Objective FP7-ICT-2009-5-257448/D-2.1

Future Networks

Project 257448

"SAIL – Scalable and Adaptable Internet Solutions"

D-2.1
(D-A.1) Description of project wide scenarios and use cases

Date of preparation: **11-02-14**
Start date of Project: **10-08-01**
Project Coordinator: **Thomas Edwall**
Ericsson AB

Revision: **1.1**
Duration: **13-01-31**

Document Properties[1]:

Document Number[2]:	FP7-ICT-2009-5-257448-SAIL/D-2.1
Document Title:	Description of project wide scenarios and use cases
Document responsible:	Thomas Edwall (acting Tech. Mgr)
Author(s)/editor(s):	Tapio Levä, João Gonçalves, Ricardo Jorge Ferreira et al / Johan Myrberger (editor)
Target Dissemination Level[3]:	PU
Status of the Document:	Final version
Version	1.1

Production Properties:

Reviewed by:	Benoit Tremblay

Revision History:

Revision	Date	Issued by	Description
1.0	110121	Johan Myrberger	Final version.
1.1	110214	Johan Myrberger	Minor corrections

This document has been produced in the context of the SAIL Project. The research leading to these results has received funding from the European Community's Seventh Framework Programme ([FP7/2010-2013] [FP7/2010-2013]) under grant agreement n° 257448

Abstract:

Keywords:

Scenarios, Use cases, SAIL

[1] Input of Title, Date, Version, Target dissemination level, Status via *"File /Properties/Custom"* in the Word menu
[2] Format: FP7-ICT-2009-5-257448-SAIL /<Deliverable number>
Example: FP7-ICT-2009-5-257448-SAIL / D-2.1
[3] Dissemination level as defined in the EU Contract:
PU = Public
PP = Distribution limited to other programme participants
RE = Distribution to a group specified by the consortium
CO = Confidential, only allowed for members of the consortium

Table of Contents

Executive summary

Main objectives

In order to guide and facilitate the forthcoming work within SAIL a number of scenarios and use cases is described and analyzed.

The work package scenarios and use cases are meant to guide the architecture development by providing some challenging scenarios as well as to promote the innovations expected from the work packages by providing some compelling examples.

Approach

From three dimensions (video, mobility and flash crowd), that are viewed to be of major importance for future networks, a base scenario is established.

Each work package has, using the three dimensions and the base scenario, identified in total six scenarios. From these six scenarios a total of 21 use cases are derived. The scenarios and use cases are described and analyzed from both a technical and business perspective.

Result

Based on the analysis we have a strong foundation for the continued work, where both the validity of the use cases is established and where the SAIL project has reached a common view and understanding.

The path that was followed during the work (from three identified dimensions of future networks, over a base scenario, to a number of scenarios defined in each work package and finally arriving at the use cases) have been of great value for the SAIL project.

1 Introduction

1.1 Motivation and objective of document

The SAIL (Scalable and Adaptive Internet Solutions) project will work with the architecture of future networks.

The work within the SAIL project is divided between three technical work packages, each covering one aspect of the future networks:

- *NetInf* - Network of Information (WP-B)
  NetInf covers the shift of focus from network nodes to information objects and will develop a general-purpose **information-centric networking architecture** that provides efficient communication and information dissemination by leveraging secure naming, name-based routing, in-network caching, and optimized distribution as general services to all applications.
- *OConS* – Open Connectivity System (WP-C)
  OConS deals with efficient use of multi-path, multi-protocol and multi-layer networking – over any fixed and mobile networks. *OConS* will develop and implement proof-of-concept protocols and mechanisms capable of **managing data transport flows between end-to-end and edge-to-edge**, efficiently exploiting different technologies and different traffic patterns.
- *CloNe* – Cloud Networking (WP-D)
  CloNe ties Cloud Computing and Network Virtualization together. *CloNe* will design a cloud networking architecture supporting flash **network slices**, evaluated through a large scale prototype distributed across at least three different sites in Europe.

The work package scenarios are meant to guide the architecture development by providing some challenging scenarios as well as to promote the innovations expected from the work packages by providing some compelling examples.

Most, but not all of our technical work can be described in the context of this set of scenarios building on top of each other – but we also point out that these scenarios should not be misconstrued as being the development goal of the SAIL project; we will build the technology as such to enable such scenarios, but not the necessary application software on top to realize these particular scenarios. We start from a simple scenario and add new roles, actors, and functionality into it to highlight how different parts of the SAIL architecture interact with each other to accomplish the scenario.

The scenarios and use cases described in this document will guide the forthcoming work, and ensure a cross-project harmonization and coordination.

1.2 Structure of document

The work package scenarios and use cases are derived in a stepwise fashion.

Initially three important dimensions of future networks are identified (section 2).

Next a base scenario is described (section 3). The base scenario is simple yet versatile and can be used to illustrate many aspects.

Using the dimensions of future networks and the base scenario the three work packages within the SAIL project have formulated a number of scenarios and use cases. These are summarized in section 4.

Section 5 contains a mapping of a few, selected, use cases towards the base scenario.

In section 6 a number of identified trends and uncertainties related to future networks are outlined.

The scenarios and use cases derived in each work package are further described and analyzed in sections 9 to 11. The methodology for this is outlined in section 7, and the template for the analysis is presented in section 8.

2 Three important dimensions of future networks – video, mobility and flash crowds

Already in the initial phases of the SAIL project three dimensions were identified that will influence and drive the evolution of a future networks architecture; video, mobility and flash crowds.

Table 1: Three common aspects

Dimension	*Characteristics*
Video	• High bandwidth content distribution. • Dynamic instantiation of network nodes • User-generated content
Mobility	• Mobility of networks, users, applications and nodes • Heterogeneity of accesses and network • Multipath, multihoming, multi protocol
Flash crowd	• Disconnectivity • Provisioning of online services in disconnected environments • Temporary by nature • Scaling of network configurations • End-users part of the infrastructure • Ad hoc/Emergency/Developing

These three dimensions have been the basis for the base scenario presented in section 3, as well as for the scenarios and use cases that have been developed in the three work packages in the SAIL project.

3 A base scenario common to the project

In order to put the identified dimensions of future networks into context a base scenario is described. This base scenario has been the starting point for the work package scenarios and use cases described later in this document.

3.1 Base scenario: single content provider in a single operator domain

In the base scenario Alice is streaming video from her mobile handset to her “personal content repository”. This content repository is an application that leverages the *NetInf* system for storing the video online as well as for providing a live streaming access to the video to an arbitrary number of users. Alice records or uploads the streaming content into her repository provided by her “community service” provider for possible later use without knowing if there would be any outsiders interested in her content. While still streaming the content she gets a request for her material and she accepts to make the content available for all requestors. She shares this content expecting that only a few of her friends will be interested. Initially, this traffic can be served directly from her repository (using *NetInf* techniques to locate the data store); since it is streaming traffic, there is no need to cache the data elsewhere.

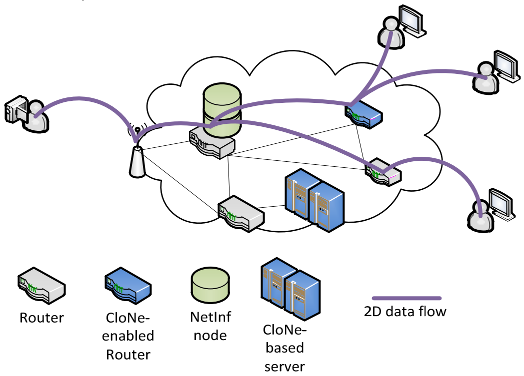

Figure 1: Alice streams to a few followers

However, it turns out that her content is exceptionally popular, and is viewed by a growing number of followers (i.e. consumers) – a ***consumer crowd*** is forming. This consumer crowd is a virtual community characterized by their relation to the source Alice or her content, but may be geographically dispersed and also heterogeneous in their technologies of being connected to the network. In addition, followers also start viewing at a later time but still want a stream from the start. Instead of serving all these traffic flows from a single server farm as would be common today, *CloNe*, *OConS* and *NetInf* will work hand in hand to cache and to transform the content at suitable places in the network, such that these flows can be served at optimal user experience and minimal use of network and transport resources.

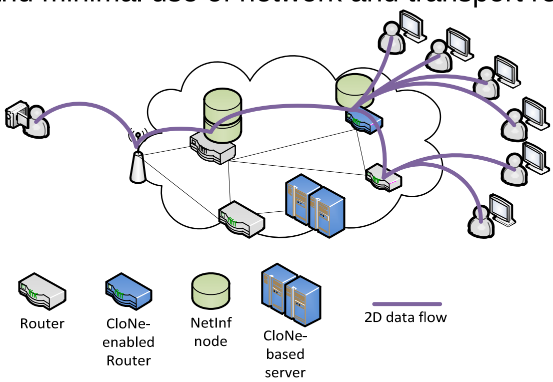

Figure 2: Alice content gains popularity

3.2 Extended base scenario: Single content provider in a multi-operator setting

After a while customers of another community service provider will join to enjoy Alice's streaming. In principle, this could be served directly by all these customers obtaining their data from the *NetInf*-based content repository in Alice's operator's network; however, this will cause considerably traffic through a cross-operator link, i.e. causing traffic across autonomous systems (AS). Alternatively, if the second operator supported *NetInf* as well, the *NetInf* requests emanating from this network would automatically cause the data to be cached in its own network and the streams to be delivered from these caches.

In the case that the new operator has not yet embraced *NetInf*, but does support *CloNe*, it will be possible to dynamically instantiate computing and storage functions inside its network. *OConS* can detect that there is potential for traffic optimization, and the new operator can use the *CloNe* service API to start up *NetInf* instances inside its network at topologically well chosen places, at some of its own network resources. Upon this, traffic flows will get rearranged such that the streams are now served from these additional *NetInf* caches (the *NetInf* name resolution scheme will adapt to take these new caches into account and will prefer these topologically advantageous sources).

In this setting, *OConS* connectivity services are used to quickly and reliably transfer the copies of the content caches between the operators (this can be a sizeable amount of already taped data, to ensure that the video stream is available completely) or the *NetInf* virtual machines to be executed or relocated by *CloNe* (which also can be quite large) as well as to locate optimal locations for the caches and *NetInf* nodes within the new operator. Indeed, *OConS* can handle the transport of the initial (previously buffered) video burst via higher bandwidth links (e.g. aggregated on dedicated optical transport flows at lambda or sub-lambda level) for inter-operator traffic, before seamlessly falling back to the cheaper connectivity that is sufficient to keep up with Alice's ongoing video stream. This optimizes and reduces the cross-AS traffic since all new members of the crowd (consumers of Alice's content) now accesses Alice's video content with traffic local to their network operator and Alice's video is only replicated once between her operator and each of the corresponding operators.

Moreover, some consumers will need transcoding of the content to better suit the type of access network and terminals they are using. A transcoding service can be dynamically deployed by *CloNe* at a place where the content is consumed or where traffic streams cross into low-rate networks. In addition, the local *CloNe* can host (and possibly even offer) other types of content personalization applications such as immersive video or location-based services (e.g., personalized advertisements inserted into the video stream) depending on user profiles and subscriptions. As some of these functions can be highly CPU-intensive, it stands to reason to only deploy them as needed.

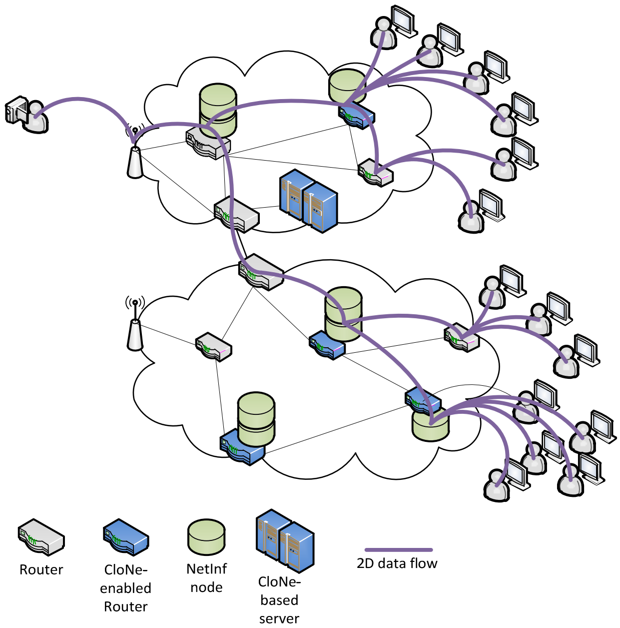

Figure 3: Extended base scenario, Single content provider in a multi-operator setting

3.3 Extended base scenario: A crowd streaming simultaneously from same event

The event that Alice is witnessing and documenting by her video post may attract a set of other people to contribute and stream their own views to the network – a ***producer crowd*** is forming as well. This producer crowd is typically a geographically co-located community but they are likely to have subscriptions with different service providers and operators, hence they are topologically dispersed from a communication network perspective. The streamers are streaming the live content into their own personal repositories located in their own home networks. So, the repository becomes a distributed repository among multiple operators.

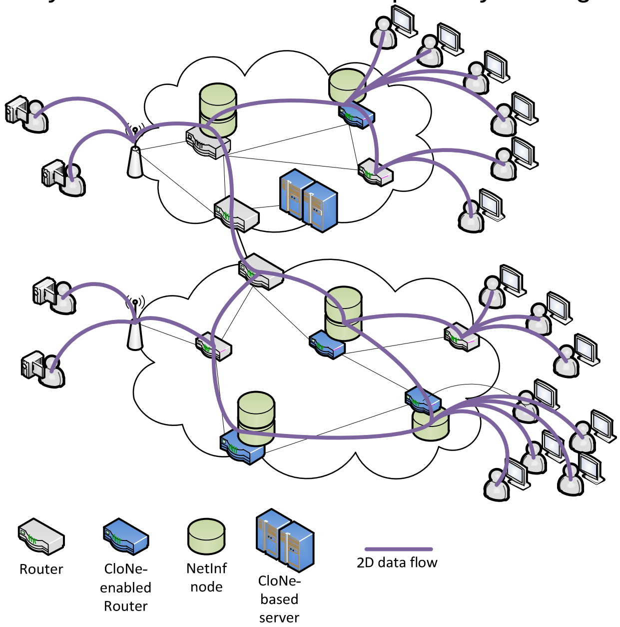

Figure 4: Extended base scenario, a crowd streaming simultaneously from same event

However, a *NetInf*-based application can group this same content into the same scope to ease the access and use of the parallel streams, exploiting meta-data provided via the *NetInf* naming scheme. *OConS* ensures efficient content transfer and mobility support and the computational resources of *CloNe* enable multi-view experience and proper adaptation of the content to match the end-user terminals – for example, *CloNe* could dynamically instantiate even highly computationally intensive services like 3D reconstruction from the individual video feeds. In addition, the various traffic flows (2D flows to the reconstruction server, 3D flows to the *NetInf* repositories and thence onwards to the consumers) are properly reconfigured by *OConS*.

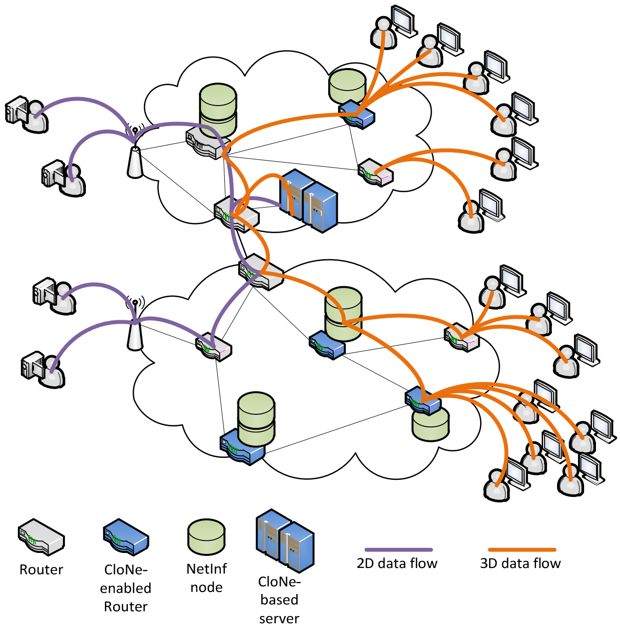

Figure 5: A *NetInf* application in the second extended base scenario

4 Overview of scenarios and use cases

Based on the three dimensions and the base scenario each work package have derived one or more scenarios, and from these scenarios created use cases.

This chapter provides an overview of the derived scenarios and use cases. In subsequent parts of the document the scenarios and use cases are described in more details and also analyzed.

4.1 Work package scenarios

A total of six scenarios have been derived by the work packages. Table 2 below provides a summary.

Table 2: Summary of derived scenarios

Scenario name	*Description*	*Further described in section*	*Work package*
NetInfTV	A scenario that highlights a heavy content distribution perspective. The scenario covers all the different types of video traffic delivered over the Internet.	9.1	NetInf (WP-B)
Next generation mobile networks	The starting point of this scenario is the current cellular networks that are still evolving, from having been primarily used for telephony to mobile data services. Such networks also evolve from a centralized and static architecture into a more distributed and dynamic one. Moreover, the mobile entities get more advanced and have the possibility to connect through different access technologies with multiple interfaces.	9.2	NetInf (WP-B)
Developing regions	Covers how *NetInf* can support the needs and conditions in developing regions.	9.3	NetInf (WP-B)
Supporting flash crowd connectivity needs	The scenario is based in a Flash Crowd; a large group of people with mobile devices that are in a location where there is an unexpected and increased demand for communications and services.	10.1	OConS (WP-C)
Dynamic Enterprise	This scenario presents and depicts the provisioning of IT/IS solutions from the cloud network ecosystem to the enterprise market, supported in concepts such of SaaS, PaaS and IaaS.	11.1	CloNe (WP-D)
Elastic video in the cloud	The scenario presents and depicts the offering of video and similar services from a cloud network ecosystem to the retail market with user perceived enhanced QoE (quality of experience), leveraged on distributed computational resources at the edge of the network architecture.	11.2	CloNe (WP-D)

4.2 Use cases

A total of 21 use cases have been derived from the above 6 scenarios. Table 3 below provides a summary.

Table 3: Summary of use cases

Use case name	***Further described in section***	***Relationship to base scenarios described in section***	***Main dimension of future networks included***	***Parent scenario***	***Work package***
NetInf as a decentralized CDN	9.1.3	5.1	Video	NetInfTV	NetInf (WP-B)
NetInf as a part of ISP multi-play solution	9.1.3	N/A	Video	NetInfTV	NetInf (WP-B)
Video with added local information	9.2.3	5.2	Video	Next generation mobile networks	NetInf (WP-B)
Mobility and multihoming	9.2.3	N/A	Mobility	Next generation mobile networks	NetInf (WP-B)
Event with large crowd	9.2.3	5.3	Flash crowd	Next generation mobile networks	NetInf (WP-B)
Mobile sensors	9.2.3	N/A	Flash crowd	Next generation mobile networks	NetInf (WP-B)
User-generated content	9.2.3	5.4		Next generation mobile networks	NetInf (WP-B)
Development and deployment of mobility-agnostic applications for all types of objects	9.2.3	N/A	Flash crowd	Next generation mobile networks	NetInf (WP-B)
Community ISP	9.3.3	N/A	Flash crowd	Developing regions	NetInf (WP-B)
Creating and sustaining the Connectivity in Wireless Challenged Networks	10.1.3	5.5	Flash crowd	Supporting flash crowd connectivity needs	OConS (WP-C)
Using multi-path/multi-protocol (MultiP) transport for optimized service delivery of heterogeneous content	10.1.3	N/A	Mobility	Supporting flash crowd connectivity needs	OConS (WP-C)
Optimising the QoE for End-users with adequate management of the (Cloud) Network services	10.1.3	N/A	Mobility	Supporting flash crowd connectivity needs	OConS (WP-C)
(Autonomous) Interoperation and Connectivity of Cloud and NetInf data centres	10.1.3	5.6		Supporting flash crowd connectivity needs	OConS (WP-C)
Media Production	11.1.3	5.7	Video	Dynamic Enterprise	CloNe (WP-D)
Remote Auditing	11.1.3	N/A		Dynamic Enterprise	CloNe (WP-D)
Business Goal Management	11.1.3	N/A		Dynamic Enterprise	CloNe (WP-D)
Virtual Desktop	11.1.3	N/A	Flash crowd	Dynamic Enterprise	CloNe (WP-D)
Elastic Live Video Distribution	11.2.3	5.8	Video	Elastic video in the cloud	CloNe (WP-D)
Distributed Gaming	11.2.3	N/A	Flash crowd	Elastic video in the cloud	CloNe (WP-D)
Elastic Video On-demand Distribution	11.2.3	N/A	Video	Elastic video in the cloud	CloNe (WP-D)
Video Conferencing	11.2.3	N/A	Video	Elastic video in the cloud	CloNe (WP-D)

5 Mapping of selected use cases to the base scenario

This section highlights the relation for a few, selected, use cases to the base scenario in section 3.

5.1 NetInf as a decentralized CDN

The *NetInf as a decentralized CDN* use-case (see section 9.1.3) is part of the NetInfTV scenario within *NetInf* (WP-B). In this use-case CDN functionality (for various TV and video services) is provided through collaboration between Internet Service Providers using *NetInf* technology without involving a specialised CDN operator.

The services considered in the use-case have wider scope than the project scenario, but do cover Alice's user-generated video streaming. The ISP-run collaborative CDN is directly matching the situation in the base scenario in the first and second extended variants where a second operator is present.

5.2 Video with added local information

The *video-with-added-local-information* use case (see section 9.2.3) extends Alice’s video content distribution with additional information. This could e.g. be information that is interesting or important in a certain geographic context. Examples of such local information are weather info, traffic jam alerts of even locally generated video such as web cam from a surrounding area.

One instantiation of this use case could be an extension of the extended base scenario of a single content provider in a multi-operator setting: The current visited network of a mobile user provides *NetInf* storage and service platforms that are enable with an “augmented content” feature. Based on meta-information of the original video stream (Alice’s video content), the visited network’s *NetInf* systems can compose a richer service that would either embed or augment Alice’s video content.

In another instantiation the original video stream would be provided by the mobile user’s home network operator who would leverage *NetInf* platforms in a visited network to provide a richer service in a more cost-efficient way, for instance by running interactive service components (e.g. translation functions) close to the user.

5.3 Event with large crowd

Examples of events with large crowds are sports events, e.g. marathons, and outdoor markets and fairs, without dedicated arenas. These kinds of events are arranged seldom enough that it is not feasible to deploy infrastructure built from current technology with enough capacity for all circumstances. The assumption is that the crowd present at the event are interested in content about the event, and possibly also produce content about the event.

The *event with large crowd* use-case (see section 9.2.3) can be viewed as a specialization of the base scenario. The crowd, or at least a substantial part of the crowd, is present at a physical event. There is likely more than one content producer, so this use-case is directly applicable to the second extended base scenario with many persons providing video streams.

5.4 User-generated content

The *user-generated content* use case (see section 9.2.3) is the *NetInf* instantiation of the extended base scenario where a set of persons are streaming simultaneously from the same event. From a *NetInf* perspective the content produced by a group of users would be published to the local *NetInf* infrastructure and then made available in other network domains. The *NetInf* system would provide the distribution and streaming on behalf of the content producers based on the *NetInf* caching and transport services.

Fundamentally, user-generated content is a first class application in a *NetInf* network, as *NetInf* would use the same content registration, name resolution, caching and distribution mechanisms as it would for commercially produced content. For user-generated content, as for all other content, there would be different scopes for publications (e.g. global, domain local, device/node local). The distribution (replication and caching) will depend on popularity and does not require active involvement of the content-generating user.

5.5 Creating and sustaining the Connectivity in Wireless Challenged Networks

The scenario is to support the spontaneous communication needs of a 'flash crowd' community of people and devices using diverse wireless technologies in an adverse environment. The capabilities of the environment may be either sparse or unreliable (e.g. in case of developing or undeveloped regions) or overloaded by the sheer mass of communication requests from the crowd to access the worldwide service and network infrastructure.

The novelty arises from the combination of different solution approaches in wireless challenged networks, such as creating and sustaining the connectivity by meshing heterogeneous technologies, bridging potential disruptions and quality degradation via new multipath transport protocols in wireless meshed networks and using network coding techniques to exploit multiple wireless paths/routes.

The *Creating and sustaining the Connectivity in Wireless Challenged Networks* use case (see section 10.1.3) addresses both the topics of 'mobility' and 'flash crowds' of dimensions of future networks.

This use case exploits the role of the *producer crowd* in the third stage of the *base scenario*, which is formed in an ad hoc manner across heterogeneous technologies, being dependent on a coordinated use of scarce resources to provide the basic connectivity bandwidth for a large number of mobile users.

5.6 Interoperation and Connectivity of Cloud and NetInf Data Centres

The focus of the *Interoperation and Connectivity of Cloud and NetInf Data Centres* use case (see section 10.1.3) is on the optimum connectivity services between the service centres of a community-supporting infrastructure (data centres acting as aggregators, repositories and proxies for a dynamic group of end-users on the network infrastructure side), which is not necessarily triggered directly by individual end-user actions. Dynamic changes of end-user groups (in number and size, in topological and geographical distribution, in involved service providers) will result in adaptation and re-optimization of the networking resources (processing and transport) utilized by the community service providers or their cloud deployment. For this case, data centre operators (as large customers of multiple network operators worldwide) will need direct access to optical transport networks in order to build, monitor and dynamically configure their own worldwide virtual network.

This allows for reaching new levels of efficiency (e.g. energy awareness on a global scale) by shifting packet routing towards optical packet switching (multi-layer routing) and addressing the traffic growth issues driven by the new scenarios like social, content and community-oriented networking.

Therefore, the *OconS* use case on Connectivity of Cloud and NetInf Data Centres addresses both the topic of 'mobility' in a new sense, such as the mobility of a serving virtual machine or a data repository (e.g. the 'video cache') to be relocated to a optimum position for serving 'flash crowds' both at the content producing and the content consuming side of the network, as well as the topic of the interconnectivity between service-integrating operators. Obviously, in the base scenario**,** this use case is triggered when *NetInf* or *CloNe* providers have the need to reconfigure and optimize their interconnecting network as pointed out in the extended base scenario of 'a single content provider in a multi-operator setting', or in the third scenario of a 'community streaming simultaneously from the same event', especially when a provider has to handle several such events simultaneously (multiple 'Alice', multiple producer crowds for independent events).

5.7 Media production

The *media production* use-case (see section 11.1.3) illustrates a new way for TV channels to subcontract the production of media content. TV channels often commission the production of pilot series to subcontractors. However, the risk of production for the contractors are high since the cost of production of pilots is large. The use-case reduces this risk by allowing the contractor to utilize media processing servers located at the TV channel premises, or at a cloud provisioned for the TV channel, connecting them with flash network slices.

Even though the use-case does not directly connect to the SAIL base scenario, in its essence it allows for specialized media processing servers to be created or utilized by third party users (the subcontractor). The same solution could be used to dynamically instantiate highly computationally intensive services in the SAIL scenario, like 3D reconstruction of images from the individual video feeds from multiple users.

5.8 Elastic Live Video Distribution

The *Elastic Live Video Distribution* use-case (see section 11.2.3) provides an operator with a way to dynamically deploy cache servers through the network in topologically advantageous points. The proposed solution moves one step beyond existing state of the art content delivery networks since those are built on static cache servers whose number cannot be easily increased. The dynamic creation of cache servers is very useful in situations where a highly popular event cannot be predicted. That is what happens in the SAIL base scenario since the popularity of Alice's video stream cannot be determined beforehand.

The Elastic Live Video Distribution use-case supports the SAIL base scenario by enabling the dynamic creation of *NetInf* nodes distributed in the network. The operator is then able to dimension the number of servers utilized to deliver the content based on the number and location of the users. This solution will not only reduce the delivery costs (by avoiding over-dimensioning and reducing utilized bandwidth), it will also provide better experience to the end-users (lower latency to reach the *NetInf* node).

6 Trends and uncertainties for the future networks

This section lists political, economic, social and technology trends and uncertainties having impact on the future networks.

The first part summarizes the key trends observed in a global inter-domain traffic measurement study active from July 2007 to July 2009 [1]. The findings present recent changes in the Internet interconnectivity ecosystem.

The second part based on expert opinions collected in Tapio Levä's master's thesis [2] looks more to the future and presents the trends and uncertainties that shape the networks during the next 5-10 years.

6.1 Key trends observed in inter-domain traffic measurements

This section summarizes the key trends that were identified in a global, two-year (July 2007 to July 2009) inter-domain traffic measurement study covering 25% of all Internet inter-domain traffic [1][3]. These trends mean significant new commercial, security and engineering challenges.

- **Evolution of the Internet core:** Over the last two years the majority of Internet inter-domain traffic growth has occurred outside the traditional ten to twelve global transit carriers. Today, most Internet inter-domain traffic (volume-wise) flows directly between large content providers, hosting / CDNs and consumer networks because of cost and, increasingly, performance reasons.

- **Consolidation of Content:** Most content has migrated to a relatively small number of large hosting, cloud and content providers due to significant economies of scale. Out of the approximately thirty-thousand autonomous system numbers (ASNs) in the default-free BGP routing tables, 30 ASNs contribute a disproportionate average of 30% of all Internet inter-domain traffic.
- **Consolidation of Application Transport:** The majority of inter-domain traffic has migrated to a relatively small number of protocols and TCP / UDP ports, including video over HTTP and Adobe Flash. Other mechanisms for video and application distribution, like P2P, have declined significantly. One reason for this is that browser is increasingly used as an application front-end. Contrary to this last finding, Anderson and Wolff [4] claim that “the web is dead” because Internet services are more and more accessed through specialized apps. Nevertheless, these apps still use the same protocols as web. Additionally, the importance of video traffic has increased, and continues to increase in the future too [5].
- Market **focus shifting to higher value services:** Dropping prices and revenue of the wholesale transit has lead to commoditization of basic transport services. Therefore the revenue possibilities are increasingly in higher value services, like virtual private networks (VPNs) and content delivery networks (CDNs).
- **Transition from focus on connectivity to content:** The market share of CDN traffic is increasing, due to which CDN providers present a strong competitor to traditional Tier-1 transit providers. To avoid becoming only bit pipes tier-1 ISPs are moving to the CDN market through vertical integration with CDN players. Different information-centric research activities (like SAIL itself) underline this development.
- **New economic models are emerging:** Old global Internet economic models are evolving while new entrants are reshaping definition and value of connectivity. This has lead to changes in peering strategies, increased peering and experimentation with new economic models like paid peering and partial transit, which complicates the interconnectivity ecosystem [6].

6.2 Key trends and uncertainties based on expert opinions

This section summarizes the key trends and uncertainties that were identified in three brainstorming sessions organized in the autumn of 2008 to provide input for creating four possible scenarios for the Future Internet. The scenarios are not presented here, interested readers are referred to the thesis [2] or the paper written based on that [7].

Each brainstorming session had 6-8 Finnish academics/industry experts who represented different stakeholders. To cover all the important macro-environmental factors affecting the future Internet, the analysis was divided into political, economic, social, and technological domains, known also as the PEST framework. Similar forces were combined and they were mapped to a flip chart matrix based on their importance and uncertainty.

