Socially-aware Management of New Overlay Applications Traffic - The Optimization Potentials of the SmartenIT Approach

Krzysztof Wajda[1], Rafał Stankiewicz[1], Zbigniew Duliński[1], Tobias Hoßfeld[2], Michael Seufert[2], David Hausheer[3], Matthias Wichtlhuber[3], Ioanna Papafili[4], Manos Dramitinos[4], Paolo Cruschelli[5], Sergios Soursos[6], Roman Łapacz[7], and Burkhard Stiller[8]

[1] AGH University of Science and Technology, al. Mickiewicza 30 30-059 Kraków, Poland,
wajda@kt.agh.edu.pl,
SmartenIT home page: http://www.smartenit.eu
[2] University of Würzburg, Germany,
[3] TU Darmstadt, Germany
[4] Athens University of Economics and Business, Greece
[5] Interoute, Italy
[6] Intracom SA Telecom Solutions, Greece
[7] Poznan Supercomputing and Networking Center, Poland
[8] University of Zürich, Switzerland

Abstract. Today's overlay-based mobile cloud applications determine a challenge to operators and cloud providers in terms of increasing traffic demands and energy costs. The social-aware management of overlay traffic is a promising optimization approach, which shows potential for improvements by exploiting social information. This paper identifies key stakeholders and their roles in the service provisioning value chain and outlines major markets and optimization potentials. Accordingly, two scenarios are developed: the end user focused scenario aming at increased QoE for end users, and the operator focused scenario targeting at the highest operating efficiency in terms of low cost and high revenue for the operator. The energy efficiency plays a major role as a key performance metric in both scenarios. SmartenIT's socially-aware management approach is illustrated based on two example mechanisms for traffic optimization: the home router sharing mechanism (HORST) on the end user side, as well as the dynamic traffic management mechanism (DTM) on the operator side. The paper is concluded by a first sketch of SmartenIT's architecture and its mapping to the two scenarios.

Key words: application-layer traffic optimization, economic traffic management, social networks, QoE, energy efficiency, inter-cloud communications

1 Introduction

In the current phase of Internet development we are observing significant coexistence and mutual stimulation of cloud computing and on-line networking. Fundamental solutions elaborated for traditional network management are probably not adequate for observed new situation. For evolving Future Internet concept and for contemporary overlay applications, we need novel characterization approaches, new business models, stakeholders characterization leading further to design of innovative network and traffic management mechanisms.

SmartenIT consortium has defined social awareness, QoE awareness and energy efficiency as the key targets for the design and optimization of traffic management mechanisms for overlay networks and cloud-based applications. Such three main targets, although defined in wide meaning, can be used as design goals for emerging proposals, e.g., to establish content distribution systems minimizing energy consumption by using energy awareness of its architecture elements and using social awareness translated into efficient structure of caches which minimizes volume of transferred data.

The remaining part of the paper is organized as follows. In Section 2, we introduce relevant definitions and terminology used to describe stakeholders in current service layer of networks. Section 3 provides descriptions of two different segments of the value chain of service provisioning, i.e, network-centric and user-centric cases. Section 4 aims at presenting the optimization potential for joint consideration of cloud and overlay, illustrated by description of HORST and DTM solutions. System architecture specified and being under practical development within SmartenIT consortium is presented in Section 5. Finally, Section 6 draws conclusions worked out after specification phase of SmartenIT project and briefly overviews next steps to come.

2 Definition and Terminology

In today's Internet marketplace, the Internet service layer stakeholders are those who buy and sell Internet services, namely Connectivity Providers, Data Centres Operators (DCO), Cloud Service Providers, Information Providers, and End-users. The Connectivity Providers operating at Layers 3 and below include: End-User Network Provider, Access, Transit, and Backbone Provider. The Information Providers are commonly referred to as Over-The-Top providers (OTTs). End-users may be classified as residential and business (small, medium, and large). A Connectivity Provider, commonly referred to as Network Service Provider (NSP), normally owns its network and is responsible for the provisioning of its functionalities. An Internet Service Provider (ISP) is an NSP also offering Internet services. DCOs own and manage a set of resources, including servers, storages, security devices, power management devices and the network infrastructure responsible for both, internal network traffic and external connectivity to other stakeholders.

