Fernando Boavida, Eduardo Cerqueira
Universidade de Coimbra
Departamento de Engenharia Informática

Robert Chodorek, Michał Grega, Mikołaj Leszczuk, Zdzisław Papir, Piotr Romaniak
Department of Telecommunications
AGH University of Science and Technology

Carmen Guerrero
Universidad Carlos III de Madrid

Benchmarking the Quality of Experience of Video Streaming and Multimedia Search Services: the CONTENT Network of Excellence

Ocena postrzeganej jakości usług strumieniowania wideo oraz wyszukiwania plików multimedialnych: Sieć Doskonałości CONTENT

The paper presents selected issues on benchmarking the Quality of Experience (QoE) for video streaming and multimedia P2P search services being investigated in the VIFP NoE CONTENT *Content Networks and Services for Home Users* under participation the Department of Telecommunications (AGH University of Science and Technology Krakow). The first section of the paper presents the general approach used in the project including the state-of-art of Quality of Experience. The QoE of video streaming services is analysed in the third section while the QoE of P2P search services is described in the fourth section.

Artykuł przedstawia wybrane problemy oceny postrzeganej jakości usług (QoE) strumieniowania wideo oraz wyszukiwania plików multimedialnych w sieciach P2P badane przez Sieć Doskonałości VIPR *Content Networks and Services for Home Users* z udziałem Katedry Telekomunikacji AGH. Pierwsza część artykułu przedstawia metodykę zastosowaną w badaniach z uwzględnieniem stanu wiedzy w zakresie QoE. Trzecia część artykułu dotyczy oceny jakości usługi strumieniowania, a w części czwartej zostały omówione zagadnienia zaawansowanego wyszukiwania plików multimedialnych w sieciach P2P.

1. CONTENT NoE Overview

Due to the technological developments, production of multimedia content is no longer restricted to first tier producers. Furthermore, many citizens live and work along with the newly established “always connected” paradigm. They want to use their connection for sharing content they produced. It is the goal of the CONTENT Network of Excellence [11], to enable end-user communities to efficiently share, distribute, manage, and use audio visual content via these networks. At the network level CONTENT addresses issues in the delivery path and develop on top Peer-to-Peer (P2P) based overlay solutions for content services. While these concerns are reflected in a three layer architecture comprising community networks, overlay networks, and content service networks, we believe that it is of importance to integrate those different concerns and consider at the same time cross cutting issues, like monitoring, adaptations, and routing.

1.1. Motivation

The variety of audio-visual (AV) capable devices is increasing, ranging from High-Definition TV sets and PCs with large high resolution monitors to PDAs and mobile phones. Most of these devices can be connected to some kind of communication network, like cable, GSM, Wi-Fi, Bluetooth, or power line transmission. Consequently, in the near future end-users will demand services requiring access and handling of AV content seamlessly by all of their digital devices in their normal environment, i.e., home, neighbourhood, and work place. Future AV content networks for residential end-users face several new challenges due to the following developments: increasing number of users, dynamics of user demand, the vanishing differences between live and stored content distribution, and the move from a single-source provider with static content to many providers with dynamic content. However, this development does not only represent a challenge, but also opens up new opportunities. Traditional research in AV content distribution, such as Video-on-Demand, has focused on the large scale distribution of AV content from the so-called first tier content providers, e.g. Hollywood movie producers. However, despite the large research effort and substantial results, its impact is rather low as evidenced by the lack of commercially successful video-on-demand services on the market today. The two main reasons for this are non-technical. First, efficient traditional distribution infrastructures with low costs are available, like video rental stores and TV broadcasting. Second, unsolved legal issues, like ownership, digital rights management, as well as unclear business models, have discouraged the big players on this market, such as the first tier providers, from sufficiently supporting video-on-demand services.

However, nowadays digital AV content can be produced at relative low cost and can be also distributed at low cost via the Internet. In the future, more content providers are expected to step in, such as regional TV broadcasters and newspaper editors, schools and local organizations, groups of people and even individuals. Thus, end-users may eventually become both providers and consumers, and require new services to support all tasks that are related to production, provision, distribution, sharing, adaptation, personalization, searching, and consumption of AV content. This is a large opportunity for innovative services related to content networks that address the needs of all end-users that handle AV from second, third and fourth tier providers. The chances in this area to have a break-through with innovative services are higher than for large scale video-on-demand service, because the legal and financial issues are in most cases simpler or non-existent. Moreover, there are no well established low cost distribution infrastructures that can compete with the prospects of IP based communication networks. The heterogeneity of the available network technologies introduces many new technical challenges. In order to address and leverage the opportunities of new innovative services, it is necessary to address these challenges. End-users may eventually become both providers and consumers, and require new services to support all tasks that are related to production, provision, distribution, sharing, adaptation, personalization, searching, and consumption of AV content. However, the development of new services alone would not be sufficient, because the delivery path for content and also the access to content services has changed fundamentally. At the networking level, community networks are expected to play a central role in the immediate future. In this context we understand community networks as the sum of all networks that interconnect devices in the homes and the homes in a neighbourhood, like Bluetooth and Wi-Fi, and their combination into multiple-hop networks and mesh networks. Furthermore, overlay networks comprise more and more end-users or peers as overlay nodes that provide certain resources and services: see for example the overwhelming success of P2P based file sharing networks.

1.2. The CONTENT Approach

CONTENT addresses concurrently the new research challenges for AV networks and services for the end-user at home within the following three system planes, which directly map to three technical activities (TAs):

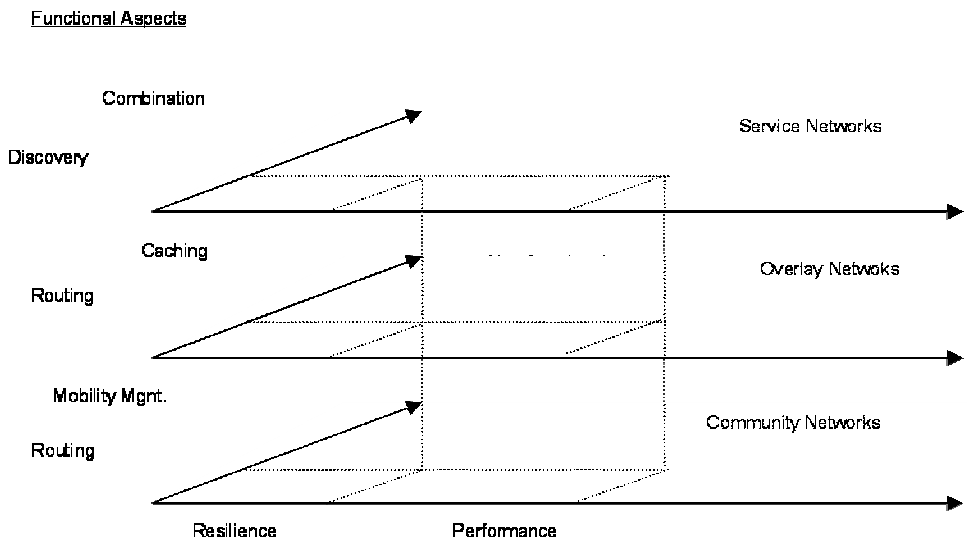

Fig. 1. Relationships between CONTENT Planes

TA1 Community networks: Community Networks (sometimes also called civic networks, Free-Nets, community computing-centres, or public access networks), form a networking infrastructure confined to a relatively small geographical area and serving locally based (community) or visiting users; it consists of four main components: the users, the community-network software, the computer hardware and the delivery channels through which users can access the available services. Current technologies adopted by community networks include Wi-Fi, UMTS, WiMAX, XDSL, etc. As social networks, the primary aim of community networks is to support the local community. Since AV content is usually distributed over such networks, several new appealing research issues come up, as for example, mobility, nomadicity, measurement and monitoring, resilience, resource assignment, user required/perceived *Quality of Service* (QoS), topological robustness, network protection, etc.

TA2 Overlay networks: Overlay networks provide an abstraction that hides the irksome details in the underlying physical networks, such as community networks, but must also be aware of the basic properties of the underlying (community) networks, to fulfil the non-functional requirements, such as resilience and performance, of the content services. Typical functional aspects of overlays are caching and request routing, and can be solved through networks of proxy caches or distributed hash tables that interconnect peers directly.

TA3 Content Service Networks: A set of innovative services for handling audio-visual content. These services should support the entire life-cycle of audio-visual content and should also be able to interoperate, such that complex services can be created by combining several simpler ones. Typical services are for instance automatic analysis and indexing services for content classification and content abstracts, watermarking services for content protections, trans-coding services for format adaptation, as well as search services to support the users to find the content of their interest.

The concept of planes and layers is used for complex systems to split up the complex system into more “manageable” parts and achieve a separation of concerns. However, the transparency that is introduced by planes must not be absolute, because there are many interdependencies between the planes and also many cross layer issues.

Fig. 1. illustrates the three planes in CONTENT and indicates the dependencies and cross layer issues in three dimensions:

1. Functional aspects within a plane: (a) community networks: the management of mobile nodes in community networks, for example, strongly influences the routing of data in community networks; (b) overlay networks: self-organizing caching and adaptive overlay networks, for example, might perform counter-productive adaptations when they are not coordinated; (c) content service networks: to support the combination of simple services into more complex ones requires descriptions of the corresponding services, which will have to be supported within the service discovery.

2. Tradeoffs between non-functional requirements: can be found in all planes, like for example, performance vs. resilience.

3. Coordination between planes: besides the well-known problem of mapping non-functional requirements from the application to content services, overlay networks, and community networks, it is necessary to: (a) avoid redundant functionality, e.g., typically monitoring is done in the community network plane and the overlay network; (b) enable overlays and content services to be context-aware, resource-aware, and location-aware; (c) coordinate adaptations that are performed independently in the different planes, like routing in the community network and routing in the overlay.