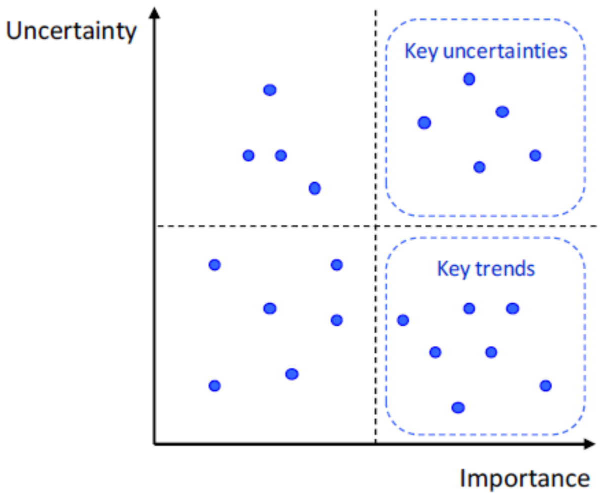

Figure 6: Flip chart matrix used in evaluating the importance and uncertainty of identified forces [7].

6.2.1 Key trends

Key trends are important factors that are certain or very likely to realize and have significant impact on the future networks. They are assumed to be valid with a reasonable probability for the next 10 years. The presentation is divided into four categories based on the PEST framework.

Political/Regulatory Trends

The society will be increasingly dependent on the Internet

Economy, administration, education and industry move their operations increasingly to the Internet and manual fall-backs in problem situations are disappearing. This raises governmental interest in regulative control and re-regulation.

The world (and the Internet) is moving from unipolar to multipolar

The U.S.-centred western world loses its dominant role since the rise of China, India, and other developing nations scatters the power around the globe. Additionally, the next two billion Internet users come mostly from the third world and developing nations.

The usage and allocation of spectrum will be more market-based

Increasing mobile Internet usage channels more spectrum for Internet access. Spectrum usage will be more effective and spectrum auctions are used in most countries.

Environment and energy will be more important

Environmental awareness increases and energy consumption is controlled and regulated stricter.

Economic/Business trends

The world is moving from products to services

The money is on the services because producing goods is highly competed on the global space. The Internet speeds up this development.

Hyper-competition between Internet services promotes hyper-usability

Hyper-competition between content service providers over end users leads to hyper-usability, *i.e.*, high-quality services. This is due to close to zero switching costs meaning that end users can easily switch from one content provider to another.

Using ICT becomes low-cost compared to manual alternatives

Cost reductions and possibility for rationalization of business processes drive adoption of ICT in every field of economy.

Power consumption becomes a cost driver in ICT

Awareness of the environmental effects of ICT increases at the same pace with improving performance and power consumption of devices. Therefore, energy efficiency becomes an important design criterion. “Green ICT” is also seen as having marketing value.

Globalization continues

This old trend continues to hold true since countries depend more and more on each other and borders disappear. However, in the future globalization will be stronger in service and knowledge industries than in manufacturing industries.

Social Trends

The Internet is integrating deeper into everyday life

Mapping between the real and virtual worlds tightens and people are increasingly able and willing to use Internet services. Tighter integration creates need for improvements in security, trust, and privacy.

Desire for all around availability increases

People are used to being reachable all the time with their mobile phones and now the same level of accessibility to email, social networking sites, and instant messaging is generating a demand for mobile data services. This is supported by the increasing use of location and context information.

Social networking will be faster and stronger

Social networking services gain importance and affect how people communicate and consume. For example, the increasing usage of ratings and suggestions from other consumers changes buying behaviour.

Content creation will be more user-driven

The easiness of creating and sharing content in the Internet drives to YouTube and Wikipedia style of services where users are active participants and not just passive consumers.

Internet generation continues to drive Internet usage

Young people are eager to adopt new services whereas old people are not able to do that. This preserves the generation gap between the Internet generation and older people.

Technological trends

Mobile always-on Internet connectivity increases

The Internet will be used more and more with small, portable devices like mobile phones, tablets, and ultra-portable PCs. Additionally, for many new users, especially in developing regions, mobile connectivity will be the first and only access method.

Performance continues to improve

Processing power improves, optical transmission boosts transfer rates and storage capacity increases. These improvements can also be seen in better price-performance ratios.

Complexity of software, services and architectures increases

Patch-on-patch tradition and new requirements increase the complexity of networks. At the same time usage of new applications is still too complex for most users. This raises usability and reliability questions to a new level.

Diversity of networks and devices increases

The *Internet of things* spreads ubiquitous computing quietly and increases the amount of hosts significantly. The diverse device base is connected to the Internet with a variety of access technologies. Furthermore, machine-to-machine communication brings new requirements for networking.

Remote management of network and home devices increases

Managing prolific and more complex device base will be carried out more and more remotely. This will happen both in households and in the core network.

6.2.2 Key uncertainties

Key uncertainties are important factors with uncertain direction and impact on the future networks. All the key uncertainties are listed in Table 4 and introduced in-depth in the following sections.

Most important key uncertainties

Two uncertainties rose as more important than the others, and thus they were chosen to be the key determinants of the scenarios created in the Master's thesis.

Network structure

The future Internet may either remain one whole network or it may fragment into many networks. The characteristics of these two extremes – one network vs. fragmented network – are presented by relating questions listed below.

- Will there be free connectivity in the Internet?
- Will the Internet be able to scale up?
- Will the (single) Internet be suited to all purposes of use?

Although the Internet consists of many different networks they still form one Internet where, at least theoretically, every host is able to connect to every other host only by knowing their IP addresses. The flexibility of the Internet protocol suite has allowed an all-IP trend meaning that the IP technology is used for various networking needs including telephone and video services. This development underlines the possible cost savings that the economies of scale enable when only one network infrastructure is used. All the traffic flows in the same wires and diverse requirements of different traffic types can be taken into account at the network level. Fundamental prerequisite of one network to be possible can be expressed as a slogan "one size fits all".

Fragmentation would mean that free end-to-end connectivity would be questioned. Extensive usage of NATs, firewalls and other middle-boxes disturb already today the end-to-end connectivity. Due to the importance of connectivity complete separation of networks does not seem feasible but the connectivity may be heavily restricted so that all the traffic between networks travels through gateways. The fragmentation does not need to happen in the physical level but it can as well – or even more probably – happen in the service level through overlay networks. These overlays borrow only the connectivity from the Internet and use their own, possibly proprietary protocols to fulfil requirements that the core Internet architecture is not capable to satisfy. These solutions, however, break the Internet architecture intentionally and thus increase the complexity of the Internet ecosystem.

Table 4: Key uncertainties

	Network structure	
Single network	o o o o o	Fragmented network
	Openness of applications, services and hosts	
Open	o o o o o	Closed
	Will Internet face a larger collapse?	
No	o o o o o	Yes
	Where will the intelligence be located?	
	Client vs. Server	
In clients	o o o o o	In servers
	End points vs. Network	
In end points / edges	o o o o o	In the network
	What will be the dominating business model in the Internet economy?	
Ad-based model	o o o o o	Other models
	How will solutions for trust, security, and authentication be implemented?	
Openly in the architecture	o o o o o	Somewhere else in closed form
	Will the traffic be treated neutral?	
Yes	o o o o o	No
	Amount of standardization: standards vs. proprietary solutions?	
Standards	o o o o o	Proprietary solutions
	Where will the standardization happen?	
IETF	o o o o o	Industry-driven forum

Scalability (from a technical viewpoint meaning a large enough address space, fast enough routing protocols and algorithms, and small enough energy consumption) is one issue that can be solved either in the level of the Internet architecture or by building separate networks. The applicability of the Internet to every imaginable and non-imaginable purpose of use is another type of scalability issue that affects substantially the level of fragmentation. For instance, end-to-end multicast and end-to-end quality of service cannot be supported well by the best effort type of service. Increasing real-time (video) traffic is one of those applications that have brought demand for specialized network fragments called content delivery networks (CDNs).

Openness of applications, services and hosts

Applications, services and hosts may either be open like PC's and their open source software or closed like Apple's iPhone with proprietary software. These both worlds – open and closed – can be explained by relating questions listed below.

- Are the hosts freely programmable?
- Are users willing to be dependent on a single actor?
- Do users prefer bundling or buying separately?

The world of open applications, services and hosts is the world of PC-like multipurpose devices. A single device is used to access various kinds of applications and services and is able to suffice most purposes of use. Successful and open standardization, particularly in the application level, and high availability of open source software mean that everyone has in principle the possibility to program own applications. Closed applications, services and hosts, for one, are optimized for some usages (or even for a single use). Specialization may enable better usability and fewer bugs, since all the use cases are predictable, but it restricts versatility. Security is another issue that is much easier to take into account in closed systems. Actually, Zittrain [8] sees security nuisances of open or as he expresses, generative systems, one of the most important drivers for the closed world.

All the causal factors relate to the question: who has the control over users' actions? In the open world user is the king of the hill. The Internet offers wide selection of services and user has the freedom of choice. He can install whichever applications he wants and is not locked in to some service for a long period of time. Thus the open world is naturally competitive. Respectively, in the closed world, user has handed the control to some actor. Providers can most easily acquire this kind of control position through end user devices (*e.g.*, Apple's iPhone and Microsoft's xBox 360) that cannot be changed to new ones as often as applications and services. Bundling devices, applications, services and even networks together is another method through which a stakeholder may try to get customers locked in and dependent on a single actor. From user perspective bundling reduces the amount of purchase decisions and may thereby be an easier choice especially for technology non-enthusiasts.

Other key uncertainties

The other key uncertainties are presented briefly in this section.

Will the Internet face a larger collapse?

A larger collapse is defined here as an incident that would black-out parts of the Internet for a short period of time, cause severe economic losses and reduce people's trust on the Internet. In this respect it has significant analogies with the financial crises. A collapse is seen as a possibly disruptive event that wakes up the Internet community, especially decision-makers, and thus disrupts Internet's evolutional development by speeding up the implementation of new technical or regulatory means and by increasing regulative control. Fear for collapse drives for pre-emptive actions like improving the resiliency of networks and services.

Where will the intelligence be located?

Originally the Internet was a dumb network connecting smart hosts. The hosts were equal in their capabilities and roles. Client-server model used widely in the Web differentiated the roles of the hosts. High level of intelligence in clients indicates more important role of peer-to-peer model, whereas significant amount of intelligence in servers speaks for client-server model. It is also constantly questioned if any intelligence should be inserted to the network. Thus the question here is divided into two questions: 1) client vs. server and 2) end points vs. network.

What will be the dominating business model in the Internet economy?

Simple and "free" ad-based business model has been by far the most successful revenue model when Internet services are considered. Transaction-based business model, like paying with PayPal or credit cards, has been mostly used when physical goods are sold through the Internet. Additionally, subscription-based model would be highly interesting to companies and simple enough for users. Thus the big question here is, will the Internet business be mostly ad-based or do other models break through?

How will solutions for trust, security and authentication be implemented?

Lack of trust, security and authentication is a recognized challenge that needs to be tackled somehow, at least in the case of mission-critical applications. Universal, open solutions built in the architecture are a reasonable option, but closed solutions relating for example to separate network or provider-controlled solutions in a closed architecture are other choices.

Will the traffic be treated neutral?

Principle of net neutrality requires that all content, sites, and platforms are treated equally [9]. In a neutral network traffic flows related to for instance e-banking, video streaming, peer-to-peer file sharing or emailing are not treated differently but they all have same priority level from the network perspective. Blocking content and communication is one of the things that violates net neutrality

Amount of standardization: standards vs. proprietary solutions?

The Internet architecture relies heavily on open standards called RFCs (Request for Comments). On the application and service level, proprietary solutions have, however, an important role. For example, some important network overlays, VoIP network Skype and peer-to-peer file sharing network BitTorrent, are based on proprietary solutions. Standards allow competition, while proprietary solutions enable emergence of monopolistic pockets.

Where will the standardization happen?

Internet-related issues have been traditionally standardized in the IETF. After the commercialization, other forums have emerged including W3C (The World Wide Web Consortium) concentrating on Web standards and 3GPP (The Third Generation Partnership Project) working for the third and fourth generation mobile phone systems. Standardization could also be done in the industry-driven forums that would be open only for part of the Internet industry.

6.3 Implications to SAIL

Labovitz et al.'s [1] measurements concretize the common wisdom of Internet industry about the increasing value and power of content (companies) and the changes in the interconnectivity ecosystem. These trends had already impact when the research topics of the project where chosen, but it's necessary to keep them in mind when the actual architectures are designed. Many technological solutions of SAIL are planned to be provided by ISPs but in order to succeed, they need to be accepted also by content providers (and other actors) whose power position seems to continue to grow.

Trends and especially the uncertainties identified by experts [2] relate closely to the technological choices that will be made when the SAIL solutions are designed. Avoiding choices that exclude the other outcome of an uncertainty is recommended to allow the solutions to match the dynamically changing world. Besides, the wide scope of these trends and uncertainties emphasizes the importance of political, economic and social factors in addition to technical enablers.

7 Introduction to scenario and use case business analysis

7.1 Terminology

The key terms related to the business analysis of use case scenarios are explained below.

- **Scenario** (or use case scenario) is a wider application area, where a particular technology can be useful. A scenario describes the environment in which a set of use cases can be defined.
- **Use case** describes how a particular technology can be used to solve a problem or satisfy a need.
- **Industry architecture** (see [10] and [11]) results as actors take on roles and establish business interfaces among each other.
    - **Actor** is any person or organization, which has an interest in, provides resources for, or is affected by the use case. An actor can take multiple roles.
    - **Role** is a set of activities and technical components, the responsibility of which is not divided between separate actors. A role can contain multiple technical components of technical architecture.
    - **Business interface** presents a business relationship (*e.g.*, contracts and transactions) between actors.
- **Technical architecture** describes the technical implementation of a use case.
    - **Technical component** is a collection and realization of technical functionalities, including the technical interfaces to other technical components.
    - **Technical interface** presents the technical relationship (*e.g.*, protocols or radio interfaces) between components.

Figure 7 presents the notation used in combined technical and industry architecture figures [12]. The technical and industry architecture are combined into the same figure to show the mapping and interdependencies between them. A single technical architecture may allow multiple co-existing industry architectures.

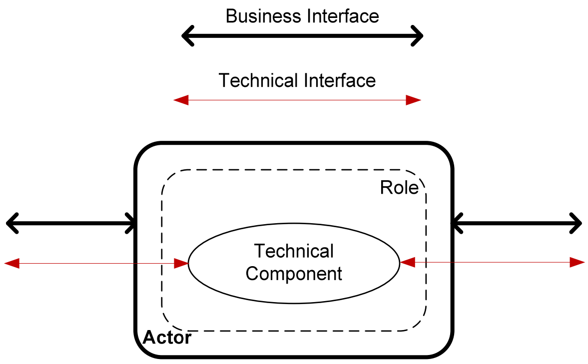

Figure 7 Notation used in the combined technical architecture and industry architecture figures (from [12]).

7.2 Structure of the use case scenario business analysis

Each technical work package (*NetInf* (WP-B), *OConS* (WP-C), and *CloNe* (WP-D)) has identified relevant scenarios in their own technical area. The identified scenarios and use cases are described technically, and then one use case per each scenario is analyzed from the business perspective. The goal of this business analysis is to understand the incentives of all the stakeholders concerning the use case. To achieve this, the stakeholders and their relationships, i.e. technical and industry architecture, are identified, the proposed SAIL solution is compared with competing solutions, and the pros and cons of SAIL solution per each actor are studied. Section 8 introduces the template, which is then applied to the WP-specific scenarios presented in sections 9-11.

8 Template for analyzing a scenario and its use cases

This section is a template for WP-specific sections 9-11.

8.1 Scenario 1: <Scenario name>

<Give a **technical** overview of the scenario that sets the scene; it describes the area of concern with its main characteristics and features.>

8.1.1 Challenges

<List challenges that are encountered in this context, and where SAIL solution can help.>

8.1.2 Market potential

<Analyze the market potential and/or (business) relevance of the scenario.>

8.1.3 Use cases

<Give a **technical** description of each use case under the following subheadings.>

Use case 1: <Use case name (list the business analysis use case as first use case)>

Use case 2: <Use case name>

8.1.4 Business analysis of use case 1: <Use case name>

<Analyze one of the use cases from business perspective using the suggested tools. Here in the beginning, describe the business relevance of the use case.>

Actors and roles

<Identify actors and roles in the use case. Start from roles since they are the key building blocks. In defining roles, separation is a virtue.>

Technical and industry architecture

<Describe both the technical architecture and the industry architecture of the use case, and their mapping to each other using suggested notation and figure template >

Comparison to competing solutions

<Evaluate the use case comparing it to the other technical solutions solving the same problem. Use the suggested table template below.>

Criteria	<Comp. tech 1>	<Comp. tech 2>	<Comp. tech 3>	<Comp. tech 4>	<Comp. tech 5>
<Criteria 1>	Medium	High	Medium	Medium	Low
<Criteria 2>	Medium	Low	Medium	Medium	High

Pros and cons for key actors

<Evaluate the attractiveness of the use case from the perspective of each relevant actor. Use the suggested table template below.>

Actor	Pros	Cons
<Actor 1>	• <pro 1>	• <con 1>
<Actor 2>	• <pro 1>	• <con 1>

9 Network of Information (NetInf) scenarios (WP-B)

WP-B has three scenarios: 1) NetInfTV, 2) Next generation mobile networks, 3) Developing regions.

9.1 Scenario 1: NetInfTV

This scenario represents a heavy content distribution perspective of NetInf. The scenario focuses on video distribution even though NetInf itself is not content type specific. The scenario covers all the different types of video traffic delivered over the Internet. The video content is divided into three categories:

1. Live video delivery (*e.g.* IPTV, WebTV, …)
2. High quality (managed) video delivery (*e.g.* Video-on-Demand, catch-up TV)
3. Best-effort video clip delivery (*e.g.* YouTube). Content is partly user-generated.

The variety of video delivery applications involves a wide range of actors and business relationships. Video delivery also comes with a significant cost to these actors because of need for large storage capacity and high bandwidth. Technically, video delivery scenario is also challenging from at least two perspectives: 1) quality-of-service, and 2) access right management.

9.1.1 Challenges

There are technical challenges that lead to expensive solutions for IPTV delivery. The video streaming is a consumer activity and the time-shifting aspect of user behaviour requires the IPTV provider to support individual video streams from the back-office data centre servers to IPTV devices. This requires high bandwidth throughout the network. *NetInf* can offer alternative sources for video distribution and utilize in-network caching to lower network load, while enhancing end-users' quality-of-experience. From the content providers perspective *NetInf* eases video publication and access right management.

9.1.2 Market potential

This section describes the current market situation and the market potential of Internet video traffic. Cisco [5][13] has studied the current demand and forecasted the future trends of video traffic in the Internet in its Visual Networking Index. Some of the key findings are listed below:

- By 2014, the various forms of video (TV, VoD, Internet Video, and P2P) will exceed 91% of global consumer traffic. Internet video alone will account for 40% of all consumer Internet traffic by the end of 2010, and 57% by 2014.
- Advanced Internet video (3D and HD) will increase 23-fold between 2009 and 2014. By 2014, 3D and HD Internet video will comprise 46% of consumer Internet video traffic.
- Video communications traffic will increase sevenfold from 2009 to 2014.
- Real-time video is growing in importance. By 2014, Internet TV will be over eight percent of consumer Internet traffic, and ambient video will be an additional five percent of consumer Internet traffic.
- Video-on-demand traffic will double every two and a half years through 2014.
- While the overall application mix is shifting toward video, video is undergoing internal shifts of its own. In particular, video delivery with real-time constraints is growing in importance. This can be seen in Figure 8 below, which shows the proportions of different types of video traffic of the total video traffic (excluding P2P downloads). Please check the explanation of the terms below the figure.

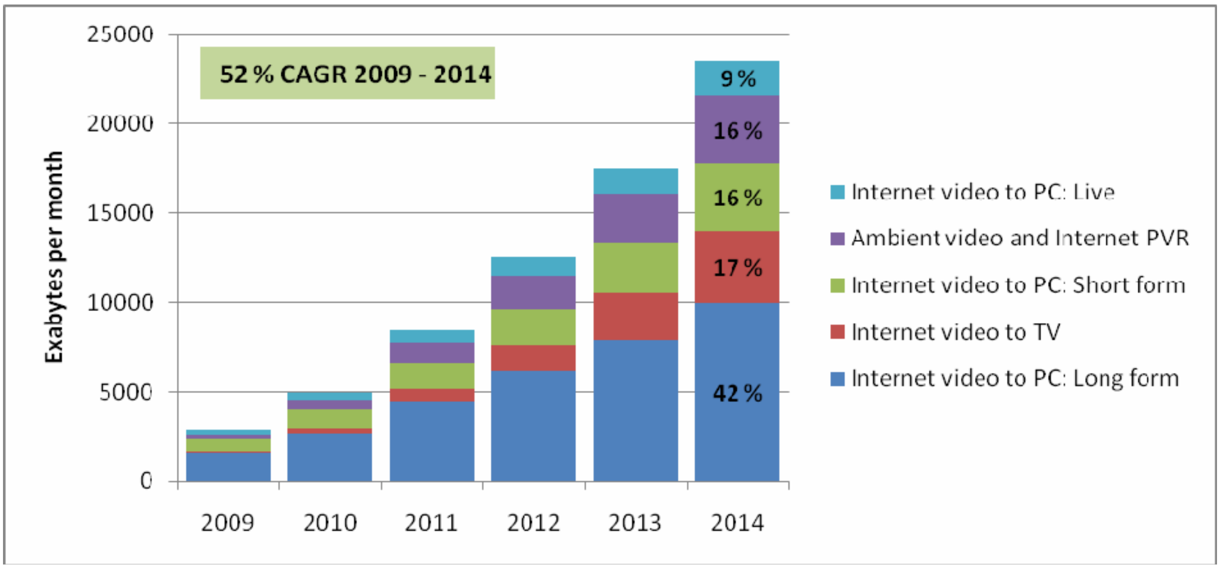

Figure 8. The proportions of different types of video traffic of the total video traffic (explanations of the categories below). Data from [13].

- **Internet Video to PC**: online video that is downloaded or streamed for viewing on a PC screen (*e.g.* WebTV, VoD). It excludes P2P downloads.
- **Internet Video to TV:** video delivered via Internet to a TV screen, using an Internet-enabled set-top-box or equivalent device (*e.g.*, IPTV, WebTV, VoD).
- **Short form:** User-generated video and other video clips shorter than 7 minutes in length (*e.g*. YouTube video clips).
- **Long form:** Video content generally greater than 7 minutes in length (*e.g.* movies, TV programs).
- **Live TV:** Peer-to-peer TV and live television streaming over the Internet (*e.g.* IPTV, WebTV).
- **Internet PVR:** Recording live TV content for later viewing.
- **Ambient video:** Nanny-cams, home security cams, and other persistent video streams.

If we move closer to the market potential of NetInf, or content-centric networking in general, the proportion of cacheable content limits the potential of any technology which is based on content caching. “Cacheability” can be understood from the technical perspective covering, *e.g.* the distribution of the content popularity and the dynamism of the content, but it should not limit on technical aspects only. Caching typically reduces the control of the content owner and makes it more difficult to deliver accurate information about the usage of the content. Therefore business considerations can significantly reduce the attractiveness of caching, the extent of which should be studied to understand the significance of this possible challenge.

9.1.3 Use cases

Two different use cases are identified in the NetInfTV scenario. In use case 1 *NetInf* caches are used to create a decentralized, content type independent, CDN. Use case 2 is a single-stakeholder use case where an Internet Service Provider uses *NetInf* to optimize the usage and video distribution performance of its own network. Later in Section 9.1.4, use case 1 is analyzed from the business perspective.

Use case 1: NetInf as a decentralized CDN

This use case presents a competing solution for legacy CDNs where the cache ownership, intelligence for redirecting requests (*e.g.* DNS), and cache selection are typically centralized, *i.e.* controlled by single stakeholder. In the *NetInf* CDN the cache servers are not owned and operated by a single stakeholder, that responsibility is shared between multiple actors.

Although this role can be played by any type of actor, for example data centre providers or even end-users, the focus here is on the case where the role is taken by Internet Service Providers. This allows the content to be served as close as possible to the end-users, which decreases the latency and thus improves end-user's quality of experience.

Use case 2: NetInf as a part of ISP multi-play solution

Use case 2 can be seen as a particular instance of use case 1, where the ISP acts as the *NetInf* CDN provider (so called NetInfTV provider). The use case may focus on managed video delivery services such as IPTV, Web TV, VoD and Catch-up TV in order to limit actors, but may also include video delivery services offered by over-the-top players to the ISP's customers, in the context of a business agreement. In the latter case, footprint extension needs to be considered (*e.g.* in the form of an alliance between "local" ISPs at least), as the content provider may be reluctant to have to set up several business agreements for its content delivery across a given region (*e.g.* country), as discussed in the sequel in the context of use case 1.

9.1.4 Business analysis of use case 1: NetInf as a decentralized CDN

Using *NetInf* to build a decentralized CDN seems like a promising application area because the large size of video files offers an immediate chance to save in transport costs if the content is delivered close from the end-user. However, decentralization and multistakeholderism of the use case present significant challenges. Let's try to understand these challenges better by analyzing the technical and industry architecture of the use case.

Actors and roles

Here we present a comprehensive list of actors and roles in this use case. They are used in the following technical and industry architecture figure.

Actors

- End-user
- Internet Service Providers, including
  - Internet Access Provider (IAP)
  - Internet Backbone Provider (IBP) (national, international)
- CDN Provider
- Content Provider
- NetInf CDN Provider

Roles

- **Usage:** watching video content.
- **Content provisioning:** providing the video content requested by the end-user.
- **Access provisioning:** providing mobile and/or fixed Internet access to end-users (customers).
- **Backbone connectivity provisioning:** offering global Internet connectivity.
- **Content management:** aggregating, indexing and securing the content, also acting as a single point of contact for the content providers.
- **Content caching:** owning and operating NetInf cache servers.
- **Location information resolving:** resolving the nearest location of the content.
- **Search:** resolving the NetInf ID of the content (not visible in architecture figures).

Technical and industry architecture

Figure 9 shows the technical and industry architecture of this use case. The new actor, *NetInf* CDN provider, acts as a single point of contact for content providers who want their content to be distributed using a *NetInf* CDN. The *NetInf* CDN provider has only the directory of content itself so it negotiates agreements with Internet Access Providers who have *NetInf* caches in their networks. Although content management role is not completely necessary, it dispenses content providers from separate negotiations with all the *NetInf* cache owners in the targeted footprint and thus makes the use case more interesting from a content provider's perspective. It could actually be interesting to study if content management and content provider business interface could be provided by an alliance of ISPs through a lightweight technical solution.

Based on the contract terms with *NetInf* CDN provider, the IAP caches that (video) content for which the content provider has paid to be cached. The paid content can be identified by the hash of publisher's public key (content for priority treatment gets published with a specific public key, an alternative solution could be using the metadata of the content). IAP also controls the resolution server which defines the best location from which end-user content requests are to be served.

An end-user pays to his IAP only for the Internet access whereas the content provider charges end-user for improved end-user experience depending on its revenue model, *e.g.*, directly in form of monetary payment in subscription or transaction model, or indirectly through end-user's attention in ad-based model. The revenue models of content provider are not studied further here, the reader is referred to [14] for further discussion.

The figure also includes the backbone connectivity provider role connecting access networks and cache servers to each other. This is important because of its gatekeeper-type of position in the middle of traffic flows. If local caching reduces the profits of those who are operating backbone networks, they may see the use case as a negative thing. On the other hand, in-network caching may provide incentives for new business models for transport that could be more sustainable for carriers that current which may be problematic due to high ratio asymmetry.

The presented architecture excludes caching in backbone networks due to simplicity, even though it could be desirable at interconnection points.

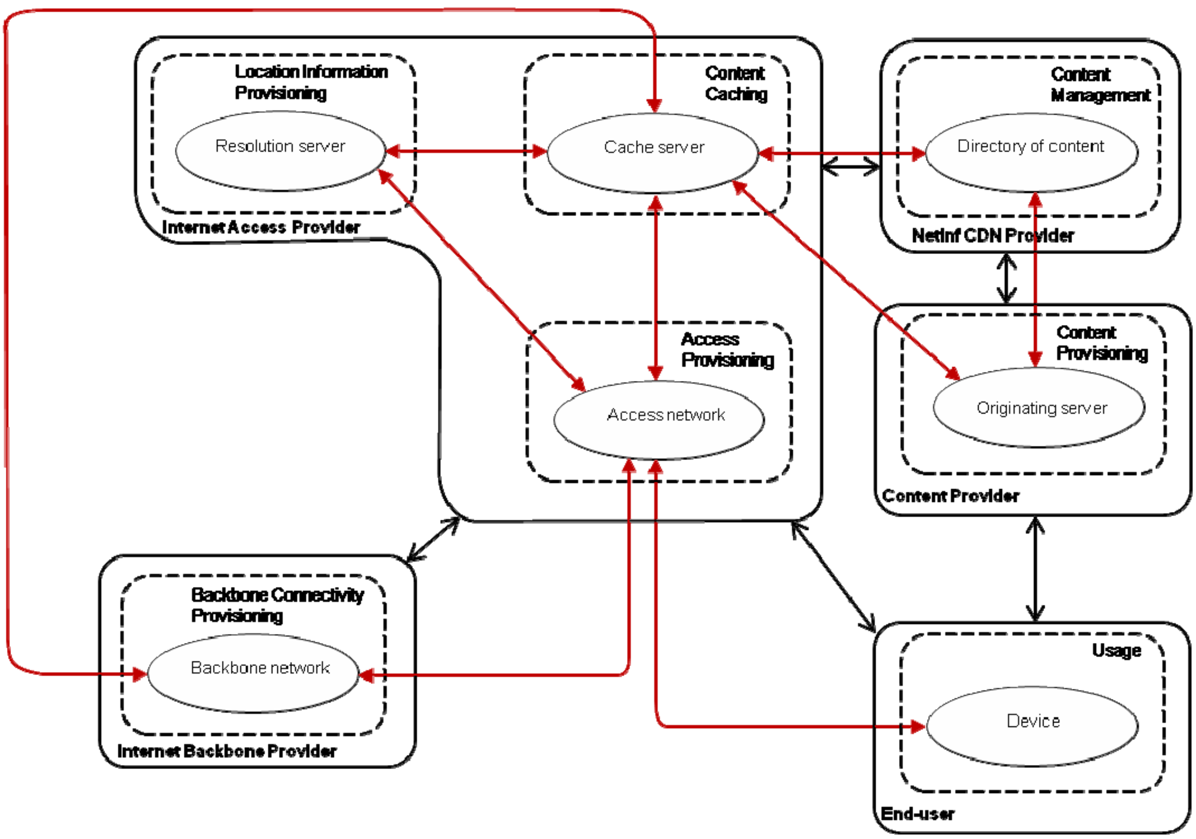

Figure 9: Technical architecture and industry architecture of NetInf CDN use case

To demonstrate the usefulness of this notation, Figure 10 shows the corresponding technical and industry architecture of the legacy CDN model. The high-level technical architecture (*i.e.*

roles needed) is almost similar but the industry architecture changes. In legacy CDN model, the role of the CDN provider is bigger, whereas the role of IAP is much smaller. Nevertheless, also here the content provider interface is handled by the CDN provider, not by IAP.