The role of data centers (DC) is rapidly growing. Although the Internet is forecasted to reach the 'Zettabyte' era in 2016 [2], data centers have already entered it. While the amount of traffic crossing the Internet and IP WAN networks is projected to reach 1.3 ZBs (i.e., 10^9 TBs) per year in 2016, the amount of DC traffic is already 1.8 ZBs per year, and by 2016 will nearly quadruple to reach 6.6 ZBs per year. This represents a 31% Compound Annual Growth Rate (CAGR). The higher volume of DC traffic is due to the inclusion of traffic inside DC (typically, definitions of Internet and WAN stop at the boundary of DC). Factors contributing to traffic remaining in DC include functional separation of application servers, storage, and databases, which generates replication, backup, and read/write traffic traversing DC. In the cloud systems era Cloud Service Providers are becoming the main direct customers of DCs. The resources provided on demand by DCs allow those providers to build new innovative services addressing requirements, like scalability, mobility, availability and reliability. Also, the flexibility of Cloud Service Providers may introduce new business models which result in cost reduction on the end-user's side. DCOs have been convinced that the use of cloud platforms to manage their resources is profitable for their business. Using systems like OpenStack they can apply the Infrastructure as a Service (IaaS) model to offer dynamically resources like CPU, memory and storage. According to the NIST definition [3], the consumer in this cloud model does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of selected networking components which are in fact virtual items.

The SmartenIT approach provides consistent sequence of terms and concepts starting from most generally defined scenarios, providing more functionality-oriented use cases, up to detailed solutions. We foresee an environment where both user-driven and operator-driven actions lead to a mutually accepted solution. Therefore a proper architecture is required for this. The Internet services provisioning is enabled by means of commercial business agreements among the Internet stakeholders, thus creating a multi-provider and user-centric environment, presented in the next section.

In the remainder of the paper we focus our analysis on the End User Focused Scenario (EFS), pertaining to the services and interactions visible to the end-users and the Operator Focused Scenario (OFS) reflecting the wholesale agreements among network and overlay service providers - that though not visible to the end users, greatly affect the Internet services. These two scenarios are also important from a technical point of view since they allow different potential for optimizations and respective traffic mechanisms that operate in different time scales and over different granularities of Internet traffic under a unified framework, the SmartenIT architecture.

3 The multi-provider and user-centric environment

There exist two different segments of the value chain of service provisioning. On the one hand, there is the wholesale market with large time scale agreements regarding traffic aggregates of providers who interact in order to provision services to the users. On the other hand, users exhibit their demand in small time scales by means of flows. Consequently, two different scenarios are defined, one addressing the end-users' view of the system (EFS) and the other addressing the operators' view of the system (OFS).

The EFS is based on the involvement of the end-user and his devices as active instances in the system, e.g., for prefetching data to be consumed in the near future. As opposed to that, the OFS is focusing on the interactions among the different operators acting in SmartenIT framework, namely Cloud Providers, Service Providers, and Internet Service Providers. One key feature of Cloud Computing paradigm is the fact that cloud services are offered to the end-user in a transparent way, e.g. the user is not aware of physical location

The SmartenIT scope clearly highlights the fact that multiple stakeholders are involved, each of them having its own interests, strategies and business goals. It must be noted that a SmartenIT solution would inevitably have to accommodate the possibly conflicting goals and needs of the stakeholders in an incentive-compatible way. Otherwise, the SmartenIT propositions would have limited or even no chance of being adopted in practice, resulting in limited or zero impact on the market.

3.1 End User Focused Scenario

The end user focused scenario has the goal of providing increased QoE and energy efficiency for end users by distributing and moving content and services in an energy and social aware way, while taking provider's interests into account. The scenario definition adds a number of innovative aspects to former definitions of service and content mobility, which are described in the following. First, besides mobility of services and content on the same level of aggregation (e.g., the migration of virtual machines between DCs or the caching of content in a CDN), to which we refer as horizontal mobility, we add an additional infrastructure layer to the existing solutions consisting of user-owned Nano Data Centers (uNaDas). uNaDas are WiFi routers, which provide extended functionality such as increased computation and storage capabilities. The devices are located at end-user's premises and enable vertical mobility, i.e., a movement of services and content into the user's home, thus making him part of the content and service delivery chain.