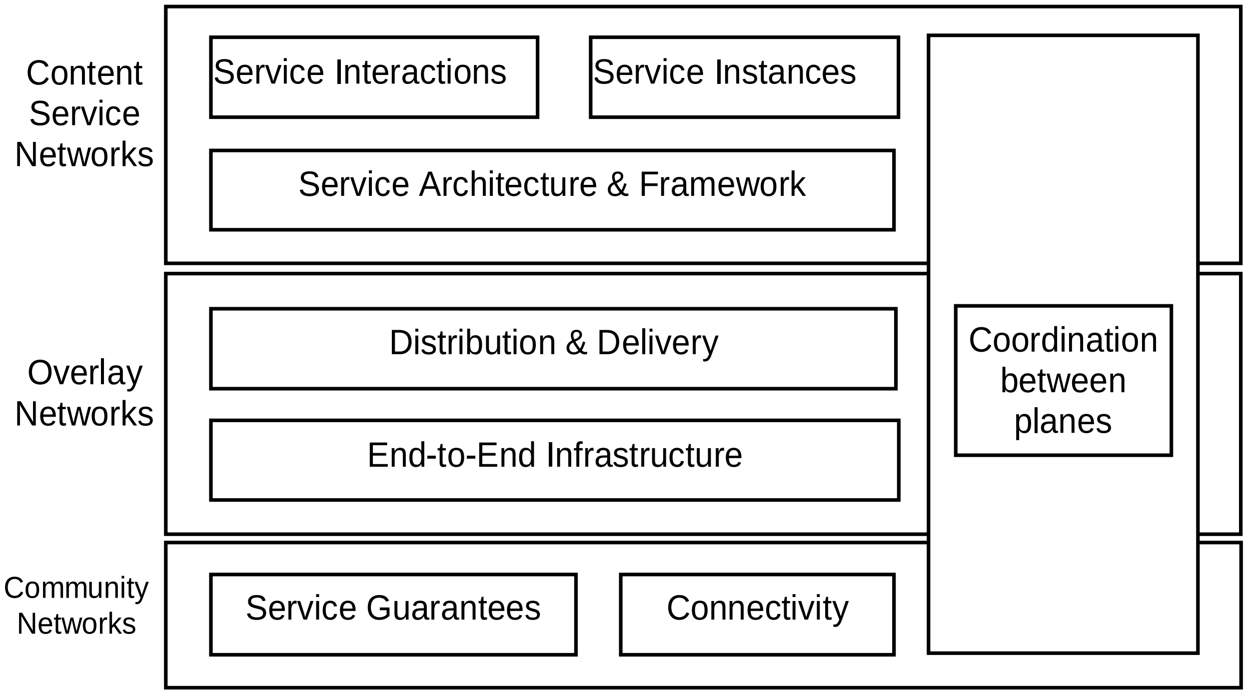

Fig. 2. CONTENT Architecture

The resulting architectural framework is illustrated in Fig. 2.

1.3. Community Networking

While the term "community network" is intuitively well understood it is worthwhile to analyze the concept of community networks. Rosson and Carrol define in [30] community networks as follows: "*A network community is a group of people whose communication and collaboration over networks strengthens and facilitates their shared identity and goals. The emergence of network communities is a striking example of what might be called grassroots technology development. (…) A community network is a special case of a network community in which a physical community coextends with the network community.*" According to this the community is not only formed by people collaborating through the network, but also by people contributing with their own resources (like in civic networks and neighbourhood networks). Community members mainly provide the access network in form of several kinds of wireless network technologies, which are connected to the Internet via one or several Internet Service Providers. Since a (substantial) part of the content delivery in community networks can be done within the physical community networks without any ISP involvement, there is no evidence that communities might be a larger threat to the Internet than classical Content Delivery Network (CDN) and P2P users, quite the contrary.

With respect to content delivery the most important insight is that the "grassroots technology development" in community networks is driven by "people", i.e., the average end-users, which might not have any particular education and skills in computer and network administration, software development etc. Thus, decentralization of content delivery must be combined with self-configuring, self-organizing, self-managing, and self-adapting solutions at all technical layers to minimize the need for human intervention.

Furthermore, Cowan et al. [12] have already in 1998 identified that content services play a central role: “In fact, communities are repositories of large amounts of heterogeneous information that need to be searched, read, explored, acted upon, updated, and that offer opportunities for collaboration and other forms of two-way communication.” In 1998, multimedia content was not central to this insight. However, we argue that the technological developments in consumer electronics and Information Communication Technologies enable the easy use of multimedia content and create by this a strong demand for various kinds of content services in community networks. Community members do not only want to consume content, but they want to share it, to search for particular content, to combine artifacts, and to edit complex multimedia objects.

Within the concept of community networking multiple networking technologies come together such as mobility with Mobile IPv4 and IPv6, multihoming, network mobility (NEMO), mobile ad-hoc networks (MANETs), wireless mesh networks (WMNs), and even wireless sensor networks (WSNs) and wireless multimedia sensor networks (WMSN). Usually, this interworking of different networking technologies is not pre-planned nor is it managed by operators. Hence, self-configuration capabilities as addressed by autonomic networks are required. In summary, community networks exploit a wide range of network technologies and techniques resulting in a challenging research environment.

As the distribution of multimedia content includes real-time delivery, QoS becomes a key aspect in community networks. QoS provision is still an open issue in wired networks, but it is even more complex in wireless environments. In this context, the evolution of the IEEE 802.11 extensions to provide QoS is crucial for the deployment of Multimedia Wireless. Also, contributions for QoS in MANETs and WMNs are of utmost importance for content delivery in community networks.

Content delivery and usage is special in the context of community networks for two major reasons: first, autonomic network and overlay solutions are needed to establish and maintain proper CDNs over physical community networks; and second, arbitrary and complex content services (e.g. content adaptation, transcoding, indexing, storage) are needed that go far beyond the simple transfer and consumption of content.

1.4. Overlay Networks

Overlay networks are virtual communications infrastructures implemented “on top of” an underlying physical network such as the Internet. Routing of packets in an overlay is usually performed at the application level. This routing function can be implemented either in the user terminals or in application-level gateways located in strategic places of the network. Two are the main reasons to build an overlay: either to force some special routing in the network (e.g. multicast routing, or QoS routing) or to organize application-level entities in order to efficiently provide some form of service to large scale communities (e.g. P2P networks for file sharing or video streaming) [5,29,41,42,67]. In this sense, building large scale overlays is an evolution of the classical client-server model.

Typical supporting services implemented by means of overlays are for instance request routing and actual content delivery. These services can either be implemented with the collaboration of end systems alone, or with support of specialized proxies. Systems based on overlay of peers are inherently distributed and, if properly designed, more robust and scalable, due to the decentralized nature and to the absence of single points of failure.

P2P networks can be classified in two groups: structured and unstructured. Unstructured P2P networks are characterized by the absence of any kind of control on the topology of the overlay.

Flooding is the predominant search technique in unstructured P2P networks. An alternative technique is random-walk. More recently, structured P2P networks have gained great interest among researchers. These networks organize their peers according to some topological criteria, usually by means of Distributed Hash Tables (DHT). Each node is responsible for a given set of keys (identifiers) and lookup of a key is achieved by routing a request through the network toward the current peer responsible for the desired key.

Considering the key building blocks of the widely deployed P2P based content delivery networks, three basic elements can be distinguished, viz. the *P2P overlay network*, a specific *content delivery strategy* and a *caching strategy*. The overlay network is responsible for connecting the participating peers, management of joining and leaving peers, and routing of queries and other messages. The content delivery strategy is responsible for delivering the required content from the source to its destination. The last strategy increases the availability of the content in the P2P system and its efficiency.

The enormous potential and advantages of decentralized infrastructures has already become apparent in the days of Napster. Since then, significant research effort has been invested in designing self-organized, scalable, robust and efficient overlay networks. However, it is crucial to note that the performance of a P2P overlay depends on various factors (e.g. application, resources of participating peers, user behaviour, etc.) that are less relevant in centralized systems. For example, a specific overlay design can perform well in the case of low churn rate whereas in the case of high churn its performance may decrease to average. Furthermore, content delivery systems pose certain requirements on overlay networks, like finding users that are sharing the demanded files, incentive mechanisms or enabling efficient inter-peer communication at low costs. Thus, there are many research initiatives to study the direct or indirect influences and dependencies between P2P overlay networks and the underlined networking strategies in a content delivery system.

Considering content delivery strategies, many aspects have to be taken into account separately alongside of interdependencies that might exist. Their influence is crucial for the overall efficiency and performance of a content delivery system. One of the most important aspects is choosing a scheduling strategy for the files to be transmitted. Download strategies as the one used by BitTorrent or network coding are proven to be very efficient for long and large scale downloading sessions [6,7]. However, with the current trend of content delivery technology, such as Podcasting, new challenges are arising. Therefore, it is necessary to investigate if the aforementioned state-of-the-art strategies are still appropriate give the requirements of emerging content sharing and delivery strategies.

Not only file sharing, but also the use of live streaming applications is growing fast in community environments. These applications and many others relying on continuous data flows, from IPTV to massive multiplayer online games, have special needs. They are delay sensitive, need group communication and QoS support. Many solutions have been proposed, but none has been adopted on a wider scale. Nowadays, protocols designed for continuous data flows do not rely exclusively on the classical client/server model, but can also organize the receivers into an overlay network, where they are supposed to collaborate with each other following the P2P paradigm.

Many recent proposals related to Live Audio/Video Streaming using P2P overlays are derived from initial work that extended application-level multicast to the end systems [68]. The first generation control-driven approach focuses on building an initial overlay corresponding to the control plane and is usually implemented as a mesh or a tree. A second overlay, usually a spanning tree, is then created and managed for the actual data transmission. Peer-cast [31] is the most famous example with a popular implementation and a large audience. A lot of work has been carried out to improve the control plane in order to cope with the high dynamics of the P2P overlay. For example, Nice is using a sophisticated clustering scheme [52]. More recent work tries to improve robustness using a hybrid tree/structure. An example for this is Bullet [13]. A new generation, data-driven approach stresses the need to cope directly with data. Peers exchange data availability and then they choose their neighbourhood according to the data they need [52]. Further, epidemic algorithms are currently being proposed in systems such as Donet [66] to improve the data delivery.

P2P Live Streaming is already reality. However, so far little has been done to demonstrate their efficiency on a very large scale. Simulation is one way to validate the feasibility of such dynamic infrastructures [58]. An alternative approach is to study proprietary applications in real testbeds, like Planet-lab [48]. The largest P2P Live Streaming deployments are related to IPTV applications and are only associated to proprietary protocols and architectures [44,48,49,61]. Thus, only their behaviour but not the protocols itself can be analysed.

The behaviour of peers in a community network plays a key role. At the one end of the scale are altruistic peers that provide resources without expecting any return. At the other end there are so called free riders who only consume but do not provide any resources, which is a rational behaviour in systems without any sharing incentives. Therefore, it has become clear that some kind of incentive scheme is necessary to achieve an optimal utilization of system resources in a system context as well as for individual peers. This is currently an active research area.