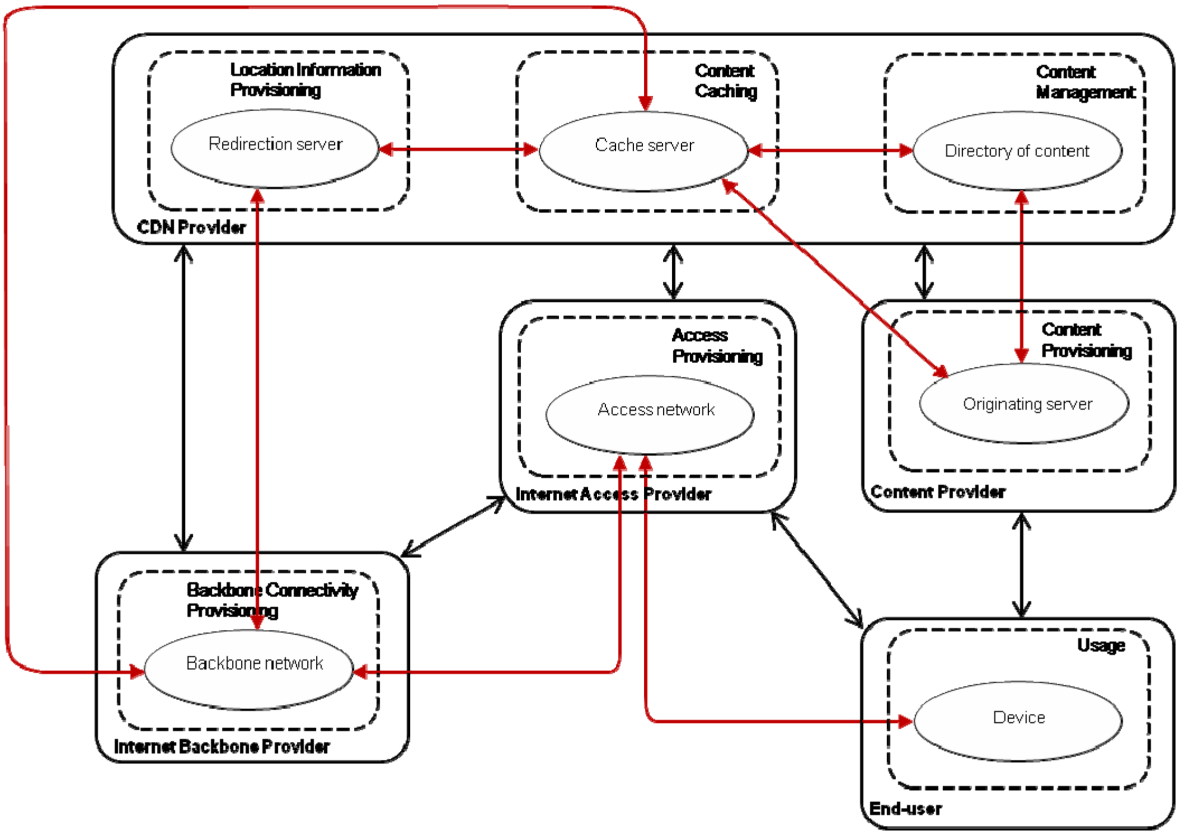

Figure 10: Corresponding value network configuration of the legacy CDN model (*e.g.*, Akamai)

Comparison to competing solutions

In video delivery the final choice of content delivery method is done by content providers who have at least four high-level alternatives for NetInf CDN:

1. Traditional CDN (Akamai-style)
2. Classical client-server model with own or rented/outsourced servers (pure transit used)
3. Do-It-Yourself CDN , *e.g.*, Google with Global Google Caching (combination of transit and peering agreements, and caching)
4. P2P distribution

The content provider's choice of the content delivery model can be based on multiple parameters that may be valued differently by different providers. The comparison parameters are collected into Table 5 and explained in the text below that. The traffic light colour code is used to visualize the goodness of the values (green = desirable, red = undesirable).

Table 5: Comparison of NetInf CDN to competing solutions

Criteria	NetInf CDN	Traditional CDN	Client-Server	DIY-CDN	P2P
Cost for content provider	Medium	High	Medium	Medium	Low
Cost for Internet access provider	Medium	Low	Medium	Medium	High
Scalability	High	Medium	Low	Medium	High
Latency	Low	Medium	High	Medium	Medium
Application-independency	High	Low	High	High	Medium
Availability of controllable QoS	Medium	High	Low	Medium	Low
Level of expertise needed	Low	Low	Medium	High	Medium
CP's control over content	Low	Medium	High	High	Low
Copyright protection	High	Medium	High	High	Low

The cost efficiency, *i.e.* cost level compared to achievable quality, is a very important parameter when video content providers choose the content delivery model. Bill Norton [15] has studied the costs of the presented four alternatives to *NetInf* CDN. Based on Norton's study, the biggest cost component for video content providers is transit expense. Therefore P2P which moves almost all of the distribution costs to the Internet access providers is clearly the cheapest one for content providers. The cost of *NetInf* CDN for content providers should be lower than the cost of traditional CDNs because distributed ownership of caches allows multiple *NetInf* CDN providers to use same cache servers which increases the competition compared to the traditional CDN market.

In general, popular video content can create scalability problems if the content requests are served from small number of locations. Thus *NetInf* CDN and P2P scale better because the amount of storage locations is higher than in the other models. One of the key benefits of *NetInf* CDN is that cache servers are located in IAPs' networks, which means that the content can be served closer to end-user than in any other model. This decreases the latency of video distribution leading into better end-user experience.

The inherent benefit of *NetInf* compared especially to traditional CDNs is that cache servers are application-independent. Thus the cache servers used for video traffic can also be used for other purposes. This decreases the costs of *NetInf* CDN both for the cache server owners and users.

Maybe the biggest uncertainty in the comparison table is how well the quality of service can be controlled and thus guaranteed in *NetInf* CDN. If quality of service cannot be guaranteed, content providers may not be willing to risk their customer-relationship with end-users by choosing unreliable solution. Currently QoS differentiation is possible in *NetInf*, but only at the portal level.

Pros and cons for key actors

The key actors for this use case are end-user, content provider, IAP, IBP, CDN Provider and *NetInf* CDN provider. The pros and cons for each of them are collected into Table 6.

End-user does not care which content delivery model is used, as long as the quality of end-user experience is good and the cost level low.

Content provider is interested in *NetInf* CDN if they can get comparable or better quality of service to CDN or DIY-CDN with lower costs. The biggest question mark is if *NetInf* CDN provider can control and guarantee service level. If that is not possible, content providers may be willing to stick with a solution where they can better control both content distribution and content itself.

Internet access provider may be interested in installing *NetInf* caches because in-network caching can enable transit cost savings. This use case, however, introduces also a way how IAPs could monetize their investment and get larger share of revenue from the growing content delivery market. On the other hand, purchasing *NetInf* caches is a big investment that may turn out to be unprofitable if *NetInf* does not fly in the market. Nevertheless, the risk of investment is actually smaller because *NetInf* cache servers can also be used for IAPs internal purposes like, for example, in multi-play use case (use case 2).

Local caching, direct peering between content providers and IAPs, as well as traditional CDNs all decrease **Internet backbone provider's** transit revenues. Thus IBPs may oppose deployment of NetInf, even though they are also in excellent position to move into *NetInf* CDN business.

Because **NetInf CDN provider** is a new actor, the analysis focuses here on the pros and cons for a new company to enter into the market and taking the content management role (possibility that established actors take that role has been discussed inside the analysis for those actors). *NetInf* CDN provider resembles the virtual mobile network operator, because also there the actor's main responsibility is to handle customer relationships, and the actual technical resources are owned by other actors. Thus entering to the market can be fast and does not require heavy upfront investments but the service quality can be controlled only indirectly through service level agreements.

Table 6: Pros and cons of NetInf CDN use case from the perspective of key actors

Actor	Pros	Cons
End-user	• Improved quality of end-user experience	• Installation of browser add-on or other piece of software may be needed
Content Provider	• NetInf CDN provider's market power is probably smaller than traditional CDN provider's • Cheaper than traditional CDN due to competition and standardization • NetInf CDN provider negotiates deals with IAPs on behalf of content provider	• Possibility for guaranteed QoS is questionable • Worse control over the content as in CDN model • Less control over content distribution than in client-server and DIY-CDN models
Internet Access Provider	• New business opportunity • Possibility to take larger share of revenue from content delivery market • Transit cost savings possible • NetInf cache servers can be used also for internal purposes (*e.g.*, in triple-play use case)	• Risk of investment • Relationship to content providers is handled by another party (NetInf CDN provider) • Removes the competitive advantage of low latency from IAPs "own" content
Internet Backbone Provider	• IBP can move to CDN market by building its own NetInf CDN	• Local caching at IAP can decrease IBP's transit revenues
CDN Provider	• Can get closer to end-users than with their own cache servers • Is in natural position to become NetInf CDN provider because of existing content provider customerships	• NetInf CDN is a direct competitor to traditional CDNs • Changing to NetInf CDN would reduce CDN provider's market power because of decreased control
NetInf CDN Provider	• Market entrance does not require large CAPEX	• Service quality can be controlled only indirectly through SLAs

9.2 Scenario 2: Next generation mobile networks

The starting point of this scenario is the current cellular networks that are still evolving from having been primarily used for telephony to mobile data services. Such networks also evolve from a centralized and static architecture into a more distributed and dynamic one, for example with femtocells and more dynamic routing. Moreover, the mobile entities get more advanced and have the possibility to connect through different access technologies with multiple interfaces.

In this context, content and service delivery to mobile users should be reconsidered. To optimize the user experience the content and services may need to be adapted to the user terminals and access networks. Since users are mobile the content and service placement should be more dynamic than in the fixed Internet and adapt to the variations of the geographical user population and service popularity. This allows more efficient usage of the network resources and higher QoE for the users due to lower latency. In particular this is important for more complex services than basic content delivery, where a service consists of programs running within the network and the interactions between terminal and servers may add multiple RTTs of latency. The mobile operators may use their knowledge of the network state and user locations to dynamically move services and content in order to optimize the user experience and network resource usage.

The concept of mobility can be extended to cover mobile objects which are able to move between mobile hosts while maintaining their reachability and in some cases also continuity of ongoing communication sessions. Examples of such objects are content files, real-world objects, sensors, environmental parameters that can be measured by sensors, and persons. A multi-hop scenario can be considered, where a mobile object attaches to a mobile host, which in turn attaches to a mobile network. Alternatively, a mobile object can be detected by a mobile sensor, see the mobile sensor use case below.

9.2.1 Challenges

Dynamic service delivery with mobility of both services and users

Mobile networks are successful in providing seamless mobility for moving users, but support for moving services and content remains a challenge. Ideally seamless delivery should be provided both during user mobility and during service or content mobility.

Variety of access technologies

As terminals get equipped with multiple interfaces there are possibilities to benefit from multi-homing and multipath routing. With the current addressing schemes and the various middle-boxes present in the network it is difficult to provide solutions to multihoming and multipath transport for mobile users.

Variety of Services

The next generation mobile communication network should not be limited to certain services or applications. Although efficient content distribution should be supported, other applications such as interactive communications should be supported as well. Also, the network should be flexible to allow new applications and innovations from different players.

9.2.2 Market potential

This scenario is really broad covering the whole field of mobile communication. Therefore the market potential is not analyzed on the general level but on the level of more specific use cases.

9.2.3 Use cases

This scenario is a very broad one, due to which it has six use cases which are presented below.

Use case 1: Video with added local information

Imagine the following scenario: a video is distributed from a central server to a number of mobile users. However, the video is enriched with some information that is generated locally, close to the users receiving the video. Examples of such local information might be weather info, traffic jam alerts of even locally generated video such as web cam from a surrounding area.

A straightforward solution to realize such a service is to send the local information to the central server, process it there, i.e. create a set of video streams customized for specific users or areas and then send those back to their corresponding users.

A better way to provide the described service is as follows: Consider the available locations in the network providing processing units capable of integrating the locally generated information with the centrally originated video and place the respective service components at those locations which are beneficial for execution with respect to some performance metrics.

Both flexible placement of service components at instantiation time and dynamic re-arrangement of service components onto fitting resources at runtime should be possible to achieve best possible service quality under varying service conditions. This requires functionality in the network which derives appropriate processing resources and coordinates instantiation or migration at runtime based on information like which resources are available at which locations, service requirements and current network status.

NetInf aspects

An information centric approach can be used to address several challenges. A first issue is the management of the video processing units in the network. This resembles the management of caches in content delivery networks, which means that the best locations for servers should be determined under the constraints on cost and service quality and choosing the best location for each service instance.

To make the service delivery adapt to changes in the user locations and service demand requires handling of large amounts of state information to optimize the placement of the services in the network. Both the traditional network equipment and the processing units need to be monitored, for example with respect to where service software is running and what the load on different network and processing equipments are. *NetInf* may provide an efficient solution for collecting and distributing this information.

One aspect which is different compared to content delivery is that there is an additional need to check that the service can be executed in a specific processing unit. The requirements on the runtime environment have to be fulfilled by the server. Metadata indicating the requirements should be provided for the service software objects.

The security aspects also get more critical when software processes are being moved and executed in different elements in the network. One risk is that harmful code could be spread to the processing units thus the processing units need protection. On the other side the code moved should be protected from harmful processing units and be treated as requested by the code owner. A different risk is that sensitive information may be contained in the services, hence some means of protecting the information also when moving it to other providers has to be included.

The network should also support seamless connectivity between the user and the service when both may be moving. Information centric name resolution and routing solutions may be one way of solving this problem, but this requires a solution which can support mobility of both ends of the connection.

The added local information may be user provided content or derived from mobile sensors. Automatically using the right local information may be supported by the NetInf naming solution.

Use case 2: Mobility and multihoming

Mobile users today typically have multiple access networks to choose from, such as wide area networks (UMTS, LTE) networks and local area networks such as WLAN networks, For a given access technology there are potentially multiple networks, operated by different operators. The set of available networks can change according to user mobility, but also according to operator decisions, *e.g.*, powering down base stations for energy saving.

Given technical characteristics of mobile devices, service contracts etc., a mobile device can thus utilize multiple interfaces. As the user moves, previously available access opportunities disappear while new opportunities emerge. The different available networks at a time may provide different characteristics with respect to the fundamental communication service (bi-directional Internet access vs. broadcast), performance (capacity, delay), current utilization (congestion) and cost (tariffs). An additional characteristics may be content availability (not all content may be available in all networks).

For these dynamically changing access alternatives, the mobile device has to make a decision, which network to use for which application. Different ways of using the network are conceivable, for instance:

- always select one optimal network
- use multiple networks simultaneously for enhancing "throughput" and reliability
- use a subset of networks simultaneously according to dynamic cost/congestion indications and/or content availability

NetInf aspects

- Mobility management on mobile devices: managing a dynamic set of access opportunities to the NetInf network
- Strategy layer (a CCN term) for selecting interfaces
- Selecting interfaces for querying/subscribing to content according to content availability

- Concepts of congestion/cost in NetInf networks
- Seamless connectivity and hand-over: what needs to be considered for a NetInf system?

Use case 3: Event with large crowd

For sport events, royal wedding ceremonies, music festivals etc. a large crowd in a geographic location is often interested in the same content, such as live broadcasts of the event itself and/or related information. For example, during the recent wedding of the Swedish princess, there were a lot of people in the city waiting for the cortège that wanted to also watch the TV broadcast of the event on their mobile handsets using 3G, resulting in total overload. The broadcasting company planned for distribution capacity to support demand elsewhere in the world (which wasn't needed), but there was no resources to handle the local (mobile) interest.

Especially when such events cannot be planned in advance, it is difficult to provide the requested content using the existing network infrastructure. Here, it would be desirable

- to allocate resources dynamically and in an application-independent way.
- to use appropriate networks and transmission mechanisms on demand. This could pertain to topics such as multicast/unicast scheduling for hot content.
- to redistribute/share content locally, *e.g.* over WiFi/Bluetooth -- after it has been received over the mobile communication infrastructure.

NetInf aspects

An information-centric network provides the basic functionality for user-provided caching and possibility to retrieve information from multiple interfaces. Some resource management function is needed to organize local redistribution of content over WiFi or Bluetooth.

If the downlink to the cellular base-stations has less capacity than the maximum radio channel capacity, information caching at the base-station would improve performance.

Other NetInf aspects:

- Dynamic scalability of *NetInf* networks
- Multi-interface
- Network-infrastructure to adapt distribution and caching resources dynamically
- User-provided caching and forwarding

Use Case 4: Mobile sensors

Wireless sensors measuring environmental data can be deployed massively on vehicles in the public transportation system. When data for a specific geographical location is desired the sensor currently closest to that location is looked up and polled for data. The use case can be extended to cover "mobile" entities of interest, such as areas with icy road conditions, which have an extent that varies with time.

NetInf aspects

The entity of interest is the geographical location that the mobile sensor is close to, and the associated environmental data. This data can be packaged as content with a specific URI. Therefore, this is a use case of information-centric networking, where the content name is a network-independent and sensor-independent URI. Based on this URI, information-centric mechanisms are used to retrieve the desired environmental data, i.e. the content.

Naming issue is very interesting in this use case because it is not certain that the use case is directly supported by the naming schemes we know. CCN is based on publisher prefixes, and *NetInf* on publisher keys, so how do we construct publisher-independent names? Therefore publisher-independent names are an interesting problem statement. Nevertheless, in this particular use case this problem could be avoided if a specific actor takes the publisher role. In case of the icy area, the public transportation company or anyone else who discovers this area could take the role as the publisher.

Use Case 5: User-generated content

User generates and publishes a new content, which will become available via the global *NetInf* name resolution service. Additionally, the name resolution info is added in the name resolution system of the local domain. The content is located in the user's device. Once the content is subscribed by other users in the same domain, the content's "popularity" increases and the content is (automatically/without human interactions) replicated in a network storage of the local domain that can serve all new subscriptions. The local *NetInf* name resolution information is updated accordingly. The network storage could be part of the access/service provider's infrastructure.

When the content becomes even more popular, another operator decides to replicate the content in its own cache servers. This means that the content is published again in this new operator's domain. This publication is intended only for operator's own users and therefore it is not published in the global scope, *i.e.* only the local *NetInf* name resolution system is updated.

NetInf aspects

- Native resolution scopes for publications: global, domain local, device/node local. This means that multiple *NetInf* instances are possible.
    - If the content is not found in node local, then domain local is checked, which would contact global scope if needed. This resembles a recursive DNS query from signalling point of view.
    - Publication scope defines in which scope *NetInf* subscriptions are available.
- *NetInf* provides a control interface through which a Rendezvous system can maintain its mappings.
    - *NetInf* IDs and resolutions (global, local domain, device/node local) are used to access the information available via a Rendezvous system
    - Caching and other information replications visible at Rendezvous should be hidden from *NetInf*.
- This use case presents a two-step approach, where *NetInf* is only involved in a resolution phase. The second step, *i.e.* information retrieval, is non-*NetInf*.

Use Case 6: Development and deployment of mobility-agnostic applications for all types of objects

This use case describes development and deployment of applications that require interaction between a variety of object types, where the objects are nomadic or mobile. The application development time is reduced due to an API which can be used for all types of networked objects, and which hides mobility and nomadicity issues, and supports reachability functions. Deployment is simplified compared to vertically integrated applications since there is a common infrastructure for name resolution, reachability, and mobility. This lowers the barrier to entry for new applications and new application providers.

Note that this use case describes how an application developer and service provider uses the NetInf system. This perspective is as important as the end-user perspective, which is described in several other use cases.

NetInf aspects

- *NetInf* reachability and mobility system which allows for reachability, nomadicity, and mobility of all types of objects.
- *NetInf* API which relieves the application developer and the application deployer from reachability, mobility and nomadicity issues regardless of object types.

9.2.4 Business analysis of use case 1: Video with added local information

Video with added local information is studied here from the business perspective. The use case could also belong to the NetInfTV scenario, due to which the actors and roles are similar in this and *NetInf* CDN use case. However, the focus of this use case is on where to process the video stream to add information and on mobility aspects in cellular networks. Actually the case studies only the roaming use case, *i.e.* when end-user visits other than own home operator's network, but it could be extended to allow local processing when end-user uses her own home network, for example processing done in the base station end-user is connected to. Such extensions would primarily have technical consequences while the business aspects are more relevant in the roaming case.

Actors and roles

Several different roles and business models are possible in the future mobile network scenario. The roles include access providers, content/application providers, device providers, content delivery network providers, cloud computing providers etc. Of course, one company may have several roles simultaneously, and the interactions between the different roles are likely to change compared to the current practice. One topic of specific interest in the mobile network scenario is the roaming agreements between mobile operators, which traditionally have built on providing services from the home network. These are likely to change if the operators succeed in expanding their role in service and content delivery, how remains to be analyzed.

Actors

- End-user
- Mobile network operator
  - Home network provider
  - Visited network provider
- Content provider
  - Local information provider
  - Broadcast content provider

Roles

- **Service/content usage:** watching the video with added local information.
- **Home network operation:** managing business relation with end-user and data about location, services, etc.
- **Visited network operation:** providing connectivity to users where the home network lacks coverage.
- **Content provisioning:** providing video content, can be further divided into
  - **End-user content provisioning:** The end-user produces and provides content.
  - **External content provisioning:** The content is provided by a party separate from the mobile network providers, *e.g.* a weather forecasting institute
  - **Central content provisioning:** The content provision is centralized, *e.g.* broadcast TV.
  - **Localized content provisioning:** Content is generated from locally available information, for example gathered through sensor networks.
- **Content processing:** integrating local information into the centrally provided video.
- **Subscription and billing:** contracting with end-user related especially to connectivity.
- **Access provisioning:** providing (mobile) Internet access.

Technical and industry architecture

Figure 11 describes the technical and industry architecture for the current situation with services provided from the home network whereas Figure 12 presents the use case with the *NetInf* enabled service mobility. The difference between the two cases is that the *NetInf* version allows the processing server to be located in the visited network. Overall, the location and ownership of the processing server defines the architecture. In addition to home or visited network providers, processing server could be located (and owned) in broadcast content providers' premises.

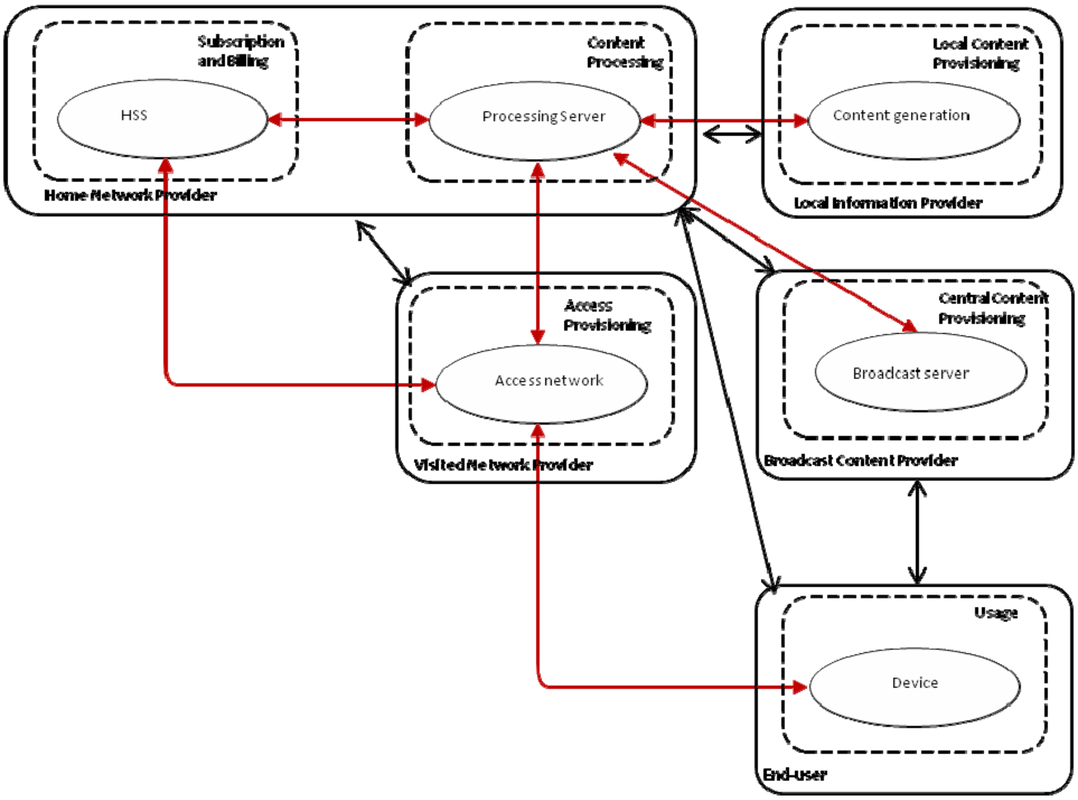

Figure 11: Current architecture without NetInf

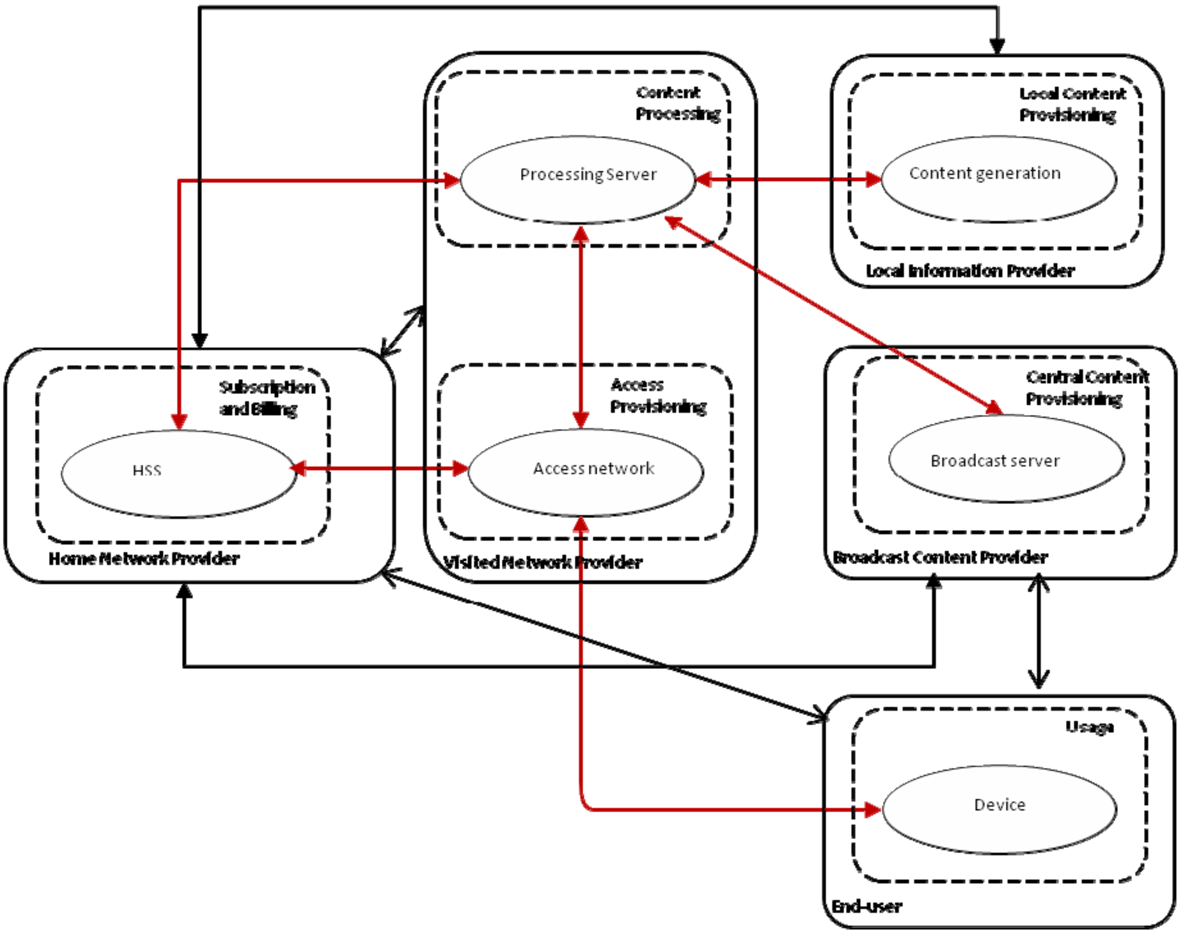

Figure 12: Architecture with NetInf

Comparison to competing solutions

The key question in this use case is where the content processing, *i.e.* integrating local information with centrally distributed video, is done. Figure 11 present an alternative where this service is provided from the home network of the end-user. Another solution is that external service providers provide the service from their own servers. Both of these alternatives are shortly described below and compared with *NetInf* solution (providing services from the visited network, or on the edge as close to the end-user if end-user is located in home network) in Table 7.

1. Providing services from the home network

For services provided by the home network operator the current practice is to run all service components on servers located in the home network. *NetInf* would provide the support for placing/finding service components in the visited network instead.

2. Providing services over-the-top

For external service providers the main alternative is to provide services from their own servers and only relying on the mobile network for the connectivity. The *NetInf* solution would provide servers within the mobile operator network which can be used for delivery of the services.

The latency of the over-the-top services depends on the location the services are delivered from, so unless the content provider has a wide-spread infrastructure high delay can be expected. In general it is difficult to estimate the cost for the different actors since limited data is available, but if the total amount of services delivered by the home and visited mobile network remains the same the cost for the operators will be reduced by the *NetInf* solution which is more efficient. For the content providers the cost may not differ significantly for providing the services from own servers or letting the mobile operator take care of it.

Table 7: Comparison of NetInf solution to competing solutions

Criteria	NetInf solution	Providing services from the home network	Providing services over-the-top
Latency	Low	High	High
Network resource usage	Low	High	Medium
Cost for content provider	Medium	Medium	Medium
Cost for mobile operator	Medium	High	Medium
Availability of controllable QoS	High	High	Low

Pros and cons for key actors

Table 8 lists the pros and cons for end-user, mobile network operators including both home and visited network providers, and external service / content providers.

Table 8: Pros and cons of Video with added local information use case from the perspective of key actors

Actor	Pros	Cons
End-user	• Improved quality when the service-related processing is moved closer to the user due to lower latency and lower risk of bottlenecks • Improved battery life time by offloading processing from the terminal to processing nodes in the network • Possibility to support additional services without support on the terminal by service integration in the network	•
Mobile network operator (both home and visited network providers)	• Lower resource usage, especially less traffic generated in the core network and more efficient usage of processing servers by dynamic placement of service processes • Opportunity to offer service platform to external service providers	• Additional cost for deploying the service mobility mechanisms
External service / content provider	• Better service delivery using mobile infrastructure support for QoS, location of services close to users, etc. • Contract with a single mobile operator gives global delivery for the customers of that operator, which makes the use case good for regional providers	• More detailed SLA with the mobile network provider • Need to specify the runtime environment for the service software in addition to connectivity • To reach full coverage of users/countries multiple counterparties are needed, which may be impractical for global providers compared to over the top delivery • Lost control over the final product delivered to the end-users

9.3 Scenario 3: Developing regions

Internet technology has often been seen as a tool and enabler for equality and empowerment of people - something that gives everybody equal opportunities to communicate, socialize, and create business. There are, however, great differences between the real possibilities for using the Internet as wanted in different parts of the world and between different social groups. As more and more of our daily activities move online, this discrepancy risks increasing the digital divide rather than reducing it.

