Project acronym: CONCERTO

Project full title: Content and cOntext aware delivery for iNteraCtive multimEdia healthcaRe applications

Grant Agreement no.: 288502

Deliverable 2.2

Specification of cross layer system architecture

Contractual date of delivery to EC	T0+16
Actual date of delivery to EC	07/04/2014
Version number	2.0
Updated sections	2.3, 2.8, 2.9, 3, 4.1.2, 4.3, 5.2

Lead Beneficiary	BME
Participants	TCS, SIEMENS, VTT, NTUK, CEFRIEL, CNIT, BME, UOS, KU, UNIPG

Dissemination Level:	PU
Nature:	R
Total number of pages	72

Keywords list:
System architecture, architectural requirements of telemedicine services, main functional blocks, cross-layer optimization and adaptation

Executive Summary

The provision of telemedicine services requires the usage of advanced and reliable communication techniques to offer medically acceptable Quality of Experience (QoE) in the delivery of biomedical data between involved parties either using wired or wireless access networks. To overcome the restrictions of conventional communication systems and to address the challenges imposed by wireless/mobile multimedia delivery and adaptation for healthcare applications, consistently to the standards, specifications and trends for the Next-Generation Mobile Networks (NGMNs) as 4G mobile systems and beyond, CONCERTO project proposes a cross-layer optimized architecture with all the critical building blocks integrated for medical media content fusion, delivery and access.

This document is the first deliverable of Task 2.3 focusing on the general architecture design of CONCERTO and presenting the main components together with a sophisticated access network and inter-domain signalling framework with a distributed decision engine (with extensions to cover the whole Internet) for advanced cross-layer communication management and adaptation addressing the control and requirements of quality demanding interactive multimedia healthcare services. For this reason, the main goals of the document are to provide a state-of-the-art review of the context and the reference architectures, summarize the identified use-cases and related architectural requirements, as well as to give the first overview of the considered functional entities by describing them and their integration/cooperation in a simple but clear and comprehensive manner. The further elaboration and research of the presented building blocks and the cross-layer adaptation and signalling framework will provide the basis for the future and more practical work of Task 2.3, always in accordance with rules, standards and trends put in force or simply foreseen in the field of telemedicine and alike.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

Telemedicine applications play an increasingly important role in healthcare and provide tools that are indispensable for preventive care, remote patient monitoring, disease management, and addressing emergency. Advances in Information and Communication Technology (ICT), in particular multimedia processing, wireless networks, mobile devices, and information security provide practitioners, medical centres and hospitals emerging tools for managing patient care, electronic records and for the billing of medical services.

In the near future telemedicine can be applied even more widely, allowing ubiquitous telerehabilitation and therapy, personal healthcare monitoring, self-diagnosis, or even remote participation in life-saving operations by renowned experts. As mobile healthcare (m-Health) rises in parallel with the rapid adoption of mobile communications and computing, sensor networking, and advanced radio technologies, wireless telemedicine will play an important role also as a leading contributor for mobile internet evolution.

The CONCERTO project works to foster this expansion of telemedicine in a wider spectrum of applications and use cases. Our researches and proposed technology aim to be applicable to a wide range of use cases, providing the necessary techniques and algorithms to advance the use of ICT and multimedia solutions in the medicine field. In order to achieve this objective, several studies are on-going within the project to cover the technical challenges identified in the different use cases going from secure QoE-aware image/video coding to transmission techniques, mobility management, application layer optimization and QoE definition. The use cases of interest identified by our previous documents cover a vast range of possible employments of telemedicine applications in the medical world. The variety of the identified scenarios is reflected by differences among the use cases in terms of requirements and challenges to be solved. This wide spectrum of problems to solve implies strong constraints in terms of architectural requirements. The problem is further complicated by the restrictions and limitations of currently applied conventional communication systems.

Nowadays, more and more applications exploit the Internet and the underlying network connection for providing an enhanced user interaction. Upper layer services can rely on a variety of application layer resources such as data storage, computation capacity, specialized server capabilities, large data sets, and can make dramatic demands on network resources, such as bandwidth, and Quality of Service (QoS) guarantees. However there is a lack of layer interaction between networked applications and the underlying telecommunication infrastructure during service provisioning and consequently many upper layer services make poor use of network resources or do not achieve their overall QoS objectives, including being wrongly denied of requested resources actually available.

Currently, if an application client can obtain a desired large data set (file, video, database, etc.) from a server selected between many different options, the application service will take into account the current status and load on the possible servers but only minimal network considerations, such as topological proximity, connectivity, ping latency (i.e. the issued Round Trip Time), rather than actual link bandwidth utilization or other relevant QoS parameters (e.g. reliability, delay and jitter). The lack of communication between the application and lower layers (from physical to network strata) across the Internet does not allow cross-layer optimization to be utilised in several functionalities such as:

- coordinated application and network requests for available resources of all layers (including transmission ones)
- efficient recovery from failures based on policy related to all levels of the protocol stack
- monitoring and provisioning processes relying on synchronized information about application and lower layer resource availability
- consistent operations on information coming from different layers, devices and domains independently from employed technology (e.g. optical or electrical, wired or wireless)
- detailed knowledge and control on expected and actually provided end-to-end QoS

Without communication and optimized management across multiple layers, poor service provisioning, resource exploitation and non-optimal usage of available resources is likely to happen. Therefore cross-layer signalling is a necessary building block for content and network adaptation.

In this deliverable we present the first outcome about the definition of the full-comprehensive architecture with all the relevant building blocks and functional entities proposed in the project to satisfy the range of challenges posed by emerging wired and wireless/mobile telemedicine services. In the focus stand the associated cross-layer adaptation strategies and the cross-layer signalling subsystem devised to enable emerging multimedia e-Health and m-Health applications with medical level Quality of Service (QoS) and Quality of Experience (QoE). The document concentrates on the last bullet of the list above (signalling solution) by characterising the preliminary requirements for cross-layer signalling using the requirements of the overall CONCERTO architecture, identifying the functionalities which need or could benefit from the cross-layer adaptation and defining how different cross-layer signalling messages could be collected and made available to the relevant entities appropriately for optimization purposes. Based on this information together with the state-of-the-art studies we propose the first solution for the overall cross-layer system architecture.

2 Context and reference architectures

2.1 Goals and context of CONCERTO

Telemedicine applications play an increasingly important role in health care and provide tools that are indispensable for home healthcare, remote patient monitoring and disease management. Information and Communication Technology has reached a degree of maturity allowing us to envisage setting up a virtual hospital in a patient's home, relying near-real-time tele-consulting and diagnosis from a remote location and transmitting clinical data and multimedia medical content from one location to a large number of geographically dispersed locations. Advances in technology including wireless connectivity and mobile devices will give practitioners, medical centres and hospitals important new tools for managing patient care, electronic records and for the billing of medical services, in order to ultimately enable patients to have more control over their own well-being. While multimedia applications, such as for example 2D/3D imaging of anatomy, have become a routine part of doctors' everyday work, by contrast, remotely participating in life-saving operations by renowned experts remains an elusive future application – perhaps apart from a few fixed-network-based hospital-to-hospital communications.

The goal of CONCERTO is to propose solutions for speeding up diagnosis as well as therapeutic care delivery also from remote and on the move, allowing healthcare providers to receive continuous assistance from specialized centres.

CONCERTO studies and technology aim to be applicable to a wide range of use cases, providing the necessary techniques and algorithms to foster the use of ICT and multimedia solutions in the medicine field. In order to achieve this objective, several studies are on-going on the project to cover the wide technical challenges identified in the different use cases, going from transmission techniques to video and image coding and application layer optimization.

2.2 Reference architecture models

2.2.1 The cross-layer optimization paradigm

This chapter provides evidence of the need for cross-layer optimization (and communication) in order to achieve an effective support of value-added services and applications in the Future Internet, which can be seen as a confederation of autonomous systems [1], having each its own architecture, topology, technology and facility, over an IP infrastructure. In practice, the overall optimization of application layer and network resources, including transmission ones, requires the interaction and exchange of up-to-date and consistent information between the lower and upper strata across multiple devices, domains and technologies.

In this chapter, we will first introduce the limitations of current Internet for an effective support of networked application as a basis for the problem statement and objectives definition. Second, main challenges and necessary capabilities are discussed together with the inadequacy of already deployed control planes and proposals in the field. Then, design options for addressing the identified issues are considered in detail, as basis for the specification of an architecture supporting cross-layer communication and hence optimization over the Future Internet.

Later in the document we will generally describe a framework for cross-layer signalling to be used in the overall architecture of CONCERTO solution. Just to introduce it, the cross-layer signalling framework is based on event dissemination and it is somewhat similar to the architecture proposed in OPTIMIX project. Cross-layer events are made available for other network entities interested in those, through the services of Distributed Decision Engine (DDE), which is introduced to the CONCERTO system architecture. DDE provides dynamic event dissemination between different entities such as system layers and/or network equipment. DDE framework is an evolution of Triggering (TRG) framework, which has been introduced for cross-layer signalling purposes in [33]. In addition to the different signalling framework design in CONCERTO compared to OPTIMIX, DDE improves several issues of TRG framework and redesigns the overall architecture of it. The main design goal of DDE has been to facilitate decision making in cognitive networking. DDE provides an interface for algorithms making different kind of decisions regarding, for example, network load balancing, network changes, and traffic adaptation. Moreover, the message format is much improved from TRG. DDE messages support, for example, data validity field and digital signatures for data integrity and authentication. Event caching is a new feature, which improves accurate and timely decision making by making all valid previous events available. Dynamic information collected from each protocol layer enables various ways to timely and effectively adapt network traffic to varying link and network conditions. We also introduce a network information service, providing information about networks and their access points that change rarely. The DDE is discussed in more details in chapter 5, together with extensions for a wide spread over the internet in a scalable manner, leveraging already standardized protocols and mechanisms of IPv6-base networks.

2.2.1.1 Applications and issues related to cross-layer information

In the following, the issues related to popular applications and their effective support are presented with the purpose of underlining the need for cross-layer communication and clarifying the ultimate optimization objectives. It is worthwhile to point out that the use-cases addressed by CONCERTO project will be built on services that can be seen as a proper combination of the ones reported below or having the same nature. The applications are mainly considered from the Internet-wide perspective focusing on networking issues, while access network specific aspects are not included in details as discussed in depth later on in the document when appropriate.

2.2.1.1.1 File distribution systems

Internet content and file distribution systems have been set up as overlays on existing network infrastructures. The way to distribute files is an important point to consider also in telemedicine applications, where large amount of files with additional security issues may need to be exchanged (for example in medical education use case). Commonly encountered optimization problems with network implications include cache and mirror placement, efficient transfer of content to servers and client to server assignment. The cache placement problem concerns what content to allocate to which cache servers based on their proximity to clients and their load [2]. Mirrors differ from caches in that a client is only directed to a mirror if it has the desired content [3]. The mirror and cache server placement problem concerns where to place them given a fixed number of possible sites [3][4].

The employed algorithm necessarily works with knowledge of application topology and some type of network topological information, such as relative link cost models. Synchronization of application and network data (in general, of every layer) is the key to optimization. Actually, exact network models are not always necessary to achieve significant performance improvements [5]. The efficient transport of the original content to the replication servers becomes more and more important for optimization with the increase of the amount of material to transfer. When dealing with a large set of replication servers and quantity of data, the delivery benefits from point-to-multipoint concurrent communication path selected on the basis of the current network load conditions. Therefore, the cross-layer optimization process must have visibility into the underlying network resources and operating on up-to-date and consistent (hence, synchronized between layers) information.

In the assignment or selection of a content server for a given client, both server load (application layer information) and transfer latency between the client and server [5] should be taken into account. This highlights the need for synchronized multi-layer monitoring and configuration.

2.2.1.1.2 Content streaming systems

Basically, streaming services can be either live or on-demand. However, many variants in between these two extremes are created when pause or replay functionality is included in a live streaming service. Such services are quite different from file downloading ones. First, the beginning of content consumption does not require an entire file to be transferred. Second, minimum bandwidth and possibly other QoS requirements are to be guaranteed when delivering content between the server and client. Both live and on-demand streaming services are important to enable telemedicine applications and are key features in most of the identified CONCERTO use cases. This includes both context and medical data transmissions.

In live streaming, the client is willing to receive the content at its current play out point rather than at some pre-existing start point. A key network issue is whether multicasting takes place at the application or network level. Both the options are being adopted. For example, in carrier operated IPTV telecommunication infrastructures IP multicasting is spreading while in the case of an independent live video distribution service, overlay networks of servers provide the multicasting facility. In the end, optimization problems for a live streaming service are: either server selection (application based multicast) or leaf attachment (network based multicast) [6], and either server placement (application based multicast) or tree construction (network based multicast).

On-demand services entail additional technical challenges. From one hand, long start up delays should be avoided to retain customers, but batching together requests to save on server costs is desirable, on the other. Therefore, further optimization decisions and problems typically arise in the on-demand applications in addition to those seen in live streaming, such as client stream sharing technique and batch or Multicast Server selection.

Actually, the on-demand streaming service concerns are similar to those in file distribution: data allocation (when and where to pre-stock video files), on-demand server placement (where to put and how much capacity) and efficient (cost effective and timely) transfer of content to servers. Therefore, the need for cross-layer synchronized data is obvious.

2.2.1.1.3 Conferencing and gaming

Conferencing and gaming imply further complexity with respect to the cases above, in the related application connectivity and the need for cross-layer synchronization of monitoring, configuration, Operation Administration and Management (OAM). First, the issues associated with streaming services are also applicable to conferencing and gaming. Second, bi-directional and asymmetric bandwidth connections between the server and the user host are concerned. Third, gaming requires multipoint-to-multipoint communications with hard QoS constraints on latency.

This increased complexity over the point-to-multipoint case of streaming content distribution brings additional problems. Firstly, multipoint-to-multipoint data path formation and re-formation can be very inefficient without considering the underlying network resources. Secondly, addressing QoS constraints on latency and bandwidth guarantees for multipoint-to-multipoint connectivity requires coordination across the layers in terms of both path selection and reservation. Finally, massively multi-player online games (MMOGs) have the multipoint nature and QoS needs of conferencing but with additional concerns on scalability. For this [7] and due to constraints imposed by player preferences 0, an optimized server selection can be more complicated than in streaming content distribution.

2.2.1.1.4 Grid computing

Grid computing supports extremely large transfers of files and data-streams (live or on-demand). The volume of the traffic makes it critical to synchronize changes at the application and network. In addition, grid computing may have delivery requirement similar to those of streaming content distribution systems. Therefore, optimization objectives related to grid computing include effective instantiation of connectivity with high data rates and/or data set size, as well as ability to efficiently control very high speed networks.

2.2.1.2 Challenges and needed capabilities

In the previous section, problems that occur when resource allocations for both application and network are not synchronized in their action have been outlined together with the associated optimization objectives.

Hereafter, a more detailed discussion about the challenges and needed capabilities for achieving cross-layer optimization is provided with the aim of understanding the key issues to address in the design of an architectural solution supporting efficient communication and management across multiple layers, devices, domains and technologies.

Synchronized reception of multiple real-time topologies and Traffic Engineering related databases
As explained in the previous section, the processes of server selection and content placement can have dramatically better outcomes if current network and application topologies are known at the same time. It is critical to have quite detailed data about where the application clients and servers are located and how they are connected at network layers (from physical to IP ones).

The ability to capture up-to-date and consistent information allows planning during rapidly changing conditions or short traffic burst. For example, location selection for servers and clients requires that the performance estimates about the network and application layers align, in particular when stringent QoS parameters are to be guaranteed.

As 90% of traffic is represented by short flows [8] [9], out-of- date information or inconsistent between different strata, will provide an inaccurate representation of the network traffic and status in general. This can cause the selection of non-optimal paths across multi-layer network topologies. The key point is that the data needs to be synchronized at all levels.

Cross-layer cooperative load and traffic monitoring
Load and traffic monitoring can be facilitated having available information from Cross layer Optimization and management entities in all layers. Indeed, it is critical to have up-to-date and synchronized) data about QoS and load at every level.

Furthermore, as conditions change or problems occur, it may be important to adjust the granularity of these measurements. For example, bandwidth used and allocated/reserved, network delay statistics, existing client-server relationships and data regarding the allocation of clients to servers should be made available with different levels of detail, as needed.

Cross-layer synchronization of configuration changes
Re-optimization over the whole protocol stack end-to-end requires synchronized configuration at all levels. Otherwise, flows at a given layer may fall outside the planned network traffic patterns at other layers.

Cross-layer provisioning

The cross-layer connectivity entails the ability to provision additional communication paths and resources at some layers. For example, in MPLS-TE [10] and GMPLS [11] networks, the responsible IP entity is required to initiate connection setup across multiple-layers on behalf of the application entities (such as clients and servers).

2.2.1.3 Inadequacy and limitations of current control-planes and proposals

In the last few years, there has been a lot of research in the field of cross-layer communication and optimization over wireless networks. Many solutions have come out with the aim of adapting physical and MAC layers together with the application one (e.g. joint source-channel (de-)coding). Various approaches can be followed:

- application layer-based [12], where the application considers the lower layers as a black box and modifies the transmission according to a small set of feedbacks it can receive from them;
- Layer-Centric, where the application directly [13] or through the mediation of an intermediate layer [14] can access the internal protocol parameters of the lower layers and uses them for optimization purposes;
- Centralized [15], where the layers are connected to a middleware or are monitored by a single optimizer that estimates resource availability and coordinates the resource allocation, adapting each protocol's algorithms and parameters;
- Foresight oriented [16], where each layer decides autonomously for the most appropriate operation mode, using its own data and information gathered from messages exchanged with other layers.

All the above strategies do not take into account the IP and upper levels because the main issues related to the transmission on wireless links concern the MAC and PHY layers. However, in a general scenario there are a lot of feedbacks and information pieces that the network and upper layers can provide in order to improve the service provisioning, as clearly shown in the previous section. This is true indeed in CONCERTO, where all the layers should be jointly optimized in order to address the requirements of the considered use-cases, and enlarging the picture, to allow all critical application of NNGNs to efficiently coexist.

Actually, some attempts have already been made to address this point in the research world. CONCERTO cross-layer signalling architecture mainly follows the centralized approach, however some aspects from other approaches are used when utilising the cross-layer information in different functionalities.

Application Layer Traffic Optimization (ALTO) IETF WG [18] has been focusing on overlay optimization among peers by utilizing information about topological proximity and appropriate geographical locations of the underlying networks, including related resource usage and availability. With this method, an application may optimize selecting peer by location, but with a limited view and awareness. It is oriented to Peer-to-Peer (P2P) applications, Content Delivery Networks (CDNs) and in general services requiring a mirror selection, though it may be useful also in client-server environments. Anyway, the current scope of ALTO work does not cover multi-layer data synchronization and communication aspects for a full optimization (actually, the information exchange is only from IP to upper layers and not vice-versa), and it considers slowly time-variant information only.

Proactive Portals for Providers Participation (P4P) DCIA WG [19] has defined a portal for synergic collaboration between providers purely oriented to P2P applications. The P4P framework consists of a control-plane and data–plane components. The former introduces an entity (the iTracker) that supports portals for communication between P2P overlays and network providers so that it is possible to optimize the traffic management. Network providers allow routers on the data plane to give fine-grained feedbacks to P2P applications, enabling a more efficient usage of the network resources. It is worthwhile to underline that P4P is quite ahead in the development and implementation phases, but it is not of general validity and does not yet concern highly time-variant information.

A non-IETF BOF, namely Infrastructure to Application Information Exposure (I2AEX) has been extending the scope of ALTO IETF WG, considering that in controlled or partially controlled environments, such as Data Centres (DCs) and private CDNs respectively, it is safe and effective to make available a wider set of information to the application layer with respects to ALTO solution. However, the dynamism issued by wireless networks is not still addressed.

Yet, cross-layer strategies aiming at specific application type or network scenario have been developed into some FP7 EU projects: Alicante [20], Envision [21] and Adamantium [22]. The former provides Content-awareness to the Network Environment, Network- and User Context-awareness to the Service Environment, and adapted services and content to the End-User by relying on advanced middleware and network components that indeed, do not have a full multi-layer view. The latter ones follow a cross-layer approach where the problem of efficiently supporting multimedia services is solved cooperatively by service providers, ISPs, users and the applications themselves by delivering both content-aware networks and network-aware applications. Envision has focused on the cross-layer exchange of information and capabilities between the application and network layers (by the introduction of a newly developed interface), as well as on optimization strategies to be applied. The Envision cross-layer approach is able to delivers both content-aware networks, network-aware applications and media types similar to the ones issued by CONCERTO project. However, the supported facilities proposed also within ALTO WG do not cover any status information or feedbacks related to wireless interfaces, which are critical for optimization purposes in 4G networks. Adamantium has been based on a user/customer-centric approach and not on an engineering one. It proposes extensions to be applied to IMS-compatible architectures [23] specifically. In the end, the synchronization aspects related to extremely time-variant data exchanged across multiple layers are not explicitly tackled by either of the mentioned EU projects.

Regarding the existing network management solutions, significant limitations are present and critical extensions should be conceived and applied. SNMPv3 [24] [25] provides the idea of a SNMP context, which is defined as a collection of management information accessible by an SNMP entity. An SNMP entity potentially has access to many contexts. The missing piece is context with a management information view that allows synchronization of actions across multiple layers, devices and domains for read-view, write-view, notify-view and actions. Netconf [26] supports an XML based access to SNMP MIB data. The same concepts found in the SNMP access models of context for viewing data are implemented; therefore, the same lack of functionality for synchronization across all levels still exists. MPLS [12] is regarded as one of the underlying network transport technologies that could enable cross-layer optimization with application layer. However, current scope of MPLS OAM [27] does not encompass any non-MPLS device for its configuration and provisioning functions. In addition, UNI interface for GMPLS-controlled telecommunication infrastructures [28] is currently defined for network equipment only, not including direct interaction with the application layer services (i.e. precluding a bi-directional exchange of information between the upper and lower layers). ITU-T Y.2011 NGN and Y.2011 Resource and Admission Control Functions [29] discuss the service and transport strata separation. ITU-T Y.2012 [30] defines an Application-Network Interface (ANI), which provides a channel for interactions and exchanges between upper and lower layers. However, it does not address any issues on the synchronization of exchanged data.

2.3 Wireless access network architectures

As identified from the use–cases of D2.1, wireless networks are essential to support the expected telemedicine services. Those services are requiring both a local wireless network, and a wide area wireless network.

Wireless LAN or Wireless PAN are expected for supporting the connection of the end users wireless terminals like tablets and smartphones that are becoming more common in the hospital and also on emergency area. They can also be used to support the interconnection of medical equipments and in particular medical sensors. The use of wireless access network is expected to support many use-cases requiring an access to remote medical teams in mobility situation.

Some use cases identified could likely be realised using consumer wireless networks based on 3GPP LTE technologies. For more secured and critical scenarios, the use of dedicated network is necessary in order to ensure the reliability of the communication. Several years ago, for such a critical scenario, the use of dedicated network was expected to be based on WIMAX technology. Due to the rapid expansion of 3GPP LTE network around the world, the situation seems to have recently changed. A major movement started with the decision of the USA to deploy a national Public Safety network based on LTE technology in 2011. 3GPP started then to enhance the LTE technology to meet public safety application requirements, by adding a direct communication between devices and group communication services. Cooperation has been established between 3GPP and other PPDR standardisation of lobby groups such as ETSI TC TETRA, TCCA and US National Institute of Standards and Technology (NIST) to ensure broad representation of the public safety community. 3GPP’s objective is to preserve the considerable strengths of LTE while also adding features needed for public safety. A further goal is to maximise the technical commonality between commercial and public safety aspects to provide the best and most cost effective solution for both communities.

In light with this recent business evolution, CONCERTO assumes the use of 3GPP LTE technologies for accessing to the wide area network and the use of WLAN technologies like 802.11 and 802.15 for local area network.

2.3.1 3GPP EPS Architecture

Figure 1: illustrates the overall 3GPP network architecture [107] expected LTE. It is marked by the elimination of the circuit-switched domain and a simplified access network. The functional entities depicted in this figure are highly flexible, and can be physically co-located, or reside in dedicated hardware according to the network operator’s needs.