1.5. Content Service Networks

Within the CONTENT architecture, the content services network provides an abstraction of how different services related to content handling and delivery can form an infrastructure of value added services. These provide support for various tasks and processes, e.g. to offer a wider variety of formats, provide easier access or introduce interactivity. The idea is to use so called *content services* in conjunction with the underlying network infrastructure to provide a network of content services and by doing so forming a *content network*.

Apart from user services there can be also services for optimizing content delivery and monitoring the performance of the content network. One such service can be an Objective Video Assessment service located at strategic positions in the content network that monitors the quality of the delivered video and locates where problems are caused. In addition to these functional aspects of services the actual content services network architecture has to deal with specific aspects associated with service based architectures in general and issues related to service discovery and service description. In order to represent all the tasks related to a content services network within the overall architecture they have been split into three distinct areas of concern, viz. Content Service Network Architecture and Services Framework, Service Interaction, and Service Instances.

1. Content Service Network Architecture and Services: The aim of building such a services network is to integrate, in an open way, tools and mechanisms that would enable the “curation” of multimedia assets and the subsequent access to assets for the benefit of the communities of users.

In order to achieve this, a suitable model and architecture is necessary that allows to easily “plugging” such content services into the services network. Therefore a *service based architecture* is required that provides such a framework into which each of these services can be integrated. The concept of Service Oriented Architecture (SOA) has been introduced to achieve optimal support for business processes through the underlying IT architecture [50]. SOA is an architectural concept for enterprise-class, distributed, IT systems. Services are loosely coupled but independent location transparent components which together represent an application environment [38].

The architectural model is reflected in the service framework that defines the form of service interfaces and interaction which can take place between the services. The framework does not only provide support to easily deploy services within a home network infrastructure but also has to address deployment issues.

2. Service Interaction: Service interaction describes the way services within the architecture can interact. A service description is necessary to insure easy access to services by users and a simple management of them. In order to invoke appropriate services to meet specific goals, a good matching between a service request and services is essential. Therefore the request must be expressed in a sufficiently rich formalism, compatible with the description of services. A formalized knowledge description is necessary, ideally based on standards. Examples of such standards are for instance those defined by the W3C for the Semantic Web. Several formalisms have been proposed, at various expressivity levels, from simple semantic mark-up syntaxes (e.g. RDF [18]) to ontologies (e.g. OWL [32]). An OWL-based Web Service Ontology, OWL-S, has been proposed specifically for Web services, in order to describe their properties unambiguously [15]. A recent initiative defined a Semantic Web Services Framework (SWSF) [53], which includes the Semantic Web Services Language. These various formalisms differ in the richness of the description they provide and in their reasoning capabilities.

Content services share many general properties with Web Services. Therefore, it seems that Web Service discovery mechanisms could be used to discover content services. However, Web Service discovery mechanisms were initially designed for classical Internet environments where the network topology and availability of hosts is relatively static. Within Content Networks, however, more dynamic environments akin to P2P networks are encountered. Clients, services, and service registries may appear and disappear randomly on the network, making it important to ensure that information residing in service registries is up-to-date.

3. Service Instances: Services in the context of CONTENT can be for example tools that adapt content formats, automatically analysis and indexing content, create visual abstracts for easier search and navigation, but also a watermarking tool to protect content and the associated IPR is possible here. The focus at the moment is on content adaptation services, scaling and transcoding services, and video summarization and indexing services. The goal of looking at these services is to investigate how specific services can be represented and integrated into the service architecture. This is not a closed set of service instances and the goal of the content service network architecture is to support all kinds of different services.

1.6. Cross Layer Issues

It is generally accepted in the research community that layered system architectures have besides their advantages also clear disadvantages. In order to enable, for example, resource aware distributed applications, access to network layer information is necessary. Cross layered approaches are used to achieve this kind of awareness beyond layer interfaces, but they are designed for particular solutions. Thus, understanding and developing a better architectural solution than strict layering is an important research challenge in general. However, cross layer issues are especially important in the context of future content aware community networks since autonomic solutions, like self-adapting functions, need to be applied. As mentioned earlier, independent adaptation of different functions might influence each other since they share resources. For instance both might have an impact on network traffic. The first step towards addressing this challenge is to identify a set of metrics for each layer, including QoS parameters and resource consumption parameters and to model their dependencies between the layers. This first step seems trivial, but to carry it out successfully, this set of metrics and their definitions need to be accepted and used by the entire research community working in this area. Nowadays, many different and incompatible metrics and definitions are used. Modelling the dependency among parameters needs also to include the understanding of the functional behaviour of the system elements. To provide the proper tools for this challenge, the CONTENT NoE investigates the development of a generic benchmarking suite for content networks following a modular approach in which the different levels of a content network might be considered as the system under test and the other levels represent the environment and the workload.

The rest of the paper presents background information about *Quality of Experience* (QoE) as well as the two approaches towards benchmarking the above-mentioned QoE of two cases studied. The presented services are video streaming and multimedia search services.

2. Background on QoE

The efficient management and distribution of multimedia services, such as multimedia search services, video streaming, mobile IPTV and other kind of multimedia applications, over an all-IP system is a major requirement to the success of **next generation networks**. The quality level control of multimedia services aims to maximize the user's satisfaction and the usage of network resources as well as to keep and attract customers, while increasing the profits of network providers [35].

Traditional techniques that aim to maximize the quality level of multimedia services in a networking system are focused on QoS aspects. QoS-based schemes define a set of network level (and packet level) measurement and control operations to guarantee the distribution of multimedia content, in wired and wireless networks, with an acceptable quality level [40]. Existing QoS metrics, such as packet loss rate, packet delay rate and throughput, are typically used to indicate the impact on the video quality level from the network's point of view, but do not reflect the user's experience. Consequently, these QoS parameters fail in capturing subjective aspects associated with human perception.

In order overcome the limitations of current QoS-aware multimedia networking schemes regarding human perception and subjective-related aspects, QoE approaches have been introduced [4]. QoE measurement operations can be used as an indicator of how a networking environment meets the end-user needs. The QoE applicability scenarios, requirements, evaluations and assessment methodologies in multimedia systems have been investigated by several researchers and working groups, such as *International Telecommunication Union – Telecommunication Standardization Sector* (ITU-T) [21], *Video Quality Experts Group* (VQEG) [64] and *European Technical Committee for Speech, Transmission, Planning, and Quality of Service* (ETSI STQ) [16].

The results of QoE research can be used as an extension to the traditional QoS in the sense that QoE provides information regarding the delivered multimedia service from the user's point of view. Hence, QoE procedures can be explored to improve the accuracy of **QoS control plane operations** and to ensure smooth transmission of audio and video over all-IP networks [62]. The advances in QoE-aware systems will allow the deployment of new QoS/QoE-sensitive services as well as provide new **paradigms** for the creation of new protocols, routing approaches and overlay networks, such as the deployment of QoE routing schemes.

Nowadays, QoE operations are not fully implemented in end-to-end networking systems due to the high CPU and memory consumption required by current QoE schemes, as well as to the lack of accuracy of in-service quality assessment methods. Usually, only QoE out-service measurement procedures are performed to evaluate the quality level of multimedia services in the Internet.

The quality of processed multimedia services from an end-user perspective is analysed according to a set of QoE metrics and methods. The existing evaluation QoE metrics and procedures can be divided into quantitative (objective) and qualitative (subjective) ones. Metrics of the first type refer to the broadly understood and objectively measured system performance. The latter refer to the quality of the system. Subjective methods are performed to acquire information about the quality level of multimedia services based on human opinion score schemes, while objective methods are used to estimate the performance of multimedia systems by using models that approximate results of subjective quality assessment. In the other words, QoE metrics can be classified in terms of their objectiveness and, thus, range from qualitative to quantitative. In addition, objective QoE measurements can be classified based on the amount of available reference information during the multimedia service quality assessment process, namely *Full Reference* (FR), *Reduced Reference* (RR) and *No Reference* (NR).

Subjective metrics assess how audio and/or video streams are perceived by users, i.e., what is their opinion on the quality of particular audio/video sequences, as described in ITU-T recommendation BT.500 [22]. An example of the most popular qualitative (subjective) metric is called a *Mean Opinion Score* (MOS) scale, which was initially standardised by the International Telecommunication Union [24]. In this metric the quality of the system is subjectively assessed by the users in a five-grade scale, where 5 is the best quality and 1 is the worst, as presented in Tab. 1. Another example of a qualitative metric is an R-factor, which may be utilised in a way similar to MOS. R-factor is used for subjective evaluation of speech quality in the voice transmission systems [23].

Tab. 1. Mean Opinion Score

MOS	**Quality**	**Impairment**
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

The MOS values are achieved based on subjective tests and methodologies performed with a set of viewers. For instance, the *Single Stimulus Continuous Quality Evaluation* (SSCQE) tests allows viewers to dynamically rate the quality of an arbitrarily long video sequence using a slider mechanism with an associated quality scale. The drawback of subjective metrics is the fact that they are neither practical nor scalable for real-time multimedia environments.

Subjective approaches assume human experience as the only grading factor. Objective procedures are performed without human intervention and give more stable results, but do not necessarily reflect the user quality perception. For example, for benchmarking picture quality, examples of objective metrics include *Peak Signal to Noise Ratio* (PSNR), *Mean Absolute Error* (MAE), *Mean Square Error* (MSE), and *Root Mean Square Error* (RMSE) [45,59,60]. The methods for assessing the perceived video quality objectively do usually not take the *Human Visual Senses/System* (HVS) sufficiently into account. The human senses cover many errors quite effectively. Thus, objective measurements may not reflect the user perceived quality. Other methods that also consider HVS are therefore required (as for instance discussed in [8,56,71]). A detailed analysis of benchmarking picture quality can be found in Subsection 3.1.

3. Benchmarking Video Streaming Services

Previous research efforts towards the assessment of end-user perceived quality are mostly adequate for MPEG-2 videos only. The goal of this work is to provide QoE assessment as a service within the Content Services Infrastructure. Obviously, many of the QoE parameters inherit directly from QoS parameters of the underlying network. Both QoE and QoS issues, for benchmarking video streaming services, have been addressed in this section. Both unicast and multicast[1] approaches have been analysed.

3.1. Quality of Real-Time Video Streaming Experience

Perceived video quality assessment is a massive and challenging task. There are many different factors affecting perceived video quality. The examples are screen size, screen illumination, video content, application type, viewing distance, user profile, and many others. However, the one, permanently addressed quality factor, is video fidelity considering distortion level introduced by the content production phase, by codec (lossy compression) and network during the transmission (all related QoS parameters like PLR[2], delay, delay jitter or throughput).