In the developed world, network connectivity is ubiquitous, reliable, fast, and cheap. Most households can get fast wired (or wireless) broadband connectivity to their homes, and 3G cellular coverage is ubiquitous.

In many developing countries, and in remote regions of developed countries, on the other hand, this is more problematic. Many users in these countries still rely on dial-up connectivity to the Internet, if they have any at all. If broadband connectivity is available, it is very expensive. Often, the price per megabit can be comparable or higher than that in Western countries, and if you also factor in the lower average income in the developing world, the cost becomes astronomical. Even when connectivity is available, there is often a problem with reliability of the service. This is even more apparent in many regions where the power supply is very unreliable, with frequent power outages, which means that you cannot rely on connectivity to always be there. Cellular network coverage is becoming better in many developing countries and the number of subscribers is growing rapidly. This still has problems of reliability, capacity, and coverage areas. The population density is often very high in these regions, so as popularity of the cellular networks grows (and in particular, if popularity of data traffic over them grows), there is likely to be a great capacity problem - either in the wireless spectrum, or in the access networks.

We believe that *NetInf* can provide a better, more robust and efficient solution to the communication needs of these communities than the existing Internet protocols. The information-centric approach and the inherent caching features as well as its improved resilience to disruptions make *NetInf* a good option here.

First of all, caching can obviously help cope better with the low bandwidth access links. As content can now be cached locally, less traffic will have to be sent over these links, which improves the user experience (one will be able to get more data to the local network, and content that is already available in the cache can be served quickly). This is obviously dependent on the level of data reuse and content popularity. Many communities in this part of the world are however more homogeneous than in the Western world, so it is likely that one will be able to get some good use of the caches.

The information-centric approach to networking can also help to better manage disruptions in the communication with the global Internet. This can either be in terms of shorter disruptions where it is possible to just wait for the service to be resumed (the request just takes a bit longer to be delivered to the end-user), or for longer disruptions where you might want to consider using some DTN style transport to deliver the content.

Including disruption tolerance as an integral part of the *NetInf* system together with its globally-connected mode of operation is vital and has many different benefits. First of all, it can obviously be used to serve the "classic DTN scenario" of completely disconnected villages/users where there is no connectivity except through data mules. Even though this is often true for only a small part of the population, it is important that the system can handle this. The need to deal with temporary outages (that still can be on the time-scale of days to weeks) is likely to be more common. This can be situations where connected operation is usually available, but due to some event (such as the monsoon destroying the phone lines), connectivity is lost. In such situations, it is useful to be able to fall back on a DTN mode of transport for content (*e.g.* using the local taxi company as data mules). The inherent disruption tolerance will also improve the user experience in cases of spotty and bad coverage of the cellular network, and it can further extend the range of a wireless network to, for example, cover an entire village. You will no longer need to go to the top of the hill or the centre of the village to send your data request, but you can do it whenever you want, and content will be delivered when you are in range.

Yet another possibility to use DTN techniques in order to alleviate the problem of the low bandwidth access links is to use DTN transport for bulk data transfers. Often data transfers are asymmetrical. There will usually be enough bandwidth available to send a request for some large piece of content (*e.g.*, a movie), but downloading it could be very slow, and it could also be difficult to maintain a stable connection for the duration required for the data transfer. In situations like this, the capacity could be increased by using DTN style transport (i.e. data mules) to deliver the bulk data. This will save a lot in terms of utilization of the access network, and can often be a faster method of delivery as well (if regular buses, taxis, etc, can be used as data mules).

NetInf may also provide a nice abstraction above standard lower layer transports (TCP) and disruption-tolerant transports (BP) that could be very useful for this scenario.

9.3.1 Challenges

- Access networks are often low-bandwidth.
  - Even if you get good cellular coverage, you will have a capacity problem to get all the data out there.
  - Cellular data service may also be pre-empted by voice calls, leading to disruption of the data channel.
- Power outages and other disruptions are common.
  - Wired infrastructure (both power lines and communications lines) are vulnerable to weather conditions and other disruptions.
- Broadband bandwidth is **expensive**.
  - Cost per Mbit can be comparable or higher than in Europe/US.
  - Given the low incomes of people, the relative cost is even higher.
- Often more than one user per computer.
  - Computers are shared by a family/school/village.
  - There may be tens of thousands users per host in a town.
- There are still areas without any connectivity whatsoever.
  - There may be a "core" (*e.g.*, a town) that is relatively well-connected, though still subject to, *e.g.*, power disruption, surrounded by a geographically large periphery with much more challenging networking requirements.
- Privacy
  - You might not want your neighbour to know (or be able to infer) what content you have been accessing.
- Directory services
  - In N4C (http://www.n4c.eu/) experiments, users often did not remember email addresses to which they wanted to send mail. This makes directory services valuable, but nonetheless difficult.

9.3.2 Market potential

The population and Internet penetration rate statistics in Figure 13 reveal large regional differences. The Internet plays the most significant role in North America, Oceania and (Western) Europe where the penetration rates are over 50%. However, the largest growth potential is in developing countries in Asia, Africa and Latin America, which have large population but modest penetration rates. Asia for instance has already now the most Internet users but this is not due to high penetration rate but large population (56.3% of the world's population).Therefore developing countries have potential of 4.5 billion Internet users (even though roughly half of the population lives in urban areas that are only the secondary target of this scenario) that can be served by sustainable Base of the Pyramid innovations.

As already described, the low income levels present a challenge. Making money is not the primary driver for local communities who are willing to provide Internet connectivity for their members in developing regions. Internet access is seen primarily as an enabler of new or more productive businesses, a tool for education and a way to improve the standard of living by narrowing the digital divide. Making profitable business can still be possible, but this requires innovative business models. Indian community ISP AirJaldi is an existing company in this area, and therefore validating our model of this scenario with them would be beneficial in order to get real feedback.

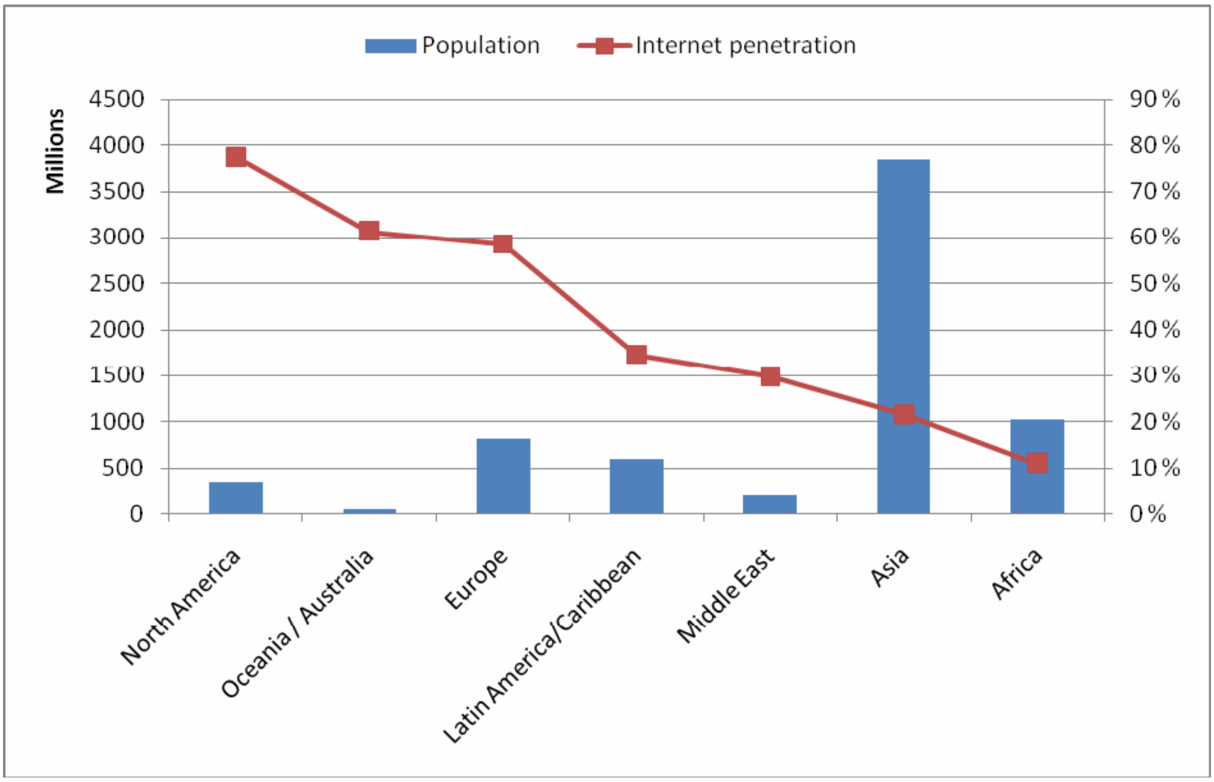

Figure 13: Estimated population (in 2010) and Internet penetration rates (June 30, 2010) in the world by geographic regions (data from [15]).

9.3.3 Use cases

This scenario has only one use case which is presented below.

Use case 1: Community ISP

In areas where it is not economically attractive for larger telcos to deploy infrastructure, or where the services provided by them are not attractive/sufficient (not enough capacity, bad coverage, or too expensive for a large portion of the population), it should be possible for a local company, non-profit organization or co-op group to set up a NetInf network and offer services to the local population with a model that is more appropriate for that area.

Much information in these regions is highly local, so there is no need for external network connectivity for content that is both produced and consumed within the area. It will, however, be desirable to provide real Internet connectivity as well. Thus the community ISP should purchase some "big" (for some definition of big - depending on price and need) pipe of bandwidth to the outside world. These are often very costly and unreliable in the developing world, so it will be crucial to use *NetInf* in order to use the bandwidth as efficiently as possible and/or to use alternative methods (e.g. “data-mule”) for transporting data to and from the Internet. Such ways of using *NetInf* to improve the usage include:

- Use of caching to reduce traffic required from the global Internet.
- Less problems of temporal differences in bandwidth demand by shifting content retrieval to off-peak periods
  - Pre-fetching of predicted content (e.g. based on previous accesses) or content that is known to be of interest (e.g. news/weather).
  - Delayed retrieval of content that is not delay-sensitive (possibly within some certain deadline)

- Potential for "centralized" processing (compression of images, ad-blocking, torrent-blocking, etc.)
  - This can be controversial.
  - Can we deal with content transformation like this with persistent *NetInf* names?

Network building/expansion

It should be possible to set up a *NetInf* network with only two *NetInf* devices and a link connecting them, *e.g.*, Bluetooth. This network should then be able to grow seamlessly/organically to a global network of billion of devices. In a situation like this, it is necessary to have functionality for initial trust building in a small disconnected network, while still being able to eventually connect this network to the global network with maintained trust. As the network grows (or if parts of it are lost due to partitions or other outages), there will be a need for dynamic election procedures to elect certain nodes to perform network functions, for example name resolution service (NRS) and storage. Depending on the scale of the network, different NRS solutions will need to be used (local broadcast, dedicated resolution servers, distributed p2p solutions, etc).

The community ISP can use a combination of different cost-efficient infrastructures to build the network, using whatever resources are easy to acquire and deploy. This could for example include local and long-range deployments of WiFi mesh networks and local wired networks. Such organizations are also in a good position to negotiate with the local community or local businesses such as taxi operators to cooperate to provide data mule services to be able to cope with outages and provide capacity for bulk data transfers.

Since the extra cost to further extend a network will be rather low, if a long-range link to a remote village has already been set up, this creates an opportunity for people at the edges of the network to extend the network themselves. As discussed in the business section below, this could involve incentives through cheaper access for themselves.

Technical challenges

- Scaling a network from a small user-operated network to an operator-operated network
- Selection of nodes to run common services on (relation to *CloNe* (WP-D) work)
- Selecting the appropriate transport depending on:
  - type of content, *e.g.*, requesting content over GSM, but get it delivered over something else
  - priority of content (*e.g.*, emergency data: select lowest latency transport, also prioritization within that transport may be needed)
  - current availability
- Cache and bandwidth management
  - What data to pre-fill caches with?
  - How to make use of spare off-peak bandwidth?
- Adopting applications like banking, e-mail, and e-commerce, which normally require online connectivity to work, in a DTN environment. This could be done by combination of bulk transfers and low bandwidth connectivity. Also reserving a certain amount from a bank account to be used in offline local e-commerce is a possibility.
  - Feedback to the users - how do we let the users know what kind of service to expect? Will the content arrive in 1 minute or 1 day?

Applications

- Interpersonal communication (mail, chat, texting, both connectionless and connection-oriented)
- Bulk transfer (*e.g.*, IPTV/video - YouTube)
- Live Internet services which were requested by N4C trial users in summer 2010 [17].

- Banking (transactional, requested by ~50% of adults)
- Facebook (requested by almost all kids)
- Weather/News: no specific service providers needed here though, almost any source is fine.
- The problem here relates to mapping "live" web content to the challenging region, even what appears to be static web content can be problematic and applications like Facebook much more so, both technically and from a business perspective.

Reality check

In order to make sure that the challenges and properties that we describe above are real ones and not simply made up problems that do not exist in reality, we intend to communicate with people that operate in at least one environment such as the target one. This should validate this scenario, and increase the possibility to use work based on this scenario in a real-life situation.

9.3.4 Business analysis of use case 1: Community ISP

As backbone bandwidth can be more efficiently used, it is possible for the community ISP to offer its services at a more affordable rate. It will also be possible to organically grow the network. If a long-range link to a remote village has been set up (at somewhat higher cost), the required extra investment to offer services to others in that village will now be cheaper. This also provides an opportunity to give people at the edges of the network an opportunity to extend the network further themselves, possibly offering a discount on their service cost for acting as relays in extending the network.

The cost of infrastructure can be a big hurdle here. N4C village router nodes currently have a component cost of about €2500 in EU; a low-cost equivalent would still cost some hundreds of Euro which is still very expensive for such markets. It is not clear how to solve this problem if the community ISP is to own the infrastructure.

Actors and roles

Actors

- End-user
- Local community ISP
- Traditional fixed ISP
- Traditional cellular operator (national/international)
- Transport company (taxi/bus)
- Global NetInf network – for NetInf to be of use in the challenged-communication region, we assume that there are one or more sources of NetInf content on the well-connected Internet that can be used from gateways (or even possibly directly)

Roles

- **Service usage:** using voice services and local and global Internet services.
- **Local connectivity provisioning:** providing and operating the local community network.
- **Local caching:** caching global Internet content locally.
- **Local content production:** producing and providing local content and services.
- **Data mule service:** transporting content and service requests from local community to the location where better connectivity to the global Internet is available, and returning
- **Content provisioning:** providing content and services in global Internet.
- **Voice and data traffic provisioning:** providing mobile voice and data services.
- **Backbone connectivity provisioning:** providing connectivity to the global Internet.

Technical and industry architecture

Figure 14 presents the technical and industry architecture of the use case. The local community ISP provides end-users access to the local content caches and local services, and also to the global Internet through its connection to the traditional fixed ISP. Therefore end-users do not need separate business agreement with fixed ISP. Cellular voice and data services may still be provided separately by the traditional cellular operator. It is however also possible that a business agreement may be set up between the local community ISP and the cellular operator such that subscribers to the local community ISP can access the cellular network when coverage from the local network is not available. This access will on a technical level be done directly through the cellular infrastructure and not going through the local community ISP.

The global *NetInf* network provides content to end users, which may require a business interface between them to regulate access to content that has access restrictions. Similarly, the local community ISP may wish to cache some of the content in order to improve the user experience for its subscribers. This may require permissions from the content providers in the global *NetInf* network, but it may also be a service that the local community ISP can offer to the content providers to make the content more easily available to end users.

The importance of the local community ISP is significant because it controls a bunch of important roles and is most probably the driving actor, even though traditional fixed ISPs may also encourage local communities to build their own ISPs in hopes of additional revenue due to increased coverage.

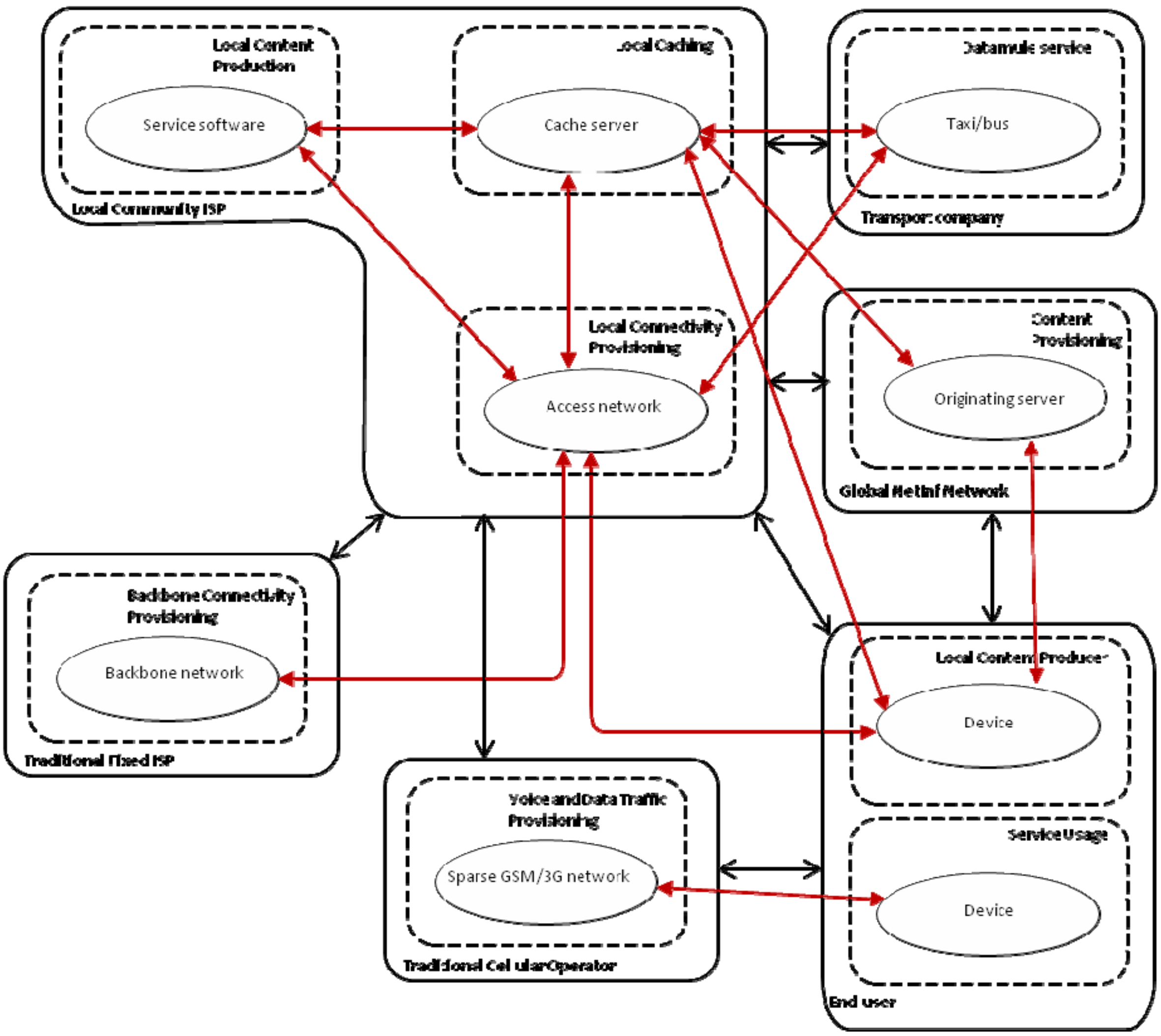

Figure 14: Technical and industry architecture of NetInf Community ISP use case

Comparison to competing solutions

NetInf Community ISP competes with traditional ISPs offering both dial-up and broadband connections. Broadband covers fixed DSL and cable connections but also mobile broadband since many developing countries are bypassing the fixed Internet phase due to missing cabling. Table 9 collects the analysis which is further discussed below the table.

Table 9: Comparison of NetInf Community ISP to competing solutions

Criteria	NetInf Community ISP	Traditional ISP, dial-up	Traditional ISP, "broadband"
Cost for end-user	Low	Medium	High
Perceived level of service and performance	High	Low	Medium
Robustness to failures	High	Low-Medium	Low
Cost for ISP to provide a certain level of service	Low	High	High
Cost of Organic network expansion	Low	Low-Medium	High
Bulk data transfer speed outside of local community	Very low (Can use alternative data transport for large data items, *e.g.*, data mules)	Low	High-Medium (Depends on bandwidth and network reliability)
Interactive services	Problematic (local instantiation of services may help)	Can be slow	Works good

The key selling points of *NetInf* community ISP are low cost, high service level and robustness to failures. Local communication is handled effectively, but the challenges stem from communicating outside of the local community. Bulk data transfer is typically not too sensitive for long latency, due to which the low bulk transfer speed can be solved using data mules. Interactive services, especially in less connected areas, present a bigger challenge, where local instantiation of services may be help. Live communication remains still difficult.

Pros and cons for key actors

There are multiple key actors in the use case, including end-user, traditional ISP interesting in making an entry to community ISP market, local community ISP, and local (non-connectivity) service providers offering, e.g., local content or data mule services. The pros and cons for each of these actors are collected into Table 10.

Table 10: Pros and cons of NetInf community ISP use case from the perspective of key actors

Actor	Pros	Cons
End-user	• Perceived capacity increase or performance enhancement • Better user experience in situations with spotty coverage • Better robustness • Knowledge transfer may educate people, create jobs, and enable further network expansion	• High-latency for services that require use of data-mules • Inability to use existing interactive service accounts (eg Gmai, Facebook) in general
Traditional ISP	• Capacity increase / cost reduction because more users can be served with the existing access network • Increased coverage/better service →potential for more subscribers	• Requires interaction with new actors (transport providers) • Community ISP may be seen as competition if traditional ISP tries to expand coverage to areas served by community ISP
Local community ISP	• Easier for someone to deploy a local network and provide service to a community. • Possibility to be acquired by a bigger, more traditional telco operator	• Local people involved in community ISP may be seen as privileged, causing bad feeling, or the community ISP may re-enforce existing privileged positions
Local (non-connectivity) service providers	• Opportunities for local service providers, like: • Remote/catalog shopping a la "last mile solutions" • Local taxi/bus companies can get extra revenue from delivering data	• Not clear how to fairly reward data-mules • High learning curve, these service providers may not be at all familiar with technology

10 Open Connectivity System (OConS) scenarios (WP-C)

There is a single scenario being developed within *OconS* (WP-C). It is named Supporting Flash Crowd Connectivity Needs.

10.1 Scenario 1: Supporting Flash Crowd Connectivity Needs

The scenario is based in a Flash Crowd; a large group of people with mobile devices that are in a location where there is an unexpected and increased demand for communications and services. The flash crowd may be a group of people who are at a sports event or an open-air concert, they may be travelling together on a train, they may be pedestrians who are in a station or any other public space during, *e.g.* a rush hour, or even a flash crowd in an emergency situation. Their requirements for communications, services and content are dynamically changing, but these have to be available for everybody and need to be provided with the appropriate quality. People who are closer to the centre of the event of interest may upload videos, photos or blogs to their personal or social network pages, others may wish to see this on-spot generated content or other videos and news of what is currently happening. Most people would wish to communicate, either with people who are on-site or outside of the flash crowd area and this can be the case for social or business purposes. Therefore, the service providers, including the Over-the-Top providers, need suitable solutions to deliver their services to both end-users and businesses at the appropriate quality levels.

The available access networks may include different technologies and access points, such as 3G/anyG or mobile WiMax, or Wi-Fi in infrastructure mode. Likewise, a self-organized community mesh network in a pedestrian / public space location could complement the available communication networks. Finally, ad-hoc networks, *e.g.* using Wi-Fi Direct, may also be available. The nature of the flash crowd creates immediate requirements for either server based or user generated content, with various services and communications constraints. For example, these will likely require higher bandwidth than what may be available for both access and core networks. Moreover, the service platforms or data centres may also experience a higher demand in regards to access and/or processing requirements.

Flash crowd ‘consumer services ‘may be offered by local self-organized community networks or be supported by service providers with a global presence in the internet. Services in this respect may include and extend social networks, like Facebook, ‘(personal) content distribution’ networks like YouTube, ‘information management’ services like Google, or service integrators that operate services on top of these (i.e. OTT) like Animoto.

From a provider/business-to-business view, flash crowd connectivity service providers will have the need for special edge-to-edge communication across the core networks as well. That is to say that the *OConS* user in this case is a complete network, e.g. a data-centre cloud or service cloud and thus creates a different demand for connectivity between such (business-, provider) users and the *OConS* provider (in the role of a network operator or operator of a sub-network/domain). Here, the focus is on the autonomous behaviour between the (edge-) domains (represented by service or data centres) which may not necessarily be triggered directly by end user actions but more dependent on the type of service, its connectivity requirements and the aggregation of the traffic of users demanding this service. Service examples are data & transmission intensive workflows between several service providers, e.g. video creation, production and distribution by service providers like media companies, agencies, broadcasters. Other cloud service (co-)operation requiring reliable and cost-effective connectivity across network domains include the migration of data bases and servers, processes for backup, load balancing, energy efficiency, etc. Large distributed data centres with variable processing, storage and networking resources create a global challenging interconnectivity and transport demand to be served by cooperating network operators as *OConS* providers, providing their (physical) resources (e.g. optical transport networks) to build a provider-to-provider service.

10.1.1 Challenges

There are several challenges to support this scenario that need to be addressed:

1. To provide content and services to users, virtual operators, service providers, OTT service providers or any other parties. Avoid bottleneck in both access and core networks.
2. Use multi-path multi-protocol transport to optimize the quality of heterogeneous services and applications in a dynamically changing context; *e.g.*, upload video of the event or the accident, look at news about it, i.e. download by both people on site and by others who may be interested in uploaded content; VoIP , video-chat or other applications may also be used simultaneously. *In line with Multi-path Multi-protocol transport from Use-Case 2.*
3. Address cooperation, self-organization and opportunism in various types of networks, from infrastructure to spontaneous ad-hoc and wireless mesh. *In line with Use-Case 1 Creating and Sustaining the Connectivity in Wireless challenged networks.*
4. Use virtualization and dynamic allocation of resources in order to provide bandwidth isolation and support for services with stringent requirements. *In line with Use-Case 1 and Use-Case 3.*
5. Optimize the use of available resources (including wireless accesses, core networks and enablers for services) e.g., to provide the right level of Quality of Experience (QoE), minimise power usage and reduce the cost. This can be seen from user, operator, or any other player point of view, or be seen as a global optimisation. *In line with Optimising the QoE with adequate management mechanisms from Use-Case 3.*
6. Enable operators to provide services to OTT service providers (and their customers) in an optimized way. This may be achieved by, *e.g.*, moving the (cloud) services and the networking services closer to users for a period of time; *e.g.* move the Facebook or YouTube service; *e.g.*, move the caches of OTT service providers to provide better service to customers. *In line with Use-Case 4 Autonomous Interoperation and Connectivity.*
7. From the network operator side the main overarching security objectives are network availability and control integrity. The scenario and all the derived use cases question how users/customers are authenticated and to what extend their authorization is sufficient for network attachment, service discovery and network domain usage.

10.1.2 Market Potential

Nowadays the world is facing a constant need of information. One very important need for all of us is to have access to information, especially about a particular special event. As an example, if we think that most people have cameras in a cell phone, and they witness something unusual, they will start to take pictures or make movies and send them “online” starting to consume network capacity, and the network may not be prepared for such a capacity increase in that particular location. That problem has to be solved.

The news websites also experience somehow this problem during major world events. This kind of bottlenecks in the network can be solved in both access and core networks. The end results of flash crowds are usually very poor performance at the server side and a significant number of unsatisfied clients [18]. The reason for this to happen is that the only support from the network is made by the infrastructure provider. This can change if a Content Distribution Networks (CDNs) can be created with a straight collaboration between business partners. The provider of the information, or data, will unleash the information to other points of the network decongesting the end point where the provider holds the data. After this other network user become a network provider giving the information to the rest of the user shortening the path where the network is used and gaining capability for the rest.

This does have potential and the content provider does want a faster and reliable way of delivering the data to the users without crashing or delaying the network to do so. If this does happen we will have all the stakeholders to benefit from it. We can list some benefits that are envisaged for the major players:

- From a user point of view:
  - Enable services in an unexpected situation where normally they would not be available, provide increased QoE, optionally reduce cost, reduce power consumption
- From an operator point of view:
  - Optimised way for operators to provide services to OTT service providers and their customers. This may be achieved by, *e.g.* moving the (cloud) services and the networking services closer to users for a period of time. *E.g.* move the Facebook or YouTube service, move the caches of OTT service providers or deploy CDN solutions to provide better service to customers.
  - Facilitate the implementation of operator policies and priorities for moving users to different networks, technologies or services if there is congestion or, in case of an emergency, a failure of technology.
- Cloud network provider:
  - Manage capacity to satisfy SLA agreements with users without incurring additional investment or operations costs.
- Cloud network user:
  - Achieve ubiquitous service access from any location at any time.

But how does an infrastructure provider benefit from the peeks in network capacity needs?

The infrastructure provider can win the costumers support if it is able to deliver good quality on these occasions. So if they can use the solutions proposed in this scenario to increase capacity and deal with these events, the bottlenecks will disappear.

Other benefit that can be obtained through this is the storage capacity within Storage Area Networks (SAN) [19] that is increasing to huge levels and usually provided by the infrastructure provider. The main idea for the infrastructure provider is that the data that is accessed with frequency for a specific period of time can range from very hot, i.e., accessed extremely frequently, to very cold, *i.e.* not accessed for long periods of time such as the archival data. If hot data can be identified and separated from the cold data, using the capacity of the providers it can be assigned to storage technologies with faster response time and higher throughput compared to servers in order to improve the overall performance.

10.1.3 Use Cases

Based on the global scenario defined in *OconS* (WP-C) there are 4 different use-cases that are being used to study different aspects of the work being developed in WP-C.

Use case 1: Creating and sustaining the Connectivity in Wireless Challenged Networks

Within the Flash Crowd, imagine that several (heterogeneous) wireless nodes are willing to build a multi-hop network in order to provide the end-users with the connectivity between them and towards a fixed Internet infrastructure. This communication environment is often under adverse conditions, e.g., expectations of connectivity between certain nodes no longer holds, or congestion is experienced on some links because of the multiple simultaneous requests from the crowd.

Subsequently, for these Wireless Challenged Networks, use case 1 needs to employ innovative techniques that explore the resources and communication conditions in the best way, creating and sustaining thus the Connectivity. Depending on these specific conditions (e.g. requirements from the application service, or the need to prioritize critical messages in case of a disaster, or the lack of physical paths among end-devices, etc.) several approaches will be studied aiming at designing and experimenting novel connection mechanisms.