Consequently, optimizing placement means optimizing along the two dimensions of vertical and horizontal placement. Finding the right placement is an optimization problem to be solved in a socially aware and energy efficient manner. For this purpose, social data is either provided by the end user offering his local view on his social network account or from a direct cooperation with online social network providers given a suitable business model exists. The inclusion of social data allows predicting the necessity to move content, e.g., content consumed by friends can be prefetched to a user's uNaDa, in order to provide better QoE. Moreover, social data can be used to derive trust relations between users in a highly distributed and untrusted environment, which is the base for cooperation among users and their devices. In order to optimize the placement with respect to energy efficiency, information on energy consumption is needed. This data can be derived from energy models providing an energy estimate of the placement. Energy models which estimate energy consumption from network measurements are existing in literature, e.g., [10]. Moreover, modelling the energy efficiency of placements allows also for the prediction of future energy consumptions.

Besides providing a platform for placing content and services, the uNaDa layer accounts for the increasing importance of mobile devices such as smart phones and tablets. Therefore, means for offloading from cellular technologies to the uNaDas, local WiFis are offered. The offloading capabilities are combined with possibilities to access the content stored on the uNaDa, thus the uNaDa may act as a wireless cache for relevant content.

The stakeholders in this scenario are Cloud Application Providers offering the content and services (e.g., YouTube, Vimeo or Amazon EC2), the Internet Service Provider, and the end user. Cloud Application Providers and end users have a clear incentive to collaborate in this scenario, as it will increase the QoE of the end user, which will in turn increase his willingness to pay for services and content [11]. Moreover, the cloud application can reduce the provisioning of own resources. On the contrary, the ISP can benefit from the solution, if it is implemented in an underlay aware way, which will reduce traffic over peered links and can relieve the cellular infrastructure.

3.2 Operator Focused Scenario

The Operator Focused Scenario (OFS) has the goal of achieving highest operating efficiency in terms of high Quality of Service, low energy consumption, low operating cost and highest revenues (where applicable). The scenario focuses on a certain number of well-known aspects that mixed together lay the foundation to drive potential SmartenIT innovation. First, the impact of inter-cloud communication on other involved stakeholders, i.e. the ISPs, and the reaction of those to the decisions of the DCOs which constitute the cloud layer are taken into account. Second, new coalition schemes between DCOs are considered, i.e. the so-called Cloud Federation, which enable collaboration so as to achieve both individual and overall improvement of cost and energy consumption, while simultaneously achieve footprint expansion by addressing a geographically larger set of users. Other key topics (to be intended as triggers of inter-cloud communication process among DCOs) of the selected scenario are content placement (to provide QoE enhancements), data replication (to achieve geographical fault-tolerance) and VM migration (to give support service mobility). Moreover some social awareness information (even if at a different aggregation level w.r.t EFS) can be seen as potentially impacting the actual scenario.

Whichever the trigger of inter-cloud communication process a trade-off rises due to the competing interests of the DCOs which constitute the source and destination end-points for each transmission flow, as well as the interests of the ISPs providing inter-connectivity to the DCOs. In particular, the DCO is interested in performing often data replication/migration in order to provide high QoS/QoE to its end-users, or the Cloud Service Providers that employ its (storage, computing) service to offer other services and applications to end-users. Moreover, data migration/re-location may often be imposed by the need to reduce overall energy consumption within the federation by consolidating processes and jobs to few DCs only.

On the other hand, the traffic generated by the data replication or migration performed by the DCOs significantly burdens the networks of the ISPs, which implies increase of operating cost for the ISP mainly in terms of transit inter-connection cost. Consequently, the ISPs would like to employ certain mechanisms and pricing models that will enable efficient traffic management, as well as sharing of revenues obtained by cloud services and applications.

Therefore, the OFS scenario defines a series of interesting problems to be addressed by SmartenIT, specifically:

– the interactions between the members of the Cloud Federation (from technical and business point of view)
– the interactions of DCOs and ISP in terms of traffic crossing ISP WAN links, fair optimization of resource allocation and sharing among the federated members, and
– energy efficiency for DCOs, either individually or overall for all member of the federation.