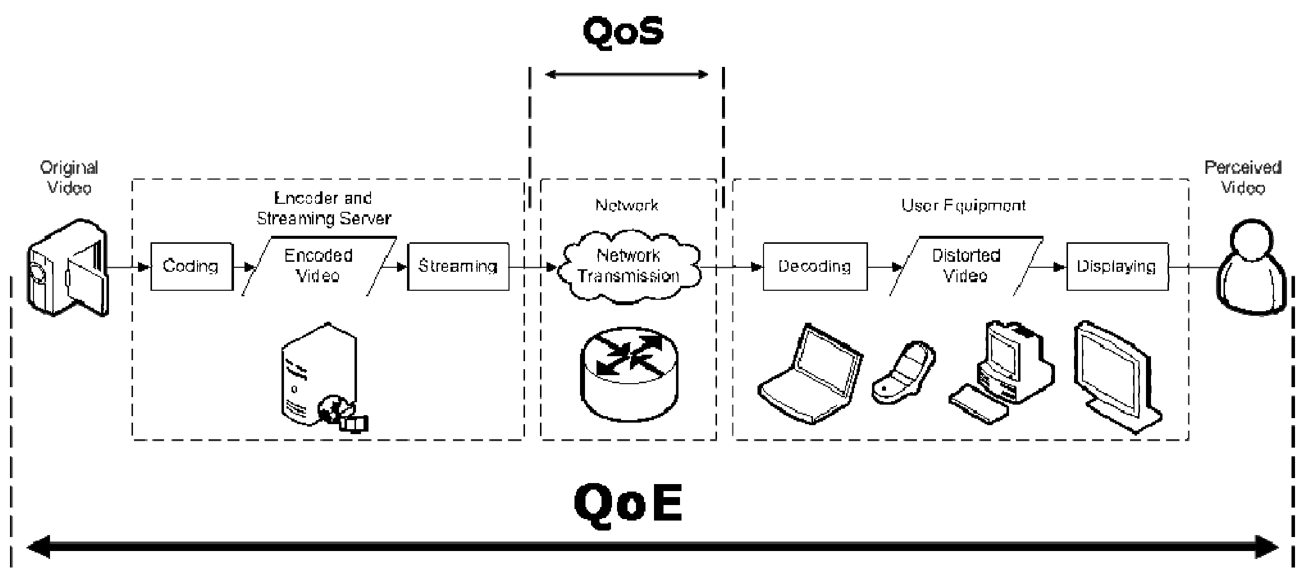

Fig. 3. Quality of Service vs. Quality of Experience

End-to-end system for video content delivery is illustrated in Fig. 3. As it was stated before, video quality degradation may have three main sources. During the production phase video content is captured by a device which performs analogue to digital domain signal conversion. In the second step video content is encoded, usually using lossy compression algorithms. Afterwards, encoded video is transmitted through a network using real time streaming protocols such as RTP/UDP. In the last step, video is decoded by the end user terminal and finally displayed. After each stage, a reduction in the quality of the original video sequence may occur. Low performance and image processing capabilities (e.g. mobile phone with digital camera) of the video capture device can

[1] Please refer to the Annex A for more information on multicast applications for real-time multimedia communication.
[2] *Packet Loss Ratio*

significantly reduce the video quality in the very first step of the delivery chain (the most common image/video artifacts are noise and poor spatial and temporal resolution). Lossy compression will result in spatial artifacts (e.g. blockiness or blur), while transmission will cause both temporal and spatial artefact (e.g. unnatural motion or parts of/whole frame skipped).

In the traditional approach, quality of video services is measured from the network perspective on the packets level, using simple QoS parameters like PLR or delay. However, as it was proved recently [14,69], the same level of artifacts or QoS parameters can have completely different impact on the visual quality. It implies a need of a more comprehensive approach towards perceived quality assessment that would reflect the end user's preferences and aim for the overall experience assessment. In order to account for this problem, the concept of QoE of video services has been introduced to address the issue concerning the assessment of how well a video service meets the customers' expectations. The relation between QoS parameters and QoE concept is presented in Fig. 3.

Nowadays, video quality assessment in terms of user satisfaction level (QoE) is a topic of high interest for telecommunication services providers and researchers, being under rapid development. Miscellaneous video quality metrics were developed over recent years, the most appealing works are presented in [33,46,51,54,57,70]. As the quality assessment becomes more and more standardized, a need of quality assurance and optimization is emerging. Providers are looking forward solutions capable of permanent video quality monitoring, degradation prevention and quality optimization at the same time.

The remainder of this section is organized as follows: Sub-subsection 3.1.1 presents objective video quality assessment metrics. Sub-subsection 3.1.2 introduces QoE measurement approaches based on a reference-based classification. Sub-subsection 3.1.3 describes several video quality evaluation tools used to acquire information about the quality level of video services.

3.1.1. Objective Real-Time Video Streaming Metrics

There are several objective methods to measure the quality level and detect impairments (blocking, blurring and colour errors) of multimedia services. Several objective QoE metrics have been developed to estimate/predict the quality level of multimedia services according to the user's perception. Among them, the PSNR is a traditional objective metric used to measure, in decibels, the video quality level based on original and processed video sequences. Typical values for the PSNR in lossy videos are between 30 dB and 50 dB, where higher is better. The PSNR of a video is defined through the MSE metric. Considering the luminance (Y) of the processed and original frames and assuming frames with $M \times N$ pixels, the MSE is obtained using the Eq. 1.

$$MSE = \frac{1}{MxN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left\| Ys(i,j) - Yd(i,j) \right\|^2 \tag{1}$$

In Eq. 1, while $Y_S(i,j)$ designates the pixel in the position (i,j) of the original frame, the $Y_D(i,j)$ represents the pixel located in the position (i,j) of the processed frame. Based on the MSE definition and on 8 bits/sample, the PSNR, in a logarithmic scale, is achieved using the Eq. 2.

$$PSNR = 20 \log_{10} \left(\frac{255}{\sqrt{\frac{1}{MxN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left\| Ys(i,j) - Yd(i,j) \right\|^2}} \right) \tag{2}$$

The MSE and PSNR metrics only provide an indication of the difference between the received frame and a reference signal, and do not consider any other important aspects which can strongly influence the video quality level, such as HVS characteristics (a detailed analysis of HVS can be found in [72]). The PSNR can also be used to map MOS values as described in Tab. 2.

Tab. 2. PSNR to MOS conversion

PSNR (db)	MOS
> 37	5 (Excellent)
31 – 37	4 (Good)
25 – 31	3 (Fair)
20 – 25	2 (Poor)
< 20	1 (Bad)

The *Structural Similarity Index* (SSIM) metric improves the traditional PSNR and MSE, which is inconsistent with HVS characteristics, such as human eye perception [69]. The SSIM metric is based on frame-to-frame measuring of three components (luminance similarity, contrast similarity and structural similarity) and combining them into a single value, called index. The SSIM index is a decimal value between 0 and 1, where 0 means zero correlation with the original image, and 1 means the exact same image.

The *Video Quality Metric* (VQM) method defines a set of computational models that also have been shown to be superior to traditional PSNR and MSE metrics [65]. The VQM method takes as input the original video and the processed video and verifies the multimedia quality level based on human eye perception and subjectivity aspects, including blurring, global noise, block distortion and colour distortion. The VQM evaluation results vary from 0 to 5 values, where 0 is the best possible score.

The *Moving Picture Quality Metric* (MPQM) evaluates the video quality using HVS modelling characteristics [9]. The input to the MPQM metric is an original video sequence and a distorted version of it. The distortion is first computed as the difference between the original and the distorted sequences. The original and the error sequences are then decomposed into perceptual channels segmented using uniform areas, textures and contours classification. After that, HVS-based contrast sensitivity and masking parameters are considered for each perceptual channel in detection threshold calculation. Finally, data from channels are gathered to yield a single figure and to account for higher levels of perception, which is called pooling. Due to the MPQM's purely frequency-domain implementation of the spatio-temporal filtering process, this complex metric requires huge memory consumption. The final quality measure can be expressed either using a Masked PSNR (MPSNR) equation or can be mapped to MOS scale as detailed in [9].

The *Perceptual Evaluation of Video Quality* (PEVQ) provides MOS values of the video quality degradation as a consequence of end-to-end communication [46,47]. The PEVQ approach is based on the combination of spatial and temporal artefacts measurement with human visual system behaviour. PEVQ provides MOS scores of the video quality, from 1 (bad) to 5 (excellent). In addition, PEVQ also provides information about the perceptual level of distortion in luminance, chrominance and temporal aspects of the evaluated video.

The previous QoE methods are based on a set of user/service information about the original and processed video. In order to reduce the system complexity and the amount of available reference information, a packet-based method, called *Media Delivery Index* (MDI), was proposed in IETF RFC 4445 [28]. The MDI metric is not the most accurate video quality level method and does not provide a good characterization of QoE, but can provide an indication of the video quality in a cost effective manner. The MDI scheme provides an indication of traffic jitter, a measure of deviation from nominal flow rates and a data loss at-a-glance measure for a particular multimedia service. According to MDI values, the overall video quality level through an end-to-end communication path can be estimated.

As presented in this subsection, several objective multimedia quality assessment methods have been proposed. A comparison of the different schemes regarding performance, accuracy, feasibility, scalability and flexibility is very difficult and is still a challenging research topic. A usual manner to aggregate and classify objective video quality methods is based on their dependence on the amount of available reference information during the video quality assessment process. This will be presented in Sub-subsection 3.1.2.

3.1.2. Metrics by Reference-Based Classification

Three different approaches are used to classify video quality assessment methods, based on reference-related video procedures, namely FR, RR and NR.

The FR approach assumes unlimited access to the original multimedia sequence. This approach uses the video reference to predict the quality level (degradation) of the processed video, by comparing the difference of every pixel in each image of the distorted video with its corresponding pixel in the original video. As consequence, it provides, in general, superior quality assessment performance. The FR method is difficult to implement in real-time networking systems (QoE-aware equipment/monitoring agent) because it always requires the original sequence during the evaluation process (common for offline experiments). Examples of metrics based on an FR approach are PSNR, SSIM and MPQM.

For in-service video quality measurements, RR and NR approaches are generally more suitable. The RR approach differs from the FR approach only selected multimedia parameters (or characteristics) are required during quality evaluation process, such as motion information. The set of reference parameters can be transmitted piggy-backed with the multimedia flow or by using a secondary channel. The objective of RR is to be as accurate as the full reference model, although using less network and processing resources. An example of an RR scheme is *Video Quality Model* (VQM), developed by the *National Telecommunications and Information Administrative* (NTIA) and reported in [19].