Use case 2: Using multi-path/multi-protocol (MultiP) transport for optimized service delivery of heterogeneous content

The users within the Flash crowd have mobile devices with a number of wireless technologies, e.g. 3G/LTE, mobile WiMAX, Wi-Fi including infrastructure, mesh or ad-hoc. The heterogeneous content and services (i.e., with different requirements) provided on a single web page could include: videos of sport or music events (i.e. live events), video news (i.e. pre-recorded video), video conferencing, chat and VoIP (e.g. Google talk), text and pictures, or any other content.

Moreover, the content and services may come from various sources and locations, again with different impact on the end-user experience. Likewise, the services used by the end users may be based on HTTP/HTML5, in line with the current developments where HTTP is becoming a full application standard for delivery of multiple types of data.

We need thus to deal with the discovery, selection and usage of advanced transport mechanisms (transport/connectivity services) which exploit the diversity of multi-path multi-protocol transport and are benefiting from various available wireless/wired networks. These *OConS* functions support and optimize (autonomously or in cooperation with service providers) the delivery of services with heterogeneous content between, e.g., NetInf providers and end users, especially under the conditions of a dynamically changing environment such as the flash crowd.

Use case 3: Optimising the QoE for End-users with adequate management of the (Cloud) Network services

Diverse services are provided to the Flash crowd end-users through several available networks, such as 3G/anyG, mobile WiMax, and Wi-Fi infrastructure mode where the APs may dynamically become available and then disappear. For some of the end-users, another possibility for having the connectivity is through the access provided by a self-organised community (i.e. mesh network) in a public location, and which could complement the available communication means before-mentioned. Furthermore, ad-hoc networks (e.g. using Wi-Fi Direct) may also be available, and technology failure or power shortage may sometime disable one or more technologies or devices. Obviously, these access alternatives might be operated by different entities, with which the End-user could (or not) have an agreement.

Accordingly, we will deal with decision making mechanisms to "optimally" choose the interfaces, the access networks, or the paths (and hence to schedule/map/route the traffic flows accordingly) in order to achieve the highest possible quality-of-experience (QoE) for an End-user. Besides, this will likely imply appropriate management of connectivity and optimisations (i.e., trade-off and/or tuning) at the Cloud/Network providers/operators level, as well as at the Service/Application level. Likewise, these decision mechanisms need to cope with (sometime) contradictory goals of the involved actors, because various policies from network providers/operators and end-users, as well as requirements from service providers (which might include e.g. *NetInf*) need to be considered.

Finally, whatever the decision made, we should also provide some means to enforce it; therefore, we should at least interfacing (i.e., choosing, activating, steering) with the corresponding execution protocols and procedures (e.g. dynamic mobility anchoring, multi-path/protocol transport services from Use-case 2, virtualized slices from *CloNe*, and so on).

Use case 4: (Autonomous) Interoperation and Connectivity of Cloud and NetInf data centers

Consumer services to the Flash crowd end-users are often offered by service providers with a global presence in the internet. Services in this respect may include social networks, like Facebook, '(personal) content distribution' networks like YouTube, 'information management' services like Google, or service integrators that operate services on top of these like Animoto.

As opposed to end-user and access oriented *OConS* use-cases 1, 2 and 3, this last use-case is based on a provider/business-to-business view and will focus on edge-to-edge communication across the core network. That is to say that the *OConS* user in this case is a complete network, e.g. a data-centre cloud or service cloud and thus creates a different demand for connectivity between such (business-, provider-) users and the *OConS* provider (in the role of a network operator or operator of a sub-network/domain).

The focus is on the autonomous behaviour between the (edge) domains (represented by service or data centres) which may not necessarily be triggered directly by end user actions but more dependent on the type of service, its connectivity requirements and the aggregation of the traffic of users demanding this service. Traffic examples are data & transmission intensive workflows between several service providers, e.g. video creation, production and distribution by service providers like media companies, agencies, broadcasters. Other cloud service (co-)operation requiring reliable and cost-effective connectivity across network domains include the migration of data bases and servers, processes for backup, load balancing, energy efficiency, etc. Large distributed data centres with variable processing, storage and networking resources create a global challenging interconnectivity and transport demand to be served by cooperating network operators as OConS providers, providing their resources (e.g. optical transport networks) to build a provider-to-provider service.

10.1.4 Business analysis of Use-Case 1: Creating and Sustaining the Connectivity in Wireless Challenged Networks

Concerning the Wireless Mesh Networks (WMN) business models several attempts were made. Some of them are still in use in several distinct modes of functionality. We have an example of a municipal community such as the Seattle city area that explores it in concern of the population or FON (http://corp.fon.com) that explores the business in an economic way.

Actors and Roles

The identified actors and roles are represented in Figure 15.

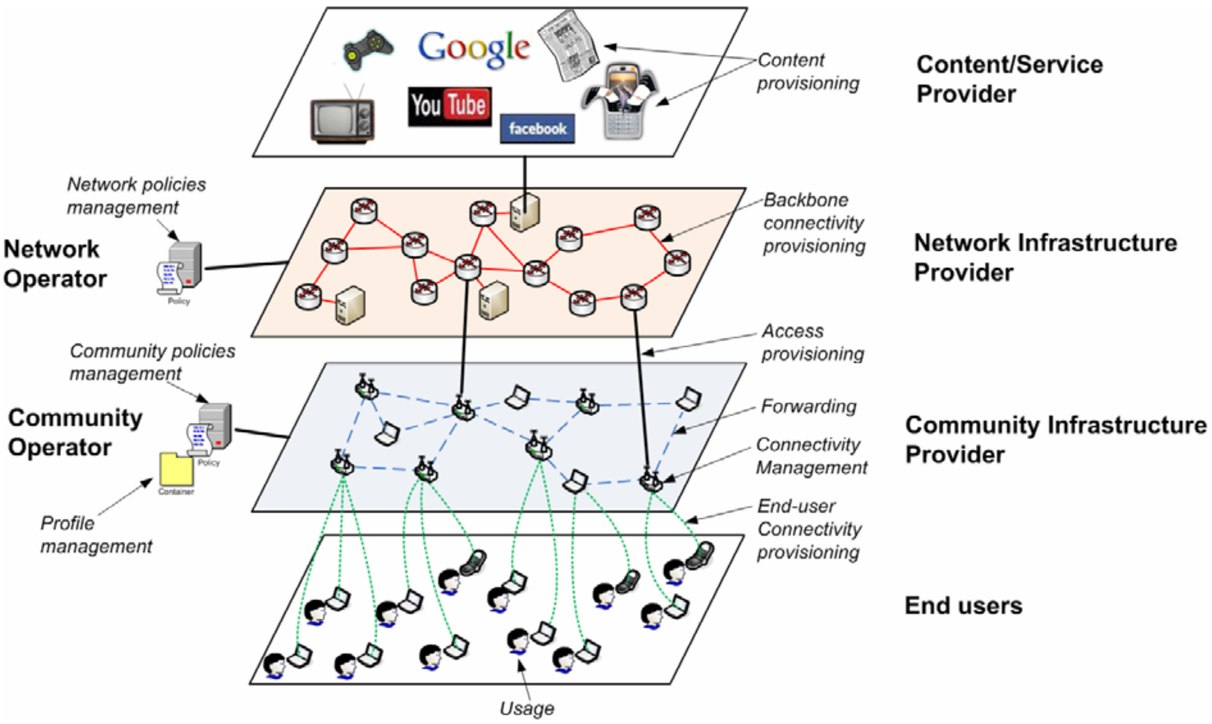

Figure 15: Actors and roles of the Wireless Challenged Networks use case

- **End-user**: The end user is the source or destination party that initiates (connects) or receives/terminates service requests to content/service providers through the network. An end-user may have forwarding capabilities, participating in the wireless network.
- **Community Infrastructure Provider:** The party elements with forwarding functionalities that build a spontaneous and self-organized wireless network

community of cooperating nodes. It can be an ad hoc or a wireless mesh community network. In the case of an ad hoc network, it is composed by end-user nodes with forwarding capabilities that generate traffic and forward it between them. In the case of a wireless mesh network, mesh nodes forward traffic mostly towards gateways to the Internet (mesh nodes with gateway connectivity to the Internet). Mesh nodes may be end-users that also generate traffic (as ad hoc nodes) or may be mesh access points that provide infrastructure wireless connectivity to end-users. This is the most important actor of this use case, as it's where developed connectivity services are implemented.

- **Community Operator:** The party that provides, authorizes and manages the community infrastructure resources. It has defined a set of policies on which the community is ruled in order to build a spontaneous and self-organized network. Incentives and rewards encourage participating nodes to share their resources and cooperate.
- **Network Infrastructure Provider:** The party that owns and uses the access and core infrastructure resources to provide global network connectivity. It exchanges traffic with the WMN through gateway mesh nodes.

  Note: the "infrastructure provider" and "network provider" actors are not separated in this use-case, as no work on virtualization is foreseen that would justify such need.
- **Network Operator:** The party that authorizes and manages the network.
- **Content/Service Provider:** The party that provides services, applications and content.

Roles

- **Usage:** use different communication services (e.g., video-conference), consuming content, etc.
- **Forwarding:** Traffic forwarding functionality. It may exist in end-users (in ad-hoc mode) or/and mesh nodes, constituting a wireless network.
- **Connectivity Management:** Implements and manages the developed *OConS* connectivity services. This is the key role where innovation is introduced (network coding, delay tolerant networking, radio resource management strategies).
- **End-User Connectivity Provisioning:** Providing local access connectivity to end-users (access point).
- **Access Provisioning:** Providing Internet access to the WMN. Gateway connectivity to the backbone network (through, e.g., fiber, cable, ADSL, satellite, 3G or LTE).
- **Backbone Connectivity Provisioning:** Offering global Internet connectivity.
- **Community Policies Management:** Manages applicable policies to the ad hoc/mesh (community) network.
- **Network Policies Management:** Manages applicable policies to the access/backbone network.
- **Profile Management:** Manages the end-user/mesh APs profiles according to the (community) policies and connectivity usage.
- **Content Provisioning:** Provisioning of content to end-users.

Technical and industry architecture for Use-Case 1

The most important concepts in this use-case are:

- **Wireless Mesh Networks** (WMN): They recur to cooperation and self-organization to explore the available resources in an opportunistic way, building a spontaneous network that maximizes connectivity to end-users, Figure 16 b. The main characteristics are:

  - WMN spontaneous deployment is an intermediate step towards ad-hoc, user-centric environments.
  - A user’s physical neighbourhood offers a friendly environment for this type of communication.
  - Neighbourhood WMNs need users to contribute to their creation and operation.
  - Cooperative strategies for self-organization enable users (nodes) to pool their resources to support the creation and operation of the underlying communication network, participating at all physical, access, and network layers, but also for service provision on top of it.
- **Delay Tolerant Networks:** nodes do not have permanent physical paths to other nodes in the network. It can happen that node Z wants to send a packet to node X, but temporarily Z doesn't have a path to reach X, so Z takes the decision of (i) store the packet and wait some time or (ii) forward the packet to node Y hoping that the future probability for Y to reach X is higher than its own, for instance (there are also other alternatives). Considering self-learning and/or historical patterns of nodes, different useful approaches could be incorporated to the routing strategy of a DTN.
- **Network Coding**: the main idea here would be to analyze the possibilities of NC over WMN. The goal would be to study the interaction with upper layer protocols (*e.g.*, TCP) and to analyze the enhancements which could be achieved when using multicast transmissions.
  - Improve performance and reliability
  - Tackle congestion.
  - Regarding DTN transport, NC techniques added to the routing strategy for 1-to-N or N-to-1 transmissions, could lead to a better throughput or minimizing congestion.

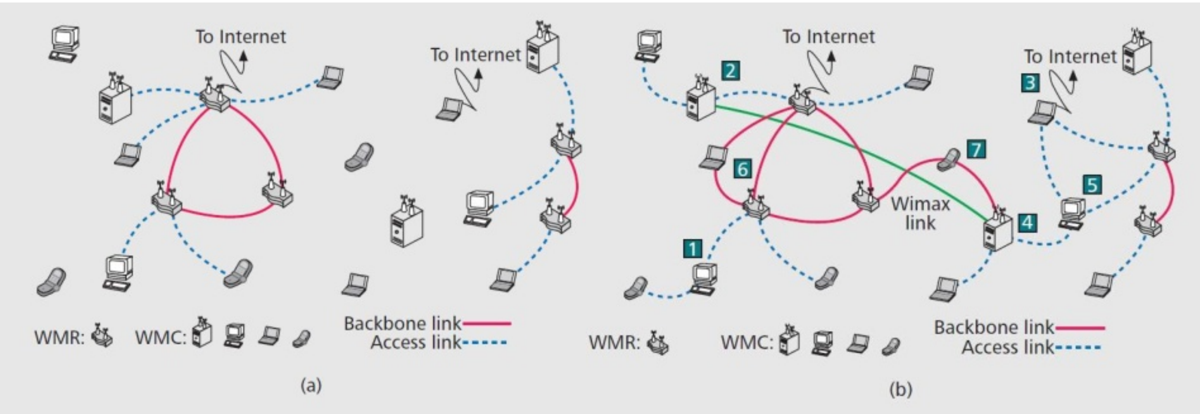

Figure 16: **a) spontaneous network with limited connectivity; b) spontaneous network after optimization of connectivity thanks to innovative techniques.**

Architecture

Below in Figure 17 the architecture for the "Challenging Wireless Networks" use case is represented, presenting the technical and business interfaces.

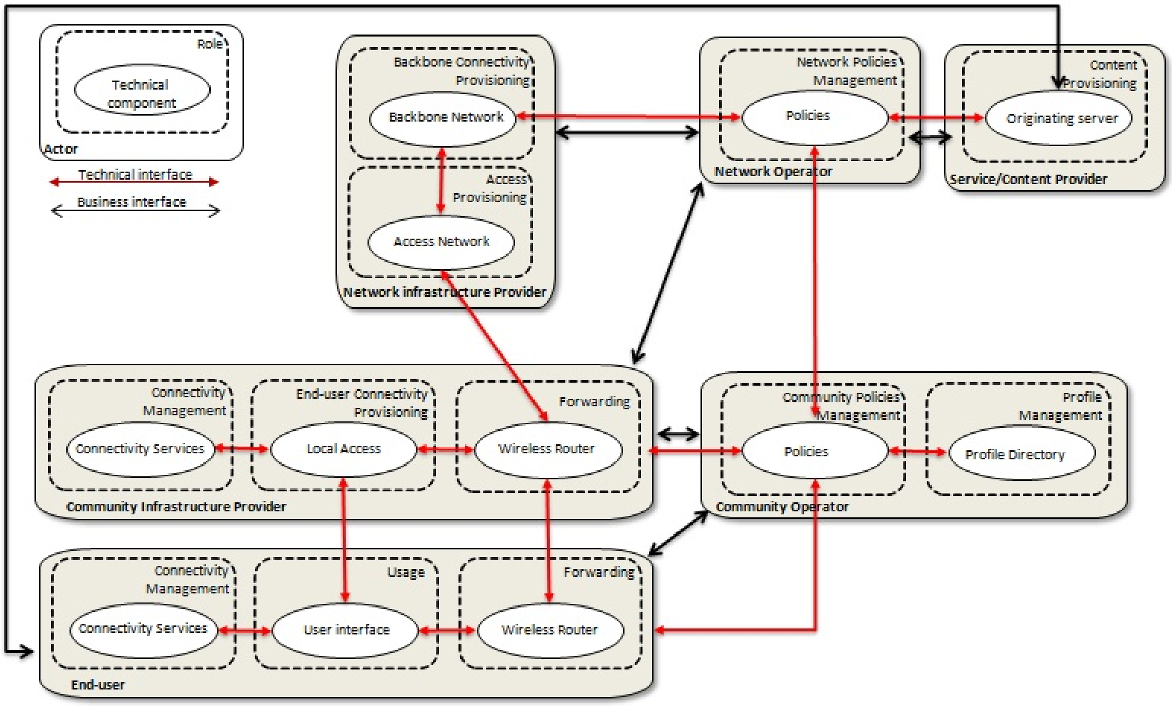

Figure 17: Architecture of the Wireless Challenged Networks use case with identification of actors, roles, technical components and technical and business interfaces

Pros and cons for key actors

Table 11: Pros and cons of Wireless Challenged Networks use case from the perspective of key actors

Actor	Pros	Cons
End-user	• Extended end-user connectivity; • Share of Internet Access costs between multiple users of a communitarian Wireless Mesh Network (WMN). • A group of persons (e.g., neighbours) can build a communitarian WMN, dimensioned for the type of usage they are willing. • The end-user can participate in the network creation and operation (contributing with resources and capabilities that in classical networks are not explored), and benefiting from incentive and reward mechanisms.	• Internet access through a community network with self-established rules. • Quality of service depends on the intensity of utilization of the WMN, as well as on amount of resources (MAPs, gateways) shared by the community members. • More energy consumed by mobile devices that become gateways in the WMN
Community Infrastructure Provider	• Communitarian self-organizes WMN that provides extended coverage and Internet connectivity to end-users. Some end-users may participate in the multi-hop wireless mesh network, than to forwarding capabilities and their operation in ad-hoc mode. A set of Mesh Access Points (MAPs) (that can belong to some of the users, or can be bought by the community) interconnect automatically, building an infrastructure access network to "simple" end-users. The Internet connectivity of a small subset of MAPs gateways is shared to provide Internet connectivity to the entire WMN. • Robust and adaptive network, with self-healing and self-configuration properties, where novel mechanisms (multi-radio resource management, network coding, delay tolerant networking) can be applied. • Cheap, fast, easy and self-organized deployment, not needing to use expensive fixed communication infrastructures.	• Communitarian network that depends on the cooperation of individual members for its efficient operation. • Network scalability dependent on the resources shared by the community members. • Complexity of forwarding nodes, in terms of equipment (multi-radio mesh nodes), processing (network coding) and protocols (delay tolerant transport protocols),
Community Operator	• New business opportunity of communitarian Internet Access operators. • Possibility to share Internet access costs within neighbourhood community member; • Maximized exploitation of a small set of Internet access resources, shared by a large set of users (currently, in residential environments 10% of capacity available to ADSL users is used). Lowers the price per community member to have Internet connectivity; • "Communitarian" connectivity operator, with self-defined rules. Cooperative network, from physical to network and social layers. Recur to rating and rewarding mechanisms to incentive community users to cooperate. Punishment mechanisms are applied to members not willing to cooperate. • Users may participate in the novel communitarian WMN business model, by sharing their resources (MAPs, gateways). • Communitarian business model competing against classical Internet (fixed and wireless) providers business models of individual end-users.	• Rules and principles of network operation must be defined and agreed within the community. • Need of cross-layer incentive and punishment mechanisms to be applied in the community (up to the application and a potential "social" layer) to guarantee fairness of usage of the network resources. • WMNs are vulnerable in terms of security (e.g., wormhole attacks), requiring advanced security mechanisms (e.g. robust routing protocols). • Support of Quality of Service requires admission control, resource allocation and resource management mechanisms. • Without specific QoS mechanisms, quality of service per user depends on the size of simultaneously-active end-users and the number of Internet access gateways.
Network Infrastructure Provider	• Maximized utilization of the network resources, with WMN gateways injecting intense aggregated traffic flows.	• Network must be adapted to support more intense traffic flows (compared to classical Internet Access).
Network Operator	• Extension of service coverage provision to areas that classical Internet infrastructure would be very expensive. • Maximization of utilization of contracted resources.	• The communitarian network solution competes directly with legacy network operator business models, designed for individual Internet users with a dedicated Internet gateway.

Content Provider	• The communitarian WMN provides end-users' access to content provider services through a cheap, shared and communitarian networking infrastructure. • The communitarian WMN brings more clients of the Content Provider, thanks to the extended Internet service coverage to areas originally without access, or with expensive access. • A network based on incentives provides the means for supporting community-oriented applications, increasing the value of the network and further encouraging users to participate and cooperate	

10.1.5 Business analysis of Use-Case 3: Optimising the QoE for End-users with adequate management of the Cloud/Network services

In the Use-Case 3 of the OConS scenario we deal with the decision making mechanisms to "optimally" choose the interfaces, the access networks, or the paths in order to achieve the highest possible quality-of-experience (QoE) for an end-user. This aspect of the use case will likely imply appropriate management of connectivity and optimisations (i.e., trade-off and/or tuning) at the Cloud/Network providers/operators level, as well as at the service/application level. These decision mechanisms need to cope with (sometime) contradictory goals of the involved actors, because various policies from network providers/operators and end-users, as well as requirements from service providers (which might include e.g. *NetInf*) need to be considered.

Among the ***Requirements*** we can mention:

- ***Transparency***: the interaction between involved actors should be performed in such a way that the end-users cannot spot any significant difference in their user experiences, compared to the every-day use of the network. In normal situations, i.e., normal traffic and load conditions (no flash crowd), the end-users will already enjoy new services which are based on novel *OConS* functions, deployed within the networks and/or the user devices. Of course users can still employ the traditional services, which do not exploit the *OConS* benefits. In both cases, when a flash crowd forms the *OConS* will respond with high reactivity and cope with the changing conditions, so that the end-user will still be able to get the appropriate QoE. In other words, this highly dynamic management of connectivity, provided by *OConS*, should not impact the end-user. This will also be the case when handling the mobility, i.e., the mobility (of the users and, more generally, of any "mobile entity") should be provided in a seamless and transparent way, whatever the access network(s) used.
- ***Flexibility***: this means to cope with the (sometime) contradictory goals of the involved actors, i.e. the QoS requested by the user and the QoS available in the network during the flash crowd could be different. This situation is solved by *OConS*, by dynamically allocating resources where the users demand, e.g. exploiting multiple radio access networks or allocating a higher number of gateways for Internet access. But there could be some cases, when available resources are anyway not sufficient to satisfy the users' needs, e.g., if they all require "high-level" of QoE/QoS. In those cases, *OConS* should dynamically find a new balance between allocated resources and user/service requirements, in terms of QoS.
- ***Security***: in the heterogeneous environment depicted in the flash crowd use-case, many entities can interact, which can also have weak or no trust relationships between each other. Moreover, security policies can be different from actor to actor. In this context, it is of outer importance to verify the trustfulness of the corresponding sources. Moreover, it is also required to protect user information exchange across the different actors, e.g., especially sensitive information such as the identity.
- ***Manageability***: all the actors, in particular network operators/providers, should be able to optimally manage (also in the sense of "personalization") the connectivity services. This requires the use of expressive policies, by means of which end-users, operators

and other actors can dynamically configure the network and the devices to customize services and related QoE. This more dynamical management also requires (possibly new) monitoring features, which are able to react to network changes and to provide information to effectively enable the decision making process in *OConS*.

And as ***Challenges (from OConS perspective)*** we can cite:

- Collecting the necessary information from several available sources (e.g. access networks, end-user, operator, core network with partial and/or e2e path, etc.), dealing with the fluctuations in the availability, trustfulness and validity of this information coming from the network infrastructure (antennas, access gateways, core network node, and other nodes).
- Minimizing the overhead induced when carrying the information, using smart correlation and timing of information updates (e.g. improving IEEE 802.21) and dynamically change the information collection policy, e.g. to increase the querying rate.
- Storage of information, e.g., fully distributed, partially distributed, or centralized.
- Capitalize on the end-user behaviour and contextual information, e.g., the routes the users usually take, services they usually request, etc. (yet alleviating the Privacy and Security impact, see next)
- Privacy and Security for collected information: some information (e.g. IP addresses of an end-user) could be advertised to other end-users, for example to optimize routing among them, or traces could be formed presenting history of moving users; challenges here are to first to minimize the collection of information and second to protect sensitive information that has been collected and will be used (operators should be able to handle the data according to applicable legal obligations).
- How to define QoE and how to measure it.
- Collecting information from the application provider in order to select the path that better fit the requirements for the service requested, according the service request, the network can decide which network resource should be reserved.
- Which entity takes the decision, and how to use novel distributed decision mechanisms to taken into account several actors.
- What cooperation strategies between the involved actors and entities? (Can the decision be dynamically be made on different entities depending on the context?)
- Develop of efficient, yet secure, protocols to enable such cooperative decision making.
- Develop comprehensive, yet lightweight, algorithms to make the decision (i.e., selection of the interface, the access, the anchor, or the partial or end-to-end path, and so on).
- How to update (or change, override) normal operators policies and preferences to cope with an abnormal change of mobility patterns, traffic patterns, radio network load.
- Appropriate Naming and Addressing schemes, allowing the necessary separation between user/data/host ID and Locators.
- How to use virtualization and dynamic allocation of physical resources in order to provide bandwidth isolation and support for services with stringent QoS Requirements.
- How to benefit from the available resources and capabilities? *E.g.* exploiting networks' heterogeneity, taking advantage from the simultaneous use of various paths (Multi-P), with efficient distribution of uplink and downlink traffic, dynamic support and activation of distributed mobility schemes, and so on.
- Quickly enable and control an efficient distribution of traffic load among different available networks (depending on the device types, user needs and operators preferences), and avoiding "dangerous" bottlenecks on some links or parts of the network.

- Selection and execution of mobility management mechanisms. For example, Gateway selection to locate the best gateway, tuning of the Mobility management parameters, to avoid signalling packets "storm" in the core network.
- Efficient use of virtualization techniques.

Actors within use case 3

- **End user:** The End-user can be a human, a machine, a business user (such as cloud and/or service provider), an abstract "higher-layer application", or an information object (content). The End-user can be the source, the destination, or both for the communication services and/or the content. The End-user is one of the main actors within this use case, since the decision to be made will likely have impact on her/his QoE.
- **Infrastructure/Resource provider:** The Infrastructure/Resource provider is the stakeholder which provides the means for communication (e.g. ducts, towers, cables, base stations, access points, switches, routers, etc.).Thus it owns and manages the physical and virtualized resources of the infrastructure and it offers them to the Network providers.
- **Network provider:** The Network provider uses the subjacent infrastructure and resources to build and provide networking services (i.e. *OConS*), such as Ethernet connectivity, Radio bearers or Wireless service sets, IP services, managed Wavelengths, Dark-Fiber/SONET transport, and so on. They usually maintain a business relationship with the network operators to get paid for the right-of-use/leasing and for the maintenance of their networks. A Network provider can be also seen as Cloud/Virtual Network provider where it requests, uses and consumes shared network services and resources provided by the Network providers; thus, the Cloud/Virtual Network provider combines several service and resource offerings of various network and infrastructure providers (both horizontally and vertically) and offer them to a third party.
- **Network operator:** The Network operator provides Connectivity Services to the end-users, it maintains the adequate level for these services, and it performs overall operation/management of these communication services. The Network operator may or may not own the underlying infrastructure or other networking assets (e.g. radio spectrum). Likewise, the Network operator benefits from the *OConS* capabilities to offer better services to the end-users (who will be more satisfied) and to cope with the traffic increase from service/content providers (who will be able to expand their services). The Network operator has a business relationship with the end-users (e.g. billing), network providers, and with the service/content providers.
- **Service/Content Provider:** This actor will benefit from the optimum connectivity of the end-users, e.g. being able to adapt the quality of the service; it does not have a direct role on the whole process, rather than the specific parameters which should be considered when taking the decision.

Roles within Use-Case 3

- **Usage/Generation:** i.e., the end-user can employ different communication services (e.g. video-conference), consuming content, generating content etc.
- **Content management:** at the service/content provider side, e.g., generating, publishing/distributing, storing, and so on.
- **Subscription management (and Billing)**
- **Context Information Management**
- **Connectivity Management**

- **Advanced OConS (Connectivity) Provisioning**
- **Connectivity Broker**
- **Access/Backbone Provisioning**
- **Resource Management** (including the Virtualized ones)
- **Cloud Slices Management**

Technical and industry architecture for Use-Case 3

Compared with current architectures (e.g. like those described within 3GPP and IETF), we propose a quite similar approach. Nonetheless, we have enriched the technical architecture with an additional role (i.e., the Connectivity Broker), we have distributed the Connectivity Management between the networks and end-terminals, and we have kept separated the control/management and the data/connectivity (i.e., à la OpenFlow or Evolved-Packet-System). When it comes to the industry architecture, even if the actors are well known from the current approaches, we have proposed more clearly relationships among them, so as to keep a high degree of flexibility for the business models as well.

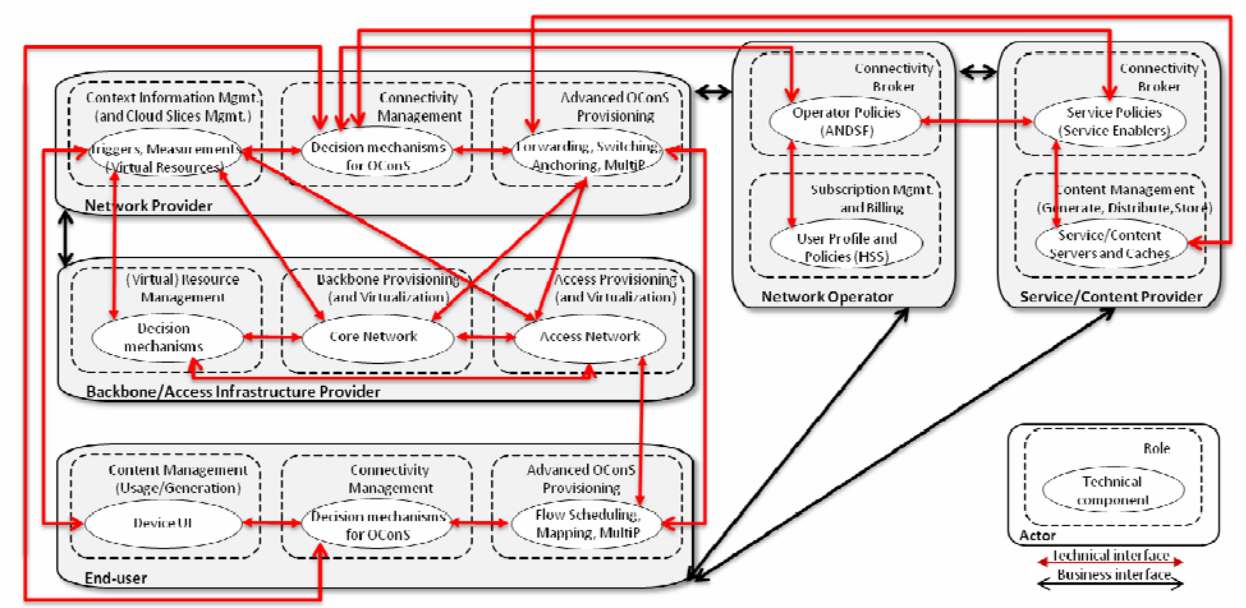

Figure 18: Technical and industry architecture of use case 3

Comparison to competing solutions

Table 12: Comparison of OConS to competing solutions in the use case 3

Criteria	OConS	3GPP Evolved Packet System	IETF Multi-Interface	OpenFlow
Resiliency	High	High	Low	Medium
Manageability	High	Low	Low	High
Complexity	Medium	High	Medium	Medium
Energy Efficiency	High	Medium	Medium	Medium
Performance	High	Medium	Low	Medium
Flexibility	High	Low	Low	Medium
Scalability (for the signalling)	High	High	Medium	Medium
Scalability (for the data)	High	High	Medium	High

Pros and cons for key actors in UseCase 3

Table 13: Pros and cons use case 3from the perspective of key actors

Actor	Pros	Cons
End user	• Transparent and access-agnostic connectivity to services and content and better/personalized end-to-end Quality of Experience (QoE) • Is always served with the most appropriate access alternative; depending on the factors used to make the decision, the goodness is perceived by means of, e.g., better performance, lower price, lower power consumption • If Multi-P is allowed, reliability and performance enhancements might be brought about by using (in parallel) various interfaces	• Terminal complexity • Dependence on the availability of particular technologies/techniques/procedures • Increase of energy consumption (if using various interfaces at the same time) • Privacy considerations (information flow between providers, with which the end user might not have agreements)
Infrastructure/Resource Provider	• Efficient use of the available resources, e.g. load balancing • Increase of available capacity • Improve user (client) satisfaction	• Need to cooperate with other providers (leading to potential privacy concerns) • Increase complexity • Depending on the particular conditions of networks, some providers might get their traffic reduced (by more aggressive counterparts) or overwhelmed (causing their clients to be unsatisfied) • Need for a tight interoperability between heterogeneous technologies
Network provider	• Provide optimal application-aware connectivity services • Lower OPEX by using shared infrastructure from Infrastructure providers and/or the capabilities of other network providers • Optimize network resources for delivery of content (e.g. connectivity management/intelligence distributed in the right places within the network)	• Cooperation schemes between peer entities should be available • Increased complexity
Network operator	• Increase customers' satisfaction and prevent churn • Greater productivity for Business users (e.g., rescue team, professionals on the move, etc.) • Generate more revenues per-user by enabling customized/personalized offerings • Enable closer partnerships with Content/Service providers	• Need to establish cooperation mechanisms with other operators • Increase complexity in the needed procedures • Privacy issues (in terms of the information which should be made available)
Service/Content Provider	• Possibility to increase the quality of the offered service (thanks to the optimum connectivity) • New business opportunities from open APIs with the network side	• New interface (business and technical) needed towards Network operators

11 Cloud Networking (CloNe) scenarios (WP-D)

WP-D main focus is on the development of a complete and flexible architecture for Cloud Networking (*CloNe*), with flash network slices capabilities, which will operate as a reference model for deploying complex applications over heterogeneous virtualized networks. The cloud networking work package will integrate distributed processing and network provisioning capabilities to provide an enhanced platform for the execution of applications with specific demands.