The EPS as Evolved Packet System, is composed off 2 two main parts: the E-UTRAN and the Evolved Packet Core (EPC). The result is a system characterized by its simplicity, a non-hierarchical structure for increased scalability and efficiency, and a design optimized to support real-time IP-based services. The services can be realized by the IMS connected to the PDN Gateway and the PCRF.

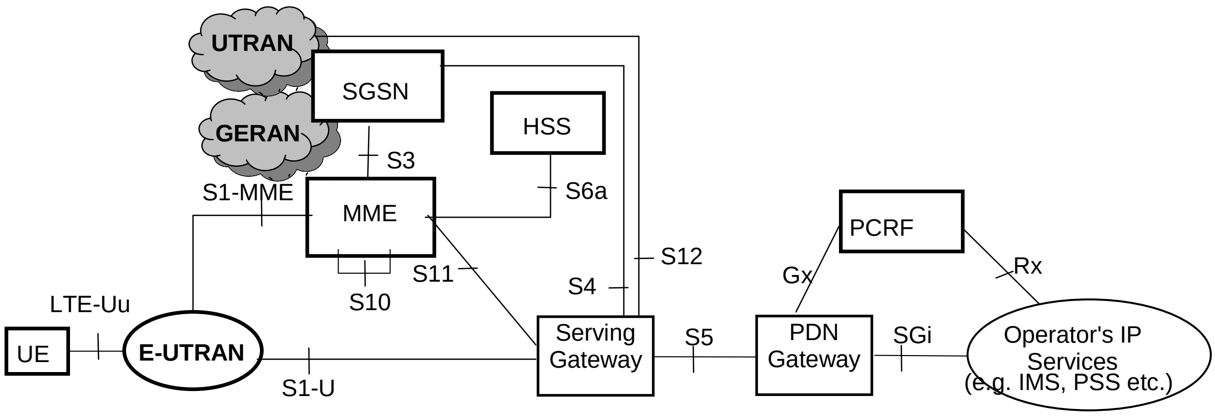

Figure 1: EPS Reference Architecture

The 3GPP network architecture [108] integrates the use of non 3GPP access technology that could be based on 802.11. Figure 2: present this architecture.

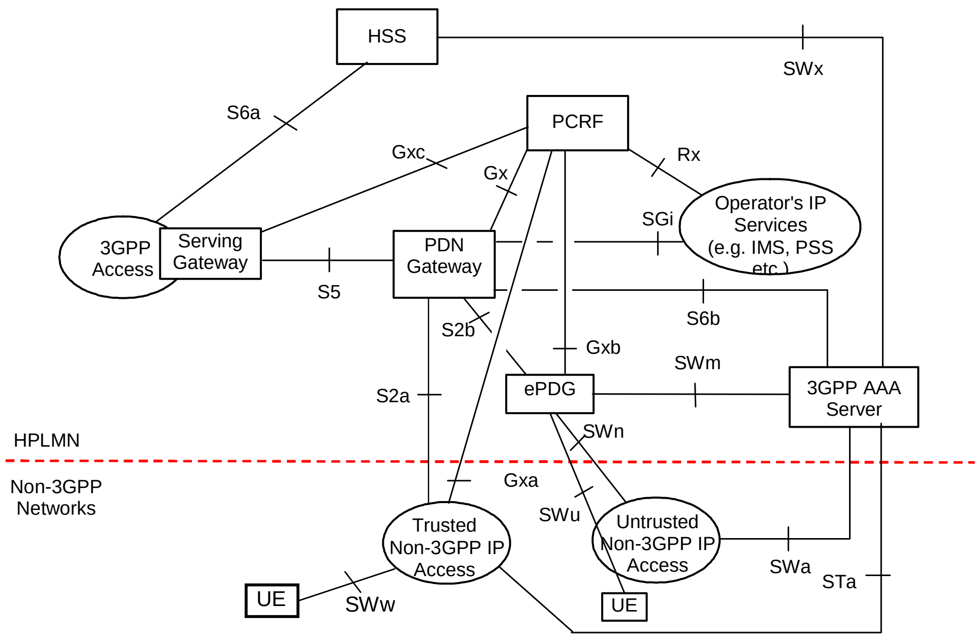

Figure 2: EPS Reference Architecture, Non-Roaming Architecture

2.3.2 Evolution of 3GPP EPS to support Public Safety features

Public safety networks allows communication between the first responders like the police, fireman, ambulance. As presented in the chapter 2.3, the 3GPP has started an evolution in order to support Public Safety features. The expectation of the first responders is to support group communication, and to allow communication between mobile devices even without infrastructure. Those functionalities will be introduced in the 3GPP release 12, expected to be finalised by June 2014. The group communication for public safety users requires mainly to improve the existing broadcast communication mechanism defined as part of the MBMS (Multimedia Broadcast Multicast Service). The enhancement of communication without the infrastructure requires the support of direct communication between mobile devices and also the use of mobile devices under infrastructure coverage, as relay for mobile device out of infrastructure coverage.

In term of business, the future LTE based public safety network will be able to fully or partially replace TETRA or P25 networks that are currently deployed.

In term of business roles, the current Public Safety networks are operated by institutional or private operator. The support of public safety features by LTE access network open a variety of possible roles for telecom operators, including public network operators. In 2013, the TCCA, TETRA + Critical Communications Association, a major association in charge of defining the Critical Communications, has produced a document [109] describing various business option for deploying broadband public safety network. In its conclusion, it clearly stated that future public safety network will likely not be dedicated to the first responders, but will partly rely on public networks.
It can also be foreseen that for many non-critical situations, the first responders will use commercial network in complement to the public safety networks. In section 3.3 some insights are provided on the possibility to integrate public commercial networks in public safety network offers.

2.4 Wireless and mobile telemedicine

The emerging concept of wireless and mobile telemedicine represents a natural evolution of eHealth (Electronic Health) from the conventional telemedicine to the progressing applications of mobile and wireless telecommunication. Being part of modern telemedicine – which generally offers higher diagnosis and treatment quality standards and reduces medical costs, – wireless and mobile telemedicine rises in parallel with the rapid adoption of mobile communications and computing, sensor networking, and advanced wireless technologies into our daily life. Recent years have witnessed an increasing and remarkable advancement of such systems and their applications: new, immersive, multimedia-driven, pervasive and interactive health services are rising to provide timely and prompt medical attention even through wireless access and mobile use-cases. The key areas are how wireless and mobile technologies can facilitate wearable sensor system based personal healthcare monitoring, self-diagnosis, early detection and preventive care [80], mobile-assisted telerehabilitation and therapy [81], monitoring of soldiers in tactical environments [82], medical care in emergency and mass casualty situations [83], intelligent home care [84], pandemic management [85], mobile robotic systems for controlling medical devices in isolated areas [86], personal and mobile wellness/dietary management [87], and other specialized services like drug intake reminder, raising awareness of health issues, performing health surveys, maintenance of personalized medical records, and medical advice provision/teleconsultation.

However, like other services running in mobile and wireless environments, the efficiency and usability of wireless and mobile telemedicine applications are substantially impacted by the continuously varying environmental characteristics, scarce network resources, sparse radio bands and bandwidth, battery and computational power of mobile devices, fluctuating delay, jitter and other QoS parameter values. This clearly implies the need of cross-layer based, context- and content-aware mechanisms making applications, service provision and delivery procedures adaptable to the extremely diverse wireless and mobile environments. In the following sections we provide a brief overview of the most important existing wireless and mobile telemedicine architectures.

2.5 Short and mid-range wireless telemedicine architectures

Short-range wireless telemedicine architectures rely on communication techniques like Ultra-wideband (UWB) Bluetooth, Zigbee and traditional Wi-Fi. The transmission range of these wireless solutions varies from several centimetres to several hundreds of meters; therefore medical services provided upon the architectures in this section are focusing on the capabilities of body/personal area networking or local area communication.

2.5.1 Wi-Fi telemedicine architecture

The study of [91] introduces a telemedicine architecture based on modern and widespread wireless local area networking schemes such as 802.11b/a/g/n. Authors describe the applied platform and possible Wi-Fi based local wireless telemedicine transmission services from paging nurses and voice communication through telemedicine remote sensing, bedside patient surveillance, electronic medical record/clinical information system application, drug barcode management till patient billing. A Direct Sequence UWB wireless indoor telemedicine solution has also been studied by the same correspondent author in [92], where the applied Ultra-wideband technology provides data rates of 1.320 Mbps and a transmission range of 10 m for the same indoor telemedicine services.

2.5.2 Bluetooth telemedicine architecture

In [93] authors propose a telemedicine architecture for transmitting clinical biomedical data to a mobile phone from a mobile telemedicine processor using the three performance/range classes of the Bluetooth communication standard (i.e. 100mW/100m for Class 1, 2.5mW/10 m for Class 2, and 2, 1mW/1 m for Class 3). The biomedical data consist of electrocardiogram signals of 100 KB, digital X-ray static images of 1 MB, and ultrasound video images of 30s of 10 MB.

2.5.3 BodyLAN

Another example for short-range wireless telemedicine architectures is the wearable medical sensor system called BodyLAN [94]. BodyLAN employs a wireless body area network (WBAN) based scheme using Wi-Fi, Bluetooth RF-ID or Zigbee communication techniques to transmit real-time physiological data gathered from a set of biomedical sensors distributed all around the body of the patient to a wearable computer. The proposed solution is compact, light weight, and easy to wear.

2.6 Satellite telemedicine architectures

It is an essential requirement for emergency services to provide personal healthcare for travellers on ships and aircrafts, or for people who work/live in remote locations such as high mountains and isolated islands. In current maritime telemedicine architectures this is achieved by adopting interactive video conferencing with the transmission of biomedical signals using mobile satellite communication techniques.

2.6.1 ETS-V

The Engineering Test Satellite-Five (ETS-V) maritime telemedicine architecture [88] provides medical diagnosis and emergency services to travellers on ships and aircrafts. Authors of [88] propose a solution to transmit one colour image, one audio, three channel electrocardiogram, and blood pressure information. The ETS-V telemedicine architecture includes a video camera, a microphone, a sound amplifier, a portable cardiograph, an automatic blood pressure measurement device, and a computer with a display, and employs channel coding, and ARQ retransmission to reduce the bit error ratio. The transmission data rate is 16/24 kbps for ships and 24/24kbps for aircrafts. It is used by the 8 kbps colour video signal (8-bit, 256×256 pixels image per 20s, compression rate=10:1), by the 10 kbps audio signal (6000 8-bit audio samples per second, compression rate=4.8:1), by the 600 bps electrocardiogram signal (containing 200 samples per second for each of the three channels, compression rate=8:1), and by the 0.3 bps blood pressure signal (16-bit sample per minute, compression rate=1:1).

2.6.2 TelePACS

The WWW-based TelePACS was proposed by [89] to realize online and offline real-time telemedicine applications by providing access every permitted picture archiving and communication system (PACS) server over the Internet. The architecture consists of a computer, a modem for receiving medical information over the Internet using QPSK modulation, an MPEG2 multiplexer decoder, and a PCI bus interface. The system uses DICOM format medical image provided by computer tomography (CT), and magnetic resonance imaging (MRI) devices. Users of TelePACS can browse medical image data using a simple web browser and store them over HTTP protocol. The architecture employs 20:1 JPEG static image compression and 20:1 wavelet image compression for CT images.

2.6.3 MERMAID

The MERMAID maritime telemedicine system [90] uses INMARSAT satellite access and Integrated Services Digital Network (ISDN) to provide medical emergency aid for ships and travellers in general. Its feature set includes telemedicine conference and multimedia communication, and relies on different communication media like optical fiber, coaxial cables, satellites, cellular radio, wireless local area network, ultra-wideband (UWB), and Bluetooth, with network technologies such as asynchronous transfer mode (ATM), ISDN, and the Internet. It has two operation modes. On one hand a basic mode is available with low data transmission rate and low channel bandwidth. It is used to transmit medical records. On the other hand there is an advanced mode with a high data transmission rate as well as high channel bandwidth to support interactive multimedia telemedicine conferencing and to transmit electrocardiogram signals.

2.6.4 MEDI

The MEDI maritime telemedicine architecture [95] was designed to provide medical services to users anywhere at any time via a next-generation mobile satellite system that integrates cellular wired/wireless communication networks in order to form a seamless mobile telemedicine solution. The supported applications enable using searchable medical image databases containing patient medical images, conducting X-ray image processing, participating to remote medical conferencing, observing medical diagnostic images and real-time discussions. The medical service provision within the MEDI architecture is done by a web page application contained in an HTML document, while the communication subsystem employs the Eutelsat satellite system for the downlink, and the ITALSAT satellite system for the uplink communication from the mobile device. All medical image information is stored in DICOM compatible format. The round-trip time within this system is 600 ms, the packet loss rate is around 0.014%. FTP transmissions of 10, 100, 1,000 KB files have a data transmission rate of 20 kbps. The HTTP transmission time required for digital X-files having a size of 135 KB is 15 s, and that for 12 MB video files is 95 s.

2.6.5 SWCDMA-based maritime telemedicine architecture

In [96] authors discuss a Ka band satellite wideband code division multiple access (SWCDMA)-based low earth orbit (LEO) transport architecture for mobile telemedicine. The solution adopts orthogonal variable spreading fact codes (OVSF), pseudo noise codes, unequal error protection, adaptive modulation, and among other techniques a power assignment mechanism to transmit biomedical data. Patients and doctors use microphones and CCD image sensors for interactive telemedicine conference participation. Diagnosis can be provided based on patient history data, body temperature, pulse, electrocardiogram (ECG) and electroencephalograph (EEG) biomedical signals, both transmitted via the SWCDMA-based maritime telemedicine architecture. The system is able to transmit biomedical signals including blood pressure/body temperature, and 108 kbps bit streams of the 12 channel ECG and 262.114 kbps bit streams of the 64 channel EEG signals. A G.729 encoder is used to compress 64 kbps audio signals to 8 kbps audio bit streams, while an H.264 encoder is adopted to convert the 13.15 Mbps video data into 1.13 Mbps video bit streams. For the medical image data JPEG2000 is used for compression (3.640 KB X-ray image signals will form a 128 KB image bit stream). The system also employs a model by which temporal constraints among various data that must be observed at the time of a real-time telemedicine conference can be specified.

2.7 Cellular telemedicine architectures

First generation cellular mobile communication systems (e.g. NMT-450, NMT-900) aimed to provide analog voice services within a network range from 2 km to 30 km. The second generation (e.g. GSM, IS-95) aimed to provide digital voice services and also introduced data services for mobile terminals (short message service /SMS/ and multi-media messages /MMS/), allowed much greater penetration intensity, and provided the evolutionary steps (GPRS and EDGE) toward packet switched high bandwidth mobile infrastructures. Third generation (e.g. IMT-2000, WiMAX) aimed to provide all-IP mobile multimedia services with a transmission rate of at least 200 kbit/s to several Mbit/s, and opened the door for the mobile broadband and the true mobile Internet. The fourth generation of mobile communication technology standards integrate heterogeneous access infrastructures (e.g. LTE, LTE-A, mobile WiMAX, Wi-Fi) under an all-IP ultra-broadband mobile Internet architecture that supports amended mobile web access, IP telephony, video gaming services, high-definition mobile TV, video conferencing, 3D television and Cloud Computing. This section discusses how the above cellular communication systems are applied in telemedicine architectures.

2.7.1 WAP-based telemedicine architecture

Authors of [97] introduce a telemedicine architecture that employs the wireless application protocol (WAP) version 1.1 on the top of a GSM 1800MHz cellular system. The presented mobile platform is able to store and relay biomedical data of patients, such as ECG signals, patient medical history, hospital messages and physician’s recommendations. An example scenario is when the architecture is applied to cardiogram browser. In that use-case heart beat, patient history browser, clinical or hospital messages and physician’s recommendations are provided to the WAP user. Such applications are stored in the WAP content server, and whenever required, they are downloaded and the biomedical signals are stored in the WAP component.

2.7.2 Airmed-cardio

The Airmed-Cardio [98] is a specialized architecture developed as a heart disease patient healthcare agent and biomedical data transmission system using GSM access and Internet services. A portable recording tool and a mobile terminal is used to record cardiogram medical data and transmit them over WAP, when patients are outside the range of a hospital’s healthcare and surveillance system. The proposed architecture comprises two sub-systems. The mobile patient sub-system records heart and physiological parameters, thus enabling remote surveillance and diagnosis. Multi-channel ECG signals, pulse and blood pressure is measured and transmitted to the health administration subsystem in well defined periods by a portable electrocardiograph and a sphygmomanometer, respectively. The transmission is performed through the GSM modem of the mobile terminal and processed by automatic analysis and by medical experts.

2.7.3 AMBULANCE

The AMBULANCE [99] telemedicine architecture was designed to help paramedics providing appropriate pre-hospital medical care by transmitting images and vital signs of the patients in an emergency situation to the hospital or emergency medical centre in real-time and full-duplex manner, where interactive telemedicine conference is to be performed with the experts. The AMBULANCE mobile architecture includes a camera, electrocardiograph, and a portable computer with GSM modem for transmitting medical information via GSM to the wired and wireless local area network of the hospital. The architecture employs GSM for both the encryption and for other communications. IP-based communication scheme is applied with a data transmission rate of 9,600 bps for real time transmission of three-channel ECG signals, blood sugar concentration, and pulse signal, each channel containing 200 8-byte samples per second, with a rate of 1,600 bps, and also for image transmissions for 320×240 pixel images having sizes of 2.5–3 KB (3-5 s transmission time per piece). The terminal side of the architecture includes real-time medical signal and image acquisition, user commands, system control module, mobile storage module, medical signal display, data compression/encryption and a GSM modem.

2.7.4 3G telemedicine architectures

With the continuous evolution of the data transmission rates in 3G networks and the proliferation of feature-rich 3G mobile terminals and smartphones, it became possible to manage high quality real-time voice and video interactions for pre-hospital medical care in case of emergency situations. This section introduces two examples for 3G telemedicine architectures.

2.7.4.1 Mobile teletrauma system

The mobile teletrauma system proposed in [100] realizes real-time delivery of medical care to patients over a long distance via voice and video data transmission by making use of real-time biomedical signals. Thanks to the CDMA technique, it is possible to simultaneously transmit voice and video information, medical images and real-time ECG signals. The system architecture comprises a patient and a hospital sub-system. The patient sub-system is located on the ambulance, where patient information is recorded on a portable computer with the help of different medical sensors, a portable ultrasound machine, an electrocardiograph. This information is transmitted to the hospital through the Internet accessed with the 3G network. The patient sub-system also possesses simple functions for analysing medical images and ECG signals in case of need. The hospital sub-system first decompresses the G.729 sound, H.264 video, and JPEG2000 medical image signals together with ECG signals, then the received information is displayed, observed and stored. The multimedia transmission is optimized by choosing TCP for ECG signals and clinical images, while applying UDP protocol for lower priority video data.

2.7.4.2 UMTS-based mobile emergency architecture

Authors of [101] propose an UMTS-based mobile emergency architecture to transmit medical information from an ambulance moving at high speed to the hospital. The biomedical data transmitted includes real-time audio (4-25 Kbps with tolerated packet loss of 3%), video (32-384 Kbps H.263 encoded stream with tolerated packet loss of 1%), blood pressure and electrocardiogram signals (1-20 Kbps with tolerated packet loss of 0%). These signals are measured onboard the ambulance, then encapsulated into a packet using the data convergence protocol of the UMTS standard family. The output is transmitted using UTRAN wireless communication control protocol.

2.7.5 MIThril LiveNet

The MIThril LiveNet [103] is flexible distributed mobile platform that can be deployed for various healthcare applications. The LiveNet system is built on the MIThril 2003 architecture that accommodates inexpensive hardware, flexible sensor/peripheral interconnection, a light-weight distributed sensing, classification, and inter-process communications software layer. The MIThril LiveNet architecture comprises three main elements: a PDA/Smartphone-centric platform with a wide set of mobile and wireless networking interfaces from Wi-Fi to 3G, the Enchantment software network and resource discovery API, and the MIThril real-time machine learning inference infrastructure. The architecture and its group-based context-aware healthcare services allow people to receive real-time feedback from their continuously monitored and analysed health state, as well as to communicate health information with medical personnel and other members of the patient’s social network for support and interaction.

2.7.6 m-Hippocrates

The m-Hippocrates mobile telemedicine architecture [102] provides mobile Health applications with constant and reliable communication using an application-level communication technology called Always Best Packet Switching (ABPS). The m-Hippocrates architecture comprises five main modules.

The first is a wide set of portable sensor and actuator devices worn and used by the patient.

The second is a client-side application that runs on the patient’s smartphone and manages mobile healthcare procedures by collecting data from the body area sensor network. This application also executes procedures ordered by the central healthcare system (e.g. activate actuators, administer medicine, etc.).

The third module is the ABPS subsystem, which is a distributed cross-layer networking service running on a fixed server and operating as a multipath communication channel. The ABPS module provides seamless communication between the mobile terminal and the rest of the world by maintaining an abstraction of an always available (TCP-based) connection between the client healthcare application and the healthcare service with the help of a proxy client installed on the smartphone and a proxy server installed on the fixed ABPS server. The proxy client monitors and dynamically configures the smartphone’s wireless interfaces by adaptively selecting the best access points and balances the load among the different mobile node network adapters and recovers from connection faults. The proxy server comprises a so called healthcare service interface (HSI) that takes the data stream coming from the patient device and forwards it to the healthcare centre.

The fourth module is a positioning system on the mobile device, and provides information about the patient's geographical location, so that he or she easily can be localized in case of need.

The fifth module is the server-side healthcare application. This module includes an expert system that can process biomedical signals coming from the patient, and contacts medical personnel for real-time evaluation of measurement data if the system identifies an abnormal situation. If the set of remote actuators at the patient contains a device that can administer medicine, the medical personnel might trigger it remotely.

The m-Hippocrates architecture assumes that the sensor and actuator devices are interacting with the head communication device (i.e. the smartphone) via a short-range communication technology (e.g. Bluetooth).

2.8 Emergency call management: EENA NG112

All over the world, citizens expect to be able to contact emergency services with technologies they use to communicate every day. Thus, European citizens have clear expectations about the availability of 112 emergency services with enhanced capabilities of technologies being used in daily life.

The Next Generation 112 (NG112) architecture [103] enables citizens to contact emergency services in different ways, using the same types of technology as those they use to communicate every day. It also makes possible that 112 Public Safety Answering Points (PSAPs) receive more and better information about emergencies of all magnitudes and improves interoperability between emergency services. Consequently, response time and operation cost will be reduced, while effective response will increase significantly.

NG112 addresses three major objectives:

1. Communication between citizens and emergency services: NG112 is designed to enable citizens to reach an authority (e.g., PSAP) by calls using VoIP, text messaging, real-time text, pictures and video. It could also provide emergency services with more data, such as location and health data. NG112 enables the delivery of calls, messages and data to the appropriate PSAP and other appropriate emergency entities, and adds significant value to the call handling process.

2. Interoperability between emergency services: NG112 enables several Public Safety Answering Points (PSAPs) to be part of a common emergency services IP network, providing them with redundancy and interoperability features. This network should support data and communications needs for coordinated incident management between PSAPs, and provide a reliable and secure environment for emergency communications.

3. Open Standards approach: NG112 is based on Internet Protocol (IP)-network based standard interfaces between all forms of communications components. Hence, existing off-the-shelf hardware and software can be deployed, which increases the technical commonalities between EU member states, drives TCO and fosters the European public safety eco-system.

The core concept and practical infrastructure of the NG112 architecture are built on the Emergency Services IP Network (ESInet). The ESInet is an emergency services network of networks that utilizes IP technology. ESInets are private, managed, and routed IP networks. An ESInet can serve a set of PSAPs, a region, a state, or a set of states. ESInets may be interconnected and have to be built upon common functions and interfaces making ESInets interoperable. NG112 specification [103] defines a number of Functional Elements (FEs), with their external interfaces. All PSAPs will have to be able to handle calls originated by different types of networks, in a secure way.

NG112 solution is based on key assumptions, such as:

1. A certificate authority that issues certificates to different entities in the emergency services networks has to be created. This enables proper authentication, and builds the foundation for authorization.