4 Cloud/Overlay Traffic Optimization

4.1 Home Router Sharing based on Trust (HORST)

Home router sharing based on trust (HORST) [1] is a mechanism which addresses data offloading, content caching/prefetching, and content delivery. The HORST mechanism eases data offloading to WiFi by sharing WiFi networks among trusted friends. Moreover, it places the content near to the end-user such that users can access it with less delay and higher speed, which generally results in a higher QoE. The SmartenIT solution consists of a firmware for a home router and an OSN application. The HORST firmware establishes a private and a shared WiFi network (with different SSIDs) and manages the local storage of the home router as a cache.

To participate a user needs a flat rate Internet access at home where he has to install the HORST firmware to his home router. The owner of the home router uploads the WiFi access information of the shared WiFi to the OSN application. Each user can share his WiFi information to other trusted users via the app and request access to other shared WiFis.

HORST ensures resource contribution by incorporating an incentive mechanism coupling different resource contributions of end-users to the system, namely the provisioning of storage capacity and offloading capacity, to the receivable QoE, based on the vINCENT incentive mechanism presented in [4]. In order to enhance the overall system performance, the incentive includes social networking data as a base of trust, thus increasing market liquidity.

The HORST router has a social monitor component to collect social information from an OSN about the router's owner and his trusted friends. If a user approaches the home router of a trusted friend, he is provided with access data via the OSN app to connect to the shared WiFi. Every HORST system predicts the content consumption (i.e., when and where will which content be requested) of his owner based on history and information from the OSN, such as content popularity and spreading. If a predicted content, e.g., a video, is not yet available in the local cache, it will be prefetched (e.g., first chunks of a video as proposed by the Socially-aware Efficient Content Delivery (SECD) mechanism [7]). If the user is connected to a friend's home router, a prefetch command is sent to the HORST system on the friend's router. For prefetching as well as for actual requests which cannot be served locally, HORST chooses the best source (either another HORST home router or a cloud source) based on overlay information, and fetches the desired content. In regular intervals, HORST checks if the content in his own local cache is still relevant (either for local consumption or as a source for content delivery) and decides whether to keep or replace it.

Finally, HORST federates all home routers to form an overlay content delivery network (CDN), which allows for efficient content placement and traffic management. Thereby, ISP costs can be included in the local decisions at the home router, e.g., by taking into account the location of contents in terms of Autonomous Systems (AS). For the communication between the home routers and the distributed storage of meta-information, RB-Tracker [9] is used due to its efficiency. Thus, RB-Tracker builds the basis of HORST and hides the complexity of the overlay management.

4.2 Dynamic Traffic Management

The dynamic traffic management (DTM) mechanism [5, 7] developed in SmartenIT is a solution for the OFS case. It aims at minimization of ISP's cost of inter-domain traffic by appropriate management of overlay traffic. The overlay traffic generated between DCs or between DCs and end-user devices often passes inter-domain links. DTM offers the ISP a capability to select the network path used for traffic transfer.

ISP aiming at inter-domain traffic cost reduction needs to answer two questions: when to transfer the traffic, and which inter-domain link to use for data transfer. If ISPs network is multihomed and the tariffs used on different inter-domain links differ, appropriate shifting of inter-domain traffic between them may result in lowering total costs [5]. Such traffic management is possible by creating tunnels between remote DCs (GRE or MPLS tunnels). Each tunnel should pass different inter-domain link. Appropriate selection of the tunnel results in transferring the traffic being sent between DCs via a given inter-domain link. The decision on link selection is taken by ISP dynamically according to the current situation and estimated cost of inter-domain traffic.

DTM consists of two main building blocks: (1) an algorithm to find the optimal solution to be achieved at the end of an accounting period. Using cost functions and information on the traffic distribution in the previous period, it makes a prediction for the next period and finds a better traffic distribution in which the ISP's cost is minimized; (2) the compensation procedure which determines how the traffic distribution should be influenced at a given moment to achieve the optimal solution at the end of the accounting period, i.e., decides on selection of inter-domain link. To make it possible to dynamically react to the situation in the network and predict costs periodical traffic measurements on links is needed. Finally, to make it possible to dispatch flows to appropriate tunnels SDN controller may be used.