The NR approach tries to assess the quality of a distorted multimedia service without any reference to the original content. This approach is usually used when the coding method is known. NR-based metrics can be used in in-service network monitoring/diagnostic operations, when the original multimedia sequence is not available. The drawbacks of NR metric are the following: (*i*) low correlation with MOS; (*ii*) high CPU and memory consumption; (*iii*) time limitation. An example of NR schemes is the V-Factor model [63] that outputs MOS.

3.1.3. Video Quality Evaluation Tools

This section briefly identifies some multimedia quality evaluation tools used to acquire information about the quality level of multimedia services.

EvalVid is a framework and tool-set for evaluation of the quality of video transmitted over a real or simulated networking environment [17]. EvalVid can be used to measure QoS-related parameters, such as packet loss rate and packet delay rate, as well as QoE-related parameters, such as PSNR. Regarding video formats, currently EvalVid supports H.264, MPEG-4 and H.263 formats.

The MSU Video Quality Measurement Tool (MSU VQMT) is an application for video quality measurements [43]. This application allows users to create objective comparison of video CODECs and performs filter video analysis. The MSU VQMT supports several video formats (e.g., AVI, YUV, MP4 and MPEG-4) and QoE metrics (e.g., PSNR, VQM, SSIM and MSE).

VQLab is a fast, reliable and cost-effective tool for assessing the quality of processed video [17]. VQLAB supports a wide range of video formats, such as AVI, MPEG-2, MPEG-4, H.264 and YUV. In addition, VQLab uses the PSNR, SSIM and *Czenakowski Distance* (CZD) QoE metrics in order to evaluate the video quality level.

The *Video Quality Measurement* (VQM) PC tool compares the video sequence that has been processed by the video system under test with the original video sequence [17,19]. As the result, video quality assessment is reported on a default scale of 0 to 1, where zero means that no impairment is visible and 1 means that the multimedia video has reached the maximum impairment level. The VQM tool supports AVI, MPEG, WAV and other video formats.

3.2. Quality of Real-Time Multicast Multimedia Services

Generally speaking, QoS is an ability of a service to fulfil user's requirements. Because QoS (and quality at all) is a non-measurable quantity, it must be described using a set of measurable parameters. In the case of real-time multicast multimedia services, typical quality parameters are: delay, delay fluctuations (usually understood as a delay variation), and available throughput. An auxiliary quality parameter is error rate, which should be relatively small.

3.2.1. QoS of Teleconferencing Services

As was mentioned above, teleconferencing services transmit audio stream (audio-conference) or audio with associated video (videoconferences). During a teleconference, other data objects (as pictures and text) also can be sent. Audio and video streaming is performed by each end-system participate in teleconference. Multimedia content is send to the address of multicast group, allocated for this teleconference. Teleconference participants, which are interested in reception of given multimedia content, must belong to proper multicast group. During a teleconference, each participant receives audio signals from other participants and video signal from only one, active participant. Choice of received signals is possible by source filtering, which is carried out using IGMPv3 signalling protocol (IPv4) or MLDv2 signalling protocol (IPv6).

Analyzing teleconference services, special attention should be paying to transmission delay. Teleconferences are interactive services, where take place on line discussion between participants, and even small delay can cause, that the service becomes uncomfortable to operate, and sometimes lose raison d'être. Therefore, delay of teleconferencing services shouldn't exceed threshold value of delay defined for professional telephony (i.e. 150 to 200 ms).

Requirement of small delay affects the next quality parameter – delay fluctuation. Delay fluctuations are very uncomfortable for multimedia service users and, typically, they are eliminated using buffering in receiving systems. However, during a teleconference, buffering must be strictly limited, because of the imperative of small transmission delay. Thus, on account of small buffer sizes, a network should introduce small delay fluctuations.

Throughput required for voice transmission is relatively small (a few kb/s). Throughput of video encoded for teleconferencing purposes is also relatively small (of the order of 64 kb/s), because of low resolution of picture.

Teleconferencing systems, like other real-time multimedia systems, are error-tolerant, although small error rates are recommended.

3.2.2. QoS of Internet radio

Internet radio broadcasts audio signals (typically: voice and music) to many recipients, via the public Internet. In the contrast to teleconferences, there is only one sender of multimedia content – radio station. Radio broadcast is no interactive in nature, and potential interactions with radio listener are performed on the basis on opinion polls (often with the use of simple voting applications) or text feedbacks (often with the use of simplified discussion forums or electronic mail).

Due to assumed lack of interactivity of radio broadcast, small transmission delays are not as important, as in the case of teleconferences, although small delay fluctuations still are required. Compensation of delay fluctuations with the use of relatively large buffers is possible. Large throughputs are not required, because audio signals has relatively small target bit rate (often less than 128 kb/s). Small error rates are recommended.

3.2.3. QoS of Internet television

Internet television broadcasts television signals (video and associated audio) to many recipients, via the public Internet. Like in the case of Internet radio, there is only one sender of multimedia content – TV station. Television broadcast is no interactive in nature, and potential interactions with viewer are performed in the same way, as in the case of Internet radio.

Transmission delays don’t play important role – delays of few seconds are accepted by a TV viewer. Delay fluctuations which should be relatively small, usually are compensated using large buffers. Television signals are characterized by relatively high target bit rates, so assurance of proper throughput becomes a key issue. Small error rates are recommended.

3.3. Influence of Network Technology on QoS Parameters

There are several network technologies, dedicated for the three main network types: local, metropolitan, and wide area networks. Local Area Networks (LANs) are intended to connect large number of end-systems, located at small area. They are characterized by relatively large throughput (e.g. 1 Gb/s in the Ethernet, up to 54 Mb/s in 802.11g), able to convey television streams. Transmission delays in LANs are small (of the order of 10 µs), because of small distances between end-systems and small buffering in active elements of network infrastructure. In loaded networks with CSMA medium access, delay fluctuations can be large. Thus, in 802.11 networks, usage of point coordination function (PCF) is recommended for multimedia transmission instead of distributed coordination function (DCF). Error rates in wired LANs are small. In wireless 802.11 LANs relatively large error rates in radio channels are corrected on the level of a network card.

Metropolitan Area Networks (MANs) are intended to connect LANs located at metropolitan area. The best known technologies of MANs are ATM, 10-Gigabit Ethernet (10GbE), and 802.16 (WiMAX). Due to larger propagation delays (of the order of 500 µs) and larger delays introduced by buffers, MANs are characterized by medium delays (larger, than delays in LANs). Throughputs available for end-user are high. In cable networks error rates are small; in wireless can be subject of fluctuation. Networks based on the ATM technology, both wired and wireless, allows an end-user for reservation of network resources according to a required QoS policy.

Wide Area Networks (WANs) are intended to connect LANs or MANs located at a wide area (e.g. country or region). The best known technologies of WANs are SDH and DWDM. Transmission delays in WANs are large (of the order of tens or hundreds milliseconds), delay fluctuations are very large. Available throughput also is a subject of large fluctuations (in congested networks, intentions throughput can collapse to zero). Error rates can be very large (during congestions, error rates can reach several percents), sometimes overstepping the boundaries of acceptability. As a result, realization of professional real-time multimedia services in WANs is very difficult.

4. Benchmarking Multimedia Search Services

The P2P overlays are gaining on popularity as a mean of access to large amounts of multimedia data. A P2P system is a self-organising system consisting of end-systems (called "peers") that form an overlay network. Peers offer and consume services and resources and have significant autonomy. Services are exchanged between any participating peers. Such networks are gaining more and more popularity and attention both from users and researchers. This growing interest can be explained, on one hand, by the numerous P2P based applications, ranging from simple files sharing to more sophisticated services such as Voice over IP (VoIP). On the other hand – P2P networking is a challenging topic for researchers due to its distributed architecture, the need of cooperation of the peers and the lack of the central authority (in some of the network architectures).

A study performed in 2004 by the CacheLogic company gave a conclusion that *"Traffic analysis conducts as a part of an European Tier 1 Service Provider field trail has shown, that P2P traffic volumes are at least double that of http during the peak evening periods and as much as tenfold at other times"* [2]. The same study shows, that in 2006 P2P services were responsible for 70% of the global traffic.

Multimedia is now rapidly moving into the every-day life. Users are not only able to access media via radio, television and the Internet but now also create their own media content. Digital cameras, video recorders, media enabled mobile phones and the wide spread availability of content creation and editing software which were previously available only to professionals, all contribute to this data volume growth. The traditional division of users has to be extended by a new kind – the "prosumers" who are, at the same time the "producers" and "consumers" of the multimedia.

The search mechanisms are also evolving. It has been observed, that the traditional, text based, search methods are failing to deliver satisfactory results in case of multimedia. The most common approach is to perform a textual search based on the media file name. This is extended, in the case of more advanced systems by textual search based on user-provided tags or the context, in which the media file is published (e.g. text surrounding media in case of media search in the WebPages).

One of the difficulties in designing good mechanisms for distributed autonomous systems is a lack of a unified process for evaluating the efficiency of mechanisms, both in the research community and in the industry. As content distribution systems have widely recognised standard overlays to compare with, in search there is no such base.

Another problem affecting the state of the art media search solutions is the distinction between the relevancy of the answer to the query and the relation of such relevancy to the users' satisfaction with the performance of the given service. This issue is well known to anyone, who tried to use a multimedia search service such as Google Image Search[3]. The answers to a given query are often relevant, but not satisfactory.

The solution to those problems is creation of a search benchmarking system for the distributed media delivery system, which would cover all critical aspects, including the user perceived quality. Such solution is now under research performed by a community of researchers within the CONTENT project. The research group consists of senior researchers and PhD students from Germany, Poland, Portugal, Spain and the Netherlands. The proposed benchmarking mechanisms allows, on one hand, to evaluate the performance of the search algorithms, and, on the other hand to make research and tuning process easier.

A benchmark can be defined as *"a standardized problem or test that serves as a basis for evaluation or comparison (as of computer system performance)"* (according to the Merriam-Webster English dictionary). The goal of benchmarking is to assess the quality of the benchmarked system and to allow comparison to other, similar systems. Researchers, especially in the computer science, are used to perform benchmarking according to either official or unofficial standards (such as e.g. [36]).