2. All calls entering the ESInet are Session Initiation Protocol (SIP) [104] based. Gateways, if needed, are outside of, or on the edge of, the ESInet. IP services that are not native SIP based, have protocol interworking to SIP prior to being presented to the ESInet.

3. Access Network Providers (e.g., DSL providers, fibre network providers, WiMax providers, Long Term Evolution (LTE) wireless carriers, etc.) have installed, provisioned and operated some kind of location function for their networks. Location is critical for 112 calls because it provides the ability for PSAP call takers to make decisions based on location and to dispatch first responders without undue delay to the person in need for help.

4. All calls entering the ESInet will normally have location (which might be coarse grained, e.g., cell site/sector) in the signalling with the call. This will allow for location based routing.

5. The Location Validation Function (LVF) and Emergency Call Routing Function (ECRF) are available. The LVF ensures that entered civic location information had been validated prior to its usage and ECRFs allow dynamic call routing based on location, and on additional policy information.

6. 112 authorities have accurate Geolocation Information Systems (GISs), which are used to provision the LVF and ECRF.

7. Civic location will be validated prior being used in an emergency call. This ensures that civic location that is incorrect can be detected early in the process.

8. Since the legacy circuit-switched TDM network will very likely continue to be used for the foreseeable future (both wireline and wireless) the NG112 architecture defines a Legacy Network Gateway (LNG) to interface between the legacy network and the ESInet.

The functional elements in the NG112 architecture are shown in Figure 3.

The left side of the picture shows various originating networks with a range of devices being able to trigger emergency communication. The originating networks include over-the-top (OTT) VoIP provider, IP Multimedia Subsystem (IMS) operators, enterprise networks, as well as legacy circuit switched networks. The standardization of the communication of the left-side of the figure is outside the scope of NG112, although proper integration of IP-based emergency services functionality helps to ensure correct working of the emergency services operations.

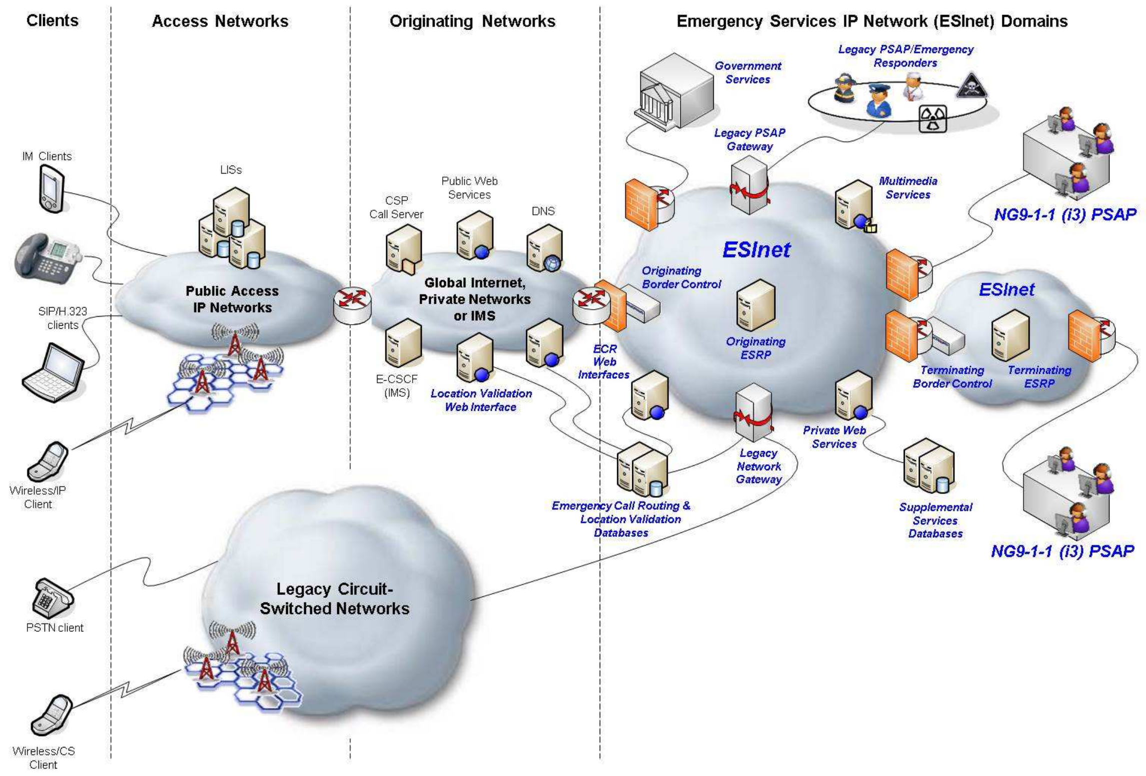

Figure 3: NG112 High Level Architecture

The right side of the figure represents the ESInet domain(s). It utilizes IP technology to perform emergency call routing and related functionality. The borders of the ESInet are secured using network level firewalls, and application layer firewalls, the so called Border Control Functions (BCFs). These devices are used to authorize every IP-based communication attempt entering the ESINet. For interworking with legacy technology, LNGs are utilized. In addition to their security function they also perform protocol translation from and to IP-based SIP signalling, and IP-based data exchange.

Location information is a crucial aspect for the ESInet in two ways. First, the Emergency Services Routing Proxy (ESRP) is a SIP entity that makes decisions about the call routing and uses location information for that purpose. Location is, however, not the only information used to determine call routing. Overload situations at PSAPs, time of day, skills of call takers, available technological features at the PSAPs, special needs of the emergency callers, etc. may influence call routing. Second, precise location information is also needed to dispatch first responders.

For this purpose various ESInet components need access to the caller's location information. In NG112, location information may be provided with the call, or accessed during the call by use of a reference to location information.

The most important components in the ESInet are the PSAPs, which are used by call takers to interact with the emergency caller.

ESInets are private, managed, and routed IP networks. An ESInet serves a set of PSAPs, a region, a state, or a set of states. The ESInet has a service area, defined by a (set of) polygon(s). ESInets are interconnected to neighbouring ESInets so that traffic can be routed from any point in the ESInet to any point in any other ESInet. The ESInet must be connected to the Internet through the BCF to accept calls. Access to ESInets must be controlled. Only public safety agencies, their contractors and service providers should be connected directly to the ESInet, preferably by VPNs. For Quality of Service (QoS) support, IP traffic handling devices within an ESInet must implement DiffServ. Many ESInet services depend on discovery of services via DHCP [105].

A BCF sits between external networks and the ESInet and between the ESInet and agency networks. All traffic from external networks transits a BCF. The BCF comprises several distinct elements pertaining to network edge control and SIP message handling, such as for firewalling, session control, security provisioning, emergency call detection and priority assignment (QoS management in general, including performance monitoring), forwarding to an ESRP, signalling protocol support and VPN bridging.

The ESRP is the base routing function for emergency calls, ESRPs are used in several positions within the ESInet. The ESRP routes a call to the next hop. The Originating ESRP routes to the appropriate intermediate ESRPs (if they exist), intermediate ESRPs route to the next level intermediate ESRP or to the Terminating ESRP, i.e., the appropriate PSAP. The Terminating ESRP routes to a call taker or set of call takers. In NG112, emergency calls will be routed to the appropriate PSAP based on the location of the caller. In addition, PSAPs may utilize the same routing functionality to determine how to route emergency calls to the correct responder.

ESRPs typically receive calls from upstream routing proxies. For the originating ESRP, this is typically a carrier routing proxy. For an intermediate or terminating ESRP, this is the upstream ESRP. The destination of the call on the output of the ESRP is conceptually a queue (i.e. emergency calls waiting for being processed), represented by a URI. Typically, the queue is maintained on a downstream ESRP, and is most often empty. However, when the network gets busy for any reason, it is possible for more than one downstream element to "pull" calls from the queue. The queue is usually First In First Out, but in some cases there can be out-of-order selections from the queue.

The primary input to an ESRP is a SIP message. The output is a SIP message with a Route header (possibly) rewritten, a Via header added, and in some cases, additional manipulation of the SIP messages. To do its job, the ESRP has interfaces to the ECRF for location based routing information, as well as for various event notification sources to gather state, which is used by its Policy Routing Function (PRF).

For a 112 call, the ESRP performs the following sequence of operations.

1. Evaluates a policy “rule set” for the queue the call arrives on.

2. Queries the location-based routing function (ECRF) with the location included with the call to determine the "normal" next hop (smaller political or network subdivision, PSAP or call taker group) URI.

3. Evaluate a policy rule set for that URI using other inputs available to it, such as headers in the SIP message, time of day, PSAP state, etc.

The result of the policy rule evaluation is a URI. The ESRP forwards the call to the URI (which is a queue as previously mentioned).

The NG112 functional element responsible for providing routing information to the various querying entities is the ECRF. The location information used (as provided by the Location Information Server), when in civic form must be proved sufficient for routing and dispatch prior to the call being placed. The LVF (Location Validation Function) is in charge of validating the gathered location information. Furthermore, a base database for NG112 is considered in the architecture, namely the Spatial Information Function (SIF). Nearly all location related data is ultimately derived from the SIF.

A PSAP provides several interfaces towards the ESInet. These are for key communication and exchange, such as for: SIP Call, location to Service Translation, Location Information Service, Special Information Function, bridging, policing and logging.

Finally, all the elements in NG112 architecture identified by hostnames must have corresponding Domain Name Service (DNS) records in the global public DNS. All elements connected to the ESInet must have local DNS resolvers to translate hostnames they receive to IP addresses.

A device, network or service provider presenting calls to an ESInet must support the interfaces for the followings: SIP Calls, Location Server, Call Information Database and PSAP Callbacks. Indeed, the originating call interface to the ESInet is a SIP call interface supporting also specific (extension) headers carrying the needed information (e.g. Geo-location and Call Info headers). The SIP call is routed through the networks depicted in Figure 3, crossing the issued border elements, possibly including gateways, contacting the relevant servers and being routed within the ESInet up to the issued PSAP. At this point, the call is then managed according to its nature and dispatched to the appropriate call taker(s) (e.g. the medical staff within a hospital).

2.9 NG112 and Concerto scope

In CONCERTO, we are addressing the need for a global delivery platform by combining the key actors in the different components of the end-to-end wireless multimedia content delivery: from multimedia content digital generation, to transport over professional networks or LTE wireless system towards the final destination.

The aim of the project is to build a media delivery platform that, by means of appropriate video coding and intelligent adaptation, allows the interactive transfer of high quality health-care related videos and images, in particular for transmissions of first responders from emergency situations to a hospital, to a doctor, etc.

NG112 relates, instead, to the control plane expected to manage this call and provide interoperability to manage the emergency calls of citizens to the emergency number 112 and to provide the necessary interoperability to manage this call. Although this is currently not detailed by NG112, there should be some adaptation of multimedia flow in gateway that could be perform to adapt the media flow to SIP/SDP/RTP protocols expected to be used in the ESInet.

CONCERTO relates to the data plane of multimedia, and its improvement between end users, eventually by using cross layer signalling. As presented in chapter 2.3, this data plane can be transported on various networks infrastructure that can be realised by a networks using LTE or Wi-Fi technologies, and that can be either a private Public Safety network (i.e., an LTE access based network dedicated to the first responders), a public network, or even mixing private and public networks.

According to such propositions, NG112 provides a service chain and a network architecture to the management of emergency calls between citizen and PSAPs and between PSAPS. The PSAPs can coordinate with first responders that could define a mission to manage CONCERTO scenario like the “Ambulance and Emergency”, “Emergency area with multiple casualties”, “Surgical assistance”, “Ubiquitous tele-consultation”. The communication between the first responders themselves and in particular the network architecture required for these communications is out of the scope of NG112. TCCA is a group of network vendors and operators that are in charge of defining such network architecture. TCCA coordinates with ETSI-TETRA and with 3GPP.

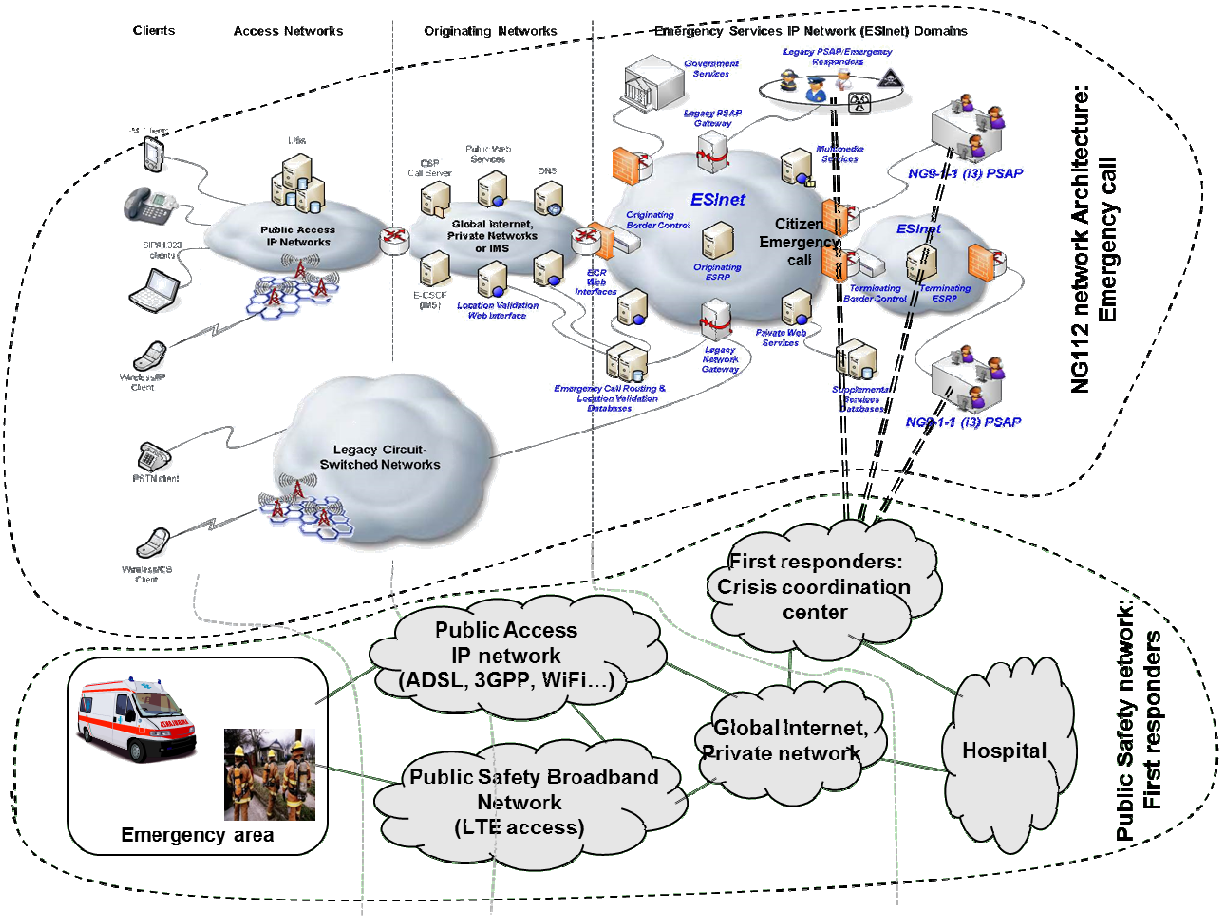

Figure 4: NG112 & Public Safety network for the Ambulance and Emergency scenario

Figure 4 represents the considered networks architecture relevant, for the NG112 network architecture to manage the emergency call, and for a public safety network architecture to allow the communication between the first responders. It shows at the bottom, the Public Safety networks used between the first responders, and on top, the NG112 considered network architecture. It should be noted, that for simplicity of the diagram, the Public Access IP networks and the global internet/private networks have been duplicated, but could in reality represent either similar or different network.

As presented, NG112 considers the management of emergency call between citizen and PSAP and between PSAP. CONCERTO solution covers the efficient delivery of the relevant data by adaptation and cross-layer optimization, , before the media consumption by final terminal equipment have been carried out. CONCERTO solutions can be applicable to various network architectures, as long as there is significant multimedia traffic to be considered, and as presented it is essentially applicable on communication network involved between the first responders.

3 Use-cases, architectural requirements and business perspectives

The use cases of interest identified by CONCERTO project cover a wide range of possible employments of telemedicine applications in the medical world, going from solutions not requiring significant improvements of technologies currently in use or particular modifications of current medical procedures, to use cases with a more long-term perspective that need more substantial evolutions in technology and emergency handling workflow.

The variety of the use cases is reflected by differences in terms of technical and functional requirements among the different use cases, which also means differences in architectural requirements.

In the following section a summary of the identified use cases is provided. Moreover a vision on how CONCERTO solutions can be integrated in commercial offers is depicted.

3.1 Summary of use-cases and considered scenarios

Seven use cases have been identified by CONCERTO project as being potentially beneficiaries of telemedicine applications. Their full description, including usage models, technical and functional requirements and potential benefits, is presented in deliverable D2.1. A short summary of each of them is provided here below.

1- Ambulance and emergency areas

This use case considers the transmission of image and video data to/from an ambulance in an emergency situation. Different types of image/video data, both medical (as live ultrasound image data streams) and not medical ones (ambient videos…) have to be transmitted from the emergency services to the hospital while on the move. Moreover, the transmission to the hospital of vital medical signals (such as Electrocardiography data) could be required as well. The availability of this information allows medics to decide the most suitable hospital / department which should treat the patient, to prepare the hospitalisation, and/or to guide medical or paramedic staff during the operation to be performed on the patient. This scenario involves the employment of both high-quality lossy as well as lossless compression methodologies and presents challenging requirements in terms of the overall reproduction quality and delay, according to transmission resources available.

2- Emergency areas with multiple casualties

This use case considers emergency scenarios and accidents, where multiple patients are involved and where there is a need to transmit and exploit/fuse multiple data flows to/from an emergency area. The fusion operation can involve heterogeneous and static (or with very limited mobility) sources of information present in the emergency area: the various devices can communicate wirelessly with the infrastructure (for example LTE/LTE-A). Extension of the infrastructure network can be realised by direct device communication and by some local/personal area (such as IEEE 802.15.4a, 802.11, LTE) or mesh (such as WDS, 802.11s) networks. The presence of multiple sources of information (e.g., medical devices, cameras, and smartphones) requires some coordination in the communications to/from the hospital, especially in case of limited availability of transmission resources. To this purpose, several criteria can be adopted to determine transmission policy and priority according to the available bandwidth and the information required at the emergency coordination centre of the hospital.

3- Emergency rooms

This use case considers the transmission of full Computer Tomography (CT) body scans from the emergency room to the radiologist/doctor, while on the move. This will permit remote medical diagnosis, avoiding the need of presence of the specialist in the emergency room. According to the client device, the data transmission should be adapted following a “thin versus smart client” approach. Moreover, the possibility to interact with and navigate through the image data would have an added value for medical diagnosis.

4- Ubiquitous tele-consultations

This use case intends to take benefit from telemedicine approach to help doctors to be more efficient and capable to provide consultations even when on travel or not in the hospital.

Preliminary consultation with nursing-home staff, patients and other doctors may be carried out using hand-held devices, which provide access to multiple heterogeneous wireless networks, regardless whether he/she is in. Such a system will also enable an information-aware ICT-savvy physician to download the detailed near-instantaneous patient information before arriving at a nursing-home for seeing multiple patients. Additionally, medical doctors may be travelling when their expertise is urgently required.

This use cases studies the possibility to transfer patient's data to the doctor's tablet PC enabling an immediate first diagnosis or preliminary evaluation of the patients' conditions. Images and videos that could be transmitted from the hospital to the physician are:

- ECG and other medical examinations results,
- CT scans,
- MRI images, and
- ultrasound (images and video).

5- Surgical assistance

In case of surgical operations it can happen that the transfer of a patient to the most appropriate centre for a specific surgical operation is not possible or too risky for the patient's conditions. The introduction of telemedicine techniques can help by providing a connection between the remote surgery site and the specialist surgeon at the main hospital. Thanks to this connection the specialist surgeon can help the remote surgical staff by providing assistance during the operation, or even remotely operate the patient using surgical tele-robots. The realization of this use case in its more advanced application (i.e., remote operations) is a challenging task, which requires an augmented vision environment facilitated by 3D video capturing, transmission and high-quality, low-delay remote visualisation.

6- In-hospital scenarios

This sub-set of scenarios intends to focus on improvement of existing scenarios through CONERTO technology. The idea is to improve comfort of the patients and of the physicians without revolutionising existing medical procedures. Additional in-hospital scenarios are for instance the following usage models:

- Diagnosis based on medical imaging (e.g., radiology, cardiography, etc.),
- Visits at the bed side of a patient, with access to medical records including images and videos,
- Treatment/surgery planning, and
- Digital pathology as an additional interesting usage model.

Typically, for these scenarios, the medical content is stored in an image data base and needs to be transferred to the respective viewing device. This can be a medical workstation, a regular desktop computer or laptop, or a mobile device like a tablet computer connected over a wireless link.

7- Medical education

This use case considers the possibility to use telemedicine applications in the medical education field. Attendance to surgical operations is an important part of the course of study of many medical education centres and universities, but it is not always easy to organize.

The use of multimedia solutions can be helpful in this context and support medical education around the world. Thanks to multimedia records, medical students could witness a surgical procedure without stepping foot into the operating room. Through a remote connection students in classrooms could assist the surgery in real-time or access to pre-recorded content or even follow on-line courses.

3.2 Architectural requirements

This section provides a short overview of the technical and functional requirements for all the considered use cases introduced in deliverable D2.1. The purpose is to show different requirements introduced by different use cases so that the reader will get familiar with the requirements and use cases defined in deliverable D2.1.

3.2.1 Technical and functional requirements for CONCERTO use cases

Table 1: Technical requirements

	Ambulance and emergency area	Emergency area with multiple casualties	Emergency rooms	Tele-consultation	Surgical assistance	In-hospital scenarios	Medical education
Uplink video transmission	x	x			x		
Downlink video streaming	x	x	x	x	x		x

Localisation capabilities of deployed cameras	x	x					
Bidirectional connection	x	x			x	x	
Asymmetric connection	x	x	x	x		x	X
Cellular technologies	x	x	x	x	(x)		
WiFi technology	x	x	x	x	x	x	X
Wireless personal/local area or mesh network	x	x					
Self organising capabilities of mesh devices		x					
Discovery of nearby mobile devices	x	x					
Relay-aided transmission	x	x					
Seamless connection	x	x	x	x	x		(x)
Low-complexity receivers				(x)		(x)	(x)
Large data-base	x	x				x	X
3D video coding	x				x		X
High fidelity coding					x	x	X
Content security	x	x	x	x	x	x	X
Bit rate	100 kbps - some Mbps (	<1Mbps per camera	Several Mbps	Several kbps	Several Mbps	Several Mbps	Several Mbps
Maximum transfer delay	Depending on video type	Depending on video type	Short	<3s	<150ms for surgery, <300ms for guided examination	200-300 ms	200-300 ms for real time attendance, <3s for streaming

We remind that functional requirements valid for all the use cases and thus not reported in the table below are:

- High QoE according to current performance and status of network and receiver terminal(s)
- User friendly devices
- Availability, as well as processing and querying of data about the underlying network and remote receiver terminal(s)
- Class of service differentiation and priority in the network
- Service differentiation capability on the basis of media type
- Seamless integration of heterogeneous wired and wireless technologies
- Full, fair and seamless exploitation of available radio spectrum and access modes.

Table 2: Functional requirements

	Ambulance and emergency area	Emergency area with multiple casualties	Emergency rooms	Tele-consultation	Surgical assistance	In-hospital scenarios	Medical education
Low-complexity video acquisition							X
Multiview video acquisition		x			x		X
3D video acquisition					x		X
Multisource information fusion and coordinated transmission	x	x					
Automatic switching of video type (2D/3D)	x						X
Automatic switching of video source		x					
Interaction 1: Client-side view selection		x				x	X
Interaction 2: Navigation options	x					x	X
Interaction 3: low loading time						x	
Very high guaranteed video quality	x		x	(x): for medical video	x	x	(x): for medical video
Session mobility						(x)	(x)
Multiview adaptation		x			x		X
No interruption (if coverage)	x		(x)		x		
Scalable, proactive, transparent and fine grained (i.e., even flow level) mobility	x		x	x			(x)
Transparent selection of caching server	x						

3.2.2 Cross-layer information foreseen in different functionalities

Cross-layer signalling is either needed or could be beneficial for several functionalities included in CONCERTO architecture. In Table 3, different functionalities are introduced which need or could benefit from cross-layer information. The functionalities are based on both functional and technical requirements introduced in Table 1 and Table 2 and different research topics which address these requirements.