5 System Architecture

Based on the description of scenarios, the identified properties of the SmartenIT approach and the high-level description of such solutions, this section provides an overview of the SmartenIT system architecture [8]. The component diagram in Figure 1 shows all the core components of the architecture as well as the necessary interfaces. The color-coding of the components denotes whether a component already exists in external systems (white) or if it is SmartenIT-specific (blue). In the rest of this section, we highlight the most important components and provide a short overview of their functionality.

The **Traffic Manager** is the central component of the architecture and includes all the decision-taking functionality. It encompasses the Cloud Traffic Manager and the Fixed/Mobile Network Traffic Manager components. The **Cloud Traffic Manager** makes high-level decisions, e.g., the caching of specific content to specific places, or the redirection of a user request to another (nano)data center. These decisions are communicated to the Overlay Manager which in turn informs the affected DCs. The **Fixed/Mobile Network Traffic Manager** takes low-level decisions, such as what QoS class should be assigned to a specific flow, which MPLS tags to be used, etc. These decisions are later materialized by the Switching/Forwarding component. The Traffic Manager provides a number of interfaces so as to interact with the rest of the system components and gather the required information or communicate the resulting decisions. The **Overlay Manager** allows the formation of overlays between remote (nano)data centers or even small Clouds, with DCs or uNaDas being the peers in these overlays. This component is responsible for the communication between peers so as to advertise offered resources, ask for resources, cache/move data, etc.

A number of Analyzers is included in the architecture, the purpose of which is to collect multi-dimensional metrics and process them so as to enrich the decision-taking functionality of the Traffic Manager. The **QoS/QoE Analyzer** addresses the QoS/QoE optimization aspects, while the **Network Analyzer** captures the network state and topology. The **Energy Analyzer** offers energy consumption considerations and the **Economics Analyzer** interacts with the

Billing/Accounting Systems of ISPs or Cloud Operators, in order to address economic optimizations. Finally, the **Social Analyzer** consider the social aspects that can provide estimates for content consumption and dissemination patterns.

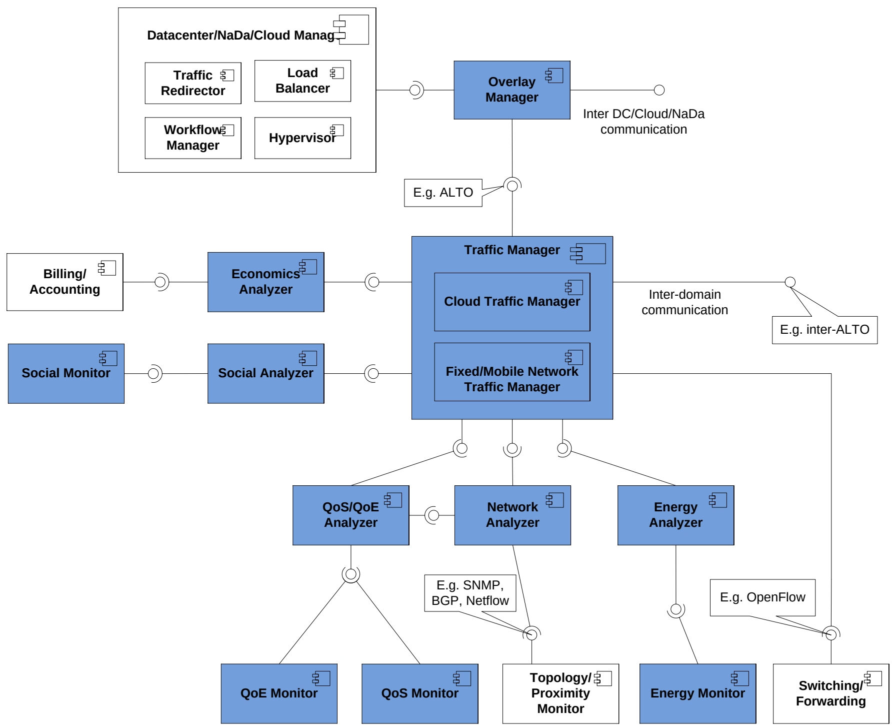

Fig. 1. The SmartenIT Architecture - Component Diagram

5.1 Adherence to the identified scenarios

In Figure 2, the topological view of the SmartenIT architecture has been used as a basis for the mapping of scenarios and the respective mechanisms. The topological view includes three domains of entities: (1) **Data Center/Cloud Layer**: This layer comprises DCs and their virtual interconnections using the Internet as a network. (2) **Core and Access Network Layer**: This layer contains components in the ISP network and the private networks of DCOs. (3) **End User Layer**: This layer covers the end-users devices as well as the users home gateway (uNaDa).