The benchmarking metrics can be divided generally into performance metrics and cost metrics. The first of the performance metrics taken into consideration in the presented framework is the Search Accuracy defined as the ability of the system to find the desired results upon the query. The second performance attribute is the Search Time. According to works describing the search benchmarking frameworks for database-based systems *"speed is not of central concern"* [10]; this is due to the high performance and locality of database systems. However, distributed P2P systems are characterised by a considerable and varying delay in communications. Therefore, the search time is also a measured factor in the developed benchmark set. The Resource Consumption is the cost metric that is also taken into consideration in the presented work, in the terms of the bandwidth consumption.

4.1. State of the Art

In order to benchmark the accuracy of the search system, it is necessary to have a ground truth, which may be defined as a full knowledge of all data stored in the system. This ground truth serves as a reference level for benchmarking accuracy. In the case of benchmarking multimedia, a ground truth is usually a collection of manually annotated media files, used to make sure that the annotations are accurate. In the case of search for images, there are several requirements for the reference collection.

There are numerous quantitative metrics for the assessment of search accuracy in the retrieval process. An overview is given in [26]. The most commonly used are *Precision* and *Recall*. Precision is the number of detections (defined as the number of relevant items detected) divided by the total number of the returned items. Recall is defined as the number of detections divided by the total number of the relevant items in the system.

The methodology for the quantitative measurement of the search accuracy gives the numeric values which describe several aspects of the system. In order to draw a conclusion about the accuracy of the system, evaluation methods need to be defined. The recommended evaluation method is the *Retrieval Effectiveness* (the comparison of precision versus recall). The best

[3] images.google.com

evaluation method will be the one that reveals the weaknesses of the benchmarked system. An overview of the recommended evaluation methods is presented in [26].

The existing visual information retrieval systems focus mainly on preparing the annotated media for benchmarking. Examples of such activities are the TRECVID system provided by the US National Institute of Standards [3] and TC-12 benchmark provided by the International Association for Pattern Recognition [37]. These benchmarking systems focus also on the media stored locally, whereas the presented benchmarking system focuses on a distributed storage.

4.2. Design of the Framework

The main purpose of a benchmarking system is to present values that allow comparing two or more similar systems. A benchmarking framework which is built with well-defined requirements allows it to be widely used and, finally, accepted as a standard. The following requirements should be fulfilled [10]:

- It should be general enough to allow measurements and comparison of different search systems. In the case of search in P2P environments, it is possible that new overlays will emerge and our benchmarking framework should be applicable also in that case.
- The benchmarking framework should be parametric [10]. It should allow for changing the parameters of the environment.
- A standard query set should be defined to allow the comparison of different benchmarked systems. According to [10] a benchmarking system should consist of approx. 20 benchmarking queries. The answer for such queries should be defined and contain more than 15 but less than 50 hits.

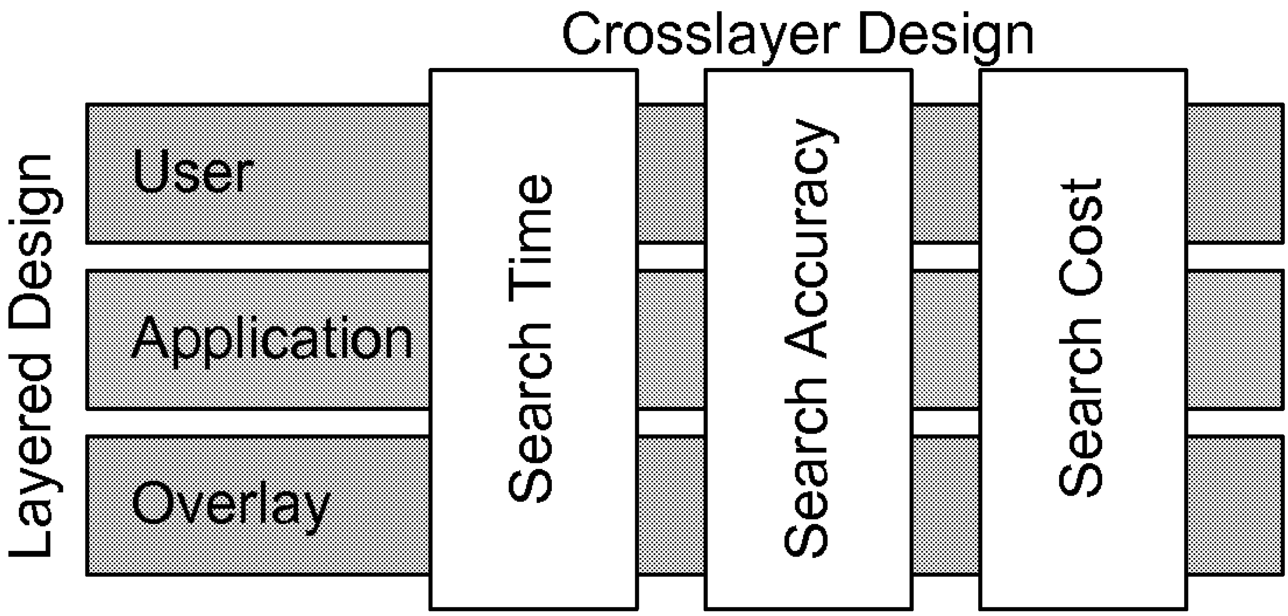

Fig. 4. The design of the search benchmarking framework

Horizontal and vertical are the two dimensions in which the benchmarking framework was designed. The layered design – the horizontal one – describes the three layers of the Benchmarking Framework and the layer dependencies and relationships. The cross layer design concept – the vertical one – describes the measurement methodology of the performance parameters, which are the search accuracy, search time and the cost of search. The overall design of the framework is depicted in Fig. 4.

Tab. 3. Classification of the benchmarking metrics

Layer	Search Accuracy Performance Metric	Search Time Performance Metric	Cost Metric
User	R-factor, MOS	R-factor, MOS	
Application	Precision, Recall	Query Preparation Time, Similarity Computation Time	Resource Consumption
Overlay	Peer Query Rate, Peer Hit Rate	Query Propagation Delay	Query Processing Load, Duplication Processing Load

The identified and proposed metrics at different layers of the system are presented in Tab. 3.

4.3. Further Work

The next step of development of the presented benchmarking system is deployment of an example measurement scenario. The planned scenario is a tool allowing content-based search in P2P overlays. Such search mechanism will allow searching for content stored in a distributed repository. The quality of the search system will be assessed with use of the presented benchmarking system. The work on the exemplary scenario is advanced. A repository of user-tagged images was constructed in order to serve as a ground truth and a methodology of effective content-based search in P2P overlays is being studied.

5. Summary

Continuous development of multimedia services like searching over P2P repositories and video streaming causes a competition between content and service providers. It is easy to envision that monitoring, assessing, and tuning quality of experience is becoming a main and a new tool used by involved market players to attract users. The paper presents an approach to benchmark the quality of experience for video streaming and P2P searching that is under development in the VIFP CONTENT *Content Networks and Services for Home Users*.

6. Acknowledgments

This work presented in this paper has been supported by the CONTENT Network-of-Excellence, which is funded by the European Commission in the 6th Framework Programme (Project No. IST-FP6-038423). It includes contributions from all project partners. Therefore, the authors would like to especially thank their colleagues from University of Oslo, Lancaster University, University Pierre and Marie Curie (Paris 6), National and Kapodistrian University of

Athens, Technische Universität Darmstadt, Delft University of Technology and University of Napoli.

The work presented in this paper has been supported partially by two national grants funded by the Ministry of Science and Higher Education: 1) *Usługi i sieci teleinformatyczne następnej generacji – aspekty techniczne, aplikacyjne i rynkowe* and 2) *Wpływ jakości usług transmisji (QoS) na poziom satysfakcji użytkownika (QoE) multikastowych usług strumieniowania w środowisku heterogenicznych sieci IP*.

A. Annex – Multicast applications for real-time multimedia communication

In this section we will describe multicast multimedia transmission, applications for multicast multimedia communication, the RTP protocol and RTP topologies.

A.1. Multicast multimedia transmission and multimedia broadcasting

The "Encyclopaedia of Internet technologies and applications" defines multicast as follows: "A *Many-to-Many (M-to-N) transmission scheme, where M senders disseminate information to N receivers. Multicast transmissions are not broadcasting to all possible receivers, but they are addressed to the group of receivers – disseminated data is received only by members of the multicast group.*" [34].

Typically, multicast is a transmission scheme, when one sender disseminates data (often: multimedia data) to many receivers. Multicast transmission is received only by these receivers, which are interested in obtaining "multicastly" distributed data (in practice: by recipients, which join specific multicast group).

This transmission scheme is a natural mapping of typical broadcast transmission[4] (terrestrial – both cable and wireless – or satellite) into the IP network (terrestrial – both cable and wireless – or satellite). In the case of multicast, the role of radio or television channels plays multicast groups which are identified by unique (in scale of specific scope – e.g. link-local, site-local, or global) multicast addresses. Users, which want to watch a television program, must adjust their television set to a specific channel. Similarly, users, which want to watch television program via IP network, using the IP multicast transmission, must join specific multicast group.

Multicast transmission is especially important everywhere, where broadcast services migrate toward the Internet (or, more generally, IP network). These services, as radio and television, in IP networks have multicast nature. Although One-to-Many multicast transmission can be replaced by set of unicast transmissions, such replacement is ineffective. Moreover, in the case of large multicast groups (i.e. in the case of thousands or millions recipients) carrying out such set of transmissions in real-time can be impossible. As a result, services which naturally disseminate he same content to many recipients must evolve toward the IP multicast or change service model (e.g. replace the client-server service model by per-to-per model).

It's worth remarking, that equivalent of broadcast transmission is IP multicast, not IP broadcast. IP broadcast is a transmission of information to all recipients in specific network – both interested in getting information and not interested.

[4] as radio and television transmission

A.2. Real-time multicast multimedia services

Real-time multicast multimedia services are such services, which distribute multimedia data to more than one user in real-time. As a result, all users obtain multimedia data simultaneously (within an accuracy of network delays) and all users are able to play multimedia data at the same pace, in which they were sent or generated.

Thus, the term "real-time multicast multimedia services" can denote such services, as slide show associated with the audio/video transmission or whiteboard for collaborative work, where changes introduced by one user are immediately redistributed to others. However, typically as real-time multicast multimedia services are understood such services, as:

- teleconferences,
- Internet radio,
- Internet television and IPTV.