Control functions and protocols, management functionalities and security solutions will be put forward on WP-D for the Cloud Network ecosystem and a test bed will be set up to evaluate the proposed architecture in large scale prototype.

From a business analysis perspective WP-D will study two distinct *CloNe* scenarios of application, each with their own case studies. The first scenario named "Dynamic Enterprise Scenario" refers to the provisioning of IT/IS solutions from the cloud network ecosystem to the enterprise market. In the second scenario named "Elastic Video Distribution on the Cloud" refers to the offering of video and similar services from a cloud network ecosystem to the retail market.

11.1 Scenario 1: Dynamic Enterprise (enterprise in the cloud)

This scenario presents and depicts the provisioning of IT/IS solutions from the cloud network ecosystem to the enterprise market, supported in concepts such of SaaS, PaaS and IaaS.

The introduction of cloud networking offerings to enterprise IT/IS provisioning provides an enterprise with additional flexibility in its operations and in the way it does business, by offering comprehensive IT/IS tools with innovative cost effective business models.

So far cloud computing has demonstrated the ability to flexibly scale services to provide on-demand and pay-per-use IT/IS solutions. With flash network slice capabilities, cloud networking introduces dynamic flexible network provisioning into the equation. An enterprise will be able to dynamically adapt its IT/IS services to include new remote locations, added functionalities and new entities within its boundaries in a swift and effortless manner in accordance with the business requirements dynamics.

This flexibility allows an enterprise to go beyond scaling IT/IS services for its core operations to provision and connect IT/IS services for short and long term projects both internal and external. As an enterprise needs to interact with its external business environment it can extend itself to include those engagements within its supported infrastructure. The enterprise can offer access to its facilities for the purpose of fulfilling its joint ventures, it can extend the scope of its facilities to support remote employees with tools to do their jobs effectively, it can adjust the scale of its facilities to meet performance and demand criteria (e.g. seasonal substantial increase in product inquiry and product ordering before Christmas).

In this scenario and related use cases we consider multiple cloud sites, implementing virtual processing and storage infrastructure, connected by an operator’s network. Some of these sites may belong to the enterprise and some may be external public cloud providers. The operator network implements flash network slices to interconnect the infrastructures implemented by the cloud sites and end users and may also provide processing and storage elements of its own. Collectively these implement a distributed virtual infrastructure service.

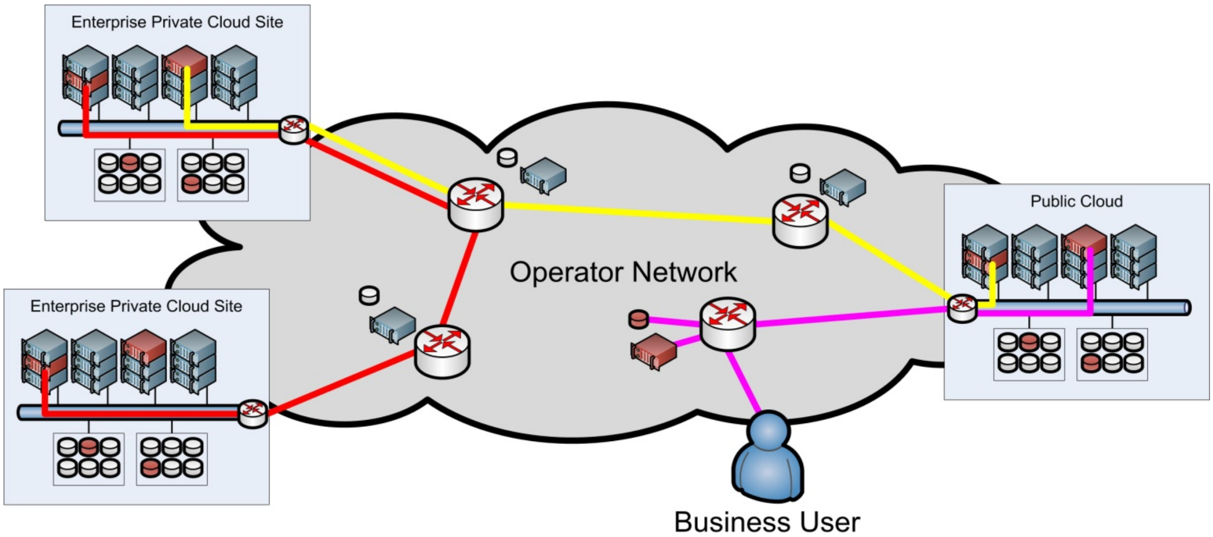

Figure 19: Enterprise virtual infrastructure of the *Dynamic Enterprise* scenario

Virtual infrastructure can be deployed on demand throughout the cloud network. In the diagram the coloured elements indicate physical resources that have been allocated and configured to implement processing, storage and networking elements of the virtual infrastructure owned, or leased, by an enterprise.

The enterprise can adapt its infrastructure and redeploy its IT/IS systems to better serve its employees and business dynamic engagements and requirements around the world.

The relevant actors for this scenario are listed below.

- **Enterprise** - an enterprise is a company organised for commercial purposes. Some requirements of an enterprise are shared by other types of organisation, but an enterprise typically has stringent security and regulatory constraints to protect its intellectual property, information privacy, and to ensure legal compliance. In the scope of this scenario, we consider an enterprise to be a large organisation with hundreds or thousands of employees, its own IT infrastructure and global presence.
- **Business partner** – a business partner is another company, customer or supplier that is involved with the enterprise in the course of its business. A business partner may be given access to some infrastructure and services of an enterprise in support of a business relationship between the partner and the enterprise.
- **Cloud site operator** - a cloud site operator is an organisation that manages a cluster of processing and storage infrastructure that implements an infrastructure service.
- **Network operator** - a network operator is an organisation that manages a network that implements communication services with flash network slices between the offices and remote infrastructures of an enterprise, its IT system administrators, its IT service users, and other cloud sites.

For this scenario four use cases were regarded as interesting to study, for they illustrate the scenario predictable usage and business related topics of concern. The use cases are (i) Media Production, (ii) Remote Auditing, (iii) Business Goal Management and (iv) Virtual Desktop.

From a techno-economic perspective this scenario introduces several security considerations due to the use of third party infrastructure, shared infrastructure and the particular sensitivities within the enterprise. These use cases share in common legal constraints regarding the storage and management of customer data, risk of unauthorised access to business data, and risk of service disruption. Technologies employed will be required to provide sufficient assurance of separation among virtual infrastructures of different providers.

11.1.1 Challenges

Some key challenges addressed by the dynamic enterprise scenario are the following:

Heterogeneity: Depending on the IT and communications infrastructure available in different parts of the world, different quality of experience can be achieved. For instance, it may be more difficult to provide the necessary performance via flash slices in developing regions as compared to highly industrialized regions. However, providing “cloud-wide” uniform guarantees (e.g. globally guaranteed end-to-end delays between any two enterprise locations) must be mapped to the overall ICT infrastructure independent of the individual underlying networks that are available.

Cloud Network Scale and Inter-Domain: Because the dynamic enterprise is a scenario of global extent, numerous inter-domain networking questions are likely to exist. The question is how the overall virtualized networks, composed of individually provided networks can be constructed. For instance, from which (virtual) network providers will the network slices be constructed? Likely, there will not be a single (virtual) network provider, but rather multiple providers that collaborate for providing global virtualization.

Network and Computational Resources for Data Distribution: The assumption is that the dynamic enterprise is dispersed over many globally distributed locations. Because the system model considered in SAIL also considers processing and storage nodes in the network, a challenge is to decide which data to carry to which location to achieve SLAs at minimum effort from the perspective of providers. In many cases, large scale cloud-based services are provided simultaneously from several, potentially distant sites. Therefore, the assignment of a specific server to a specific service request must be optimized, while taking into consideration current load of servers, and minimizing the amount of overhead. Furthermore, due to large dynamicity of cloud network services, live service migration should also be supported (i.e. the assignment of a new server to existing cloud service, in real time). Server assignments should be dynamically changed in the most efficient manner (e.g. the benefits outweigh the cost associated with the live migration), while generating minimum distortion to the service.

Security: It is expected that highly confidential and proprietary information of an enterprise will not be hosted at the public cloud, at least at this early stage of the migration. Therefore, the scenario must cope with the existence of both public and private cloud infrastructure, with services that are dependent on both clouds. Additionally, enterprises might opt to have their sensitive data hosted at specific public clouds (e.g. at the same continent of the corporate office), at least at the early stage of the migration. This means some constraints regarding resource allocations that go beyond resource optimization, and therefore, have to be addressed.

Innovative Business Models: Especially in a hybrid scenario new business models may become interesting. In particular, such models that address the migration aspects to facilitate and ease the migration of conventional or privately hosted cloud services to the public cloud. For example, with respect to security, specific packages may be offered to ease the migration in terms of security.

Innovative Business Enablers: This scenario also encourages the consolidation and evolution of specific global industrial strategic business cluster platforms and IT/IS solutions for collaborative work within specific industrial clusters. Such type of platforms shared amongst specific strategic business cluster value chain actors will enable more product and process innovation within each cluster, for the benefit of all of the stakeholders, improving the cluster specialization, the cluster market effectiveness and operational efficiency, leveraging the high cost platforms and IT/IS solutions on a large number on enterprises that collaborate in an Open System Organization Model. By Open System Organization Model we mean enterprises within strategic business clusters that allow access, or share, their IT/IS systems to/with other organizations with which they do business. The challenge will be for the identified actors to leverage on the proposed technology in order to build and deploy these collaborative suites and solutions for each strategic business clusters to be addressed.

11.1.2 Market Potential

The Dynamic Enterprise scenario market potential assessment is an undertaking far from being a trivial, secure and risk free exercise, as is, to no less extent, the broad cloud network model market potential analysis.

Cloud network paradigm is believed to bring about a shift in the way people and enterprises do business with comparable effects to the industrial revolution in the 18th century [20][21][22]. The knowledge availability, the technology push and the market pull trends will surely generate a disruptive change in the current business processes. New business dynamics, new actors and new business models will reshape the market as we know it today into something that we can only dream of; thus transforming this kind of exercise into a well informed guess at best.

At present we can envisage adoption apprehension about security, availability, performance, interoperability, systems migration strategy, flexibility and more [23], but we can also envisage intangible latent potential for the stakeholders in this scenario that we can now hardly quantify, such as: better enterprise commercial effectiveness, better operational efficiency, extra business flexibility, more process and product innovation, new micro and small enterprises benefiting from complex IT/IS platforms or services, enhanced social networking and collaborative experiences and a great deal more [20][23][24]. Clearly two of the qualitative key dimensions that emerge from the proposed scenario are the cost reduction dimension and the revenue generation dimension for the stakeholders.

In fact traditional cloud computing has already passed the "early adoption phase" according to IDC [23] and nine out of 10 IT professionals inquired by IBM [24] believe that it may overtake on-premise computing by 2015 as the primary way organizations acquire IT. Forrester [20] believes that the effect of cloud network adoption will be the transformation of traditional software and hardware market segments into IT services, and that cloud computing is a sustainable long term IT paradigm.

According to Forrester [20] the current estimated potential market size for cloud computing solutions similar to the proposed scenario is very significant, because this type of solution can impact many of the core segments of the estimated $2.4 trillion worldwide.

Unsurprisingly the effectiveness of the commercial adoption of cloud computing is very much dependent on reliable and fast communication services, thus the importance of the proposed technology, and of the proposed scenario, for trying to apprehend a share in the potential estimated market size.

In addition, and according to Gartner's definition of strategic technology [21][22] : a technology that is perceived as having the potential to significantly impact enterprise businesses within the first few years of its use, it looks as if this scenario’s technological proposal can be considered a strategic technology.

The specific dynamic enterprise scenario emerges from several trends in the way that IT infrastructure is being deployed and managed. The move to providing software-as-a-service that is accessible from anywhere independent of location or device, the move towards using infrastructure-as-a-service to manage compute resource capacity both in private clouds (managed by an enterprise for internal use) and public clouds (managed by an infrastructure service provider for commercial use), and finally the move of telecom operators towards providing value added connectivity services and cloud capability.

The growth in IaaS revenue is continuing as shown in the figure below, with The 451 Group predicting IaaS revenues to exceed $3B in 2013 [25]. This figure includes storage and compute revenue. Storage is predicted to grow at 79% CAGR with compute growing at a lower 58%; in 2010 the two are approximately equal.

The use of private cloud is less easy to monitor, but the sale of servers for public and private cloud installations shows that comparable growth can be found in both areas. According to IDC [26] the shipments of new servers for public cloud installations will grow steadily from over 300,000 in 2009 to nearly one million in 2014. Similarly shipments for private cloud are expected to grow from just over 100,000 to approaching 500,000 in 2014.

Worldwide revenue generated from software-as-a-service is also increasing with a value of £5.1B in 2007, $6.4B in 2008, and a forecast figure of $14.8B in 2012 according to Marketspace [27]. These numbers are consistent with an analysis carried out by Gartner, who says that "*in the course of the next 5 years enterprises will spend $112 billion cumulatively on software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS), combined*" [28]. Furthermore, Gartner also emphasizes that Cloud Computing should be considered within tactical projects already in 2011, and expects that Cloud Computing will gain majority by 2015 in the form of "*Service Enabled Application Platforms" (SEAPs)*" [29]. All these facts demonstrate a shift in practice towards providing services implemented on cloud infrastructure.

From the above analysis the proposed scenario appears to exhibit significant market potential both, either from a tangible point of view, either from an intangible point of view.

11.1.3 Use cases

In this section we describe four example use cases for this scenario predictable usage.

Use case 1: Media Production

In this use-case a TV channel in UK sub-contracts programme production to specialised production companies at home and overseas. Examples of such programmes include computer generated animations such as children's cartoons or documentaries with artificial content. When the channel considers the introduction of a new programme it starts by commissioning a pilot series: a small number of chapters, sometimes just one. If the pilot is a success, the channel will extend the commission; if not, the engagement will end. In this use case we consider such an engagement with an animation production company in Japan.

In the past the channel would have expected the sub-contractor to provide its own infrastructure and software, delivering only video content to the TV channel. The cost of these facilities needed for a single pilot series would exceed the revenue generated from the commission, introducing risk for the sub-contractor, adding pressure to increase the price of the commission, and so reducing the TV channels willingness to try out more experimental programmes and reduce volume.

The TV channel has moved to using cloud networking for its IT services and in so doing has the opportunity to extend its own production facilities for use by the sub-contractors. The channel reconfigures its production systems to allow access to the Japanese company. Because of the amount of data and processing involved the channel's production system is extended to a cloud site located in Japan. Secure networking connectivity is established between the extended production service and the Japanese animation company and back to the TV channel's home systems. When animation rendering demands increase the processing resources assigned are scaled. When large data volumes (film frames or animation models) are exchanged the bandwidth of these connections are increased and then reduced to minimise costs. The management of the facility is unchanged, but the location and connectivity have changed to meet the demands of the engagement.

In this use case all of the relevant actors identified in the broader scenario will play one or several roles. To implement this use case some challenges are known beforehand, such as:

- locate a suitable cloud site close to the IT service user based on cost of communication, storage and processing, and latency;
- establish a virtual infrastructure at the chosen site;

- establish a network slice between the virtual infrastructures on the home site, the chosen site and the IT user;
- configure access rights for the IT service user;
- temporarily scale flash slice bandwidth;
- remove the remote virtual infrastructure and the flash slices;
- meet the security requirements of the corporation, especially regarding the isolation of their virtual infrastructure from others hosted on shared physical infrastructure.

Also from a business perspective we expect beforehand that this arrangement will allow the TV channel to reduce the risk for the sub-contractor when accepting a pilot series commission. The channel will be able to afford a greater number of commissions, more experimental commissions, and will be able to employ a wider variety of sub-contractors, including individual animators. Besides enjoying the benefits of a pay-per-use model for its IT facilities, the TV channel will also amortise the cost of managing its IT services across all its engagements and its internal IT use.

Use case 2: Remote Auditing

In this use-case a multi-national corporation is present in different countries where a telecommunications group is present through national operators. Each office is served by its respective national operator.

The corporation has different IT systems per business unit (e.g. accounting, sales, CRM, etc.) deployed in a separate virtual infrastructures, and a user might exploit more than one of these IT systems, each with different computing requirements.

An auditor from the central headquarters needs to perform periodic audits. This employee has to move to the regional headquarters of a given national operation. There he needs access to the same IT systems he is using in his 'home' location as well as the local resources, while keeping them isolated. In order to make the quality of experience of the user as good as possible, the operator decides to extend the virtual infrastructures hosting the knowledge worker's IT systems from the home location to the visited location, thus providing better responsiveness. During the audit, the auditor also visits branch offices of the regional operation. These visits need to be optimised in time, and therefore the virtual infrastructures, and IT systems, need to 'follow' the movements of the auditor.

In this use case all of the relevant actors identified in the broader scenario will play one or several roles. To implement this use case some challenges are known beforehand, such as:

- Currently, the roaming knowledge worker will use remote access techniques which:
    - are expensive;
    - are provided over slow links which do not match the quality of service requirements to ensure an adequate responsiveness of the system.
- The Cloud networking solution has to meet following challenges provide a user experience which is comparable with the user experience at the central headquarters when accessing both remote and local clouds:
    - meet the security requirements of the corporation, especially regarding the isolation of the different clouds the nomadic auditor is accessing simultaneously;
    - the solution needs to be scalable: the network operators should not impose limits on the number of clouds their infrastructures can provide. The number of virtual infrastructures they can provide should be at least in the realm of VPNs that can be provided today;
    - the performance of the solution must be consistent, and meet the expected quality of experience. Large-scale cloud-based services are provided simultaneously from several, potentially distant sites. Therefore, the assignment of a specific server to a specific service request must be optimized,

while taking into consideration current load of servers, and minimizing the amount of overhead.

Also from a business perspective we expect beforehand that the Cloud site providers will offer a better service to their customers in terms of quality of experience. When the established network operators cannot serve the full footprint covered by the nomadic auditor, short term allocations with local network operators can ensure the service level expected by the nomadic worker. The established network operators also profit, because they do not need to establish a full-blown infrastructure in places where the revenue stream wouldn't amortise the CAPEX/OPEX to cover sporadic peaks in resource demand: they build a rationalised infrastructure and rely on alternative providers to cover sporadic peeks in resource demand.

Use case 3: Business Goal Management

In this use-case it is considered a global corporation with typically 100+ offices distributed throughout all continents. The corporation is also heterogeneous, that is, not all offices are operated at the same administrative rules, e.g. due to national requirements (e.g. GmbH versus Ltd.), and many subsidiaries may exist that are owned by the main corporation. Currently, most of IT and networking is owned by the enterprise itself, and the enterprise provides its own infrastructure and services at a single site and between sites by leasing additional lines or using VPN over the Internet. By interconnecting multiple sites, the enterprise can implement and deploy applications in a distributed manner, and users, that are mainly employees of the corporation, are able to access these applications from other sites.

Recently, enterprises have started moving individual applications to the public cloud. Technically, due to the slow process in adopting virtualization technologies, migration to the public cloud is still in its early phase (see e.g. [30]).

Adopting all features to achieve a mature virtualization environment still takes time and has been mostly considered from the computing perspective only, while networking has largely be excluded from considerations. Furthermore, security aspects are of major concern and many open questions related to public cloud outsourcing of applications remain (link to security task). For this reason, it is particularly interesting to look at mixed scenarios. In a recent survey, more than 90% of participants consider a hybrid scenario as important, with varying emphasis on internal data centre and external cloud providers (see [31]).

In the figure below, a typical mixed setup is shown where the applications in a corporation are implemented in various forms, that is, conventional forms (e.g. non-cloud-compatible applications) and based on either the private or public cloud. Altogether, all forms form a cloud infrastructure and application ecosystem that needs to be managed at any given time and also during the introduction of new infrastructure and the migration of applications between different forms.

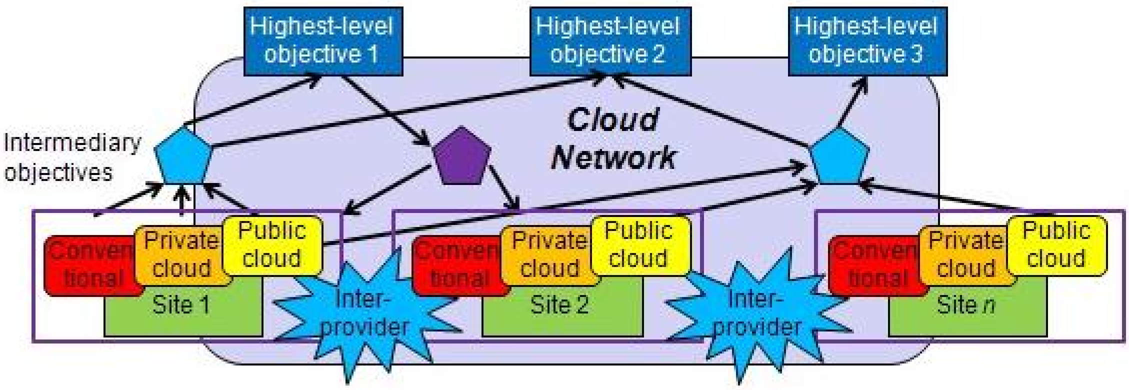

Figure 20: Enterprise heterogeneous setup for the *Business Goal Management* use case

An additional view in the system model of this use case in particular is the consideration if IT and networking resources within a single and consistent notion of a "hybrid flash slice". Such a slice allows the definition of both IT-related and network-related parameters, where network-related ones are currently not supported.

Further, in this use case, two sides can be considered separately. On the monitoring side, the corporation is constantly assessing the overall performance of the whole enterprise cloud ecosystem, and creates "snapshots" that characterize the overall situation. Performance characteristics are at very abstract levels that directly relates to business goals, but which are extracted from likely several layers of aggregation in a vertical aggregation hierarchy where the most basic parameters are obtained from individual hybrid flash slices. Based on the high-level information, the overall management system ensures that overall business goals are continually maintained by injecting management commands that are handed down vertically in a business goal refinement hierarchy, eventually re-parameterize hybrid flash slices at the lowest level. A concrete example of the whole process is the situation where an abstract, high-level KPI's threshold is exceeded (e.g. "about to run out of processing resources"), indicating that additional IT infrastructure needs to be incorporated, for example, from the cheapest IT infrastructure provider. The overall management control loop is highly dynamic and influenced also by external factors (e.g. pricing of virtual cloud operators, etc.).

In this use case all of the relevant actors identified in the broader scenario will play one or several roles. To implement this use case some challenges are known beforehand, such as:

- Heterogeneous physical network (developed vs. developing areas);
- Significant number of inter-operator issues;
- Maintenance of user experience especially in the case of failures
    - Example: maintain maximum end-to-end delays for certain transactions in any case.
- Efficiency in constructing high-level KPIs;
- Consistency in enforcing business goals on a global scale;
- Dynamics in local office scaling and introduction of new offices;

Continuity of business goal conformance is in particular challenging when the infrastructure that is being used for implementing the hybrid flash slices that are currently allocated to the enterprise are affected by challenges to the infrastructure underlying the cloud ecosystem, which may be on different scales, from local disturbances to physical disasters. Therefore, business goal management includes in particular strategies and specifications about how to initially configure and reconfigure the cloud ecosystem in case of challenges. For example, a resilient network connection may be the result of a high-level business goal that demands for certain parts of the cloud ecosystem to highly available. Furthermore, the challenge is to guarantee business goal conformance, but this may not always be possible depending on the magnitude of the challenge. Such situations need special considerations within a business goal management framework, where the enterprise is able to specify how much it is willing to invest (mostly in terms of monetary cost, both operating expenditure and capital expenditure) to maximize the continuity of business goals.

Also from a business perspective we expect beforehand, and especially in a hybrid cloud ecosystem, that new business models may become interesting. In particular, such models that target at migration aspects to facilitate and ease the migration of conventional or privately hosted cloud services to the public cloud. For example, with respect to security, specific packages may be offered to ease the migration in terms of security. Moreover, the hybrid cloud ecosystem model promises large flexibility for the enterprise to move its cloud needs between physical/virtual providers according to current pricing levels, while high-level business goals are transparently maintained. Such reconfiguration may occur on a yearly, monthly, daily, or even hourly basis. Furthermore, this last point relates to failure treatment as described above, for example, the enterprise can flexibly respond to a collapse of a complete provider and switch to another one.

Use case 4: Virtual Desktop

In this use-case a company has decided to provide its employees with "light" desktop machines that are extended on demand through the cloud. For this purpose, it has prepared a set of virtual machine images (e.g., basic browser/email image, office work image, developer images, researcher images, images for financial workers, etc.). These images are to be provisioned to the employees of the company on demand and to be removed, when they are not in use. Depending on his/her tasks at hand, an employee requests a specific VM image, the preferred virtual machine it should be run on (e.g., by specifying CPU, memory and storage) and other resources it needs (e.g., access to specific storage, network connections, etc.). Once the virtual machine is provisioned and running, the user connects to it through a remote desktop protocol.

The CEO and guests of the company always use the "browser/email" image with low CPU/low memory configuration and connectivity to the Internet. A researcher designing her simulations usually uses the "secretarial work" image or a "developer" image with medium CPU/low memory configuration. When the researcher runs her simulations, she needs a researcher image on a virtual machine with multiple CPU cores and large memory, with a high bandwidth access to the internal test bed and a large amount of temporary storage. Another researcher is involved in a collaborative engineering task with interactive 3D environments, which requires virtual machines with access to high-end CPUs. Accountants always require a machine a fast CPU/low memory and secure connection to an internal financial database.

The company has multiple offices across the globe and buys cloud services from multiple cloud providers, in order to ensure that virtual desktops are provisioned close to its employees. It also actively changes cloud providers depending on current pricing. This occurs both on a longer time scale, e.g. yearly for heavier types of long term storage and search, on a medium time scale, e.g. on a weekly basis for teleconferencing, and on a short time-scale, e.g. on an hour or minute basis for certain computations.

In order to minimize delays and improve the user experience, the location of the virtual machine where the desktop is installed should be optimized within the cloud. This service request-server assignment must also take into consideration the load of other virtual machines in the cloud, and might be migrated in real time (e.g. the employee logs on from a new remote site).

In this use case all of the relevant actors identified in the broader scenario will play one or several roles. To implement this use case some challenges are known beforehand, such as:

- Provision virtual machines with the required properties on demand;
- Provision the required network slices on demand;
- Dynamically reconfigure the current mapping of virtual to physical resources maximize efficiency and cater for user requirements;
- Interoperate with other cloud provider so as to be able to provision flash network slices on demand;
- Optimize virtual machine placement within the cloud in order to minimize delays and improve the user experience;
- Ensure that company data is kept secure and confidential, even from the administrators of the cloud environment;
- Ensure that company's data locality requirements as well as the states software export laws are respected;
- Provide a robust service even under failures of physical devices.

Also from a business perspective we expect beforehand that this arrangement will allow the enterprise to better manage its IT infrastructure, reduce hardware and software costs, increase service availability, increase the flexibility of IT resources and move load according to pricing.

11.1.4 Business Analysis of use case 1: Media Production

In this section we present the business case analysis for the *Media Production* use case.

Actors and roles

The following actors and roles are applicable for this use case:

Actors

- **Enterprise** - an enterprise is a company organised for commercial purposes. Some requirements of an enterprise are shared by other types of organisation, but an enterprise typically has stringent security and regulatory constraints to protect its intellectual property, information privacy, and to ensure legal compliance. In the scope of this use case, we consider an enterprise to be a large organisation with hundreds or thousands of employees, its own IT infrastructure and global presence.
- **Business partner** – a business partner is another company, customer or supplier that is involved with the enterprise in the course of its business. A business partner may be given access to some infrastructure and services of an enterprise in support of a business relationship between the partner and the enterprise.
- **Cloud site operator** - a cloud site operator is an organisation that manages a cluster of processing and storage infrastructure that implements an infrastructure service.
- **Network operator** - a network operator is an organisation that manages a network that implements communication services with flash network slices between the offices and remote infrastructures of an enterprise, its IT system administrators, its IT service users, and other cloud sites.

Roles

- **System administrator** - a system administrator is a person who administers the services and infrastructures owned by the enterprise.
- **Service user** – a service user is someone who has been authorized to use the services provided by a service provider. This entity could be an employee of the enterprise, residing in a branch office, or a business collaborator that is authorised to access some enterprises resources.
- **Service provider** – a service provider supplies an IT service. This may be an internal department of an enterprise or the main business of a third party such as an IT outsourcing company.
- **Virtual Infrastructure provider** – a virtual infrastructure provider supplies virtual infrastructure to run IT systems on. This may be an internal department of an enterprise or the main business of a third party such as a network operator or cloud site operator.

Technical and industry architecture

The technical and industry architecture diagram for the media production use case is shown below. In this use case the TV Channel is the enterprise.