Functionalities foreseen to use or benefit from cross-layer information are described in the table with the user/consumer of cross-layer information and the source/producer of cross-layer information. The user or consumer can defined as the system layer of OSI model which is using the available cross-layer information to make decisions during the transmission (e.g. whether to adapt the bitrate of the video to be transmitted, whether to prioritise the transmission of certain portions of the data stream). The source or producer of information is the system layer which creates this information based measurements, observations, decisions etc. In addition to the user and source of information, the type of information is described in the table to give an example what kind of information could be delivered within this functionality.

Table 3: Source and consumer of cross-layer information in different functionalities

Functionality	User of information	Source of information	Type of information
Reduction of network load by interactive view-switching	end-user application	network and physical layer	Available bandwidth, available networks (e.g. WLAN1, WLAN2, cellular…)
Adaptive HTTP streaming in cellular networks	end-user application	network and physical layer	available bandwidth (also recent history), delay
Adaptive camera set for FTV	end-user application	network and physical layer	available bandwidth (also recent history), delay
Data fusion and ranking for multi-source transmission	Application layer	lower-layers (MAC and physical)	Available bandwidth, delay, positioning information
Adaptation of frames and computation of FEC parameters	Application layer	lower-layers (MAC and physical)	Available rate at application layer, expected error/erase probability
Multipath streaming of medical video	transport layer	application layer	layer information and/or priority
Packet-level protection	transport layer	Application layer	layer information and/or priority
MAC level scheduling solutions for the downlink and the uplink	MAC-layer	application layer	layer information and/or priority
Service Differentiation in the network (Qos support for an improved QoE)	IP and lower-layers (possibly, only in the critical points of the network, as at the wireless interfaces in the access networks of the base-station(s), access point(s), mobile terminal(s))	Network Manager-Operating Centre, on the basis of the required delay differentiation guarantees	Settings for configuration, operating mode, allocation of resources, etc., consistently across multiple layers
Content-awareness in service differentiation at IP layer (based on media type and related constraints)	IP-layer (possibly, only in the critical points of the network, as at the wireless interfaces in the access networks of the base-station(s), access point(s), mobile terminal(s))	application layer (differentiating the importance of content, withing a flow and between different media flows)	Traffic type, layer information and/or priority
Proportional service differentiation at a network node	IP-layer (possibly, only in the critical points of the network, as at the wireless interfaces in the access networks of the base-station(s), access point(s), mobile terminal(s))	lower-layers (possibly, only in the critical points of the network, as at the wireless interfaces in the access networks of the base-station(s), access point(s), mobile terminal(s)). It refers to the lower layers of the same interface as the IP layer using the cross-layer information in object	Delays for each service class/queue at the lower layers
Consistent service differentiation across multiple layer at a network node	Lower-layers (possibly, only in the critical points of the network, as at the wireless interfaces in the access networks of the base-station(s), access point(s), mobile terminal(s))	IP-layer (associated with the lower layers at the same interface)	Service class identification (e.g. DSCP of AF PHB in DiffServ architecture)
Resource information collection	Collection points (i.e. associated with a DDE,	Resource information producer (e.g. user terminal	Static and slowly time-variant status,

	where a publication/notification model is employed). Roughly speaking, the so called Access Gateway (AG) as in the cross-layer signalling solution specified in D2.3	for the available uplink bandwidth at the emergency area)	configuration, etc. information
Resource information retrieval	Resource information consumer (i.e. entity that can profitably use cross-layer information, as the medical staff equipment at the hospital premises for the available bandwidth at the emergency area)	Collection point (i.e. AG associated with a DDE, as explained for the previous item of this table)	Static and slowly time-variant status, configuration, etc. information
Handover optimisation	IP-layer	Media Independent Handover Services	Available bandwidth, available networks (e.g. WLAN1, WLAN2, cellular…)

3.3 Expected service providers and business perspectives

CONCERTO multimedia platform can be seen as a suite of services and technologies that a service provider could include in a modular product to be proposed mainly to hospitals and first responders.

CONCERTO project deals with a wide range of use cases with strong differences among them. In their presentation, provided in deliverable D2.1 an analysis of the business perspectives for each use case has been done, identifying the main stakeholders and the main expected benefits for each of them.

The contribution of CONCERTO solutions to enable the selected use cases is different and goes from improvement of existing services (for example in the medical education use case) to the provision of a more complete system and of the technology to propose more advanced services (as the transmission of medical video streams from the ambulance to the hospital). The business perspective is then to integrate the CONCERTO outcomes in commercial offers of services that can be proposed, according to the use case, by different services providers.

In particular, CONCERTO has a clear interest for PMR operators.

CONCERTO outcomes provide solutions for multimedia delivery over wireless networks tailored for medical needs that can strongly benefit first responders and hospitals. Currently, the communications between the ambulance and the hospital mainly use PMR networks based on standards like TETRA.

Future PMR solutions are expected to provide broadband access and to allow the transmission of significant multimedia contents. In this context CONCERTO modules are key components to be integrated in a suite of services proposed by PMR operators to first responders and hospitals.

The business approach to the future of PMR broadband networks is still an open point as stated by the TTCA, the major association in charge of defining the Critical Communications Broadband in Europe, in [109].

For this reason, CONCERTO approach is to build a solution that can be applied on various types of networks, using 3G, LTE or Wi-Fi. In particular, CONCERTO also considers the case in which several wireless networks are available and could be potentially exploited in parallel: in this perspective, CONCERTO carries on studies and elaborates solutions to optimize the multimedia delivery through different access networks even if managed by different mobile operators, assuming that in the future current problems related to the control plan when dealing with different operators (like authentication) will be solved.

It turns out that this expectation to be correct, since recently some PMR operators went in the same direction for their broadband network proposals.

As a meaningful example, in November 2013, Thales launched the Eiji solution [110], a suite of secure PMR services which uses public wireless networks of different operators in a secured and transparent way for the end user.

4 Definition of the CONCERTO system architecture

4.1 Architecture overview

The original demand for cross-layer information in the current Internet architecture originates from the fact that legacy Internet applications and protocols – being designed with wired networks in mind – are incapable of operating efficiently over wireless links. The cross-layer signalling scheme applied in core of the CONCERTO architecture (Figure 5) aims at supporting optimization strategies for mobile and wireless health care services that can span the whole Internet in the future. Therefore our solution focuses on two aspects of the cross-layer signalling: on the access network including an inter-domain solution and on the Internet-wide aspects of signalling. leveraging the core entities, interconnections and functions dealing with cross-layer signalling and optimization (detailed in the following sections), CONCERTO architecture comprises several functional blocks for adaptive streaming, like media caching and remote viewpoint synthesis, 3D medical image/video data coding and storage, wireless access provision, and distributed dynamic mobility management (DMM).

The main goal of the CONCERTO architecture is to improve the final QoE as perceived by the end user, e.g., practitioner, surgeon or trainee, depending on the use case. The aforementioned system design aims then at providing a high experienced quality rather than just at optimizing the network QoS. The perceived quality is content and context dependent. Both subjective studies and objective metrics are being carried out and developed for this purpose with the aid of medical doctors.

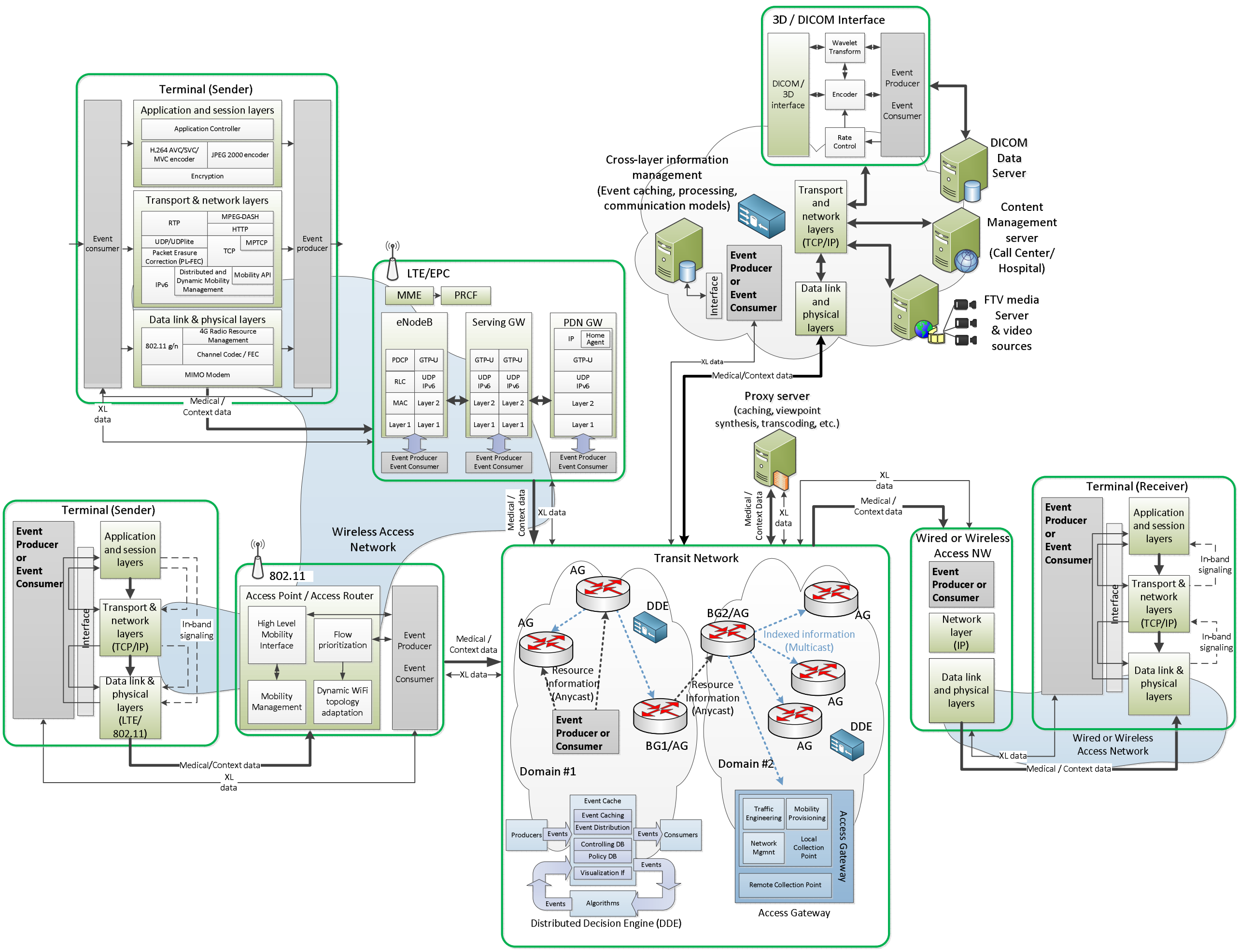

Figure 5: Overview of the general CONCERTO cross-layer system architecture

4.1.1 Terminals

Terminals that act as senders in the architecture, provide medical and context data. During the terminals' activity, their functional blocks produce and consume cross-layer information (XL data) as well. A part of this cross-layer information exchange happens inside the terminal, but that information, which is relevant for other blocks of the architecture is distributed over the network connection. Also, the terminal components consume cross-layer information originated from other points of the architecture. The destination of the medical and context data are receiver terminals and various servers, which display, cache, store, manage or process the received information. While most of these functions would traditionally require no cross-layer signalling at all, in contrast, the servers and receiver terminals in the CONCERTO architecture generate new cross-layer information while processing medical and context data or changing their connection parameters. Figure 4 shows the blocks of the sender side terminal. Blocks in the application and session layers, in the transport and network layers as well as in the data link and physical layers are all event consumers and event producers. The building blocks are detailed later in this chapter.

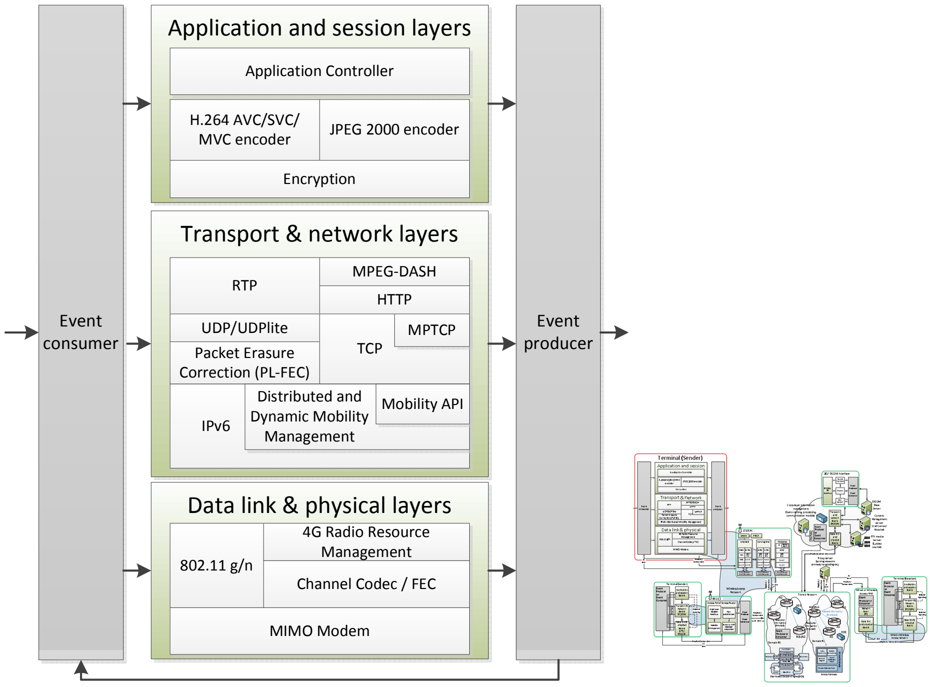

Figure 6: Sender side terminal blocks

Servers and receiver side terminals have the matching internal blocks. As an example, while we have a H.264 encoder in the sender side terminal, we have a H.264 decoder on the receiver side.

4.1.2 Networks

4.1.2.1 Access and transit networks

The transit network spanning over multiple domains and the wireless and wired interfaces provide the interconnection between terminals and servers. While they are transferring the entire medical and context data and the cross-layer signalling, they may also consume and produce cross-layer information.

Within the transit network there are cross-layer information collection points. Besides handling cross layer information, they may perform traffic engineering, network management and mobility provisioning also. The Access Gateways in different domains share the network information between each other locally and remotely. Figure 5 depicts the access gateway and its internal blocks.

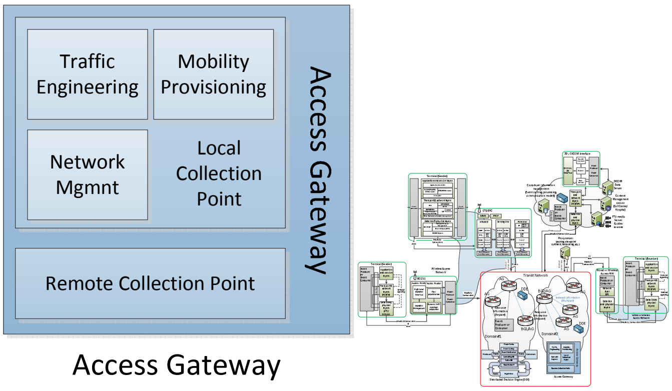

Figure 7: Access Gateway in the transit network

Access gateways and border gateways are discussed later in this chapter.

Within the overall system architecture addressed in CONCERTO, we can distinguish among three main logical areas, typically characterized by different radio access technologies:

- *Emergency Area/Local ambulance network.* Within the emergency area, as well as onboard or in the proximity of the ambulance, IEEE 802.11g/n technologies can be employed for the local transmission of videos, images and other medical data. In addition, a wireless sensor network based, for instance, on IEEE 802.15.4a may be used to realize medium and low rate connections, as well as to provide ranging functionalities among the nodes of the network aimed at supporting short-range localization of the different devices within the emergency zone (e.g. cameras and medical instrumentation).
- *Hospitals WLAN.* In the hospital area, we typically consider the presence of a WiFi local area network permitting the exchange of medical data between portable devices. Similarly, doctors' mobile terminals are connected to Internet through IEEE 802.11g/n technology.
- *Radio Access Network.* A 3GPP cellular access and Wi-Fi access are taken into account to connect the emergency area or the ambulance to a remote hospital. In particular, in the CONCERTO architecture, we consider LTE/LTE-A, 3G and Wi-Fi for the mobile links.

CONCERTO system is conceived to be capable to work with existing networks, but also propose some improvements that could be added in next generation networks (as detailed in next section).
In particular, for the access networks, CONCERTO considers not only the case in which only one network is available, but also the possibility to take advantage of several networks provided by different network operators. In both cases the focus is on data plan. Since most of the wireless scenarios are related to first responders' communications or hospital communications, it is expected that the network access (and all the related control plane issues) is guaranteed by the PMR operator both through a dedicated network or thanks to agreements with commercial operators. These aspects are better detailed in section 3.3.

4.1.2.2 CONCERTO enhancements to access networks

Wired and wireless network interfaces of the main components, as well as different components of the access network architecture may consume and produce cross-layer signalling. Functions like mobility management and the dynamic adaptation of the access radios in the IEEE 802.11 (WiFi) networks are based on this cross-layer signalling. Figure 6 presents the main blocks of the WiFi access router. The wireless access router has two separate functions. The high level mobility interface manages the mobility related tasks, while the dynamic WiFi topology adaptation block rearrange the radio transmit power and the served users according to the priority of the ongoing stream sessions.

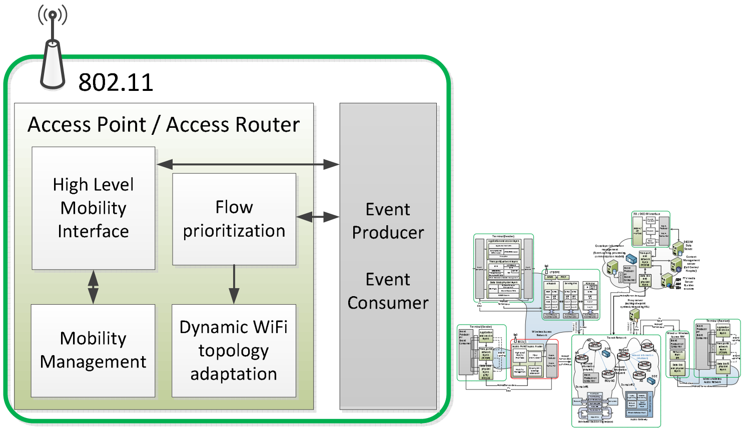

Figure 8: WiFi access network blocks

LTE is a key wireless access networking technology in the CONCERTO architecture. Different blocks of the LTE components may also take advantage of the cross layer signaling, while it is also possible that these components are event producers as well. The home agent also appears within the IP block, supporting mobility. Figure 7 shows the components and blocks of the LTE access.

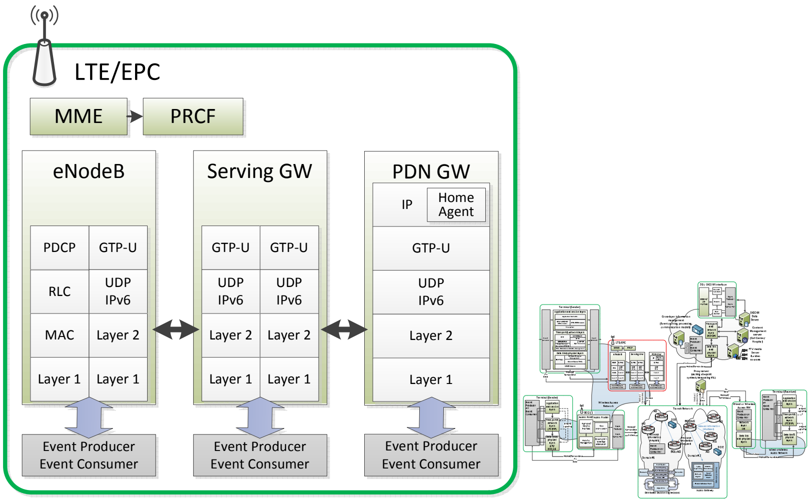

Figure 9: LTE network blocks

Further details on these blocks can be found later in this chapter.

4.1.3 Cross-layer event exchange

The cross-layer signalling is realized by cross-layer events that are produced and consumed by different system and network components residing in various layers. The cross-layer events are exchanged through the Distributed Decision Engines (DDE), which make events available for those components that require them. Basically, the DDE is a general distributed publish-subscribe message delivery system for caching and delivery of events between different entities. They can be found in the transit network, and also close to the servers. Figure 8 shows the Distributed Decision Engine.

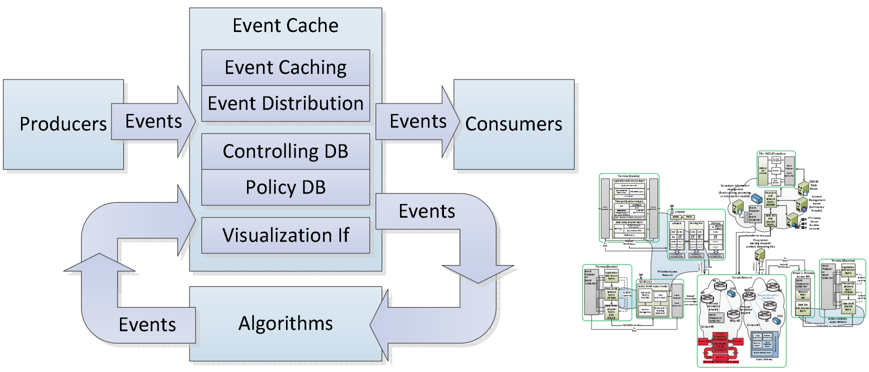

Figure 10: Distributed Decision Engine

4.2 Uses case – architecture mapping

The architecture presented in Figure 5 is sufficiently general to cover all the aforementioned use cases. For each use case, it is then necessary to map the involved actors on the correspondent blocks of the architecture. As an example, this exercise has been done for the *"Emergency areas with multiple casualties"* use-case including some functionalities identified for the *"In-hospital scenarios"*. The result is shown in Figure 11.