Teleconferencing services (teleconferences) can be divided into two groups: audio-conferences (where three or more users communicate using audio signals) and videoconferences (where three or more users communicate using video with associated audio). Teleconferencing services are similar to (video) telephony. However, the telephony is a unicast service (communication is carried out always between two persons), and teleconferences have multicast nature – one person sends multimedia data to two or more persons. Modern teleconferencing services are usually based on one of the two alternative architectures. One of them was standardized by the ITU, and the other – by the IETF. Both architectures use the RTP transport protocol for real-time delivery of multimedia data.

Internet radio is a mapping of typical radio broadcast service (terrestrial or satellite) into the Internet. As a broadcast service, Internet radio has multicast nature. However, nowadays Internet radios are very often associated with the Word Wide Web service and radio signal is podcasted using typical for the Web RSS (or Atom) technology. As a result, radio signal is often unicasted to only one user, using the TCP protocol, and broadcast character of radio transmission is emulated using a set of TCP connections. Such a situation is possible only now, when amount of service users is relatively small. In the future, if amount of users will be large (let's remind, that "traditional", broadcast radio signal is delivered to thousands or millions recipients), multicast technology will be necessary.

Like the Internet radio, Internet television and Internet Protocol Television (IPTV) are a mapping of typical television service (terrestrial or satellite) into IP networks: the public Internet (in the case of Internet television) or a dedicated operator's network (in the case of IPTV). Thus, as broadcast services, both Internet television and IPTV have multicast nature.

IPTV is a professional television service, intended for thousand or millions recipients. It is based on the IP multicast technology and use the RTP transport protocol for real-time delivery of television signals. Internet televisions are professional or semi-professional services, which also use the RTP protocol – currently, often in unicast manner. However, in the future, if the service develops, multicast distribution will be necessary.

A.3. The RTP protocol

Real-time Transport Protocol (RTP) is a communication protocol intended for transmission of multimedia data in real-time. In contrast to the best known Internet transport protocol, the Transmission Control Protocol (TCP), which poses functions, which belongs to both transport and session layers of the OSI/ISO model, functionality of the RTP locates the protocol only in the transport layer. The role of session oriented part of the TCP, plays here the Real-Time Control Protocol (RTCP), which was specified as the integral part of the RTP specification, RFC 3550 [27]. Both the RTP and the RTCP were designed as a multicast transport protocols, and their mechanisms are adjusted for transmission from one sender to many (tens as well as millions) receivers. The multicast transmission is treated as a special case of multicast transmission, where number of end systems is equal to 2.

Another difference between the RTP and the TCP is that the RTP is not able to work directly over the IP protocol. It must co-operate with other transport protocols, which assures multiplexing of transport connections. The RTP is located at upper sub-layer of the transport layer, while the lower sub-layer is occupied by a co-operated protocol, usually the UDP or the TCP. As a result, two alternative protocol stacks are used during RTP transmissions in IP networks:

- typical RTP over UDP protocol stack, i.e. RTP/UDP/IP, where the RTP is located at the top of the UDP protocol,

- RTSP interleaved protocol stack, i.e. RTP/RTSP/TCP/IP, where the RTP is located at the top of the RTSP protocol, which functions as an interface between the RTP and the TCP.

The RTP over UDP protocol stack is the most often used one. It is utilized in the case of such multicast multimedia applications, as teleconferences, Internet radio, Internet television and IPTV. The RTSP interleaved protocol stack is used instead of the RTP over UDP, if a least one firewall on the route of IP data-grams blocks incoming UDP packets. This protocol stacks is used, most often, during Internet television transmissions. More information about the RTSP interleaved transmission mode, interested Reader will found in the paper [1].

Unlike the TCP, the RTP transport protocol hasn't either flow control, or congestion control mechanisms. However, it can easy co-operate with such congestion control architectures, like adaptive coding, translators, received-driven layered multicast, receiver-driven stream replication. It is also possible to use the RTP with TCP-friendly protocols. Usage of one of these architectures – translators – is included in the specification of the RTP protocol.

The RTP transport protocol implements error control mechanism. However, functionality of the mechanism is limited, when comparing to the TCP's error control. Typically, error control consists of three stages: error detection, error signalling, and error correction.

Error detection, implemented in the RTP, is based on gaps in sequence space, what allows for detection of lost packets. If the underlying protocol is the UDP, detection of damaged packets will be possible thanks to UDP's check sum mechanisms. If the RTSP interleaved transmission mode is used, detection of both lost and damaged packets is carried out by the TCP. Because of TCP's retransmissions, in this mode RTP obtains always errorless data.

Error signalling is not performed by the RTP itself, but by the RTCP auxiliary protocol. The RTCP doesn't signal each single error, but only reports error rate, although immediate RTCP feedbacks from a receiver also are possible (see RFC 4585 [25]).

Error correction is not included in the RTP protocol specification, because error corrections via retransmissions are not recommended for real-time multimedia transmission. However, RFC 4588 [27] introduces this type of error correction to the RTP. It can be used by applications and services, which accepts larger end-to-end delays (e.g. of several seconds). Other correction techniques, as Forward Error Correction (FEC), also can be used.

The RTP is real-time protocol and some of its functions are real-time oriented. The most important one is based on a timestamp, which allows an end-user (user application) to play multimedia content at the pace of its generation. In the timestamp field of the RTP packet header is stored information about the time of generation of the first byte of data conveyed inside the RTP packet payload.

A.4. RTP topologies

RTP topology is a logical topology of overlay network, observed at the level of the transport layer of the ISO/OSI model, while the upper sub-layer of a transport layer is occupied by the RTP protocol. The RFC 5117 [39] defines eight different RTP topologies that are relevant for codec control:

- Point to Point
- Point to Multipoint Using Multicast
- Point to Multipoint Using Translator
- Point to Multipoint Using Mixer Model
- Point to Multipoint Using Video Switching MCUs
- Point to Multipoint Using RTCP-Terminating MCU
- Non-Symmetric Mixer/Translators
- Combining Topologies

The Point to Point topology consists of two end-systems. On the level of the transport layer, transmission between end-systems is carried out without any intermediate devices. On the level of the network layer, intermediate devices (here: routers) can be used (both point-to-point and chain topology is possible) and end-systems can be identified by both unicast and multicast IP addresses.

The second, Point to Multipoint Using Multicast, topology consists of N end-systems. Because every end-system can perform both sender and receiver functions, in the most general case[5], logical topology on the level of the transport layer resembles fully connected mesh – all end-systems are connected to each other[6]. This situation is typical for decentralized teleconferencing services. If only one end-system sends data (e.g. in the case of Internet television), $(N-1)$ end-systems are connected to the one, sending node.

The third and the four, Point to Multipoint Using Translator and Point to Multipoint Using Mixer Model, topologies consist of N end-systems and one of intermediate nodes defined in the RFC 3550 – the translator or the mixer. Translators are systems, which can translate both the media stream and the transport aspects of a stream. Mixers are devices, which merges audio and (or) video signals – typically for videoconference purposes. On the level of the transport layer, both defined topologies can have different structures, from the simple star topology (where the translator or the mixer is a central point of the overlay network) to very complex.

The five and six, Point to Multipoint Using Video Switching MCUs and Point to Multipoint Using RTCP-Terminating MCU, topologies are used typically in centralized teleconferencing

[5] Any Source Multicast (ASM) transmission mode is used.
[6] However, in contrast to fully connected mesh, end-systems cannot work as intermediate nodes.

services. They consist of *N* end-systems and one intermediate node – the Multipoint Control Unit (MCU). The MCU is a device, which both control teleconference and shows mixing behaviour. It selects (or merges) a media stream from all available teleconference streams (generated by conference participants) and forward it to a participant. In the case of described RTP topologies, logical topology on the level of the transport layer resembles star with the MCU as the central point of the star. The video switching MCU emulates multicast transmission between end-nodes, modifying content of RTCP reports, while the RTCP-terminating MCU establishes point-to-point RTP sessions between itself and each end-system.

The Non-Symmetric Mixer/Translators topology consist of *N* end-systems and one intermediate node – the MCU, which perform both mixer and translator behaviour. This topology is typical for mixed (centralized-decentralized) teleconferencing service, where the decentralized side use multicast transmission and the centralized side use a set of unicast connections. In the direction from centralized to decentralized side the MCU work as a translator, in the other direction it works as a mixer. In centralized side, the MCU functions as a mixer.

The last, Combining Topologies, is a hybrid topology that consists of *N* end-systems and several intermediate nodes (MCUs, mixers or translators). Combining Topologies are constructed by combining any of above described topologies.

References

1. A. Chodorek, R. Chodorek, *Streaming multimedia over TCP – RTSP interleaved transmission scheme*, KKRRiT 2008, Wroclaw 2008.
2. A. Parker, *Addressing the cost and performance challenges of digital media content delivery*, In P2P Media Summit, Santa Monica 2006.
3. A. Smeaton, P. Over, W. Kraaij, *Evaluation campaigns and TRECVID*, In MIR ’06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, 2006.
4. A. Takahashi, D. Hands, V. Barriac, *Standardization Activities in the IUT for a QoE Assessment of IPTV*, IEEE Communication Magazine, Vol. 46, No 5, 2008.
5. A. Vlavianos, M. Iliofotou, M. Faloutsos, BiToS: *Enhancing BitTorrent for Supporting Streaming Applications*, 9th IEEE Global Internet Symposium 2006, 2006.
6. C. Gkantsidis, P. Rodriguez, *Network Coding for Large Scale Content Distribution*, IEEE/INFOCOM'05, Miami 2005.
7. C. Gkantsidis, T. Karagiannis, P. Rodriguez, M. Vojnovic, *Planet Scale Software Updates*, ACM/SIGCOMM'06, Pisa 2006.
8. C. Kuhmünch, G. Kühne, S. Schremmer, T. Haenselmann, *Video-scaling algorithm based on human perception for spatio-temporal stimuli*, Technical Report Lehrstuhl Praktische Informatik IV, Mannheim 2001
9. C. Lambrecht, O. Verscheure, *Perceptual Quality Measure Using a Spatio-Temporal Model of the Human Visual System*, In Proc. of SPIE, Vol. 2668, 1996.
10. C. Leung, H. Ip, *Benchmarking for Content-Based Visual Information Search*, In Proceedings of the 4th International Conference on Advances in Visual Information Systems, 2000.
11. Content, *Home – Content*, http://www.ist-content.org/.
12. D. D. Cowan, C. I. Mayfield, F. W. Tompa, W. Gasparini, *New role for community networks*, Communications of the ACM, Vol. 41, Issue 4, 1998.
13. D. Kostic, A. Rodriguez, J. Albrecht, A. Vahdat, *Bullet: high bandwidth data dissemination using overlay mesh*, ACM SIGOPS Operating Systems Review, ACM SOSP, 2003.

