/Documents/ps4q13-20130371-ganley.pdf