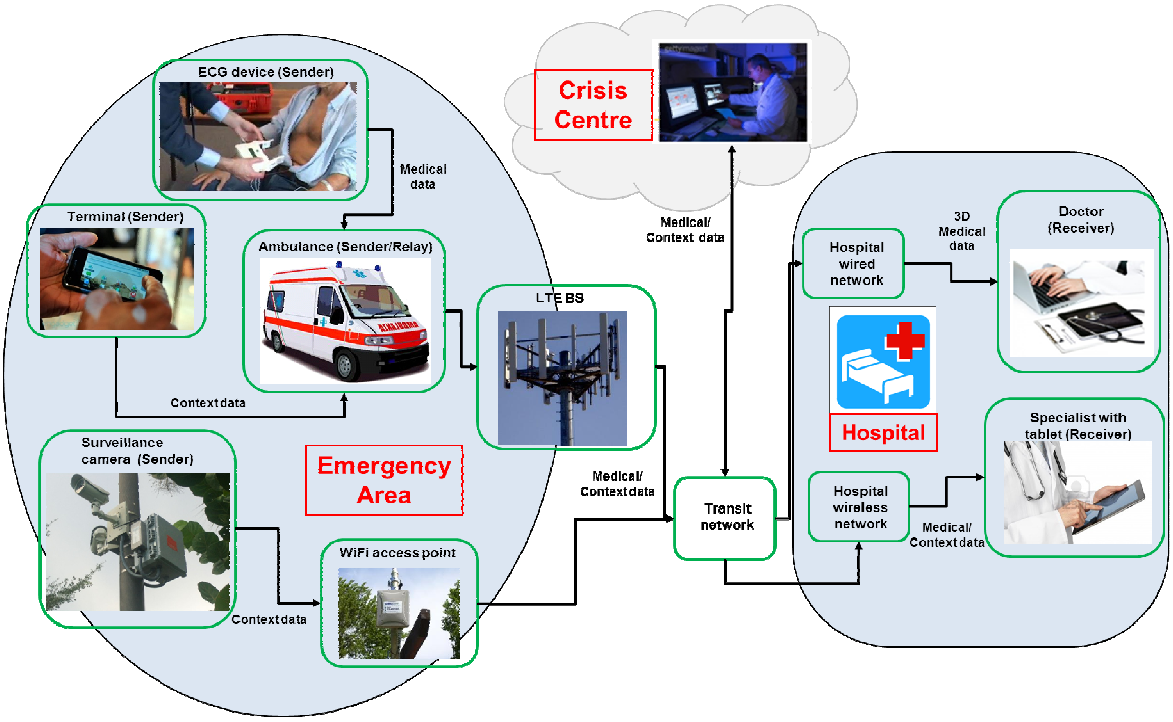

Figure 11: Use-case mapped on the CONCERTO architecture using a Public safety network

Two main areas are identified, namely the Emergency area and the Hospital. Moreover a Crisis Centre, which can be an external centre or a dedicated room inside the hospital, is also included. Basically the Emergency area contains the different sources of contents, both medical and context data. All these data can be generated by several devices, like surveillance cameras, smart phones or medical devices. The data can be transmitted through heterogeneous access networks (e.g., 802.11, LTE…) and even a relay node can help the transmission (in the figure the role of the relay is played by the ambulance). At the hospital, instead, the nodes are mainly receivers and typically correspond to doctors that can receive data from the emergency area both through the wired and the wireless networks of the hospital.

To avoid overcrowding the figure, only user data transmissions are drawn, but several fluxes of cross layer data are also present and allow the control of the different transmissions.

The Crisis Centre is responsible to coordinate all the received data and, taking into account the requests of the doctors, can prioritize different fluxes of data or ask to the senders in the emergency area to adapt their transmissions, the compression of the sent data and the kind of data transmitted. Adaptation can also be decided directly by the senders according to the cross layer information received, while not badly affecting the overall network efficiency.

Different devices can have more or less intelligence and adaptation capabilities. In the case of the presented figure, the ambulance, which relays both context and medical data, can implement more advanced adaptation techniques with respect to the smart phone or the ECG device, such as, for example, MVC encoding of data received by different sources or data protection algorithms. This also leads to an improved utilization of the network resources, as shared possibly by other critical applications.

4.3 Main functional blocks

4.3.1 Application layer

4.3.1.1 Adaptive video streaming functionalities using HTTP

As the progressive download is still one of the most popular streaming technologies, also HTTP streaming based solutions gained popularity in the past few years. The problems with caches, NATs and firewall issues in RTP/RTSP/UDP streaming are less problematic in HTTP and the evolved Internet infrastructure supports efficient usage of HTTP and Content Delivery Networks (CDN), which can also utilize HTTP streaming in video delivery to clients. Furthermore, the additional advantages in HTTP streaming include matters such as:

- No traditional streaming server is required
- Client can access and play the content instantly
- Simple setup
- Clients can seek to points which are not yet downloaded

The basic idea of HTTP streaming solutions is to split video content into segments and manifest file tells which segments are available for downloading. The client downloads and plays the short segments, usually only few seconds long, of the original stream in the order defined in the playlist. The segments are encoded to multiple different data rates. The segmented streaming solution can be used for both live and on-demand, but the main advantage lies in the capability to adapt the prevailing network bandwidth e.g. by measuring the length of client’s input buffer. For example, the client software can change the data rate based on the recent trends in measured network throughput.

Utilization of adaptive HTTP streaming for healthcare media services in wireless networks could take advantage of large support available for HTTP-based traffic in existing IP networks and could improve the overall quality of these services with adaptivity.

Adaptive HTTP streaming module will receive encoded video from the file (or directly from the encoder) and split the content into segments which are then fed to HTTP protocol module. This is the process at the sender side.

At the client side, adaptive HTTP streaming functionality is integrated at the application layer and it located between HTTP and video client (decoder). Adaptive HTTP streaming module receives data from the HTTP module and passes the data to the client in segments.

Adaptive functionalities are implemented at the client side and are mainly based on monitoring the data received. However, it is possible to improve the adaptability with additional information about the network conditions through DDE.

4.3.1.2 Cross-layer information collection points

The key components for collecting and provisioning data across multiple layers, devices, domains and technologies all over the Internet (i.e. for inter-domain signalling) are the Access Gateways (AGs), which play actually the role of collection points (CPs). The cross-layer information about network resources is to be accessible by means of AGs, locally or remotely. Resource information about the local domain (at all levels) should be collected and stored locally in databases (associated with the respective CPs of that domain. While, data related to resources of the other domains can be stored in databases associated with the respective CPs belonging to those remote domains. Extremely time-variant information can be simply notified as available to the said local CPs rather than actually delivered to them and could be provided on demand to the consumer entity. This could also apply to resource information of the remote domains, which could be simply notified to the CPs of the issued local domain.

Every AG/CP is associated with both a database and a directory, for actual information provisioning and re-direction, respectively. It is worth noting that the database of an event driven system (i.e. an event cache), used to collect and notify resource information over the access part of the network (even between different domains of the same network) can be integrated with the said AG/CP’s database, in order to have a single data repository for a given zone of the network. In CONCERTO proposal, this is operated by co-locating a DDE at the AG (refers to D2.3 for more details). For backward compatibility issue also the existing databases (e.g. for Traffic Engineering, network management, mobility provisioning) can be integrated in the architecture for cross-layer optimization by properly registering their content to the CPs of the respective local domain. CPs of a given domain can register their content to the CPs of the others, as for retrieval of information from anywhere about the whole Internet, specifying the set of indexing information (which includes the IP address) of the resources associated with the network components that they serve in the respective domain.

The goal is to efficiently obtain the needed information in relation to the delay in receiving data after a request for it, which should be as short as possible, and with respect to the employed communication model(s) that should be properly selected. To satisfy these requirements the closest CP can be reached by using an anycast address that has to be advertised beforehand all over the served network. Furthermore when a CP receives resource information by a producer entity it efficiently spread it among all the other CPs in the domain by using multicast address. If the information to retrieve is stored in the database of another domain, then the CP interrogated locally will re-direct the request to it. As already pointed out, a redirection is likely to happen as well, when issuing extremely time-variant data. In such cases, a direct communication between the consumer and the producer (or first provider when accessing a remote database) entities should be established for efficiency reasons.

Regarding the communication models, for efficient and scalable cross-layer communication, needed data should be provided on demand or when specific events happen (e.g. availability of a new value from a producer entity) in the case of subscription to a notification service (implemented by a DDE).

In practice a pool of AGs is deployed in each domain (e.g. Autonomous System) for the direct delivering of information about resources of the associated domain and the re-direction of requests for information about resources in the others.
The mentioned pool of AGs belongs to a multicast group for efficient distribution of an AG's database content to the other AGs' within that domain and an anycast address is also assigned to the said pool in order to collect resource information by the AG closest to the producer entity, as well as to provide resource information to the consumer entity by its closest AG. Therefore, every domain has associated a well-known multicast and anycast address for its pool of AGs.
The number and location of AGs In a domain (also, of DDEs) is a trade-off between several aspects, including efficiency, scalability, additional overhead, cost and complexity. The larger the number, the shorter the delay in getting the needed information, but also the higher the introduced overhead, in terms of bandwidth, storage memory and data processing.

Functionality: Cross-layer information collection and delivery

Input: Event messages of the publication/notification basis by DDE (e.g. registrations, subscriptions), (cross-layer) resource producer information (i.e. publication event messages), index information (about information available in other domains or in external databases)

Output: Event messages of the publication/notification basis by DDE (e.g. redirections), (cross-layer) resource producer information (i.e. notification event messages), index information (about information available in other domains or in external databases).

4.3.1.3 Multicast Free Viewpoint Video and Television

Free Viewpoint Video (FVV) and Free Viewpoint Television (FTV) allow users to interactively control the viewpoint and generate new views of a dynamic scene. The customers of this new service may request unique viewpoints that can lead to scalability problems, because virtual viewpoint synthesis is a resource hungry process even for sole user [77]. The streamed scene is synthetized from 2-4 real camera streams, so the original high bitrate camera streams must be delivered to the user requiring huge bandwidth resources. Besides, the independent users may continuously change the desired viewpoint requiring different set of original camera streams for the view synthesis process.
Two strategies can be distinguished for camera stream delivery:

1. Delivery of all camera streams
2. Adaptive camera stream delivery

In the first case, all original camera streams are forwarded to each user. This strategy can be accepted only if low bitrate camera streams are used and the link capacities are high enough to forward all of the streams. Although forwarding all the streams seems very wasteful, it has also some advantages, too. If the user changes his virtual viewpoint rapidly the all the necessary original camera streams will be locally available and the streaming service will be able to handle the viewpoint changes seamlessly.

The second case may provide an optimized delivery of high bitrate camera streams forwarding only the needed streams for the viewpoint synthesis. Of course a control mechanism deployed at the client is necessary to decide which streams are needed. The determination of the required camera stream set is a possible investigation topic, because it is not obvious which camera set will provide seamless viewpoint changes with the lowest bandwidth load.

Multicast delivery of camera streams is an advantageous solution to reduce the required bandwidth in both cases. Multicasting is an effective solution if high number of client must be served with the same content. Even through FTV clients may require different viewpoints, thus different sets of original camera streams, these sets may have common elements.

Using multicast for the delivery of high bitrate camera streams will minimize the required bandwidth, but the timing of join and leave messages is critical. The multicast group changes are analogous to the mobile handover process, because in both cases the change must be performed seamlessly, without interrupting the service. The client must join to the new group before the viewpoint generator will require the camera stream. The camera handover process can be optimized by predicting the viewpoint shifting.

We propose an approach to dynamically handle the camera stream handovers without interrupting the synthetized user stream. To prevent the user's viewpoint synthesizer algorithm from remaining without camera stream source, multicast group join threshold can be introduced in order to provide all camera streams that may be requested in the near future.
In the scenario, depicted in Figure 12, the viewpoint of a user can be freely changed within the zone, requesting other camera streams. Supposing that the viewpoint of *Client 1* is moving from the *purple* camera towards the *yellow* one, it will reach *Threshold 1* initiating a multicast joint message to the *yellow* camera stream group. While the viewpoint of *Client 1* is within the threshold zone, he will be member of three multicast groups (*purple*, *green* and *yellow*).If the viewpoint is moving henceforward towards the *yellow* camera position and reaches the *Threshold 2*, the client should leave the *purple* multicast group.

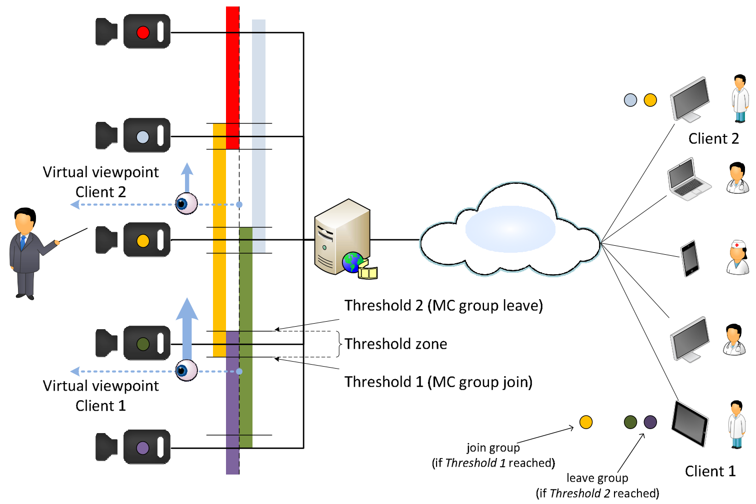

Figure 12: Multicast FTV: Multicast group join thresholds

However, Figure 10 shows a linear camera setup (1 dimensional camera topology), the cameras can be deployed in plane (2D) as well as in space (3D). In the latter cases not only two camera streams are required for the viewpoint synthesis, but three or even four that makes the threshold area determination more difficult. Our goal is to keep the threshold as low as possible to reduce the number of multicast group memberships, so the overall bandwidth, but keep it large enough to avoid playout interruption during viewpoint changes.

From architectural point of view, the proposed solution will require multicast support in the network layer. The generally used PIM-SM [78] or PIM-DM [79] protocols are applicable for the presented FTV service without any modification. Using PIM-SM rendezvous point (RP) and routers with multicast support are necessary elements of the network, while the control of group management packages must be done in the application layer.

4.3.1.3.1 FVV and FTV cross-layer signalling

New type of video streaming may appear in the future networks, called Free Viewpoint Video (FVV) and Free Viewpoint Television (FTV). The customers of these interactive multimedia services may control the viewpoint and generate new views of a dynamic scene. The uniquely generated and displayed scenes are composed from several high bitrate camera streams that must be delivered from the cameras to the viewpoint synthesis algorithm that can be deployed at the media server, at the client or distributed between proxy servers.

In order to prevent the user's viewpoint synthesizer algorithm from remaining without camera stream source even in case of fast viewpoint changes, the neighbouring camera streams should be also forwarded to the algorithm. To provide seamless camera handovers the most reliable solution is to forward all camera streams to the client, proxy server, or wherever the viewpoint synthesis is performed. If the network capacity makes it not possible to deliver these streams, only a set of camera streams can be transmitted (**Erreur ! Source du renvoi introuvable.**).

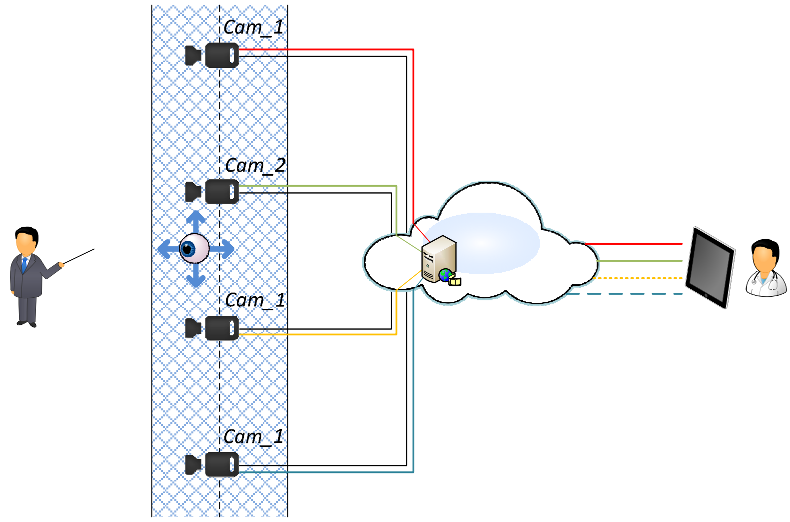

Figure 13: Adaptive camera stream delivery

The available bandwidth cross-layer information could help to determine the number of camera streams in the set. Forwarding more camera streams will reduce the chance of service interruption, but will generate more traffic load. If less camera stream is delivered, the probability of seamless viewpoint changes will be lower. Our goal is to find optima, where the congestion in the network will be avoided and the seamless viewpoint changes can be guaranteed width high probability. The bandwidth information is indispensable to control which camera streams can be delivered. Of course the selection of camera streams can be further optimized e.g., by predicting the viewpoint movement.

Functionality: Viewpoint prediction based multicast group management that reduces network load and minimizes viewpoint synthesis interruption possibility.

Input: (for each client) multicast camera streams, actual viewpoint (creating viewpoint history database)
General input: threshold values, camera positions

Output: (at each client) estimated future viewpoint, multicast group requests, client specific virtual stream

4.3.1.4 Media caching and remote viewpoint synthesis for FVV and FTV streaming

Free Viewpoint Video (FVV) and Free Viewpoint Television (FTV) are new approaches of multimedia services in order to improve the visual experience of the customers. The novelty of these new multimedia streaming solutions is that the users can interactively control the viewpoint and generate new views of a dynamic scene from any 3D position.
Real 3D video, FVV and FTV caching requires significantly higher storage capacity and computational resources compared to 2D streaming solutions, because the forwarded and displayed scene is composed from several high bitrate camera streams. The input of the viewpoint synthesizer algorithm is two, three or even four camera streams that must be delivered to the network element, where the viewpoint generation is done. The algorithm can be deployed at the media server, at the clients, or distributed in the network, although the delivery of the original camera streams is not always possible due to the limited link capacities. To solve the problem two different approaches are considered. The first one is remote rendering and view synthesis (centralized or distributed solution), while the second approach is to adaptively deliver only the required camera streams (unicast or multicast mode).

In case of client based synthesis, the camera streams - requested for the viewpoint synthesis - are delivered to the clients generating huge amount of data traffic. This approach utilizes the computational resources of the users' equipment. Using pure multicast delivery or extended with adaptive camera selection, the traffic load can be reduced.

The other method, named server based approach, uses the computational resources of the centralized media server to generate the desired viewpoint for each user and forward it with the requested bitrate. The server based solution will generate unique video streams to each customer with the requested bitrate, resolution, frame rate, etc., however control messages will be needed from user to server to set the required coding parameters and the desired viewpoint. Unfortunately, in this centralized scenario scalability problem may arise. If too many clients are served by the server, its computational capacity may not be adequate to generate all the required streams with unique viewpoints.

We propose a novel scheme to solve the scalability problem and keep the traffic load as low as possible. The idea is to distribute the viewpoint synthesis processes in the network by allowing network nodes to act as proxy servers with caching and viewpoint synthesis functionalities. These proxy servers may share their resources for viewpoint synthesis, recoding and caching purposes. Therefore, the user must not connect directly to the media server, but may ask the closest proxy server for a stream with the desired viewpoint. The proxy servers will gather the camera streams that are needed to serve the connected clients and originate the unique streams.

Basically a proxy server may cache the segments of a video stream, but in case of 3D services it is preferred to support codec functionalities and viewpoint synthesis, so we modelled the proxy element as Figure 14 shows.

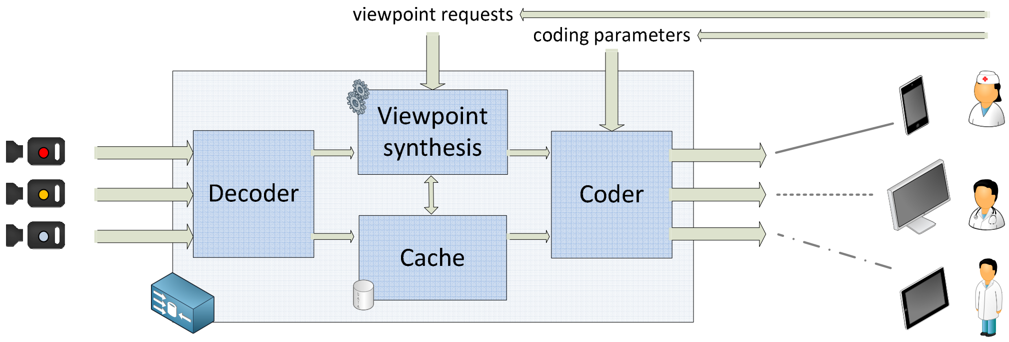

Figure 14: Cache and proxy building blocks

The received camera streams (all or only a set of streams) are processed by the decoder module and stored in the cache. Based on the incoming viewpoint requests, the viewpoint synthesis module will generate the new stream that will be coded according to users' coding setups.

It is an interesting research topic how to localize the proxy servers in order to generate the lowest traffic load, but avoid the overload of computational and storage resources. In order to provide seamless viewpoint changes, the set of required camera stream must be available at the proxy server. A strategy must be also developed to prevent the absence of a camera stream that can not be missed by the viewpoint synthesis module.

Functionality: Distributed viewpoint synthesis as a scalable solution

Input: Synthetized video stream (for each client), camera streams, camera positions, requested client viewpoint coordinates and resolutions (coding parameters) (for proxy servers)

Output: (at each client) viewpoint coordinates and resolution (coding parameters)

4.3.1.5 Delay-Conscious Caching Scheme

Video caching is an admitted method for media streaming, because it smooths the delay and bandwidth variation during the streaming. In case of transcoding of video streams or arbitrary viewpoint synthesis in case of FTV and FVV the deployment of cache is essential part of a proxy server. Media caching plays an important role in today's streaming services as well. Recent studies have shown that video streaming is responsible for 25-40% of the Internet traffic [62], but according to the forecasts [63], two-thirds of the world's mobile data traffic will be video by 2017. Of course, energy-efficiency will play a significant role in this improvement, because mobile devices typically have a limited energy budget [64], so energy efficient methods must be investigated to control the power usage of mobile devices. Media caching can be also used for energy reduction purposes, when the network interface is periodically switched on and off in order to reduce power consumption.

We propose an energy efficient caching scheme that keeps the balance between the power consumption and the acceptable delay in case of *ON*/*OFF* energy reduction scheme. Utilizing the relation between the adopted playout delay and the power consumption, the aim is to give guidelines for the playout delay determination in order to keep the delay as low as possible, but also reduce the consumed energy.

The interfaces are designed to handle different states in order to decrease the power consumption [65]. It can be described with three states: active (*ON*), idle and sleeping (*OFF*). While the network interface is down, the video stream is played from the playout cache and stored in caching servers meantime. Due to the continuous playout, the caching scheme will be effective only if the downlink bandwidth (Bw_d) is higher than the video bitrate (λ_v), because during the active (*ON*) period (t_{ON}) the cache can be filled only if $\lambda_v < Bw_d$ (Figure 12)

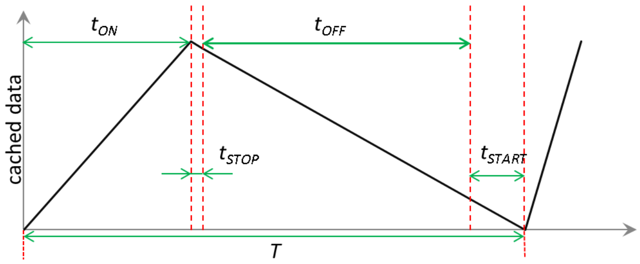

Figure 15: The caching and rendering proxy model

During the interface state changes (t_{STOP}, t_{START}) and the *OFF* period, the reception of stream packets is not possible. We propose to use control messages to inform the media or caching server that the client's interface is up and ready to receive the next portion of the video stream. Similarly, the server must inform the client that no more packets will be sent within the actual cycle and it can switch off its interface.

Of course each state has its own energy consumption, but the energy consumption can be reduced by increasing the cache size. Longer caching delay makes it possible to keep the interface longer in offline state. Unfortunately, the delay has significant impact on the experimented quality of the user, indeed the prolonged playout delay is not even acceptable for live video streaming. We have to investigate the relation between the consumption and playout cache size in order to find the optimum.

Functionality: Energy efficient video streaming

Input: Video stream and its parameters (bitrate), buffer size, acceptable delay

Output: ON/OFF times

4.3.1.6 Application controller

The application controller is the block that allows the adaptation of the videos to transmit to the real conditions of the wireless channel and of the network. It is the intelligent part of the encoding process and it is in charge to satisfy the requirements in terms of video quality and losses. The application controller may interact with the management server within the coordination server, receiving control information about the main transmission requirements and providing feedback about the implemented adaptation strategy.

It uses cross layer information - like expected throughput and loss probability - retrieved from the DDE cache and generated by different entities to compute the best compression parameters for the encoder. Moreover it controls the FEC by establishing the protection needed by the videos to satisfy the application requirements in terms of losses.

Functionality: Drive encoder and protection

Input: DDE events (available bandwidth, user preferences…)

Output: Encoding and protection parameters

4.3.1.7 Encoder/Decoder

According to the different use cases presented in deliverable D2.1, different video contents and images have to be transmitted. These multiple contents need to be encoded at the transmitter and decoded at the receiver. Videos and images can be originated by multiple devices with different characteristics and can require different kind of compression. For this reason the encoder/decoder is a key block of the CONCERTO architecture application layer. Following the different needs of the different use cases, this block has to be adapted to follow different standards. In particular it can follow H.264/SVC standard, H.264/AVC standard (including annex H for MVC video streams), JPEG standard and forthcoming HEVC standard which is still under development.