The TV channel commissions media production companies to produce computer generated content for TV programmes. In order to reduce the costs and risk for its sub-contractors it acts as service provider, infrastructure provider and system administrator to create the media production service used to perform the commissions.

The media production service is data intensive and interactive. In order to optimise communication costs and to provide a service of reasonable quality worldwide the TV channel engages cloud site operators to provide locations to run the service near to the media producers and network operators to provide secure dynamic connectivity services between the TV channel, the cloud sites and the producers.

When a new commission is established the TV channel's system administrator instructs the channel's virtual infrastructure service to deploy an instance of the rendering service near to the media producer and then administers the service to allow access to the producer and to manage capacity of the service. If local, the service will be deployed in the channels own virtual infrastructure, if remote the infrastructure service may select a remote cloud site operator and obtain a virtual infrastructure created there. It will also connect any remote infrastructure, the TV channel and the media producer with flash network slices obtained from the network operator. Collectively these represent a virtual infrastructure distributed across the participants that supports an effective framework of collaborative Open System Organization paradigm.

When the media production company has completed its commission the content will be transferred to the TV channel. Note that the rendering service is logically located in the distributed virtual infrastructure administered by the TV channel, but may be physically located in a third party public cloud site. The transfer may represent a bulk transfer from the third party to the TV channels own systems. The bandwidth required for this transfer will be dynamically obtained from the network operator for the duration of this end-of-commission operation.

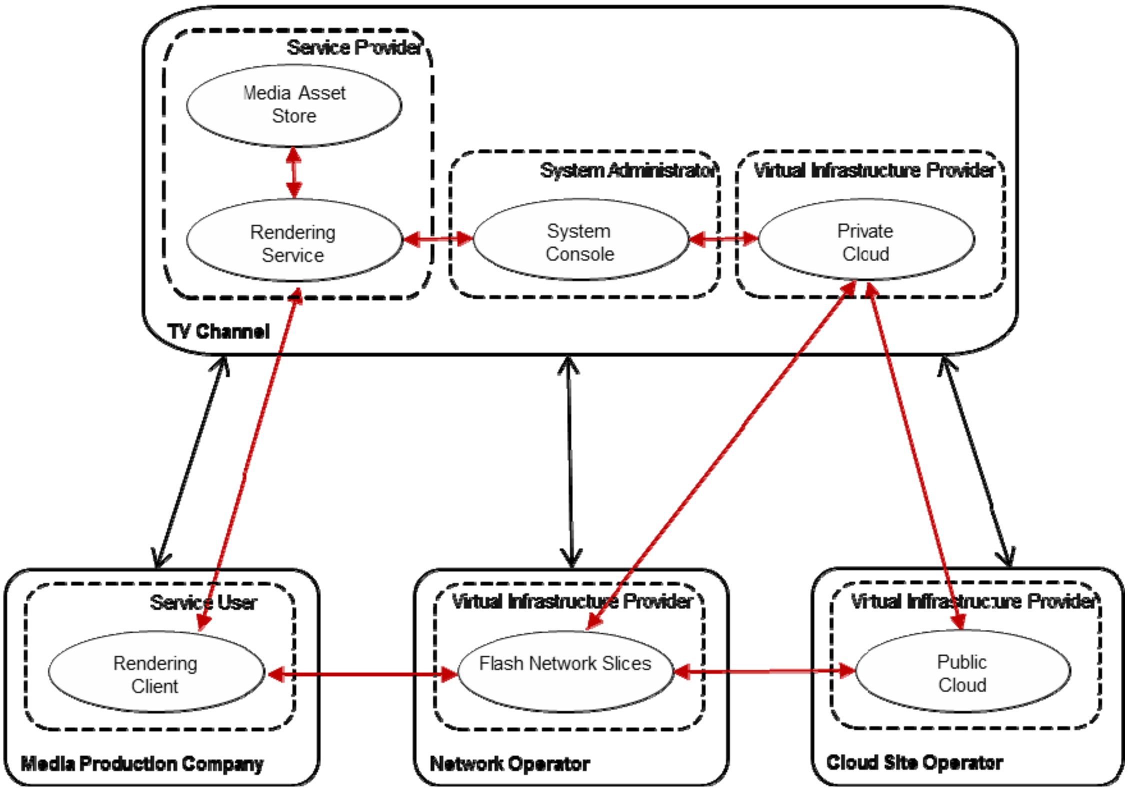

Figure 21: Industry and technical architecture for the *Media Production* use case

The above architecture diagram can be contrasted with the architecture used to support current normal practice. In this case the media production company administers its own infrastructure and rendering software. In doing so it is both provider and user of the software component.

When the TV channel enters into a commission with the media production company it has to purchase new equipment or allocate existing equipment to the engagement (although it may be able to share equipment across multiple concurrent engagements, it will need sufficient equipment for each engagement). The content generated in the production process resides in the media production company's infrastructure. At the end of the commission the content is delivered to the TV channel's asset store by either transfer across the Internet or often physically by courier.

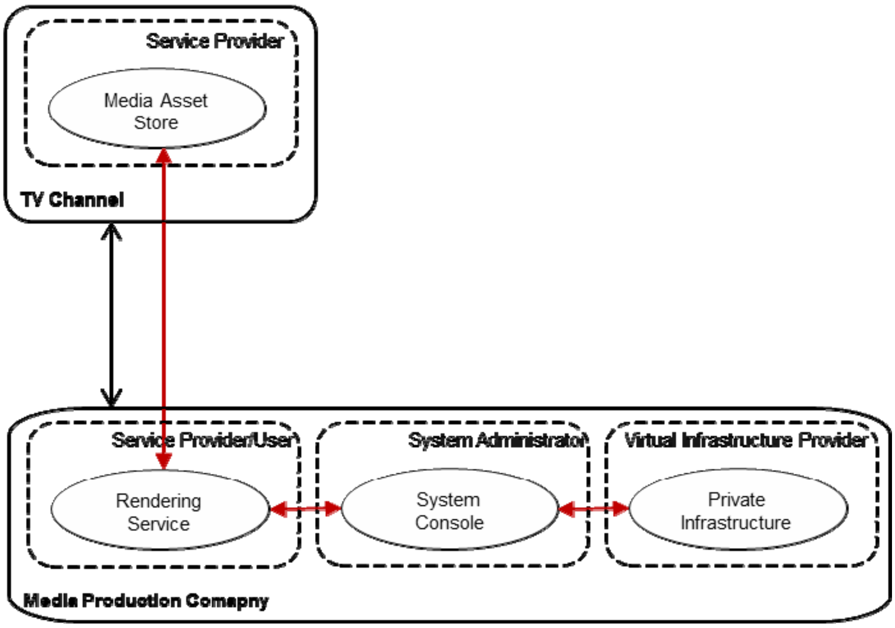

Figure 22: Industry and technical architecture for the current practice alternative *Media Production* use case

In the subsequent analysis the use case is evaluated first by comparing it to competing solutions, and secondly by analyzing its comparative attractiveness, *i.e.*, pros and con with the several other competing solutions.

Comparison to competing solutions

Competing solutions that support sub-contracting of media production include:

1. Sub-contractor uses private IT infrastructure;
2. Sub-contractor uses public cloud offering;
3. Sub-contractor uses independent service provider.

The most relevant comparison points for these competing solutions compared are highlighted in the table below.

Table 14: Comparison of Dynamic enterprise with Sail Technology to competing solutions

Criteria	Dynamic Enterprise with Sail Technology	Private IT Infrastructure	Public Cloud	Independent Service Provider
Security risk of leaked content during media production	Low	Medium	High	Medium to High
Security risk of leaked content during product delivery	Low	Medium	Medium	Medium
Time to setup	Low	High	Medium	Low
Scalability	High	Low	High	High
Management overhead and complexity for TV Channel	High	Low	Low	Low
Management overhead and complexity for media production company	Low	High	Medium	Low
Cost for TV channel	Medium	Low	Low	Low
Cost for production company	Low	High	Medium	Medium
Risk of Media Producer failing to deliver due to improper IT resources	Low	High	Medium to Low	Medium to Low

As can be seen the Dynamic Enterprise solution compares favourably in criteria except management overhead for the TV channel and costs for the TV channel. These disadvantages for the TV channel are traded against the reduced cost of the engagement due to reduced cost for the production company and the significantly reduced security risk. Although the TV channel may use a public cloud provider as part of its dynamic enterprise solution it now has complete control over the media content. It does not have to trust this security aspect to anyone else's technology decisions. The TV channel is likely to be engaged in a strategic (long term) business partner relationship with the cloud provider in which they are offered particular negotiated guarantees. A smaller company may not represent sufficient business to warrant special attention from the cloud provider.

The table also shows the costs for the TV channel to be medium although they now operate the infrastructure and production service. This is due to the fact that the cost is amortized across its engagements and so achieves better economies of scale; otherwise the costs would have been considered high.

The Private IT Infrastructure solution represents the traditional standard practice. The TV channel engages the media production company, provides input on the required content and expects the content to be delivered complete. The media production company owns and manages the hardware and software included in the production process. To cover the engagement the production company will purchases sufficient infrastructure or dedicate existing resources; it may also need to purchase licenses for software. When the media product is complete the content is delivered to the TV channel.

In this case the production company runs several risks: they are unable to scale their infrastructure to any great degree so they run the risk of purchasing too much, or not have sufficient resources to complete the contract on schedule. Ultimately they will never benefit from economies of scale. They will be required to trade-off content or quality against schedule to manage this risk. They will be held responsible for any failure in production.

The company will also be responsible for managing security of the media content. In many cases the novelty of media represents its value. Early release of all or part of the content or the potential for plagiarism can significantly impact its business value. The company will have to manage secure access to the content both through computer systems and through manual access by its employees.

The Public Cloud solution represents an emerging alternative. In this case the media production company mitigates its costs and introduces the flexibility to trade off cost and quality against content and schedule to meet its contractual commitments by using a cloud infrastructure provider. Although this option reduces management of physical infrastructure it does not reduce management of systems or reduce cost of software. It also increases the security risk as the cloud provider's security capabilities are brought into the equation. This is the main barrier to adoption and wide spread acceptance in the present market. Last but not least, the lack of use of flash slices network technology as proposed in SAIL will diminish the quality of experience perceived by the media producers and their productivity, thus increasing the overall costs of production.

The Independent Service Provider alternative represents the case in which a third party has established a service similar to the one that the TV channel would have implemented itself; this would be provided as a cloud service. Although this alternative has similar benefits in terms of cost to the production company and retains the low cost for the TV channel, it does not mitigate the overall security risks involved. In the TV and film industry the security of content is a paramount concern for new and novel material and in many cases it is likely that security will take precedence over cost. Lastly, as for the Public Cloud solution, the lack of flash slices network technology as proposed in SAIL will diminish the quality of experience perceived by the media producers and their productivity.

Pros and cons for key actors

The advantages for the TV channel and the production company are clear from the above comparison to competing solutions. In addition to simple cost benefits or technical advantages the solution opens up the media production market in a way that enables more experimentation, more product innovation and more specialized engagements. The risk and cost reductions for the production company make more contracts viable for more companies and even individual creative artists.

Table 15: Pros and cons for actors

Actor	Pros	Cons
TV Channel (Enterprise)	• Reduce cost of engagements • Mitigate security risks • Mitigate product delay risks • Increase choice of subcontractor • Low cost of opportunity for new projects • Increased product innovation	• Need to operate production service platforms • Need to manage strategic business relationships with Network operators and cloud site operators
Production Company **(Business Partner)**	• Make profit from pilot engagements • Trade-off cost/performance to meet schedule/quality • Relinquish responsible for production service issues • Mitigate security risks • Greater number of viable engagement opportunities • Low cost of opportunity for new projects • Increased product innovation	• Dependent on TV channel • Restricted production process options
Cloud Site Operator	• Potential business partner for TV channel • Increased business volume	•
Network Operator	• Potential business partner for TV channel • Increased on-net traffic • Increased business volume	•

In addition to the producer and consumer in the engagement, the dynamic enterprise includes the cloud site operator and network operator as actors. The solution allows them to participate in on-going engagements with a large customer. This relationship is likely to be along the lines of a strategic (long term) business partnership where they would otherwise have had to rely on the smaller more short term engagements with the production companies.

Conclusion of use case business analysis

The use of cloud computing with SAIL technology to deliver a rendering service to media production companies changes the cost structure and risks for the production companies in a way that allows small players to effectively participate in the market structure in equal terms with the big players. It is in the TV channel's interest to facilitate this change from the perspective of cost of engagements, variety of choice, increased innovation, risk of successful completion of the commissions and overall security. Also the overall market intangibles seem to be relevant in this use case.

This use case also opens up new business opportunities for network operators and cloud site operators that support the virtual infrastructure capabilities. Additionally there is clearly an opportunity for an independent organisation to adopt the role of service provider. The security risks and cost benefits suggest there is contention in this approach and it is likely to require a close relationship between the TV channel and the service provider. Potentially this could be an out-source relationship in which the third party actually implements the TV channel's service.

Last but not least, the provision of a secure and effective service relies on the development of a few supporting technologies, including: secure isolation of virtual infrastructures, negotiation of deployment topologies across multiple providers, composition of virtual infrastructures implemented by multiple providers, and unified management of the virtual infrastructures. The flash network slice is a key technology for this use case enabling the required distributed infrastructure.

11.2 Scenario 2: Elastic Video in the cloud

This scenario presents and depicts the offering of video and similar services from a cloud network ecosystem to the retail market with user perceived enhanced QoE (quality of experience), leveraged on distributed computational resources at the edge of the network architecture.

The scenario represents a framework of distributed resources in a cloud model, meaning that cloud resources are geographically scattered inside the operator network in a more *fine-grained* fashion than traditional centralized data centre clouds. A concrete example of what this could mean is the placement of cloud servers in operator network edge sites.

The driver for such a distributed deployment is to leverage on the closeness to the final users to offer enhanced QoE. In principle this closeness implies that latency can be reduced and network bandwidth usage patterns can be optimized, by trading off connectivity, processing and storage. Also the processing offload for resource limited user devices can be provided near the device.

Enhanced user perceived quality of experience is a driver for such a distributed type of deployment, because such a distributed deployment will result in a more scalable system as more servers are dynamically deployed when the numbers of users increase. Likewise the system will scale down when the numbers of users decrease. By using local distributed servers one will avoid the need to traverse congested core aggregation links of the network to the centralized servers, thus the higher QoE perception by the user. Also from the network operator perspective the edge servers could be feed with the relevant content at the moment when the core network is less congested thus balancing and smoothing the network traffic curve.

This type of deployment will allow multiple content providers to share the same distribution and computing infrastructure in cloud model. Content providers will no longer need dedicated physical servers but will use on-demand virtual servers. These virtual servers will be made available at the right place (network topology wise) and at the right time (when demanded and/or needed). In this way core network bandwidth should be saved since the same content will only be streamed once across specific network links.

These distributed servers in cloud architecture should not be viewed as replacements for the centralized (and typically huge) data centre clouds found today. There are important reasons why massive data centres are attractive, such as economy of scale and the ability to situate data centres at locations with low-priced power or cooling or in accordance with some other relevant criteria. This will likely not change for the foreseeable future. The distributed clouds will instead act as a complement to centralized data centres, effectively *bringing the cloud closer to the end users*.

Summarizing into one sentence, deploying distributed resources in a cloud model opens new ways of designing and implementing distributed applications and new ways to perform network adaptation in response to changing demands or irregularities in the operational environment as well as it should increase the perceived user QoE and should smooth core network traffic patterns.

To illustrate the concept a distributed cloud architecture is depicted below. Note that each one of the distributed servers could actually be a set of servers. Any type of application can be deployed on these servers through an adequate API in a PaaS model. The servers are independent and do not rely on any centralized functions. The nodes themselves will straightforwardly up or down scale when the numbers of users of the services increase or decrease.

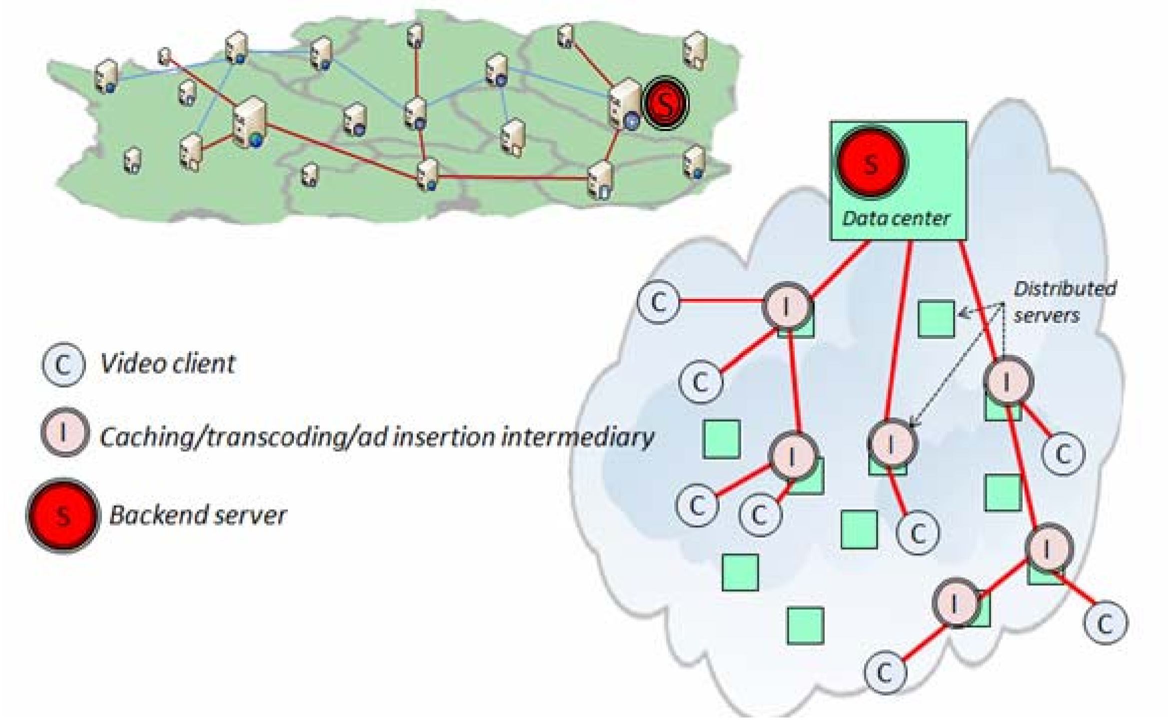

Figure 23: Distributed Cloud architecture concept illustration for *Live Video Distribution*

For this scenario four use cases were regarded as interesting to study, for they illustrate the scenario predictable usage and both technical and business related topics of concern. The use cases should be understood as particular application cases of this broad scenario above described.

The relevant actors for this scenario are listed below.

- **Network operator** - a network operator is an organisation that provides the distributed cloud within its network, either directly or indirectly. Directly means that the operator will buy, deploy and manage the distributed cloud. Indirectly means that the operator can team up with a cloud provider that wants to deploy its servers in a distributed fashion for the operator.
- **Content provider** – a content provider is an organisation that implements and deploys the applications and the content on the distributed cloud. The content provider may have specific constraints for the effective delivery of the service.
- **End-user** – an end-user can be a regular broadband user that wants to access the service. Otherwise, the end-user could be the employee of a company.

For this scenario four use cases were regarded as interesting to study, for they illustrate the scenario predictable usage and business related topics of concern. The use cases are (i) Elastic Live Video Distribution, (ii) Elastic Video On-demand Distribution, (iii) Video Conferencing and (iv) Distributed Gaming.

From a techno-economic perspective the elastic video distribution scenario brings some security considerations into play in addition to distribution over a proprietary or statically managed distribution service. These include:

- Separation of service providers (network and VM security);
- VM fairness and SLA fulfilment (computing and networking);
- Secure payment (secure accounting);
- Protection against user misuse (DRM, access control);
- Protection against denial of service.

11.2.1 Challenges

Some key challenges addressed by the elastic video in the cloud scenario are the following:

Reduced latency requirements: Perceived QoE is very much determined by the latency time between the user terminal equipment or CPE and the content delivery feeding server. In order to exhibit improved QoE the solution topology to be deployed must guarantee adequate server location at the edge network (this may imply dynamic virtual server location management), adequate access network bandwidth at all time to the deployed virtual servers, adequate core bandwidth when needed from the central data centre to the edge virtual servers in order to adequate populate them with the necessary content for distribution, and so forth. This requirement is of paramount importance in this scenario.

Fast delivery response times: For general cloud services, but specially for video and gaming services as in this scenario, the SAIL proposal must ensure that faster network response times to the final end user, in comparison with concurrent technological proposals, must be achieved, in spite of all of the management and control overheads required for the timely and adequate SAIL technology to perform.

Resource optimization: One of the relative advantages of the SAIL proposal in comparison with concurrent technological proposals is the resource optimization capabilities from the provider's point of view. Management mechanisms for effective assuring network bandwidth savings and high efficiency in the utilization of the distributed computational resources are of vital importance for this scenario.

Automatic scalability: Other relative advantage of the SAIL proposal in comparison with concurrent technological proposals is the automatic and efficient scalability of the deployed edge virtual servers in function of the end users content demands. This mechanism must be well guaranteed by the CloNe solutions.

Content distribution flexibility: In typical scenarios of video content delivery and network-based gaming the providers may feel the need to geographic segment the content in function of the end user preferences by geographical regions. Tools for providing flexible content distribution with market segmentation capabilities by several criteria must be made available in the solution to be developed.

Usage monitoring tools: Complex business intelligent (BI) tools and OSS tools are required in order to permanently monitor the network topology, resource usage, SLA KPIs, and so forth. Resource optimization will only be possible to achieve with this type of permanent monitoring. Also from the business point of view BI tools are of vital importance to successfully control the business outlook.

Business value chain cooperation: For this scenario the CPE equipments and the home peripherals that connect to the CPEs (e.g., TV set-top boxes, joysticks, consoles, etc) must evolve in accordance with the network service capabilities in order for the end users to capture the value proposition from the cloud network side and thus increase the business around this scenario. An effort must be put forward by the actors into business value chain cooperation with actors from the related business areas.

11.2.2 Market Potential

The Elastic Video in the Cloud scenario market potential assessment should be framed within a broader analysis of the general trends of the video industry, of the gaming industry and of similar entertainment and user generated content industries. In the future this scenario paradigm might constitute an alternative to some of the current commercial offerings in the industries with a relevant video component, possibly replacing some of the current offline commercial product formats for online alternatives.

The traffic in the Internet has been growing steadily since its origins as the reports of the Minnesota Internet Traffic Studies [32], which is an entity that studies the Internet traffic trends, illustrates. They estimated that the world Internet traffic for 2009 was 7,500-12,000 PB and they estimate that by 2015 it will be 1 Zettabyte in the US alone.

A very high percentage of that traffic is derived from video over the Internet. As more and more home equipments (like TV sets and game consoles) are offered in the market with native network interfaces, the gap between the end users and the content producers will narrow down and the networks that will be able to deliver high quality video, to scale up and down with demand and to proportion high perceived QoE will have exceptional opportunities to prosper, to innovate product wise and to develop new revenues.

Cisco [33] has studied the current demand and forecasted the future trends of video traffic in the Internet in its Visual Networking Index.

- Internet video is now over one-third of all consumers Internet traffic. It will approach 40 percent of consumer Internet traffic by the end of 2010.
- Real-time video is growing in importance. By 2014, Internet TV will be over 8 percent of consumer Internet traffic, and ambient video will be an additional 5 percent of consumer Internet traffic. Live TV has gained substantial ground in the past few years. Globally, P2P TV is now over 280 Petabytes per month.
- Web-based video conferencing will grow 180-fold from 2009 to 2014. Web-based video conferencing is the fastest growing sub-category (183 percent CAGR from 2009 to 2014) within the business portion of the Cisco VNI Forecast.
- HD video conferencing will account for over half (57 percent) of business video conferencing traffic in 2014, up from 31 percent in 2009. Over one-half of business videoconferencing traffic will travel over the Internet by 2014.
- By 2014, the various forms of video (TV, VoD, Internet Video, and P2P) will exceed 91 % of global consumer traffic. Internet video alone will account for 40 % of all consumer Internet traffic by the end of 2010, and 57 % by 2014. It would take over two years to watch the amount o video that will cross global IP networks every second in 2014.
- Video-on-demand (VoD) traffic will double every two and a half years through 2014. Consumer IPTV and CATV traffic will grow at a 33 percent CAGR between 2009 and 2014.

The revenues expected from streaming media are also expected to surge over the years. Insight Research forecasts that in the US, streaming content will generate almost $70 billion in revenues by 2013 [34]. In the United States, Netflix (online video rental) represents more than 20 percent of downstream traffic during peak times and is heaviest between 8-10 p.m. [35]. The growth of the Internet video traffic has put operators under pressure to find efficient ways to deliver that traffic while keeping costs down.

While the overall application mix is shifting toward video, video is undergoing internal shifts of its own. In particular, real-time video is growing in importance and complexity with HDTV, 3DTV, online gaming and so forth, requiring ever more computational and storage capabilities, either from the network side, either from the end user home equipment side.

Cap Gemini estimates that the digital electronic entertainment will amount to $10 billion USD by 2013 [35] to be shared by the value chain actors ((1)Media/content creation; (2)Media/content aggregation and packaging; (3)Media/content distribution; and (4)Consumer devices and technologies), and current trends indicate that a relevant portion of these services will be in online formats.

From this perspective the Content Delivery Networks (CDN) is another market whose growth closely relates to the potential in the video distribution case. Most of the CDN provider's main customers are content provider sites and bulks of the revenues come from delivering videos or games over the Internet. Strategy Analytics August 2009 report has the following projections in the CDN market.

- Revenues from digital content expected to reach $165 billion by 2015;
- The market for CDN services is estimated to be $5.2 billion by 2013.

Also the game console evolution from stand alone consoles to online consoles and online gaming is very relevant for this scenario. The movement towards connected consoles is by now consolidated in the industry, putting pressure on the network services to avoid underperformance.

IDC [37] predicts a growing revenue scenario worldwide for the connected console industry until 2013 and that, despite the deceleration from earlier predictions, is a clear indication of market adoption and expansion that will definitely shape this industry value chain.

It is expected that the power balance between the several industry actors will evolve in favour of a greater empowerment of specialized independent game/content producers carrying further innovation and competition benefits on game production and distribution to the end users. Not only game distribution but also video content will be supported by the connected consoles opening up a new channel to reach the households and thus animate the entire content value chain as well as posing new business challenges to the actors [36][38].

The market potential for this scenario brought about by the connected console trend should also be extended by market pull due to related product trends such as in Smartphone capabilities, tablet PCs, Kindles and iPads capabilities, the connected home trends, user-generated content trends and so forth.

From the above analysis the proposed scenario appears to exhibit significant market potential both, either from a tangible point of view, either from an intangible point of view.

11.2.3 Use case descriptions

In this section we describe four example use cases for this scenario predictable usage.

Use case 1: Elastic Live Video Distribution

This use case is about a large global/regional event where an extremely large number of users connect live. Live events include sports events, breaking news, music concerts, and similar.

Planned events allow for pre-provisioning and they have more requirements. User expectations are much higher (e.g., HD video). Unplanned events demands are lower, that is, one needs an efficient delivery of content without much expectations (it should work fairly and it should be a reactive system). In most of the cases one cannot predict how many users will be watching a specific content.

Scaling of the application is automatically provisioned by the distributed cloud according to the behaviour of the users. The scaling is controlled by the constraints set by the application provider. The distributed application in the cloud can be scaled up or down dynamically depending on application usage and set of constraints. The solution solves the near-term issues related to the use of IP multicast in a multi-domain scenario. This is a solution that allows for more flexible deployment of the servers.

In this use case all of the relevant actors identified in the broader scenario will play one or several roles. To implement this use case some challenges are known beforehand, such as:

- Reduce latency of access to the content;
- Improve the quality of experience (faster response times);
- Optimize connectivity, processing and storage as one problem;
- Optimize resource utilization: bandwidth and video servers;
- Ability to scale up and down automatically as a result of service usage;

- Continuously monitor resource usage to allow effective service delivery.

Also from a business perspective the network operator will offer a paid service to the content provider. The network operator will also benefit as it will experience better usage of its network/computing resources. The end-user will experience faster response times especially under heavy usage peak periods. The content provider does not need to build a massive server farm for video distribution or to have a specialized IT department to deal with it. The content provider can focus on media content production while the network operator will focus on scaling of the distribution network. This opens up new business opportunities for small independent content providers.

Use case 2: Distributed Gaming

This use case decentralises gaming and moves the core of the applications into the cloud. The idea is to depart from the console centric model to a distributed gaming service relying on clouds to offer 3D games to light (or thin) clients. As opposed to the traditional setting where games require the deployment of applications on the clients or the acquisition of personal consoles, the users will simply use their personal computers and phones to visualise the game in 3D through streaming, compression and decompression. These capabilities will be located in the data centres as well as incarnated in the networks.

This leads in essence to remote 3D games and rendering orchestration and management with visualisation in distributed thin clients. Note that the clients are themselves becoming sufficiently powerful to handle the visualisation and decompression requirements. The cloud and the networks would cover mostly the multiple view rendering (specialized to each user and one joint view of the entire game in the cloud), the rendering and the compression and the networks would also conduct transcoding and content adaptation as needed. The thin clients would handle the decompression and visualisation.

In this use case all of the relevant actors identified in the broader scenario will play one or several roles, but the Content Provider actor must be redefined in a broader sense:

- **Content Provider** – a content provider in this context will be a video game content editor and provider as in an organisation that implements and deploys the video game content on the distributed cloud.

To implement this use case some challenges are known beforehand, such as:

- Pixel latency (< 15 ms) to fulfil rendering requirements and synchronisation;
- Dynamic scaling with number of users and provisioning of flash slices towards end-users;
- Efficient assignment of game servers to end user requests;
- Quality of experience and minimum bandwidth requirements;
- Distributed processing (compression and transcoding) on route according to end user devices, e.g. video player and device OS;
- Joint optimisation of compute, storage, caching, processing and communications/networking.

Also from a business perspective SAIL Cloud Networking combined with cloud centric games where all the complex and intensive computations take place plays a central role in strengthening this use case in order to:

- Remove the requirement on end users to acquire an expensive personal console and software (games on CDs on uploaded from the web) by replacing the console by a game service in the cloud enhanced and augmented with capabilities (compression, decompression, caching and distribution) in the networks;
- Offer unprecedented 3D rendering and quality to end users on their terminals. The terminals will focus on visualisation and playing the streams and in few cases

decompression. The target can be at home but may as well be specialized mobile terminals whose technology improvements are removing barriers;

- End game piracy and illegal copying and use, the game and software are in the cloud;
- Reduces dependence on console technologies and barriers to focus on content, game creation and production;
- Opens the opportunity for setting up coalitions (or federations) between network providers, network technology manufacturers, 3D technology specialists, game creators, producers and vendors.

There are also some inherent advantages for end users that do not need to update their software and version as this is updated immediately and automatically in the cloud. No need to purchase yet another new game console, the cloud does the computing and the synchronisation/harmonisation. The actors can also more easily protect their assets, the software and the games. The game and video game industry gains in autonomy and improved time to market and achieves instant outreach. Networks providers strengthen their role and presence in the game industry by acting as content and video content distributors with additional capabilities in the networks.