Having identified the building blocks of the SmartenIT architecture, we proceed with the formation of various functionality stacks (by combining different components) and their placement on the envisioned entities of the SmartenIT

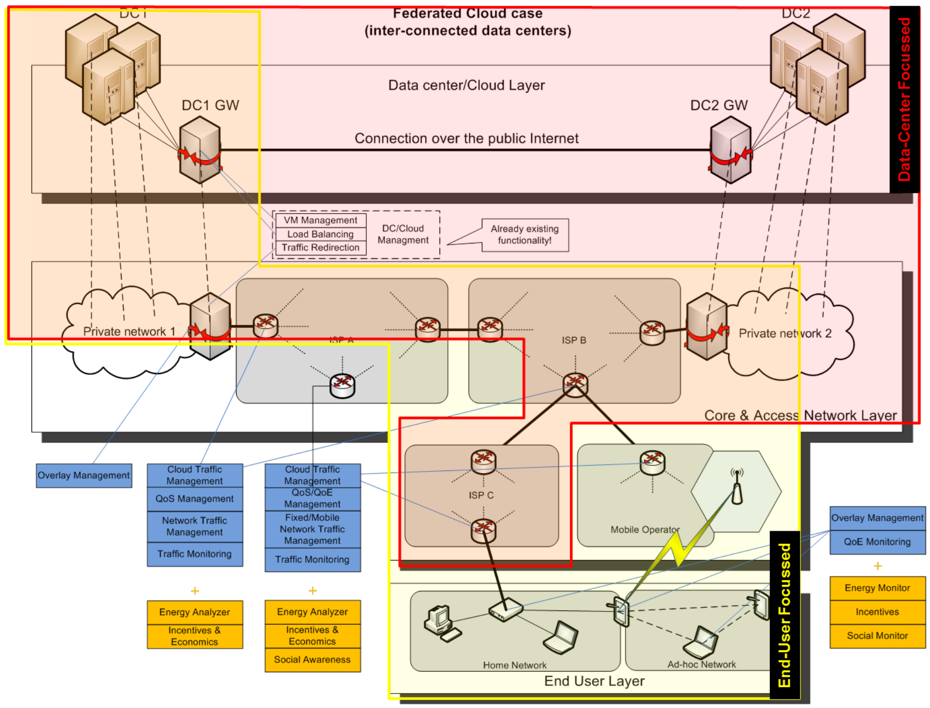

Fig. 2. Topological view of architecture with added scenario overlay [7]

deployment environment. These functionality stacks are depicted as a stack of blue/orange boxes and are associated to the respective entities in Figure 2.

In order to bridge the gap between the architecture and the mechanisms, we provide a mapping of the two scenarios to the (enriched with functionality stacks) deployment environment, as depicted in Figure 2. The topological diagram has been used to get a first overview of the traffic management solutions w.r.t the SmartenIT architecture, rather than a detailed mapping to components.

With this in mind, scenarios and functionality can be aligned along a common axis. This further allows mapping each SmartenIT proposed mechanism to a set of envisioned functionalities, while at the same time showing the mainly addressed scenario. A detailed view of this mapping of all mechanisms to the respective scenarios/functionality is attached to the appendix of [7].

6 Conclusions and Outlook

In this paper, we briefly present the social-aware, QoE-aware and energy-efficient network management approach and system architecture worked out by EU SmartenIT project.

Specific solutions, defined by two main and complementary scenarios: ISP-oriented and end-user-oriented, are defined in details as, respectively, the DTM and HORST solutions. The relevant mapping of designed mechanism onto SmartenIT architecture modules is also provided. After succesfully achieving goals of specification phase, participants of the SmartenIT project are now implementing necessary modules in order to build fully operational system.