The cross layer signalling information retrieved through the DDE engine and elaborated by the application controller is fundamental for this module since it will drive the compression level of the videos.

The encoder part can receive raw videos (or images) and have to generate the frames to transmit according to a set of compression parameters provided by the application controller.

The decoder part receives from lower layers encoded videos and images and have to decode them in order to provide to the end user videos and images in the required format (i.e. raw videos and images or encoded ones).

Functionality: Encode/decode videos

Input: Raw videos to encode/encoded video to decode; encoding parameters

Output: Encoded/decoded videos

4.3.1.8 Selective encryption of 3D medical image and video data

In 2012, the state of the art symmetric key encryption algorithm is still considered to be the AES (Advanced Encryption Standard). The selective image encryption ciphers only a small fragment of the full medical image data (Figure 13) [75], [76]. There is only a few publications studying the medical applications of selective encryption, therefore CONCERTO has to analyse the existing algorithms, in case of need to improve them or develop new ones. Not all the visual information parts of the medical image should be protected equally strongly, there could be parts of the picture which are unimportant and displaying them does not risk the patients' protection.

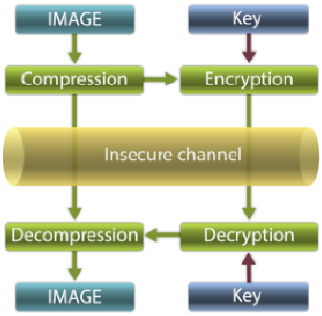

Figure 16: Mechanism of selective encryption

In real-time transmission, there are basically two strategies: whether compression takes place before encryption or vice versa. Compression applied first reduces the size data but it offers less secrecy, in addition if encryption is applied first, compression will be ineffective. Using selective encryption, image is at first compressed and then only parts of the compressed data are encrypted (Figure 14).

Figure 17: Original and selective encrypted picture

In order to implement this principle we split data into several slices. Some of the slices are left unmodified, while other slices are processed by encryption blocks. Slices are encrypted with different keys, all encryption blocks could be different. A possible architecture to be implemented is shown in Figure 15.

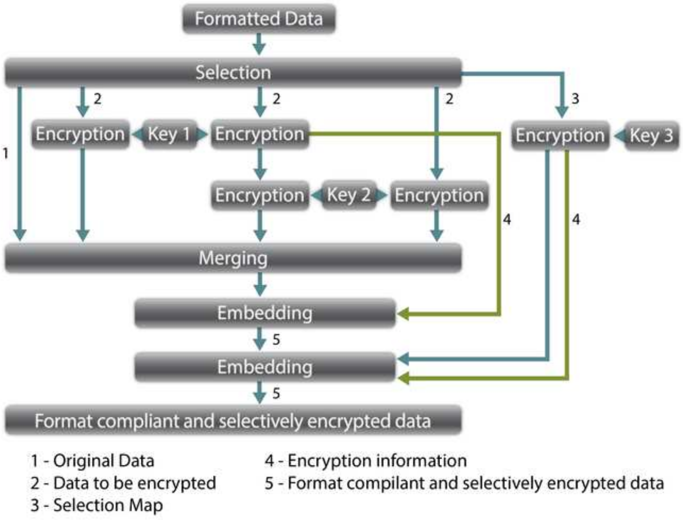

Figure 18: Selective encryption architecture

Using selective encryption with multiple keys, it is possible to provide difference access level to the encrypted media. Owning all the keys gives full access to the encrypted media, while with fewer keys the access is restricted, but still can be useful for certain purposes (e.g. learning).

Selective encryption can be useful during the transmission of video sequences as well. Initial studies within this project proved that mobile devices have enough hardware support to decrypt fully encrypted video sequences within a normal bandwidth range. Fully encrypted, 8 Mbps CBR coded video sequences were smoothly played out on mid level mobile tablets. The study also revealed that even there is the cryptographic hardware support, there are scenarios where selective encryption is still preferable. These scenarios are:

- Less decryption obviously saves on battery power. The selective encryption is a green solution.
- It is possible that the server side have resource shortage. In this case the selective encryption should be applied.

There are scenarios where the video sequence should not be fully secret. In commercial cases usually preview functionality is provided to attract positional customers. In the current medical scenarios it is possible to give out a less detailed video sequence for educational purposes, while keep the full details version in secret.

During the design of the selective encryption, one must take special care of the security level degradation. As in this case not all the bits are encrypted, there is a possibility to recover some parts of the video sequence based on the available unencrypted information. This could be a serious challenge in commercial use, however the task is somewhat easier in the case of medical use, as most of the cases only the fully detailed sequences considered useful. Still, the security of the selected selective encryption method should be studied thoroughly.

Functionality: Selective encryption of medical data

Input: Plain (non-encrypted) medical (CT, MR, Ultrasound) data files to be transmitted or stored, and ROI selection for sensitive data

Output: Selective encrypted medical data files, where only the selected sensitive data are encrypted

4.3.1.9 Encryption of non medical video data

Many of the envisioned scenarios contain video transmissions that are not transmitting medical data, but they serve as a side information channel where the events can be discussed, descriptions and instructions can be given. Some scenarios require the protection of these video transmissions as well. Initial studies carried out in this project [109] revealed that using current mobile devices, the HTTP video transmission can be switched to HTTPS video transmission without any degradation to the video quality. The analysis showed that the HTTPS streaming has better performance than it IPSec alternative. Also, HTTPS streaming adds only a small overhead to the network utilization in contrast to IPSec, where this overhead is more significant.

The analysis shows that HTTPS streaming, with the state of the art AES cipher and a 256 bit key is available on the mobile devices and is considered as highly secure. In the analysis the largest video bitrate was 8 Mbps, however the analysis also showed that the maximum achievable bitrate is the question of the networking support rather than the cryptographic capabilities.

4.3.1.10 Medical image coding and storage

The DICOM (Digital Imaging and Communications in Medicine) standard generally is used for the exchange of images and related information (Figure 16). The DICOM standard has several levels of support, such as image exchange between senders and receivers, underlying information model description and information management services [66].

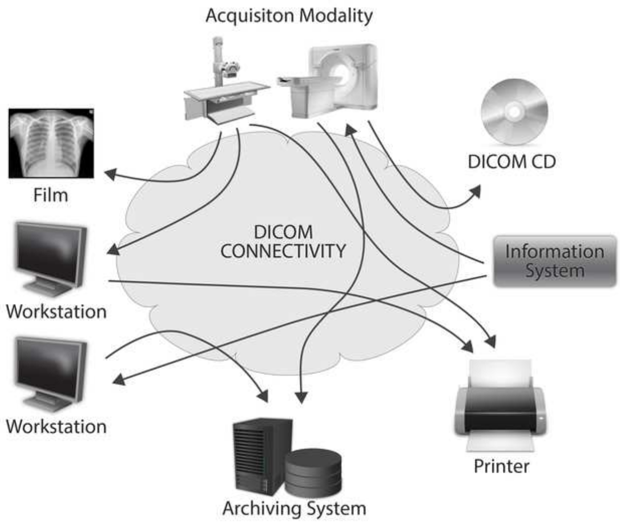

Figure 19: DICOM connections

The scope of medical imaging is wide, from a digital image of a bone X-ray (a single frame), to a set of tomographic slices through all of the body. Digital medical imaging devices generate large amounts of volumetric data sets. Advancements in digital radiology have led to a significant increase of the sample bit-depth and resolution of data sets, causing accordingly grow in size. Besides digitization of the acquisition devices, the focus of research is also shifting towards applications, transport networks and storage systems. This gives origin to new requirements in efficient transmission, storage, long-term archival and interoperability between applications in a multi-vendor environment. Both efficient storage and efficient transmission of large medical data sets request efficient compression techniques. In order to enable rapid diagnosis at a remote location, the delivery of medical data is often done by wireless network.

The JPEG 2000 standard defines a new image-coding scheme using wavelet-analysis based compression techniques [67] (Figure 17, right). The JPEG 2000 architecture is useful for many diverse applications, including Internet image distribution, and medical imaging (Figure 17, left).

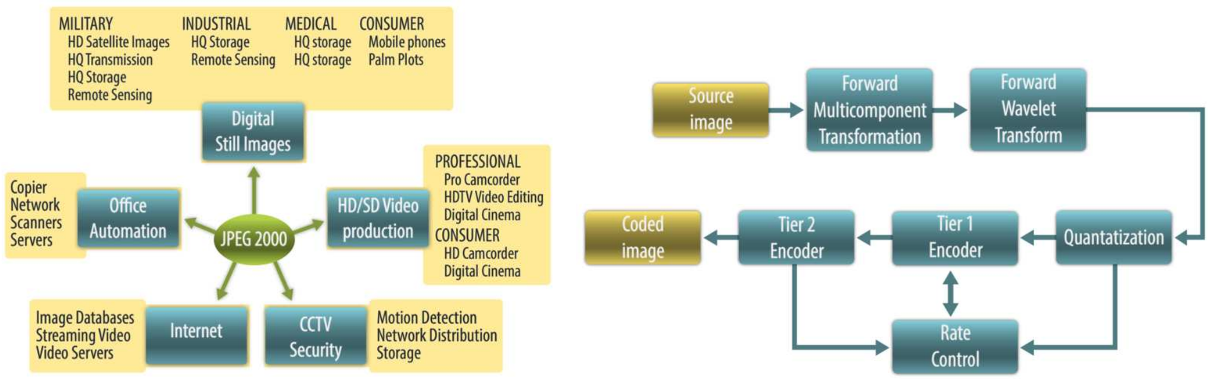

Figure 20: JPEG 2000 Applications (left) and the architecture of the JPEG2000 Compressor (right)

Version 3.0 adopted existing image compression standards from the JPEG working group. Due to the promising results of wavelet-based compression methods, DICOM adopted JPEG 2000 Part 1, so the standard is able to offer efficient compression, storage and transmission of two-dimensional medical images. Additionally, JPEG 2000 with embedded block coder by optimized truncation (EBCOT) technology delivers excellent quality and bit-rate scalability [68].

In 2005 DICOM adopted the Multiple Component Transformations (MCT) extension of JPEG 2000 Part 2 in order to improve the compression performance of medical images by exploiting multiple spectral components per image slice. Arbitrary lossy or lossless transformations on all components of an image are allowed with this extension.

Improved compression of volumetric images is also possible, by converting slices of a volumetric data set into virtual components, which allows to perform a wavelet transformation along the slice axis [69]. This approach has a significant drawback: it is unable to treat all dimensions in an isotropic way, so it will negatively affect the functionality and rate-distortion performance of the image codec [70].

In 2008 the JPEG 2000 part 10 (JP3D) was created in order to support volumetric images with optimal compression. The JP3D properly extends the wavelet transformation and the entropy coding to three dimensions, so it is able to reach better compression results compared to using Part 1 or the Part 2 MCT. The JP3D handles volumetric data sets in an isotropic way, so swapping any of the three dimension axis makes no difference in compression efficiency or coding limitations [71].

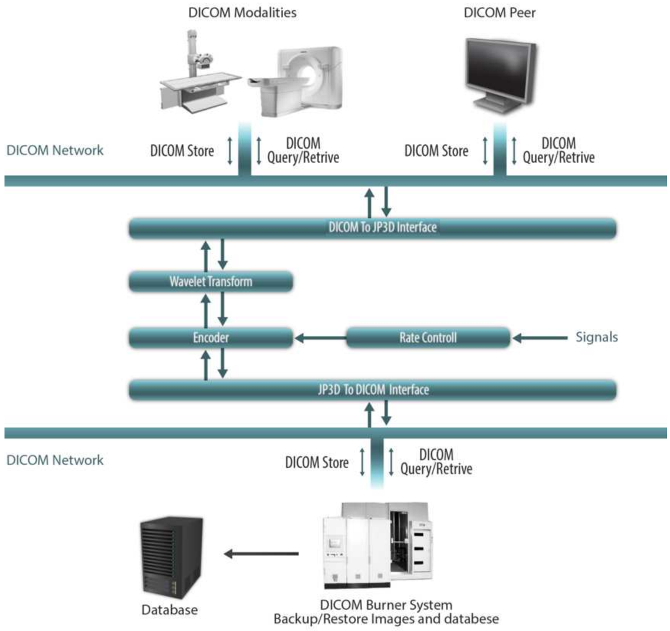

Figure 21: The proposed DICOM-JP3D interface

For near-lossless to lossy compression, test results show improvements of JP3D over 2D based compression of up to 15% at high bit-rates, and up to 50% at lower bit-rates [72]. In spite of this, even the JPEG 2000 Part 1 2D methodology is still considered state-of-the-art image compression technology. In practice, simple schemes are often employed in lossless coding of medical data. Therefore, the application of JP3D with the DICOM format will be addressed. CONCERTO aims to find solutions integrating JP3D and networking technologies with DICOM.

While in everyday use the inconvenience of having to download new software (or a new codec) to view a picture or video is generally acceptable, such untrustworthy operation is absolutely unacceptable in the practice of medicine.

Absolute accuracy is required, and both equipment manufacturers and medical staff exercise stringent control of their system configurations and make extensive use of standards. The DICOM standard has been used extensively for this ambition since it was released.

It defines a limited set of compression schemes that may be used both on the network and on offline interchange media [73]. DICOM's one main disadvantage is, that it allows a very limited set of compression methods that can be used both on the network and on offline interchange media [74].

Our aim and current work is to design and develop an interface to apply the state-of-the art 3D compression technologies in DICOM with keeping the transparency of DICOM networks (Figure 18).

Functionality: Efficient (state-of-the art techniques based) coding of large 3D medical data sets for DICOM environments.

Input: The original medical data (provided by vendor equipment) in DICOM format.

Output: Coded (with JP3D) medical data files for efficient store or transmit.

4.3.2 Transport layer

Concerning the transport layer, different solutions are adopted in the CONCERTO architecture, based on both TCP and UDP/UDP-Lite protocols. TCP is used when streaming over HTTP is adopted for media delivery, offering the advantages of a reliable connection-oriented link at the transport layer. The most promising solution for media streaming over TCP is the dynamic adaptive streaming over HTTP, also known as DASH or MPEG-DASH, whose pros and cons in the scope of CONCERTO have been evidenced in the paragraph on the application layer. Here we remark that the suitability of TCP for media delivery depends on the specific application requirements. When real-time communications are addressed and interactivity with a remote specialist is required, TCP-based streaming may not be the best solution, due to the retransmission mechanism on packet losses that may cause unacceptable latency and jitter. For this reason the CONCERTO architecture takes into account the possible adoption of non-connection oriented protocols, such as UDP/UDP-Lite, in real-time communication contexts in conjunction with RTP streaming at the application layer.

4.3.2.1 Packet-level forward error correction

When unreliable transport protocols are adopted, no guarantees are provided on the correct reception of the transmitted packets, since retransmission mechanisms typically supported at the transport layer are not applied. Traditionally, real-time streaming applications have privileged low-delays and speed on reliability. On the contrary, the MSHTM and safety scenarios addressed in CONCERTO are characterized by strict requirements in terms of both tolerable delays and packet loss rate, due to the sensitive nature of the transmitted information. For this reason, the transport layer in the CONCERTO architecture is enriched with packet-level forward error correction (PL-FEC) functionalities. At the transmitter side, a PL-FEC encoder organizes the UDP/UDP-Lite streams in proper structures called Source Blocks (SBs). By encoding the SBs, a given number of repair packets containing redundant information are generated and transmitted to the receiver in conjunction with the source packets. Dually, at the receiver side, a decoder module reconstructs the original SBs, recovering possible information losses by processing the received packets. In the CONCERTO architecture, LDPC-based PL-FEC solutions are considered at the transport layer. Non-binary LDPC codes are employed, with advanced decoding procedures, based on maximum likelihood (ML) or iterative algorithms. The co-decoding modules are capable to dynamically change the level of protection applied to the information streams, based on the controls messages received by the intelligent modules operating at the application layer.

The PL-FEC technique addressed in CONCERTO extends the packet error correction (PEC) solution developed in OPTMIX project in two directions:

- Powerful non-binary LDPC codes and co-decoding algorithms are employed to further improve the erasure recovery capabilities,
- Advanced decoding algorithms to correct possible undetected bit errors in the received packets are introduced to improve the communication reliability, which is extremely important when critical information about patient conditions are transmitted across the network.

Considering the CONCERTO architecture, PL-FEC co-decoding modules are employed across several links of the chain. In general, this kind of protection scheme is very flexible and can be applied over the entire communication, i.e. encoded by the source and decoded by the final sink of information, or just between the entities communicating across the most critical links. In the project we mainly focus on the first approach, protecting the entire communication, but other solutions may be evaluated for comparison, based on the considered scenario. For example, in the scenarios involving communications between an ambulance, which can be moving or stationary in an emergency area, and a remote hospital, the most critical link is probably constituted by the ambulance wireless access through an LTE public network and, as a consequence, PL-FEC schemes can be applied between the ambulance and the remote coordination center.

Functionality: Protection of information flows against packet losses and, possibly, residual errors within the packets.

Input encoder: RTP/UDP packets from application layer; protection level (code rate); size of the encoding source block, that, based on the source data rate, affects the maximum decoding delay; destination IP address
Input decoder: UDP source and parity packets; in-band control information necessary for decoding

Output encoder: UDP source and parity packets; in-band control information necessary for decoding
Output decoder: RTP/UDP packets to the application layer

4.3.3 Network layer

4.3.3.1 IP content-aware cross-layer scheduler

The work carried out in WP3 of FP7 OPTIMIX project [44] [45] can support a proportional delay differentiation model over a DiffServ architecture. Therefore, it is able to provide a delay differentiation between service classes at IP layer according to the mutual ratio of quality factors assigned to each queue. More specifically, if we have 4 classes of service (i.e. queues) with assigned quality factors 1,2,3 and 4 respectively for the 4 classes, the delay experienced by packets in the queue 2 is about half the delay experienced by packets in the queue 1 and in turn, twice the delay as experienced by packets in the queue 4. While, the absolute value of the delays depends on the average link load, profile of input traffic in each queue and other aspects.

The underlying hypothesis, which is the basis of IP-oriented architectural models for QoS provisioning, is that the lower layers are transparent, in other words, the bottleneck is at IP network level. This is true indeed, when wired transmission technologies (e.g. optical links) are used, but it is not verified in general when we are dealing with air interfaces. This is particularly verified in the access part of the wireless networks where the radio channel can be narrowband and highly time-variant in characteristics and hence performance. Therefore, in CONCERTO use-cases the QoS provisioning in the emergency areas needs to be properly enhanced and realized.

To cope with such an issue, a cross-layer scheduler at IP layer can be deployed. It should be able to compensate somehow the possibly highly dynamic behaviour of the MAC and PHY layers, by supporting a Proportional Differentiation Model (PDM) for QoS considering the first three layers of the protocol stack on the whole. In other words, a delay differentiation between classes is to be provided considering the cumulative latency experienced by packets in crossing the IP, MAC and PHY layers of the protocol stack at the issued interface. This objective should be achieved not only addressing reliability, robustness and scalability, but also as much as possible in a backward compatible way with already deployed technologies, mechanisms, policies and algorithms. These are general requirements and objectives of CONCERTO project in relation to the support of interactive multimedia healthcare applications, issuing QoS and to be easily/quickly deployed in the real-world, also over wireless and on the move.

Specifically, the IP cross-layer scheduler should be flexible enough to work in conjunction with existing and large variety of MAC schedulers and queuing management policies. Though, a subset of schedulers and queuing policies are likely to allow for better performance of the to be designed component. Furthermore, it could be deployed in some critical points of the network only (e.g. wireless interfaces of the access network) to accomplish backward compatibility.

Service differentiation should be applied according to the nature and content of the incoming traffic. Therefore, video flows should receive a proper treatment that could be associated with the level of importance of the delivered piece of information, for example. This leads to the design of a content-aware scheduler that is able to manage differently the packets of the same service class or even traffic flow, depending on their relative importance, such as the type of frame in a video or the type of layer in a scalable video coded stream [42][43].

Actually, the 2D or even 3D images and videos generated by CONCERTO's applications can require different and possibly stringent delay and loss guarantees, to be achieved even in highly loaded network path and with poor performance of the concerned wireless interfaces. Yet in critical conditions, mechanisms as proper scheduling disciplines should be deviced in order to provide the needed level of service.

An easier deployment of the designed solution can be achieved by working at IP layer only. In this way, a software update is enough for the introduction of the novel IP scheduler. Though, a cross-layer communication with lower layers is required. More precisely, feedbacks coming from MAC and PHY layers should provide information about the actual status of the related transmission interface (e.g. run-time performance), which we can refer as a wireless one, being the most critical case. Therefore, the deployment is needed at the wireless interfaces of the base-station(s), access point(s) and user terminal(s) in the access network (e.g. emergency area).

The IP content-aware cross-layer scheduler should allow for an easy tuning of its configuration and operating parameters. Indeed, tuning at run-time the quality factors or modifying the packet priority calculation process can result in the assignment of packet priorities aiming at supporting a proportional delay differentiation at the interface on the hole, by compensating accordingly the current delays at the lower layers. A modified version of the Advanced Waiting time Priority (AWTP) scheduler [45] is proposed. In principle, the priority of an IP packet should progressively increase when the packets in the same class of service are currently experiencing higher delays at the lower layers.

It is important to underline that a mapping between the traffic aggregates at the IP and lower layers should be defined because, a compensation of the delays for a given traffic is possible only if consistent delay feedbacks are provided. For the purpose, a suitable option is to leverage the AF PHB [47] as available in Differentiated Services architecture [46] (De Facto standard for NGNs). It can be used for in-band signalling across IP, MAC and PHY layers for consistent management (i.e. classification, queuing and scheduling) [48].

Functionality: QoS support in terms of proportional delay differentiation between service classes according to pre-assigned quality factors (i.e. support of a PDM)

Input: IP packets, DSCP of AF PHB as for DiffServ architecture in-band signalling (service class binding and content-awareness), delay feedbacks (from lower-layers)

Output: IP packets, DSCP of AF PHB as for DiffServ architecture in-band signalling (service class binding and content-awareness)

4.3.3.2 Distributed and Dynamic Mobility Management

The conventional Mobile IPv6 framework lacks of reliability, scalability and stability, due to the single Home Agent representing a single point of failure in the architecture. The high requirements of diagnostically useful multimedia transmission techniques and the thriving traffic demands pose serious research challenges for mobility architectures of mHealth. As a first alternative for eliminating this centralized way of operation researches started to implement core-level distribution procedures: anchors are distributed but still remain in the core network. A good example to this is the Global HA to HA protocol (GHAHA) [55], which extends MIPv6 in order to remove its link layer dependencies on the Home Link and distribute the anchors at the scale of the Internet. A second alternative is when mobility functions are distributed in the backhaul and access part of the network. The multi-level system of Hierarchical Mobile IPv6 (HMIPv6) [56] defines regions, in which the movement does not need binding at the Home Agent counter to the inter-region movement. It relieves the HA from the load of signalling, but it could be effective only for short-term sessions, or localized movements. A third type of distribution scenarios is the so-called host-level (peer-to-peer) distributed mobility management where once the correspondent node is found, communicating peers can directly exchange packets. MIPv6 also uses this direction when it bypasses the HA thanks to its route optimization mechanisms (e.g. [57]). Another class of distributed mobility management is based on the capability to turn off mobility signalling when such mechanisms are not needed. The so-called dynamic mobility management schemes (like [58]) dynamically execute mobility functions only for cases when Mobile Nodes (MN) are actually subjected to handover events and higher layers require address continuity. These architectures are session-based contrarily to the conventional mobility architectures, which provide the same mobility features for all of the sessions.