Use case 3: Elastic Video On-demand Distribution

In this use case caching servers are dynamically deployed in the network. Cache servers store more content than in the previous case. This use case will allow the creation of a dynamic content distribution network where the cache servers are dynamically created and released. The application provided is a network Personal Video Recorder (network PVR). Another possibility is for provisioning of pay-per-view content.

The system could be able to perform automatic recording of programs based on user profiling. In order to optimize delivery, content can be pre-recorded in the network servers. Furthermore, content can be classified based on keywords, meta-data, or titles allowing the system to make suggestions to a user based on content he/she watches. The distributed application servers will allow for a more efficient way to deliver advanced services, for example, users could be able to jump back and watch past content. The terminals do not need to be able to record material since it will be available in the distributed cloud.

The content servers distributed in the cloud can also perform transcoding if needed. This will suit the situation where the network does not have the resources to deliver a high quality stream. Instead of service disruption or no access to service at all, the system will be able to adapt the stream (video bit rate) to existing network conditions. This enhances the end-user quality of experience since video quality degradation is better than service disruption.

In this use case all of the relevant actors identified in the broader scenario will play one or several roles. To implement this use case some challenges are known beforehand, such as:

- Reduce latency of access to the content;
- Improve the quality of experience (faster response times);
- Optimize connectivity, processing and storage as one problem;
- Optimize resource utilization, namely bandwidth and video servers;
- Ability to scale up and down automatically as a result of service usage;
- Continuously monitor resource usage to allow optimization;
- Facilitate live service migration, when servers become overloaded, with minimal service disruption.

Also from a business perspective the service is essentially an enhanced content delivery network (CDN) that can adapt to flash crowd conditions by increasing the number of servers dynamically. The network operator will then be able to compete with traditional CDN providers by offering a pay-per-use service model. Besides, the distribution network can be shared by many different content providers or other services saving the need for physical deployment of new servers.

The network operator can resell this service to a traditional CDN provider. Instead of deploying physical servers the CDN provider can rent the cache nodes from the network operator for specific time periods.

This service offers the possibility for the user to record its content on network servers (network PVR). This opens up for the possibility of getting some payment from the end-users.

Use case 4: Video Conferencing

In this use case existing professional video conferencing system (like HP Halo) utilize static VPNs (or leased lines) to connect multiple sites. Flash network slices could yield more efficient use of network resources. Sites can be connected on demand only for the duration of the video call. The establishment of such a flash network slice should be performed in timely manner, emphasizing the need for rapid configuration of the network. Yet another important requirement is low latency, in order to provide natural interaction between the participants.

The video conferencing servers can also be created on demand. The servers are only provisioned for the duration of the call, being released afterwards.

The use of a flash network slices to provide connectivity to a video conferencing system will provide increased flexibility to connect to other sites and lower costs. Current state of the art is based on point-to-point static connections.

In this use case all of the relevant actors identified in the broader scenario will play one or several roles. To implement this use case some challenges are known beforehand, such as:

- Provisioning of flash network slice with low latency between participating sites;
- The flash network slice should be fast provisioned to support unplanned conferences;
- Minimum bandwidth for ensuring video quality must be provided;
- Efficient and dynamic allocation of video conferencing servers.

Also from a business perspective there will be more efficient use of network resources since the service will run over a shared network, instead of leased lines. The connectivity cost for teleconferencing provider will likely be lower in comparison to current alternatives.

11.2.4 Business Analysis of use case 1: Elastic Live Video Distribution

In this section we present the business case analysis for the *Elastic Live Video Distribution* use case.

Actors and roles

The following actors and roles are applicable for this use case:

Actors

- **Network operator** - a network operator is an organisation that provides the distributed cloud within its network, either directly or indirectly. Directly means that the operator will buy, deploy and manage the distributed cloud. Indirectly means that the operator can team up with a cloud provider that wants to deploy its servers in a distributed fashion for the operator.
- **Content provider** – a content provider is an organisation that implements and deploys the applications and the content on the distributed cloud. The content provider may have specific constraints for the effective delivery of the service.
- **End user** – an end-user can be a regular broadband user that wants to access the service. Otherwise, the end-user could be the employee of a company.

Roles

Note: In this particular use case it is relevant to also list here the internal roles or functions within the operator for better comprehension of the underling internal functions.

- **Application or content hosting:** Are the distributed entities that will compute the application. These are the distributed servers where the content is processed (e.g. rendered);
- **Application or content storage:** Are the distributed entities that will store the content to be distributed. These are the distributed servers where the content is located;
- **Application or content consumption:** Are the entities that will consume the video content. These are the end user visualization devices;
- **Network connectivity:** This is the function responsible for the transport of the content from the distributed servers to end users devices.
- **Application or content replication:** This is the function responsible for replicating the media streams inside the network.
- **Application or content deployment:** This is the function responsible for the management and deployment of the application or content in the cloud environment.
- **Dynamic scaling:** This is the function responsible for the scaling capabilities of the network regarding computing and storage resources.

Technical and industry architecture

The technical and industry architecture diagram for the elastic video distribution use case is shown below.

The three main actors in this use case architecture, (i) the end user, (ii) the content provider and the (iii) network operator, play the several roles illustrated below and commit with each other in business interfaces, by means of contractual obligations, as shown. The content provider is the owner of the content and the ultimate responsible for it, including making it available in its streaming server in a timely manner to the other actors. The network operator provides the connectivity service that allows the end user to access the content. The content provider also has a business interface with the network operator for delivering its content over the network. In this use case this business interface is not just a basic connection pipe contract, it is enriched with the required functions to guarantee the delivery of the content in an effective and scalable manner using the cloud network service provisioning provided by the network operator.

The technical interfaces between the technical components can be broken down as follows:

a) The end users device exhibits a technical interface towards the network infrastructure from the network operator, over which the content is delivered to the end user device;
b) The content provider streaming server exhibits an interface towards the application deployment service within the network operator;
c) The network operator internal processes implement the cloud network delivery service between the content provider and the end user.

When an application or content has to be deployed within the network, the deployment service makes use of the compute and storage services to distribute the application or content in the network. At this moment the end user service (video channel) can be initiated. The deployment service must also interface with the dynamic scaling decision logic that decides the initial compute requirements to use, as well as during the run time period of the application or content delivery, to provide the scalability management capabilities required. The core network connectivity requirements between different hosting sites are resolved by contacting the topology service within the network connectivity role. Once the route has been determined, the path provisioning service can be used to setup the path in the network infrastructure. The storage component interacts with the allocation and migration service if the application has to meet those constraints as well.

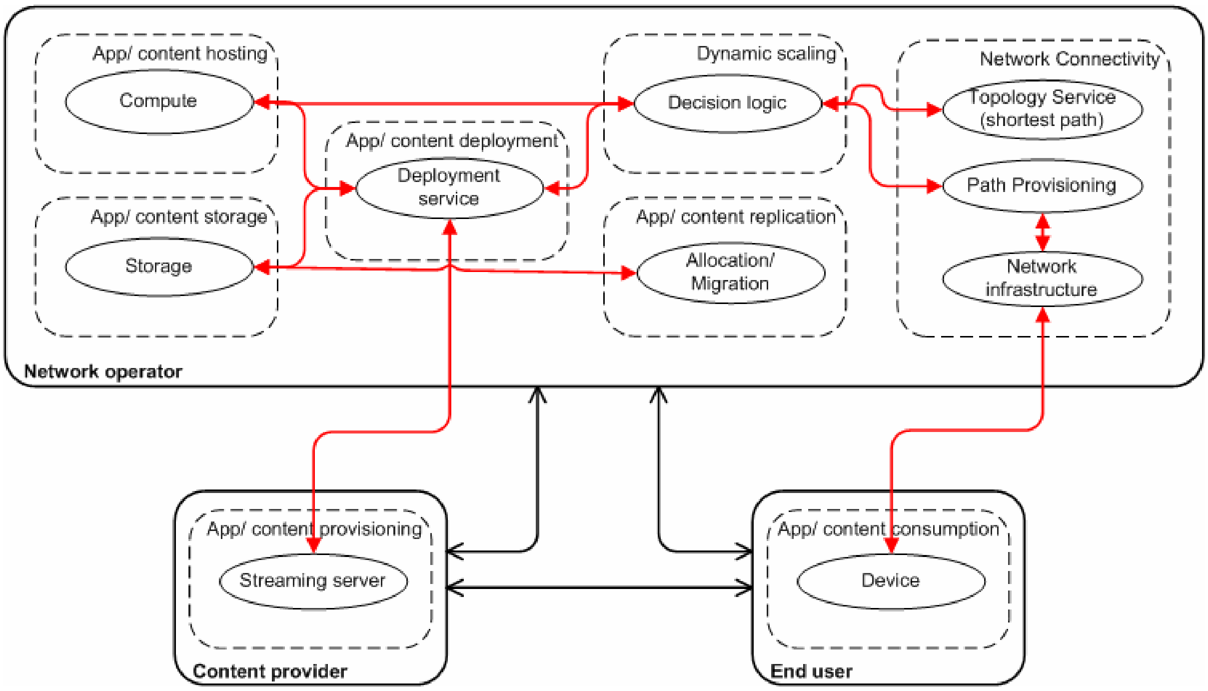

Figure 24: Industry and technical architecture for the *Elastic Live Video Distribution* use case

In the subsequent analysis the use case is evaluated first by comparing it to competing solutions, and secondly by analyzing its comparative attractiveness, *i.e.*, pros and con with the several other competing solutions.

Comparison to competing solutions

Competing solutions for the delivery of live video content include:

1. Native IP multicast;
2. Traditional CDN (such as Akamai);
3. Classical client-server (SVTPlay on QBrick server farm);
4. DIY CDN using a public cloud (Netflix on Amazon).

The most relevant comparison points for these competing solutions compared are highlighted in the table below.

Table 16: Comparison to the competing solutions

Criteria	SAIL Tech.	Native IP multicast	Traditional CDN	Classical client-server	DIY CDN using a public cloud
Scalability on the number of users	High	Medium	Medium	Low	High
Resilience to network failures	High	Medium	High	Low	Medium
Core network bandwidth consumption	Medium to Low	Medium to Low	Medium to Low	High	High
Cost for the network operator	Medium	High	Low	Low	Low
Revenues for the network operator	High	Medium	Low	Low	Low
Management overhead and complexity for the network operator	High	Medium	Low	Low	Medium
Perceived end user QoE	High	Medium	Medium	Low	Medium

SAIL technology compares well with all of the other technologies regarding scalability, resilience to network failures and perceived end user QoE. On one hand, while traditional CDN or IP multicast solutions can scale in a premeditated or planned approach, in function of end user behaviour monitoring, cloud solutions offer automatic scalability capabilities as an intrinsic characteristic thus exhibiting a comparative advantage. On the other hand, SAIL as a network solution has built-in mechanisms to secure performance when network failures occur thus its comparative advantage on QoE perception.

In addition it compares well in terms of revenue stream for the network operator, facing all of the other alternatives, by offering the possibility of further flexible and innovative business models for the network operator and alternative commercial options for the content producers to reach the end users.

Pros and cons for key actors

The advantages for the end user, the network operator and the content provider are clear from the above comparison to competing solutions. In addition to the cost benefits or technical advantages the solution will offer enhanced perceived QoE to the end users thus motivating higher usage rates and higher service adoption rates.

Table 17: Pros and cons from the perspective of key actors

Actor	Pros	Cons
End-user	• Better end-user experience QoE by means of faster response times.	• It might require the installation of a new software in the end-users' CPE.
Network operator	• Optimized utilization of video servers and bandwidth; • New revenue sources; • New innovative business models.	• Must deploy and operate the distributed cloud architecture in the network. •
Content provider	• No need to provide infrastructure for delivery of content to large audiences; • Lower cost compared to deploying delivery infrastructure by itself; • Resilience to flash crowd scenarios; • Enhanced commercial options to reach the end users.	• The content provider will not have full control over the end-users' video sessions. For some purposes (e.g., targeted advertising) that may be useful.

From the network operator view point, in spite of the superior cost and additional network management complexity, SAIL proposal offers additional revenue streams, innovating business models, as well as savings on the core bandwidth traffic, thus it seems to be a value proposition to consider for this actor.

Furthermore from the end user perspective and from the content provider perspective the enhanced QoE perception, the flexibility offered and the apparent adequacy to their requirements are solid enough arguments to justify SAIL technology adoption by these actors.

Conclusion of use case business analysis

The use of cloud computing with SAIL technology to deliver elastic live video to end users emerge after the business case analysis as a viable use case to be pursued within the industry ecosystem, since it provides overall market value from all perspectives.

11.2.5 Business Analysis of use case 2: Distributed Gaming

In this section we present the business case analysis for the *Distributed Gaming* use case.

Actors and roles

The following actors and roles are applicable for this use case:

Actors

- **Network operator** - a network operator is an organisation that provides the distributed cloud within its network, either directly or indirectly. Directly means that the operator will buy, deploy and manage the distributed cloud. Indirectly means that the operator can team up with a cloud provider that wants to deploy its servers in a distributed fashion for the operator.
- **Content Provider** – a content provider in this context will be a video game content editor and provider as in an organisation that implements and deploys the video game content on the distributed cloud.
- **End-user** – an end-user can be a regular broadband user that wants to access the service.

Roles

Note: In this particular use case it is relevant to also list the internal roles or functions within the operator for better comprehension of the underling internal functions.

- **Application or content hosting:** The distributed entities that will compute the application. These are the distributed servers where the content is processed;
- **Application or content storage:** The distributed entities that will store the content to be distributed. These are the distributed servers where the content is located;
- **Application or content consumption:** The entities that will consume the game content. These are the end user visualization and play devices;
- **Network connectivity:** The function responsible for the transport of the content from the distributed servers to end users devices.
- **Application or content replication:** The function responsible for replicating the media streams inside the network.
- **Cloud manager service:** The function responsible for the management and deployment of the application or content in the cloud environment.
- **Dynamic scaling:** The function responsible for the scaling capabilities of the network regarding computing and storage resources.

Technical and industry architecture

The technical and industry architecture diagram for the distributed gaming use case is shown below.

For the deployment of a distributed gaming service the content provider (in this case acting as the game provider) should obtain the required computing and storage resources from the cloud network service of one network operator as well as secure network resources from one or multiple network operators in order to establish an end-to-end connectivity service (i.e., between the end user and the cloud network service). The network provider can additionally provision compute and storage resources to undertake flow processing, such as video transcoding, depending on network dynamics (e.g., available bandwidth) and the client device capabilities (e.g., screen resolution).

The main technical and business interfaces between these actors are as follows:

- Cloud resource discovery and provisioning is managed and coordinated by a cloud manager service from the network provider who exposes an interface to the content provider for leasing compute and storage resources, according to the initial requirements of the gaming service deployment. Increasing demand can be matched by scaling up resources in a timely manner.
- A network operator exposes an interface to the content provider for the provisioning of the connectivity service. This service can include the lease of compute and storage resources, allowing flow processing or caching to be incarnated into the network. Information communicated via this interface will typically include network topology and

resource description, including constraints or requirements in terms of location, bandwidth, CPU capacity, and so forth.

- The network operator(s) should provide attachment points for end users, realizing the end-to-end connectivity service. The end user needs to establish connectivity with the flash slice network, which subsequently requires authentication services with the network operator. End user attachment can be achieved via existing technologies, such as VPNs.

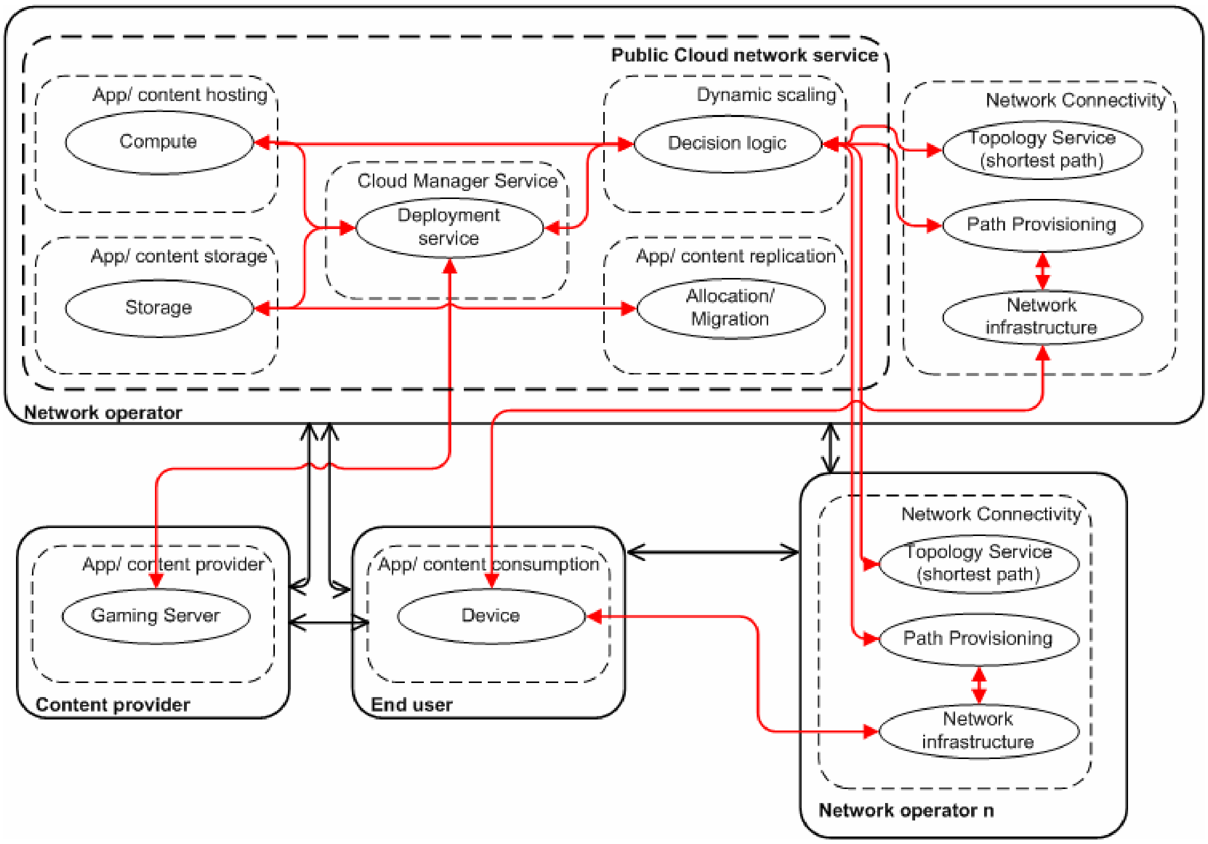

Figure 25: Industry and technical architecture for the *Distributed Gaming* use case

In the subsequent analysis the use case is evaluated first by comparing it to competing solutions, and secondly by analyzing its comparative attractiveness, *i.e.*, pros and con with the several other competing solutions.

Comparison to competing solutions

The distributed gaming option should be compared with traditional gaming alternatives (i.e., games running completely in client devices, such as consoles, PCs, PDAs, Smartphone, and so forth) that dominate the gaming market for this will be the main competitor to this type of offer.

The most relevant comparison points for these two competing options compared are highlighted in the table below.

Table 18: Comparison to competing solutions

Criteria	SAIL Technology	Traditional Gaming
Availability and diversity of games	High	Medium to High
Availability of game demos	High	Low
Game cost for end-user	Low	High
Need to upgrade equipment / client device	Low	Medium
Resilience to network failures	Medium to Low	High
Gaming externality effect (e.g. group gaming)	High	Low
New business opportunities and models for all the actors	High	Low

The table above shows that the proposed SAIL technology compares well for this uses case with the traditional gaming option for it potentially increases the game availability and diversity, including usable demos, and decreases the end user cost per use and per device allowing them to expand the game variety experience.

Additionally this solution paves the way for potential externality gains for the end users as well as for the content providers, opening up new business opportunities within the entire ecosystem.

Pros and cons for key actors

The advantages for all actors are clear from the above comparison to competing option. In the next table are highlighted the benefits and drawbacks over the traditional gaming option. These pros and cons are grouped among actors to better illustrate the business opportunities for all major stakeholders.

Table 19: Pros and cons from the perspective of key actors

Actor	Pros	Cons
End user	• Remove the requirement (on end-users) to acquire costly personal console and software by replacing the console by a game service in the cloud enhanced and augmented with capabilities (transcoding, caching and distribution) in the networks. • Offer unprecedented 3D rendering and quality to end- users on their terminals. The terminals will be only required to undertake visualization and stream playback and in few cases stream decoding. The target can be at home but may as well be specialized mobile terminals whose technology improvements are removing barriers. • End users do not need to update their software and version as this update takes place immediately and automatically in the cloud. There is also no need to purchase yet another new game console, since the cloud providers computing (which can be easily scaled up to meet the requirements of forthcoming very high-end games) and the synchronization / harmonization.	• End-user might experience impairments on visual quality during gaming, as the effect of variable delays and/or insufficient bandwidth. The gaming experience might also be degraded due to mobility issues, when the client device is a mobile terminal. In general, the experience of playing high-end games from the cloud might not match the quality level of games running in the client devices. • It is expected that the user device terminal equipment technology will evolve to accommodate the performance required for the end user to experience the necessary QoE that enables service adoption.
Content Provider	• Prevent game piracy and illegal copying and use, since game and software are in the cloud. • Protect more easily their assets, the software and the games. Game industry can gain in autonomy and improved time to market and achieves instant outreach. • Reduces dependence on console technologies and barriers to focus on content, game creation and production.	• Increased competition on the game production and deployment arena. • It is expected that the game production technology will evolve to accommodate the performance required for the end user to experience the necessary QoE that enables service adoption in a large variety of terminal devices.
Network Provider	• Networks providers can strengthen their role and presence in the game industry by acting as (video) content distributors with additional capabilities in the networks.	• Network performance requirements and SLA will be of paramount importance for the end user QoE. This will have a relevant impact on the OPEX and the internal operations management of the network operator.
All actors	• Open the opportunity for setting up coalitions (or federations) between network providers, network technology manufacturers, 3D technology specialists, game creators, producers and vendors.	•

From the above SWOT analysis it look as if the SAIL technology applied to this use case brings about concrete benefits for all of the actors that will by far outweigh the other option of standalone consoles.

A latent challenge within this use case seems to be the required evolution in distributed game production technology and in end user devices to take full advantage from the SAIL cloud network capabilities.

Conclusion of use case business analysis

In conclusion the use of cloud computing with SAIL technology to deliver distributed gaming to end users emerges after the business case analysis as a viable use case to be pursued within the industry ecosystem for it conveys new degrees of freedom to all of the actors in this market and potentially increases the overall value chain worth and attraction.

12 Conclusion

In this document we have arrived at a number of use cases that will guide the forthcoming work within SAIL. The use cases have been analyzed both from technical, architectural and business perspectives.

The uses cases as such, and the path that was followed during the work (from three dimensions of future networks, over a base scenario, to a number of scenarios defined in each work package and finally arriving at the use cases) have been of great value for the SAIL project.

Based on the analysis we have a strong foundation for the continued work, where both the validity of the use cases is established and where the SAIL project has reach a common view and understanding.

13 List of acronyms

API	Application Programming Interface
AS	Autonomous Systems
ASN	Autonomous System Number
BGP	Border Gateway Protocol
BI	Business Intelligence
CCN	Content Centric Networking
CDN	Content Delivery Network
CloNe	Cloud Networking
CP	Content Provider
CPE	Customer Premise Equipment
CPU	Central Processing Unit
CRM	Customer Relationship Management
DIY	Do It Yourself
DTN	Delay Tolerant Network
IaaS	Infrastructure as a Service
IAP	Internet Access Provider
IBP	Internet Backbone Provider
ICT	Information and Communication Technologies
IPTV	Internet Protocol TV (sometimes Interactive Personal TV)
IS	Information System
ISP	Internet Service Provider
IT	Information Technology
KPI	Key Performance Indicator
MAP	Mesh Access Point
MultiP	Multi-path/multi-protocol
NAT	Network Address Translation
NetInf	Network of Information
N4C	Networking for Communication Challenged Communities (see www.n4c.eu)
NC	Network Coding
NRS	Name Resolution Service
OConS	Open Connectivity System
OTT	Over-The-Top
P2P	Peer to Peer
PaaS	Platform as a Service
PDA	Personal Digital Assistant

PEST	Political, Economical, Social and Technological
PVR	Personal Video Recorder
QoE	Quality of Experience
QoS	Quality of Service
RFC	Request For Comments
RTT	Round-Trip delay Time
SaaS	Software as a Service
SAIL	Scalable and Adaptive Internet Solutions
SAN	Storage Area Network
SEAP	Service Enabled Application Platform
SLA	Service Level Agreement
SONET	Synchronous Optical Networking
SWOT	Strength, Weaknesses, Opportunities, Threats
URI	Uniform Resource Identifier
VM	Virtual Machine
VoIP	Voice over IP
VoD	Video on Demand
VPN	Virtual Private Network
WMN	Wireless Mesh Network
WP	Work Package

14 References

[1] Labovitz C., Iekel-Johnson S., McPherson D., Oberheide J., Jahanian, F., 2010. Internet Inter-Domain Traffic. In: *Proceedings of Sigcomm'10, pp. 75-86*. New Delhi, India.

[2] T. Levä, "Scenario Analysis on Future Internet", MSc Thesis, Helsinki University of Technology, 2009.

[3] Labovitz C., Iekel-Johnson S., McPherson D., Oberheide J., Jahanian, F., Karir, M., 2009. ATLAS Internet Observatory, 2009 Annual Report.Presentation at NANOG 47, available at http://www.nanog.org/meetings/nanog47/presentations/Monday/Labovitz_ObserveReport_N47_Mon.pdf, accessed December 12, 2010.

[4] Anderson, C., & Wolff, J. (2010). The Web Is Dead. Long Live the Internet, *Wired Magazine*, available at http://www.wired.com/magazine/2010/08/ff_webrip/all/1, accessed December 12, 2010.

[5] Visual Networking Index (VNI): Hyperconnectivity and the Approaching Zettabyte Era. Available online: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.pdf (2.11.2010)

[6] Clark, D., Faratin, P., Gilmore, P., Bauer, S., Berger A., and Lehr W., 2008. The Growing Complexity of Internet Interconnection. *Communications & Strategies*, No. 72,

4th quarter 2008, pp. 51-71.

[7] Levä T., Kilkki K., Hämmäinen H. (2009). Scenario Analysis on Future Internet. Proc. of the First International Conference on Evolving Internet (INTERNET 2009), 23-29 August, 2009, Cannes/La Bocca, France

[8] J. Zittrain, The Future of the Internet - And How to Stop It. New Haven & London: Yale University Press, 2008.

[9] T. Wu, Network Neutrality FAQ, web page. Available at http://timwu.org/network_neutrality.html, accessed 27 March 2009

[10] Tee R., Annabelle Gawer (2009), Industry architecture as a determinant of successful platform strategies: a case study of the i-mode mobile Internet service, European Management Review, vol. 6, pp. 217–232. doi:10.1057/emr.2009.22

[11] Jacobides M., Knudsen T., Augier M. (2006), Benefiting from innovation: Value creation, value appropriation and the role of industry architectures, Research Policy, vol. 35, pp. 1200–1221. doi:10.1016/j.respol.2006.09.005

[12] Casey T., Smura T., Sorri A., (2010), Value Network Configurations in wireless local area access, 9th Conference on Telecommunications Internet and Media Techno Economics (CTTE) pp.1-9, 7-9 June 2010, Ghent, Belgium. doi: 10.1109/CTTE.2010.5557719

[13] Cisco (2010a), Visual Networking Index (VNI): Forecast and Methodology, 2009–2014. Available online: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf (2.11.2010)

[14] Zhang N. (2010), Internet Content Delivery as a Two-Sided Market. Master's Thesis, Department of Communications and Networking, Aalto University, 2010.

[15] Internet World Stats (2010), World Internet usage and population statistics. Available online: http://www.internetworldstats.com/stats.htm (19.1.2011)

[16] Norton B. (2010), Video Internet: The Next Wave of Massive Disruption to the U.S. Peering Ecosystem (v1.7). Available online: http://drpeering.net/white-papers/Video-Internet-The-Next-Wave-Of-Massive-Disruption-To-The-U.S.-Peering-Ecosystem.html (2.11.2010)

[17] N4C Summer 2010 DTN Trial (Informal Report). Available online: http://down.dsg.cs.tcd.ie/s10inf/

[18] Ismail Ari, Bo Hong, Ethan L. Miller, Scott A. Brandt, Darrell D.E. Long (2003) *Managing Flash Crowds on the Internet*. Storage System Research Center, University of California, Santa Cruz

[19] I. Ari, A. Amer, R. Gramacy, E. L. Miller, S. A. Brandt, and D. D. E. Long. *ACME: adaptive caching using multiple experts*. In Distributed Data and Structures, volume 4. Carleton Scientific, 2002.

[20] *The Evolution of Cloud Computing Markets*, Stefan Ried, Holger Kisker, Pascal Matzke, Forrester Research, July 2010.

[21] Cloud-Computing Service Trends: Business Value Opportunities and Management Challenges, Part 1, Gartner Research, February 2010.

[22] Cloud-Computing Service Trends: Business Value Opportunities and Management Challenges, Part 2, Gartner Research, February 2010.

[23] *IDC Portugal Book 139*, Abril 2010.

[24] *IBM Tech Trends Survey*, Michael O'Connell, IBM, October 2010.

[25] *Cloud Heavyweights Shape the IaaS Landscape*, The 451 Group, 2010.

[26] *Worldwide Enterprise Service Cloud Computing 2010-2014* Forecast, K. Broderick, M. Bailey, M. Eastwood, IDC, May, 2010.

[27] *Envisioning the Cloud: The Next Computing Paradigm*, J. Rayport, A. Heyward, Marketspace, 2009.

[28] *Worldwide Cloud Services Market to Surpass $68 Billion in 2010*, Gartner, 2010. Available at http://www.gartner.com/it/page.jsp?id=1389313

[29] *Application Infrastructure Technologies Need Seven Years to Mature*, Gartner, 2010. Available at http://www.gartner.com/it/page.jsp?id=871113

[30] ReadWriteWeb, blog post at http://www.readwriteweb.com/cloud/2010/07/forget-arguments-about-public.php

[31] Krishnan Subramanian, blog post at http://www.cloudave.com/545/enterprises-sneaking-into-public-clouds/

[32] Minesota Internet Traffic Studies at http://www.dtc.umn.edu/mints/

[33] *Approaching the Zettabyte Era*, Cisco, 2008 available at http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481374_ns827_Networking_Solutions_White_Paper.html

[34] *Streaming media revenues set to surge*, eMarket, 2008 available at http://www.emarketer.com/Article.aspx?R=1006117

[35] *The Digital Entertainment Revolution*, Gerry Kaufhold, CapGemini In-Stat, February 2010, IN1004828WHT

[36] *US Console Games Forecast, 2007 to 2012*, Jupiter Research, Forrester Research, 2008 vol. 1.

[37] Worldwide Connected Console 2009-2013 Forecast and Analysis: Channelling the Internet onto the Family Room TV, Lewis Ward, IDC Market Analysis, 2009.

[38] *Industry Outlook 2011 9th edition*, Informa Telecoms & media, 2010