14. D. Lopez et al., *Adaptive Multimedia Streaming over IP Based on Customer-Oriented Metrics*, ISCN06 Bogazici University, Istanbul 2006.
15. D. Martin, ed., *OWL-S: Semantic Mark-up for Web Services*, Technical Overview (associated with OWL-S Release 1.1), 2003.
16. ETSI STQ, *European Technical Committee for Speech, Transmission, Planning, and Quality of Service*, available in http://portal.etsi.org, 2008.
17. EvalVid, *EvalVid: A Video Quality Evaluation Tool-set*, available in http://www.tkn.tu-berlin.de/research/evalvid/, 2008.
18. F. Manola, E. Miller, *RDF Primer*, W3C Recommandation, 2004.
19. H. Pinson, S. Wolf, *A New Standardized Method for Objectively Measuring Video Quality*, IEEE Transaction of Broadcast, Vol. 50, Issue 3, 2004.
20. H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, *RTP: A Transport Protocol for Real-Time Applications*, RFC 3550, 2003.
21. International Telecommunication Union, *International Telecommunication Union – Telecommunication Standardization Sector*, http://www.itu.int/ITU-T/, 2008.
22. International Telecommunication Union, *Recommendation ITU-R BT.500-7, Methodology for the Subjective Assessment of the Quality of Television Pictures*, Technical Report, 1990.
23. International Telecommunication Union, *Recommendation ITU-T G.107, The E-model, a computational model for use in transmission planning*, Geneva 2000.
24. International Telecommunication Union, *Recommendation ITU-T P.800, Methods for subjective determination of transmission quality*, Geneva 1996.
25. J. Ott, S. Wenger, N. Sato, C. Burmeister, J. Rey, *Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)*, RFC 4585, 2006.
26. J. R. Smith, *Image retrieval evaluation*, In CBAIVL '98: Proceedings of the IEEE Workshop on Content – Based Access of Image and Video Libraries, 1998.
27. J. Rey, D. Leon, A. Miyazaki, V. Varsa, R. Hakenberg, *RTP Retransmission Payload Format*, RFC 4588, 2006.
28. J. Welch, J. Clark, *A Proposed Media Delivery Index (MDI)*, IEFT RFC 4445, 2006.
29. L. Guo, S. Chen, X. Zhang, *Design and Evaluation of a Scalable and Reliable P2P Assisted Proxy for On-Demand Streaming Media Delivery*, In IEEE Transactions on Knowledge and Data Engineering, Vol. 18, No. 5, pp. 669-682, 2006.
30. M. B. Rosson, J. M. Carroll, *Network communities, community networks*, CHI 98 conference summary on Human factors in computing systems, 1998.
31. M. Bawa, H. Deshpande, H. Garcia-Molina, *Streaming live media over peers*, HotNets-I, pp. 107-112, 2002.
32. M. Dean, D. Connolly, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, L. A. Stein, *OWL Web Ontology Language 1.0 Reference*, W3C Working Draft, 2002.
33. M. Farias, S. K. Mitra, *No-Reference Video Quality Metric Based on Artefact Measurements*, IEEE International Conference on Image Processing ICIP 2005, Vol. 3, pp. III – 141-144, 2005.
34. M. Freire, M. Pereira (eds.), *Encyclopaedia of Internet Technologies and Applications*, Information Science Reference, 2008.
35. M. Grega, L. Janowski, M. Leszczuk, P. Romaniak, Z. Papir, *Quality of Experience Evaluation for Multimedia Services*, Przegląd Telekomunikacyjny i Wiadomości Telekomunikacyjne, No 4, pp. 142-153, 2008.
36. M. Grubinger, C. Leung, P. Clough, *The IAPR Benchmark for Assessing Retrieval Performance in Cross Language Evaluation Tasks*, In Proceedings of the MUSCLE ImageCLEF Workshop on Image and Video Retrieval Evaluation, Vienna 2005.

37. M. Grubinger, P. Clough, H. Mueller, T. Deselears, *The IAPR TC-12 Benchmark – A New Evaluation Resource for Visual Information Systems*, In the Proceedings of the International Workshop OntoImage'2006 Language Resources for Content-Based Image Retrieval, 2006.
38. M. P. Papazoglou, *Service-Oriented Computing: Concepts, Characteristics and Directions*, in Proceedings of 4th International Conference on Web Information Systems Engineering (WISE 2003), Rome 2003.
39. M. Westerlund, S. Wenger, *RTP Topologies*, RFC 5117, 2008.
40. M. Zapter, G. Bressan, *A Proposed Approach for Quality of Experience Assurance for IPTV*, In Proc. of IEEE International Conference on The Digital Society, Guadeloupe 2007.
41. M. Zhang, J. G. Luo, L. Zhao, S. Q. Yang, *A Peer-to-Peer Network for Live Media Streaming–Using a Push-Pull Approach*, Proceedings of ACM Multimedia 2005, 2005.
42. M. Zhang, Y. Xiong, Q. Zhang, S. Yang, *On the Optimal Scheduling for Media Streaming in Data-driven Overlay Networks*, Proceedings of IEEE GLOBECOM 2006, 2006.
43. MSU VQMT, *MSU Video Quality Measurement Tool*, available in http://compression.ru/video/quality_measure/video_measurement_tool_en.html, 2008.
44. Ninux.org, *FrontPage – ninux.org Wiki*, http://wiki.ninux.org/.
45. O. Verscheure, P. Frossard, M. Hamdi, *User-Oriented QoS Analysis in MPEG-2 Video Delivery*, Journal of Real-Time Imaging, special issue on Real-Time Digital Video over Multimedia Networks, Vol. 5, No 5, pp. 305-314, 1999.
46. OPTICOM, *PEVQ Advanced Perceptual Evaluation of Video Quality*, http://www.opticom.de/download/PEVQ-WP-v07-A4.pdf, 2007.
47. PEVQ, *PEVQ Perceptual Evaluation of Video Quality*, available in http://www.pevq.org/, 2008.
48. Planetlab, *Planetlab*, http://www.planet-lab.org/.
49. PPStream, *PPStream web site*, http://www.ppstream.com.
50. R. Berbner, O. Heckmann, R. Steinmetz, *An Architecture for a QoS driven Composition of Web Service based Workflows*, in Proceedings of 2005 Networking and Electronic Commerce Research Conference (NAEC2005), Riva del Garda 2005.
51. R. Dosselmann., X. D. Yang, *A Prototype No-Reference Video Quality System*, Fourth Canadian Conference on Computer and Robot Vision CRV 2007, Vol. 2007, pp. 411-417, 2007.
52. S. Banerjee, S. Lee, B. Bhattacharjee, A. Srinivasan, *Resilient multicast using overlays*, In Proceedings of the 2003 ACM SIGMETRICS international conference on Measurement and modeling of computer systems, pp. 102-113, 2003.
53. S. Battle, A. Bernstein, H. Boley, B. Grosof, M. Gruninger, R. Hull, M. Kifer, D. Martin, S. McIlraith, D. McGuinness, J. Su, S. Tabet, *Semantic Web Services Framework (SWSF) Overview Version*, 2005.
54. S. Kanumuri et al., *A Generalized Linear Model for MPEG-2 Packet Loss Visibility*, Packet Video Workshop PV2004, 2004.
55. S. Ren, L. Guo, X. Zhang, *ASAP: an AS-Aware Peer-relay protocol for high quality VoIP*, Proceedings of the 26th International Conference on Distributed Computing Systems (ICDCS'06), Lisbon 2006.
56. S. Winkler, *Digital Video Quality – Vision Models and Metrics*, Wiley 2005.
57. S. Wolf et al., *Spatial-Temporal Distortion Metrics for In-Service Quality Monitoring of Any Digital Video System*, in Proc. SPIE, Vol. 3845, pp. 266-277, 1999.
58. T. Silverston, O. Fourmaux, *Source vs. Data-Driven Approach for Live P2P Streaming*, Proceedings of IEEE ICN 2006, Mauritius 2006.
59. T. Stockhammer, M. M. Hannuksela, T. Wiegand, *H.264/AVC in Wireless Environments*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, Issue 7, pp. 657- 673, 2003.

60. T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, G. J. Sullivan, *Rate-constrained coder control and comparison of video coding standards*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No 7, pp. 688-704, 2003.
61. Tibetan technology Center, *About the Dharamsala Wireless-Mesh Community Network*, http://tibtec.org/?q=node/60.
62. U. Engelke, H. Zepernick, *Perceptual-based Quality Metrics for Image and Video Services: A Survey*, In Proc. of IEEE Next Generation Internet Networks, Trondheim 2007.
63. V-Factor, *V-Factor Quality of Experience Platform*, available in http://www.pevq.org/, 2008.
64. *VQEG*, *Video Quality Experts Group*, available in http://www.its.bldrdoc.gov/vqeg/, 2008.
65. X. Revés et al, *User perceived Quality Evaluation in a B3G Network Testbed*, In Proc. of IST Mobile Summit, Mykonos 2006.
66. X. Zhang, J. Liu, B. Li, T. P. Yum, *Coolstreaming/donet: A data-driven overlay network for peer-to-peer live media streaming*, Proceedings IEEE Infocom 2005, 2005.
67. X. Zhang, J. Liu, B. Li, T.-S. P. Yum, *DONet/CoolStreaming: A Data-driven Overlay Network for Live Media Streaming*, in Proceedings of IEEE INFOCOM'05, 2005.
68. Y. H. Chu, S. G. Rao, S. Seshan, H. Zhang H, *A Case for End-System Multicast*, In IEEE Journal on Selected Areas in Communications, special issue on Network Support for Multicast Communications, Vol. 20, Issue 8, pp. 1456-1471, 2002
69. Z. Wang et al., *Image Quality Assessment: From Error Visibility to Structural Similarity*, IEEE Transactions on Image Processing, Vol. 13, No 4, pp. 600-612, 2004.
70. Z. Wang et al., *Video Quality Assessment Based on Structural Distortion Measurement*, Signal Processing: Image Communication, Vol. 19, No 2, pp. 121-13, 2004.
71. Z. Wang, H. R. Sheikh, A. C. Bovik, *Objective Video Quality Assessment*, In The Handbook of Video Databases: Design and Applications, pp. 1041-1078, 2003.
72. Z. Wang, L. Lu, A. Bovic, *Video Quality Assessment based on Structural Distortion Measurement*, Signal Processing: Image Communication, Special Issue on Objective Video Quality Metrics, Vol. 19, 2004.