In order to have overcome the limitations of the standard solution and provide a scalable, generic, secure, transparent and widely useable mobility architecture for mHealth services, we propose an appropriate and comprehensive integration of existing Mobile IPv6 building blocks to create an IPv6-based, implementation-ready solution for the emerging mobility issues of current architectures. The scheme focuses on the context- and content-aware operations, cross-layer optimization and integration of standard components, and increases the reliability, scalability and stability of the central network thanks to the flexible multi-level distributed and dynamic mobility architecture depicted in the following figure:

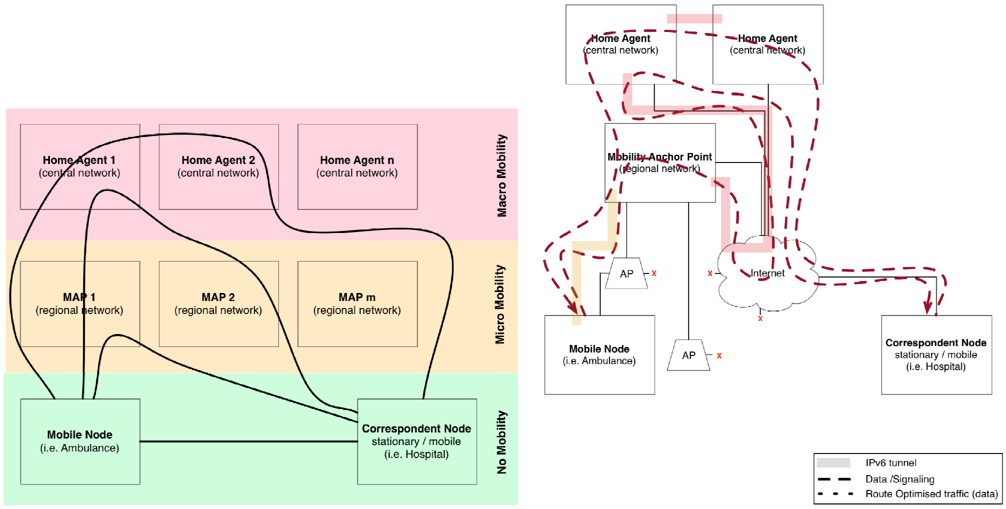

Figure 22: The proposed multi-level architecture for distributed and dynamic mobility management in CONCERTO

The simplest scenario, called No Mobility, ignores mobility support. In this case the Mobile Node communicates via its current Care-of Address. It suffers from the drawback pointed before, but it may provide higher bandwidth, and faster connection then the others. Next, we have divided the network architecture to two levels: Micro Mobility and Macro Mobility. Specified group of access points or points of attachments are grouped into regional networks, operated by 1-1 Mobility Anchor Point (MAP). The traffic inside these groups does not have to be forwarded via the Home Agent; it is forwarded by the MAP. Additionally, the movement inside a Regional Network, is handled only by the MAP, instead of the HA. Finally, the traffic between regional networks is forwarded by the Home Agent(s).

Hierarchical Mobile IPv6 introduces a middle-layer into the conventional Mobile IPv6 architecture. The movement and the data traffic inside the regional network, operated by the same Mobility Anchor Point, are hidden from the Home Agent, which decreases its load. The other major issue, the single Home Agent could to be handled by the multiplication of it. Global HA-HA defines a synchronization mechanism between the Home Agents. It makes possible to work on the same Binding Cache on all of the HAs. The Primary Home Agent will be the topologically closest one. If the Mobile Node and the Correspondent Node reaches different Home Agents, the two of them are forwarding the packets to each other. A content and context-driven, appropriate source address selection [61] is used to optimally choose between the different levels.

To make the traffic flow optimal we should apply Route Optimization as pointed before. Unfortunately the original RO specification is quite slow, it needs a long time for binding, before data communication. The Enhanced Route Optimization [57] makes the correspondent binding faster and makes the Route Optimization traffic faster. Multiple Care-of Addresses Registration [59] extends the Mobile IPv6 scheme with the ability to handle multiple parallel connections, with the same Home Address. In addition we need the Flow Bindings [60] protocol extension to load-balancing between the existing connections.

We believe that a comprehensive integration of the Mobile IPv6 building blocks introduced above could efficiently solve all the emerging mobility issues of current architectures. However, carefully designed cross-layer integration of protocols is essential. Our solution for this problem is depicted in the following figure:

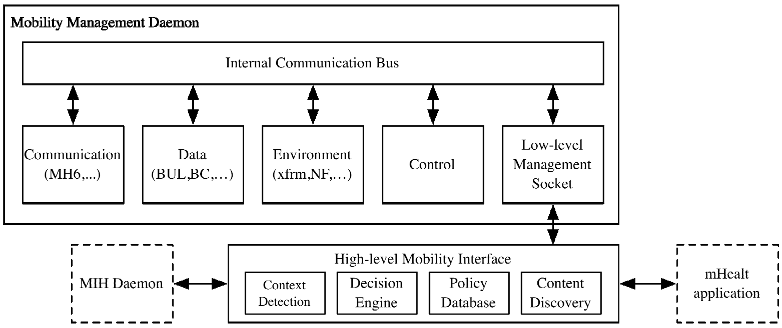

Figure 23: Integration of mobility management components

The Communication part sends and receives the MIPv6 signalling messages. In addition it handles the necessary ICMPv6 messages too. The Data plane stores and manages the Binding Update List, Binding Cache, and other data related to binding management. The Environment is responsible for setting up the routing rules. It handles all of the management tasks, which are necessary to route the data packets to the right interface with the correct source address. The Control element manages the non-signal driven functions, for example expiration of different kind of entries. These parts don't possess direct interfaces to each other. Instead, an Internal Communication Bus (ICB) connects them. Some of the commands, which are available on the ICB, are exported for third parties through the Low-level Management Socket. An external, High-level Mobility Interface, which handles more complex operations, is also available. Additionally some key features should be integrated in separate modules. More details on the cross-layer considerations of our scheme can be found in Section 5.3.

Functionality: Advanced, fine-grained (i.e., flow level) distributed and dynamic mobility management to support mHealth use-cases over heterogeneous access environments

Input: IP packets to be transmitted, event messages of the publication/notification basis by DDE (e.g. commands for handover initiation in case of network-based mobility management), (cross-layer) resource producer information (i.e. publication event messages, dynamic measurement information from the network), MIIS information (available networks and their parameters static)

Output: IP packets, event messages of the publication/notification basis by DDE (e.g. client-based dynamic measurements), (cross-layer) resource producer information

4.3.3.3 Content aware WiFi topology changes

In our in-hospital scenarios the connection between doctors and the hospital's infrastructure is a wireless networks. Considering the current reality, this wireless network is a WiFi network that has coverage all over the hospital. As the WiFi technology is a radio based transmission, it is true that close to the radio antennas, the Access Points, the signal strength is better and therefore better link quality can be achieved. One can found trivial, that based on this fact, the overall performance of a WiFi network can be increased by increasing the number of Access Points. While, at the end, this is true, actually due to the occurring radio interference, this is not trivial at all.

WiFi networks with high number of Access Points in a given area, called dense WiFi networks. Campuses, sport stadiums, conference venues, where there are lot of users, often deploy dense WiFi networks. The main reason is that the huge number of users can be served by the increased number of access points only. The deployment is usually costly, not just because of the price of the access points, but also the need of special antennas that are able to limit the interference by limiting the footprint of the serviced area. The in-hospital WiFi scenario also requires a dense WiFi network, but the purpose is different and therefore the solution could be different as well. In the hospital, thinking of the doctors as the users of the network, the number of users is low. We only need the dense deployment to give better wireless links to the doctors. This scenario can lead to a solution, where the dense deployed WiFi access points dynamically change their emission power and thus their interference with other access point. The access point that serves the area where the doctor actually resides can get a better signal strength to boost the transmission performance, while the neighbouring access points can decrease their emission to decrease the interference. Moreover, other non privileged users can be instructed to change to an other access point in order to create even less interference. The access point change can be forced by the access point itself and thus it can happen without any user interactivity. Most of the countries in the European Union, WiFi networks can utilize 3 non interfering channels, so the same area can be served by 2 or 3 access points. When the users have to change access point, they are able to continue their operations with some degradation, but without any disruption.

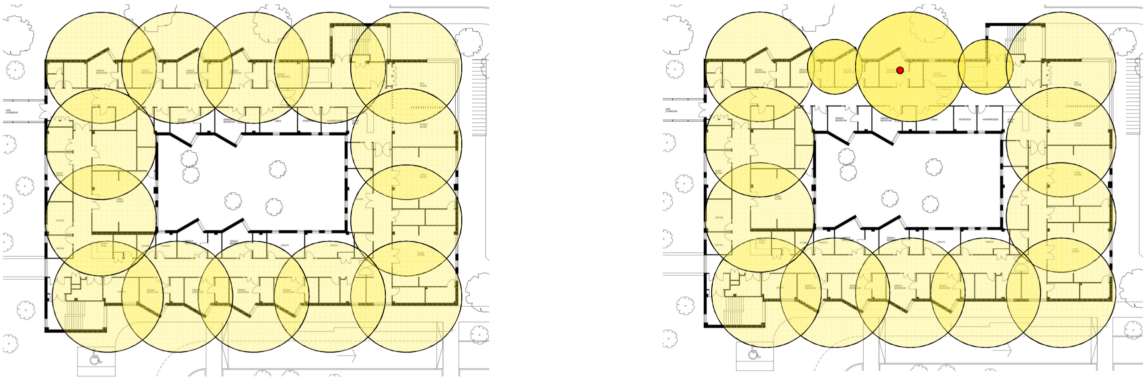

Figure 24: WiFi network with interferences (left) and WiFi network with interference free spot by changing the emission power of access points (right)

In the CONCERTO project cross-layer signalling can be used to describe those network flows that should be prioritized in the wireless network. The central controller of the access points can compute which access points should be adjusted and which users should be moved between access points. Thus the WiFi topology dynamically adapts to the required priorities.

4.3.4 Link layer

4.3.4.1 General considerations

The data link layer (DLL) is located between the physical and network layers and act as a mediator for data delivery between these two layers. In practice, the data link layer provides the functional and procedural means to transfer data between network entities and might provide the means to detect and possibly correct errors that occur in the physical layer. The data link layer functionalities are divided into data link control (DLC) and medium access control (MAC) mechanisms. The main functionalities of the MAC layer include providing and controlling the medium access for devices, creation of the transmission frame, and the delivery of the transmission frame to base band entities. The main functionalities of DLC are to provide addressing, frame synchronisation, flow control, and error correction. The actual implementation of the data link layer functionalities and protocols are, however, not access technology-independent but they differ from system to system (e.g. IEEE 802.11, 3GPP LTE, and WiMAX have their own specific data link layer mechanisms). The CONCERTO system considers heterogeneous access technologies, and to provide upper layers a unified view of the underlying access link conditions, a common solution is required for data link level information collection and management.

The data link layer plays an important role especially in the radio access network. With LTE, the link layer is in charge of providing the Quality of Service (QoS) mechanism, managing the user mobility and the radio resources. Optimizing the link layer or performing cross layer optimization across the link layer are expected to considerably improve the support of MSHTM applications. The CONCERTO architecture includes specific controlling functionalities at the DLL of LTE/LTE-A base stations, in charge of managing multiuser scheduling for both uplink and downlink, allocating the most appropriate radio resources and adaptively selecting the physical transmission schemes (modulation and coding) capable to guarantee the required QoS. The cross signalling messages could be conveyed through extension of the radio resource control (RRC) protocol or at the IP user plane, for example by extending the Access Network Discovery and Selection Function (ANDSF).

Healthcare related video traffic can be prioritized over other less critical traffic with different MAC layer QoS-solutions supported by most wireless standards (e.g., IEEE 802.11 WLAN, IEEE 802.16, 3GPP). In particular, LTE provides a standardized traffic class differentiation based on sets of predefined QoS requirements. Each class is intended to provide a desired level of service for the selected traffic across the 3GPP network. Basically, the service classes are defined in terms of throughput, delay (e.g. latency and jitter) and packet loss rate.

In LTE the data traffic generated by each device is classified into bearers. A bearer is the basic traffic element that enables differentiated treatment according to specific QoS requirements. Each traffic class is identified by a scalar value, called QoS class identifier (QCI). Basically there are two major LTE traffic categories: guaranteed bit rate (GBR) and non-guaranteed bit rate (Non-GBR). GBR are typically used for real-time services, such as conversational voice and video that have low latency and jitter tolerance. Non-GBR bearers do not need prefixed network bandwidth allocation. Non-GBR bearers are used for best-effort services, such as file downloads, email, and Internet browsing. 3GPP defines the series of standardized QCI listed in the following table.

Table 4: 3GPP standardized QCI

QCI	Resource Type	Priority	Packet Delay Budget	Packet Error Loss Rate	Example Service
1	GBR	2	100 ms	10^{-2}	Conversational Voice
2		4	150 ms	10^{-3}	Conversational video (live streaming)
3		3	50 ms	10^{-3}	Real-time gaming
4		5	300 ms	10^{-5}	Non-conversation video (buffered streaming)
5	non-GBR	1	100 ms	10^{-3}	IMS signalling
6		6	300 ms	10^{-5}	Video (buffered streaming) TCP-based (e.g., www, email, chat, FTP P2P file sharing, progressive video, etc.)
7		7	100 ms	10^{-5}	Voice, video (live streaming), interactive gaming
8		8	300 ms	10^{-3}	Video (buffered streaming) TCP-based (e.g., www, email, chat, FTP P2P file sharing, progressive video, etc.)
9		9		10^{-5}	

The LTE standard provides drop mechanisms and priority management strategies in case of network congestion. To this purpose, each bearer is associated to an allocation and retention priority parameter (ARP). For example, ARP is used for the call admission control, to decide whether or not the requested bearer should be established in case of network congestion. However medical and healthcare services are often characterized by very particular and stringent requirements. In order to ensure the desired QoS, specific mapping schemes have been proposed for e-health services on LTE [54]. On the other hand, a strict separation into multiple single flows may turn out to be inefficient, especially in case of contemporary transmission from multiple and heterogeneous information sources. For this reason, transmission solutions based on data flow aggregation at the application layer are also taken into account in CONCERTO.

The standard QoS architectures provide QoS support in per-flow basis, that is, all video packets receive equal QoS treatment from the MAC. This is insufficient for video streams that would benefit from per-packet QoS differentiation in order to ensure that the most important video packets (e.g., SVC base layer) get transmitted also under limited transmission capacity. One solution for implementing QoS differentiation within the video traffic category without hampering the QoS for other traffic types in WLAN is provided in the recently finalized IEEE 802.11aa standard which takes into account fairness towards other traffic types.

4.3.4.2 IEEE 802.11aa and beyond

The IEEE 802.11aa was approved on 2012 spring. The goal of the standard is to provide MAC enhancements for robust audio video streaming. In order to fulfill this goal, the standard introduces novel technologies into the WiFi standard. From the viewpoint of the CONCERTO project, the two most important additions are the Stream Classification Service (SCS) with intra-access category prioritization and the GroupCast with Retries (GCR).

Using the stream classification service, it is possible to use cross-layer information to put the frames of a media stream to the appropriate QoS service queue. The Enhanced Distributed Channel Access (EDCA), which was introduced with the IEEE 802.11e standard, defines 4 Access Categories (AC) for the different priorities: best effort, background, video and voice. In addition to ECDA, the intra-access category prioritization provides 6 transmit queues enabling differentiation of streams that are within the same access category. Using this scheme a finer prioritization is possible between different media streams. Base layers and enhancement layers can get different quality provisioning. It is also possible to make differentiation within a given stream, as an example packets belonging to I and P frames can be put into different transmit queues. In CONCERTO, the required information for stream and intra stream classification is provided using cross-layer techniques.

The other novelty of the standard is the GCR approach in the data transmission. Using this technique the frames can be delivered to a group of users with increased reliability compared to simple multicast transmission. In those CONCERTO scenarios, where the stream is delivered to multiple users who are sharing the same Access Point, the streaming can be more efficient.

It is out of the standard specification, how stream classification and groupcast delivery parameterized. There are already many research papers dealing with media transmission utilizing EDCA categories, however due to the novelty of the 802.11aa standard, there are only a few papers focusing on the further enhanced transmitting queues. Besides GCR there is a lot of competing approaches described in research papers. The standard gives a direction, however there is still room for optimization and new solutions.

4.3.4.3 Radio resource allocation and multiuser scheduling

Within the DDL framework described in 4.3.4.1, the CONCERTO architecture includes a specific functionality at the LTE/LTE-A base stations, in charge of managing multiuser scheduling for both uplink and downlink, allocating the most appropriate radio resources (denoted as Physical Resource Blocks- PRBs) capable to guarantee the required QoS and selecting the physical transmission schemes (modulation and coding). The radio channel capacity or transmission rate, in either direction, uplink or downlink, may vary in time from PRB to PRB and from user to user. A packet-scheduling mechanism must then be used at the base station to establish the frame-to-frame resource allocations in both directions to ensure equitable treatment of all data users, according to individual traffic type/QoS specifications and channel transmission conditions. The purpose of scheduling is two-fold: to provide the appropriate QoS measures such as maximum delay, minimum rate, maximum packet loss probability, and other QoS performance guarantees to individual users, where the propagation conditions allow this, and to ensure full and efficient use of the resources, most commonly link bandwidth or capacity.
Since at any given time some users will have better channels than others, by scheduling the users to transmit only in the best channels at any given time the system resources will result at the end allocated to the users that can best exploit them, which leads to improved system capacity and/or performance. The notion of scheduling transmissions to users based on their channel conditions is called opportunistic scheduling, which can significantly increase both uplink and downlink throughput. Scheduling transmission to users with the best channel raises two problems: fairness and delay. If user channel changes very slowly, then few users will occupy the system for a long period of time. The time between channel uses for any one user could be quite long, and such latency might be unacceptable for a given application. In addition, users with poor average channel conditions will rarely have the best channel and therefore rarely get to transmit, which leads to unfairness in the allocation of the system resources. Scheduling or resource allocation algorithms are often designed to achieve some desired performance objectives, through the maximization of suitably defined utility functions. These objectives could include the attainment of a desired QoS for each data traffic type. They would also include the attainment of maximum possible overall throughput on a given radio link, as stated above. This requires a cross-layer framework, which makes some input information available to the scheduler: channel state information or channel quality indicators from the physical layer and source priority and ranking information, target source rate, target link quality from application and upper layers. For more advanced cross-layer functionalities, such as the joint source adaptation and radio resource allocation which may be considered in CONCERTO to optimize video delivery over the network, the resource allocation unit at the air interface can send information, like available rate and delays for each service queue, to upper layers.

Functionality: allowing efficient use of radio resource at the radio access network while optimizing bandwidth provision under QoS constraints.

Input interface: Source priority and ranking information, target source rate, target link quality, wireless link quality indicators, channel state information.

Output interface: Available rate, delays for each service class/queue.

4.3.5 Physical layer

The physical layer suffers from fading and other channel-induced impairments. Both AWGN and Rayleigh fading channels are considered in this project. More specifically, thermal noise, channel fading, channel dispersion, path loss and shadowing are the main factors that can corrupt the transmitted signals. We also consider various dispersive channel models used in the LTE standard in our investigations.

In the scope of the CONCERTO architecture, we focus on the employment and design of various coding and modulation schemes for supporting reliable transmissions over various communication channels. In terms of coding, we consider turbo codes, LDPC codes, convolutional codes and Turbo Trellis Coded Modulation schemes. In terms of modulation, we consider classic M-ary PSK and QAM schemes when communicating over flat Rayleigh fading and AWGN channels. By contrast, we employ Orthogonal Frequency Division Multiplexing (OFDM) when transmitting over dispersive Rayleigh fading channels. As multiple access techniques, LTE/LTE-A supports Orthogonal Frequency Division Multiple Access (OFDMA) and Single-Carrier Frequency Division Multiple Access (SC-FDMA) schemes for uplink, while IEEE 802.11g/n supports a time division multiple access (TDMA) solution. We remark that a particular focus of CONCERTO is on the uplink technologies, since radio access from remote locations (e.g. an ambulance or an emergency area) is a critical issue of most use cases addressed in the project.

Multiple antenna systems are considered in the CONCERTO physical layer design, as nowadays supported by most wireless standards, from IEEE 802.11n to LTE/LTE-A. In this regard, when the communication channel is particularly hostile, we can invoke space-time coding schemes for achieving spatial diversity gains in order to improve the transmission reliability.

Cooperative communication techniques constitute another attractive solution to mitigate the channel fading by creating a virtual/distributed MIMO transmission scheme, which relies on user cooperation. More specifically, multiple users can serve as a network of source and relay nodes for cooperative transmission to the destination node. Both decode-and-forward and amplify-and-forward cooperative protocols are considered in the CONCERTO project. Cooperative communication is also capable of reducing the co-channel interference as well as of extending the coverage area of a cellular system, which is achieved by the reduction of the overall transmission power. We remark that direct cooperation among users is an important issue under investigation within 3GPP that is currently working on LTE extensions to support device-to-device communications in emergency scenarios. For this reason, the CONCERTO system architecture includes possible LTE-based relaying devices, as well as hybrid LTE/WiFi terminals, capable to forward the received medical information to homogenous or heterogeneous radio networks. The presence of this relaying device becomes particularly important within the emergency area, since a satisfactorily radio coverage cannot be a-priori guaranteed and, as a consequence, the employment of solutions capable to extend the service area are particularly beneficial.

Advanced radio transmission techniques not yet supported by current radio technologies are also considered in the scope of CONCERTO, to provide a first evaluation of the performance that could be achieved with next generation wireless systems. In order to provide some benchmark taking into account possible evolutions of future internet, advanced cutting-edge techniques involving the joint design of coding, modulation and cooperative communication are being developed in this project. More explicitly, Turbo Trellis Coded Hierarchical Modulation (TTCHM) and Turbo Trellis Coded Superposition Modulation (TTCSPM) are developed for cooperative communication, which are capable of approaching the virtual MIMO channel's capacity. Appropriate power-control and rate-adaptation are also crucial for the sake of mitigating the effects of channel fading. Hence, adaptive coding and modulation schemes will be conceived for reliable transmission according to the instantaneous channel capacity.

When the affordable transmission and detection complexity is limited, as in hand-held medical devices, reduced-complexity transceivers will be adopted at the cost of a lower transmission rate. More specifically, non-coherent detection can be used at the hand-held devices or at the relaying node, for detecting/relaying the source signals without resorting to complex channel estimation. Coded Differential Amplitude and Phase Modulation (DAPSK) or StarQAM are attractive non-coherent modulation schemes for the CONCERTO project.

Finally, an adaptive solution consisting of all the above mentioned techniques can be invoked according to the available bandwidth, power, device-complexity, delay-tolerance and channel capacity. The adaptation of the transceiver according to the various PHY layer parameters can be regarded as a cognitive radio system. In this project, we also aim to design cooperative cognitive radio schemes for an overall optimization of the available PHY layer resources. However, the effectiveness of this cognitive system depends on the accuracy of the cross-layer signalling and adaptation. In other words, we consider cross-layer design rather than an independent PHY layer design in the CONCERTO project.

5 Cross-layer signalling and adaptation

To address the challenges about delivering content and context information between the system layers of the CONCERTO solution, an architecture for an extensive cross-layer signalling is required. The main purpose of this chapter is to define the scope of the cross-layer signalling needed in the CONCERTO architectural solution and to introduce the cross-layer signalling architecture itself. Deliverable D2.3 will focus more deeply into the cross-layer signalling, protocols and message formats used, message flows and message exchange between network domains and how cross-layer signalling will be integrated with the overall system architecture.

The original demand for cross-layer information in the current Internet architecture originates from the fact that legacy Internet applications and protocols – being designed with wired networks in mind – are incapable of operating efficiently over wireless links. Therefore, cross-layer information today is also widely used for enabling efficient network resource usage. The cross-layer information utilised in CONCERTO will be mainly focused on wireless with optimisation of the last wireless hop. The proposed architecture focuses on two aspects of the cross-layer signalling: on the access network including an inter-domain solution and on the Internet-wide aspects of signalling. In the following sections, both of these solutions are fairly described.