Acknowledgment

This work has been performed in the framework of the EU ICT STREP SmartenIT (FP7-ICT-2011-317846). The authors would like to thank the entire SmartenIT team for discussions and providing insights on major research problems.

References

1. Seufert, M., Burger, V., Hoßfeld, T.: HORST - Home Router Sharing based on Trust. In: Social-aware Economic Traffic Management for Overlay and Cloud Applications Workshop (SETM 2013), in conjunction with 9th International Conference on Network and Service Management (CNSM), Zurich, Switzerland (Oct. 2013)
2. Cisco Systems White Paper: Cisco Global Cloud Index: Forecast and Methodology 2011 - 2016, 2012
3. Mell, P., Grance, T.: SP 800-145. The NIST Definition of Cloud Computing, National Institute of Standards & Technology, (2011)
4. Wichtlhuber, M., Heise, P., Scheurich, B., Hausheer, D: Reciprocity with virtual nodes: Supporting mobile peers in Peer-to-Peer content distribution. In: Social-aware Economic Traffic Management for Overlay and Cloud Applications Workshop (SETM 2013), in conjunction with 9th International Conference on Network and Service Management (CNSM), pp.406-409, Zurich, Switzerland (Oct. 2013)
5. Duliński, Z., Stankiewicz, R.: Dynamic traffic management mechanism for active optimization of ISP costs. In: Social-aware Economic Traffic Management for Overlay and Cloud Applications Workshop (SETM 2013), in conjunction with 9th International Conference on Network and Service Management (CNSM), pp.398-401, Zurich, Switzerland (Oct. 2013)
6. Biancani, M., Cruschelli, P., (eds.): SmartenIT Deliverable 1.2 — Cloud Service Classifications and Scenarios, (Oct. 2013)
7. Burger, V. (ed.): SmartenIT Deliverable 2.2 — Definitions of Traffic Management Mechanisms and Initial Evaluation Results, (Oct. 2013)
8. Hausheer, D., Rückert, J., (eds.): SmartenIT Deliverable 3.1 — Initial System Architecture, (Apr. 2013)
9. Lareida, A., Bocek, T., Waldburger, M., Stiller, B.: RB-tracker: A fully distributed, replicating, network-, and topology-aware P2P CDN. In: IFIP/IEEE International Symposium on Integrated Network Management (IM 2013), pp.1199-1202, Ghent, Belgium (May 2013)
10. Schwartz, C., Hoßfeld, T., Lehrieder, F., Phuoc Tran-Gia, P.: Angry Apps: The Impact of Network Timer Selection on Power Consumption, Signalling Load, and

Web QoE. Journal of Computer Networks and Communications, Hindawi Publishing Corporation, (2013)
11. Reichl, P.: From charging for Quality of Service to charging for Quality of Experience. Annals of telecommunications 65 (3-4), Springer-Verlag, 189-199 (2010)
12. Stiller, B., Hausheer, D., Hoßfeld, T.: An Overview on a Socially-aware Management of New Overlay Application Traffic Combined with Energy Efficiency in the Internet (SmartenIT), Galis, A., A. Gavras, A., (eds.) FIA Book — Validated Results and New Horizons. LNCS, vol. 7858, pp. 1–12. Springer, Heidelberg (2013)
13. Hoßfeld, T., Hausheer, D., Hecht, F., Lehrieder, F., Oechsner, S., Papafili, I., Racz, P., Soursos, S., Staehle, D., Stamoulis, G.d., Tran-Gia, P. Stiller, B., Hausheer, D.: An Economic Traffic Management Approach to Enable the TripleWin for Users, ISPs, and Overlay Providers, Tselentis, G., et al. (eds.) FIA Prague Book — Towards the Future Internet - A European Research Perspective. IOS Press Books, pp. 24–34. (2009)
14. Fiedler, M., and Hossfeld, T., Tran-Gia, P.: A Generic Quantitative Relationship between Quality of Experience and Quality of Service, IEEE Network Special Issue on Improving QoE for Network Services 24 (2), 36-41 (2010)