Taking the different scenarios and use cases that CONCERTO has to deal with into account, there are several challenges that need to be considered: synchronized reception of multiple real-time topology and Traffic Engineering related databases, cross-layer cooperative load and traffic monitoring, cross-layer provisioning and synchronization of configuration changes for joint optimisation across all layers. Therefore, the novel architecture should enable the indexing, collection, synchronization and retrieval of distributed, heterogeneous, time-variant data. Another important point is that such data can be of different type, depending on the regarded resource, and the data can be static, slightly dynamic or highly dynamic. It is necessary to understand how to get and manage this information verifying if they are available, if they have to be measured, calculated or derived considering a trade-off between complexity, signalling overhead, interoperability and timing efficiency.

5.1 Access network and inter-domain signalling architecture

In order to visualise the interconnections between the system layers and the overall architecture for the cross-layer signalling, we have summarised all cross-layer signals described in Table 3 to Figure 25. In this figure, we have described a simple network setup from the CONCERTO use cases which includes a wireless mobile stations (e.g. ambulance) streaming video via uplink to the remote client (e.g. hospital). The mobile station, remote client, and the network elements use cross-layer signalling for optimizing their network usage. To mention a few use scenarios the signalling architecture enables are network load balancing through offloading to guarantee the transmission of emergency traffic flows, to find access points and mobile stations in proximity to assist in data transmission, to find the most suitable serving network and access point, and to adapt the on-going emergency traffic flows according to the current network conditions.

Different system layers either produce or consume (or both) cross-layer events. Cross-layer events are made available for other network entities interested in those ones through the services of Distributed Decision Engine (DDE), which is introduced to the CONCERTO system architecture. DDE provides dynamic event dissemination between different entities such as system layers and/or network equipment. Dynamic information collected from each protocol layer enables various ways to timely and effectively adapt network traffic to varying link and network conditions. Making more static network information available, we introduce a network information service. It provides information about networks and their access points that change rarely. The information available through information services are used more to make longer-term decisions such as selecting the best networks and access points in case of mobility. Information services are specified, for example, in IEEE 802.21 standard [50] and 3GPP Access Network Discovery and Selection Function (ANDSF) [51]. The information services of both standards are specified to provide similar information. In IEEE 802.21, the information service is not specified for the use of mobile stations only, but all IEEE 802.21 capable network entities can make information queries. In ANDSF, the information service is designed for the mobile station use only. In the CONCERTO system architecture, DDE can also make queries on behalf of the event consumers it is serving. Thus, mobile stations do not necessarily need to make information service queries by themselves, but network resources and mobile station processing resources can be reduced when the serving DDE performs the queries. All cross-layer information exchange is visualised with thinner solid lines in the figure.

In addition to cross-layer information that is not directly linked to the actual medical data (e.g. available bitrate, signal strength, and packet loss rate), information directly related to the transmitted medical data (e.g. priority of a single packet compared to other packets) should also be transmitted. The most natural way to do signalling is to deliver it in-band together with the actual data transmission. This signalling is illustrated in the Figure 25: with dashed lines.

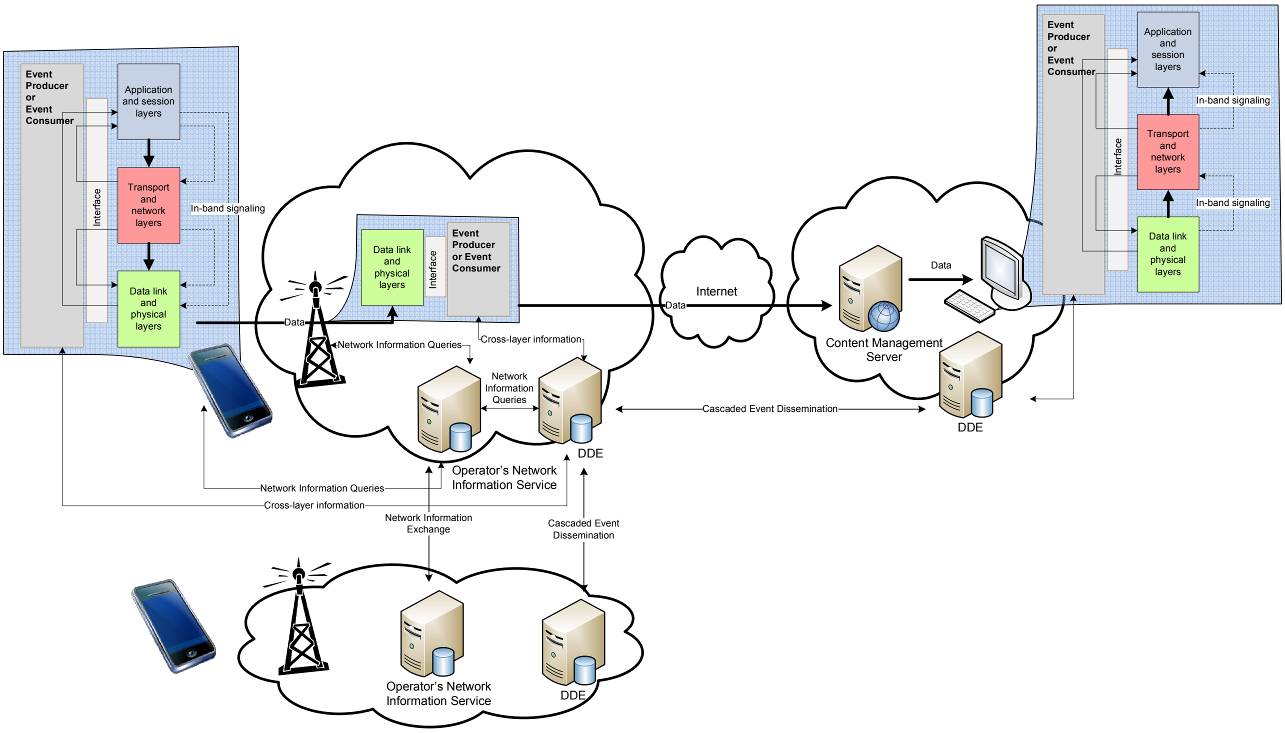

Figure 25: Cross-layer signals between the system layers and network entities.

5.1.1 Distributed Decision Engine

DDE is a collection of components for short time storage and delivery of events between different entities. It was designed to implement a general distributed publish-subscribe message delivery system, which defines a common interface for different information producers, decision algorithms and actors. The system implements an information caching functionality, in order to keep the latest information available and up-to-date, as well as information processing for minimizing the data to be sent over the network. Furthermore, DDE performs information filtering based on policies. In principle, DDE can be used to deliver any kind of events between as well as within network nodes. DDE framework is an evolution of Triggering (TRG) framework which has been introduced for cross-layer signalling purposes in [33]. The main design goal of DDE has been to facilitate decision making in cognitive networking. DDE provides an interface for algorithms making different kind of decisions regarding, for example, network load balancing, network changes, and traffic adaptation. Moreover, the message format is much improved from TRG. DDE messages support, for example, data validity field and digital signatures for data integrity and authentication. Event caching is a new feature, which improves accurate and timely decision making by making all valid previous events available. DDE improves several issues of TRG framework and redesigns the overall architecture of it. DDE framework allows interactions within a network domain but also between different networks. Interaction/information collection can also be done between/from various types of devices and technologies such as access points, switches, routers, servers etc. DDE also supports localised data processing so that only necessary information is sent over the network in order to minimise data traffic between different nodes.

DDE comprises of (a) the Event Producers, which feed DDE with information; (b) the Event Consumers, which receive information in the form of events they have subscribed; and (c) Event Caches that orchestrate this information exchange and can be interconnected in cascade fashion. For example, an algorithm would attach itself to the DDE both as an Event Consumer for their inputs and as an Event Producer for their outputs. The following figure depicts the DDE ecosystem. All DDE communication is based on the TCP protocol and socket interfaces.

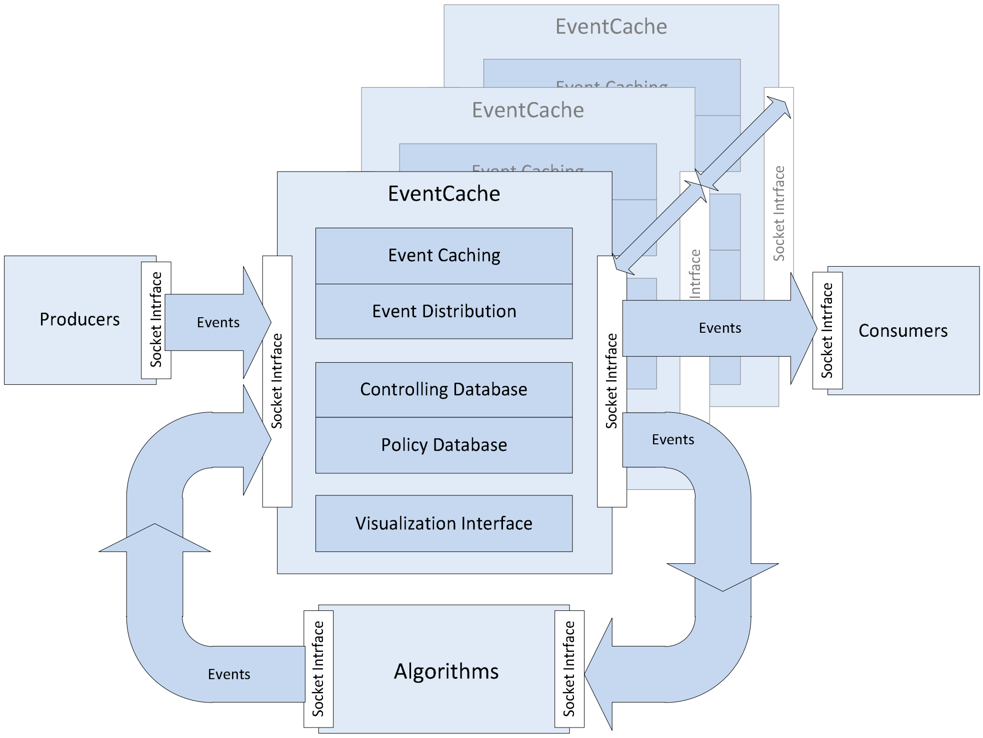

Figure 26: The DDE framework.

In the DDE framework, Producers represent the Information collection and Consumers represent the Executive entities in the knowledge model. The Algorithms connect to the architecture simply as both Producers and Consumers as they both consume and produce information. Lastly, the Event Cache forms the communication framework used by all the other entities.

5.1.2 Event Cache Components

Event Caching is responsible for storing the events for the duration of their validity, i.e. time-to-live. When an event expires or is replaced with a new event, it writes the old event also to a history file. Event Distribution handles both incoming and outgoing event traffic by managing the serving and sending socket interfaces. Controlling Database stores the registrations and subscriptions of the producers and consumers and provides lookup functionality on these for the Event Distribution. Policy Database stores the policies governing the event traffic and is also queried before sending any event. Visualization Interface provides external observer information on the actions of the event cache.

5.1.3 Basic operation

Event Producers register an Event ID with Event Cache before being able to send events to DDE for further dissemination. After the registration Event Cache accepts events sent by the Event Producer, delivers them further and retains a copy of the event for duration of events validity (determined by its Time-to-Live). The Event Consumers subscribe to the Event IDs of the events they want to receive after which the Event Cache delivers them the events with such Event IDs. Basic messaging is shown in the figure below.

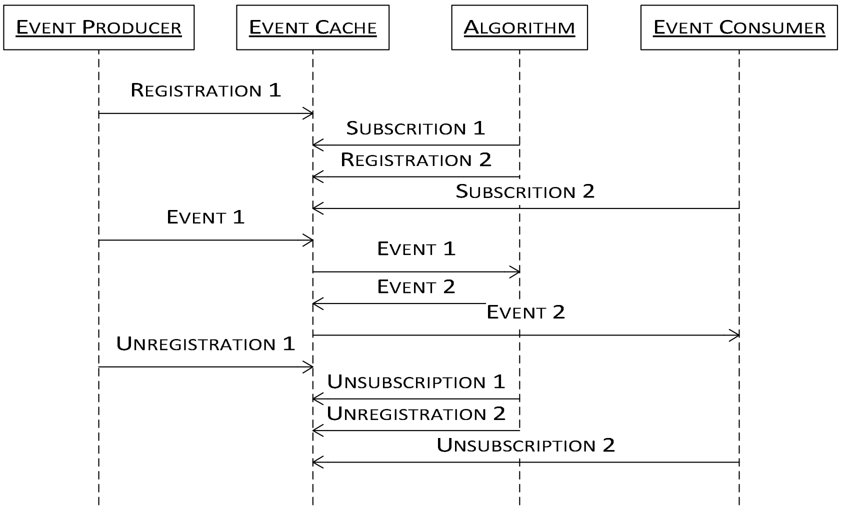

Figure 27: An example of DDE framework message flow

Event distribution can be limited via access policies. Access policies provide means for an operator to limit/restrict the Event distribution at the Consumer end. In practice this means that the operator can define, with access policies, which Consumers are allowed to get Events with certain Event ID. Policy messages are used to update these rules in the Event Cache. Access policies are stored in and implemented by Event Caches. If the access policies are enabled, they only allow an Event to be sent to a consumer if it is permitted by policies.

5.1.3.1 Improving the Scalability of DDE

In case of large number of mobile users using DDE for event dissemination per access point and network, the signalling may cause excessive overhead to the total network traffic. In order to mitigate the overhead, event aggregation bundling multiple events into one event was studied in the OPTIMIX project [52]. Proximity services are increasingly getting interest to improve the resource usage of mobile networks. For example, 3GPP has done a feasibility study of proximity services [53] and a standard specification for 3GPP proximity services has started. Proximity service for disseminating the DDE events is interesting topic to be studied. In case of large number of mobile stations connected to the same base station have subscribed to same events for their network service monitoring and mobility management, the DDE could send the events to only one mobile station and the events are delivered to other mobile stations through proximity services. This is illustrated in Figure 25. The same concept can be applied also to event producers case, where one or a few mobile stations can act as an event collector disseminating the produced events to the serving DDE as an aggregated event. The proximity event dissemination can be carried out, for example, through WLAN or, for example, also LTE connection.

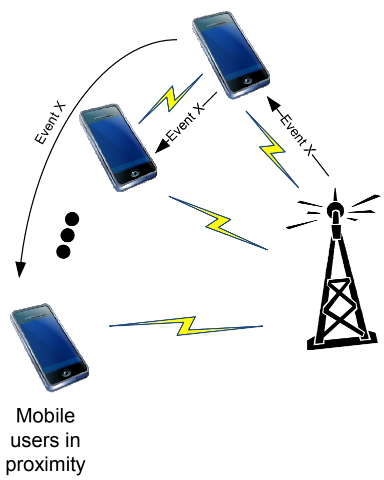

Figure 28: Proximity services to improve signalling scalability

5.1.4 Network Information Service

Information services are often specified for the purpose of facilitating seamless mobility in heterogeneous network environment by providing a variety of information about networks, and their access points and capabilities. This information is often relatively static in nature and comprises, for example, access point location, network address, supported services, frequency bands, and access technology specific information. While DDE provides dynamic event delivery capabilities, network information service acts as a database providing information that change rarely and this information can be accessed through enquiries.

Each mobile network operator can maintain their own information service database. Both mobile stations and network entities responsible for making decisions about mobility issues, and network load balancing and configuration can query information service for relevant information for their cognitive decision making. For efficient network utilization, the information services of different operators should collaborate, i.e. exchange information.

The scenarios for the network information service relate to finding suitable target networks in range in case of mobility. The information service also allows finding assisting networks for data transmission. Figure 29 illustrates a message exchange for finding access points in range to assist in an emergency data transmission. The ambulance uses, for example, LTE connection and queries for access points in range and vicinity to either assist in data transmission or to make a handover to an access point providing better Quality of Service (QoS). The frequency information the service provides can be capitalized on, for example, in proximity services, such as in the case presented in Section 5.1.3.1. The information tells mobile stations the spectrum that is highly occupied or reserved. Based on that knowledge, the mobile stations can agree on the frequency band for their mutual communication, which is available and unoccupied in their current location.

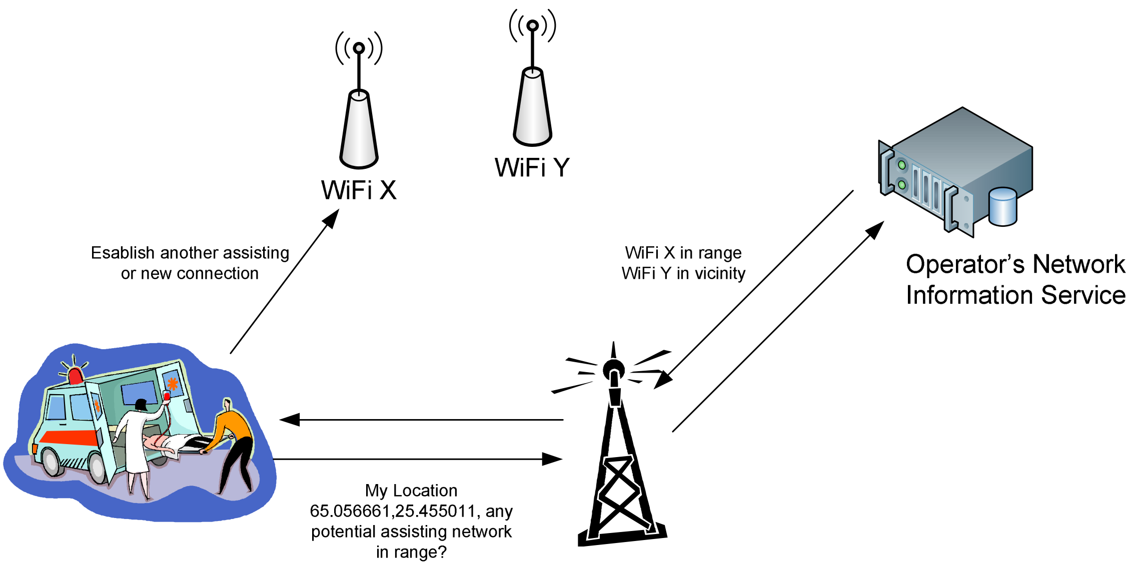

Figure 29: Use of Information Service

5.2 Distributed architecture for cross-layer information delivery

The architecture proposed in the previous section for cross-layer signalling within the access part of the network, also between different access domains introduces a solution mainly for the last hop of the session. However, a wider solution for Internet-wide cross-layer signalling is needed. In this section, we propose architecture for cross-layer information delivery based on distributed CDN-like solution where management and delivery of cross-layer information is distributed among different network domains.

Just to make an example, in adaptive HTTP streaming as foreseen in CONCERTO, multiple data rates can be used for encoding video contents and the choice of the best streaming to be sent to the client is made according to the available bandwidth. It is necessary in this circumstance that some information concerning the network status (e.g. exploitable bandwidth, delay, error-rate, etc.) are known before and made available somewhere on the network. More generally the knowledge of network metrics and the possibility to retrieve it easily is very important for supporting service differentiation technique on the network. Therefore the critical issue is the design of a scalable solution that supports the distribution by producer entities of resource information (i.e. a piece of cross-layer signalling information) for use by consumer ones; from one hand, for efficient retrieval and on the other, introducing limited overhead. Thus it is useful to define some Collection Points (CPs), within each domain (e.g. Autonomous Systems over the Global Internet) that could be Access Gateways (AGs) operating as measurement proxy with associated repository and directory server. For scalability reasons, AG could play the role of Border Gateway (BG) to efficiently interface with the remote domains (as detailed below). These Collection Points would include the functionalities of Event Caches described in the previous section focusing on the last hop of the session.

It is important that every piece of information stored within the CP is uniquely identifiable for proper management and retrieval. For this the solution proposed for access-domain in the previous section can be used. DDE framework and DDE messages provide ways to identify, manage, and retrieve cross-layer information from different entities. Another critical point is the synchronization; resource information must be up-to-date and consistent, therefore mutually synchronized, across multiple layers for optimization purposes. Furthermore, at a higher level, data regarding different videos of the same scene (i.e. multi-viewpoints) need also to be synchronized between them. For example, a time-stamp can be associated with the provided data, possibly calculated somehow, according to the synchronization constraints in place.

Basically, one of the main goals of cross-layer signalling is to efficiently obtain the needed information with respects to timing, overhead and scalability. From one hand, the delay in receiving data after a request for it should be as short as possible and on the other the regarded communication model(s) should be properly selected, addressing low overhead and complexity. Concerning the first issue, the closest (in terms of network proximity and policies in place) CP can be reached by using IPv6 anycasting [34]. Of course, the anycast address of the CPs in a given domain is to be advertised on the network (for use by all its components). When a CP belonging to a specific domain receives resource information by a producer entity, it can efficiently spread it among all the other CPs in that domain by using a multicast address [35]. Data about the network status (at all levels) should be stored locally in databases (associated with the respective CPs) located in the domain of that network. While, data related to resources of the other domains can be stored in databases belonging to those domains. For latency reduction in the provisioning, as well as for saving bandwidth and memory storage in the issued databases, the information available can be simply notified to the CPs of other domains rather than actually delivered to them.
As a cost-effective real-world deployment option, the database of an event driven system (i.e. the event cash in the DDE framework), used to collect and notify resource information over the access part of the network (even between different access domains of the same network) can be integrated with the said AG/CP's database, in order to have a single data repository for a given zone of the network.

For backward compatibility issues, it is also possible in principle to integrate the existing databases (e.g. for Traffic Engineering, associated with the implementation of MIH standard, about overlay networks, information base of deployed control and management planes) in the architecture for cross-layer optimization by properly registering their content (i.e. the related indexing information) to the CPs of the respective domains. To optimize the information retrieval, as already suggested above, the CP of a domain, by using anycasting technique, could register its content to the closest (in terms of network proximity or policies enforced by the concerned ISP) CP of the neighbouring domains whereas a multicast technique could be used for distributing the same registration information to the CPs internally to a domain. In this way, if the information to retrieve is stored in the database of another domain, then the CP interrogated locally will re-direct the request to a CP of the domain where the database is located.

For the information that is high time-variant, it is possible to consider an on-demand request from the consumer entity to the producer directly. Indeed, in different use cases of CONCERTO where a high quality video stream has to be sent on the wireless network, the information regarding the status of the network, for example in terms of bandwidth available rather than SNR, could be asked directly to the access point. While quite time-variant information to be used for a period of time in network or application operations could be delivered to the interested party every time it changes, possibly using filtering and hysteresis techniques as proposed by DDE framework, by leveraging on a subscription-notification service and avoid to explicitly and continuously querying a CP database or directly the producer entity for it which would be certainly inefficient. Finally, slowly time-variant information, such as information related to cache and mirror servers, network decoders, cellular or Wi-FI network configuration, user terminal locations, medical databases and mobility management entities, could be efficiently collected and retrieved by the closest CP.

Regarding the communication protocols, a large variety of options are available for the purpose. TCP [36] and UDP [37], together with their counterparts with security facilities (i.e. TLS [38] and DTLS [39], respectively), can be used at the transport layer, also for backward compatibility issues. Security could be addressed at application level, including credentials in the resource information request as in SNMP [40][41]) or as proposed in the IETF ALTO WG [18] for authentication, integrity protection and encryption, also HTTP could be a valid option for this purpose. Finally, hop-by-hop and destination options of IPv6 could be employed when an IP flow is already active between the producer entity and the issued CP, the CP and the consumer entity or, the producer and the consumer entities, when data delivery is required between the same end-points. The selection among the mentioned protocols is to be made by considering the resulting trade-off between aspects related to communication reliability, introduced bandwidth overhead, expected provisioning delays, backward compatibility, complexity, security and flexibility.

5.2.1 Internet-wide (inter-domain) cross-layer signalling

In Figure 27, the architecture for Internet-wide (inter-domain) cross-layer signalling is proposed as resulting from above discussions. The key components for collecting and provisioning data across multiple layers, devices, domains and technologies are the Access Gateways (AGs) that play the role of Event Caches. A pool of AGs is deployed in each domain (e.g. Autonomous System) for the direct delivering of information about resources of the associated domain and the re-direction of requests for information about resources in the others. The number and location of AGs in a domain is a trade-off between several aspects, including efficiency, scalability, additional overhead, cost and complexity. The larger the number, the shorter the delay in getting the needed information, but also the higher the introduced overhead, in terms of bandwidth, storage memory and data processing. For small domains even a single AG could be enough.

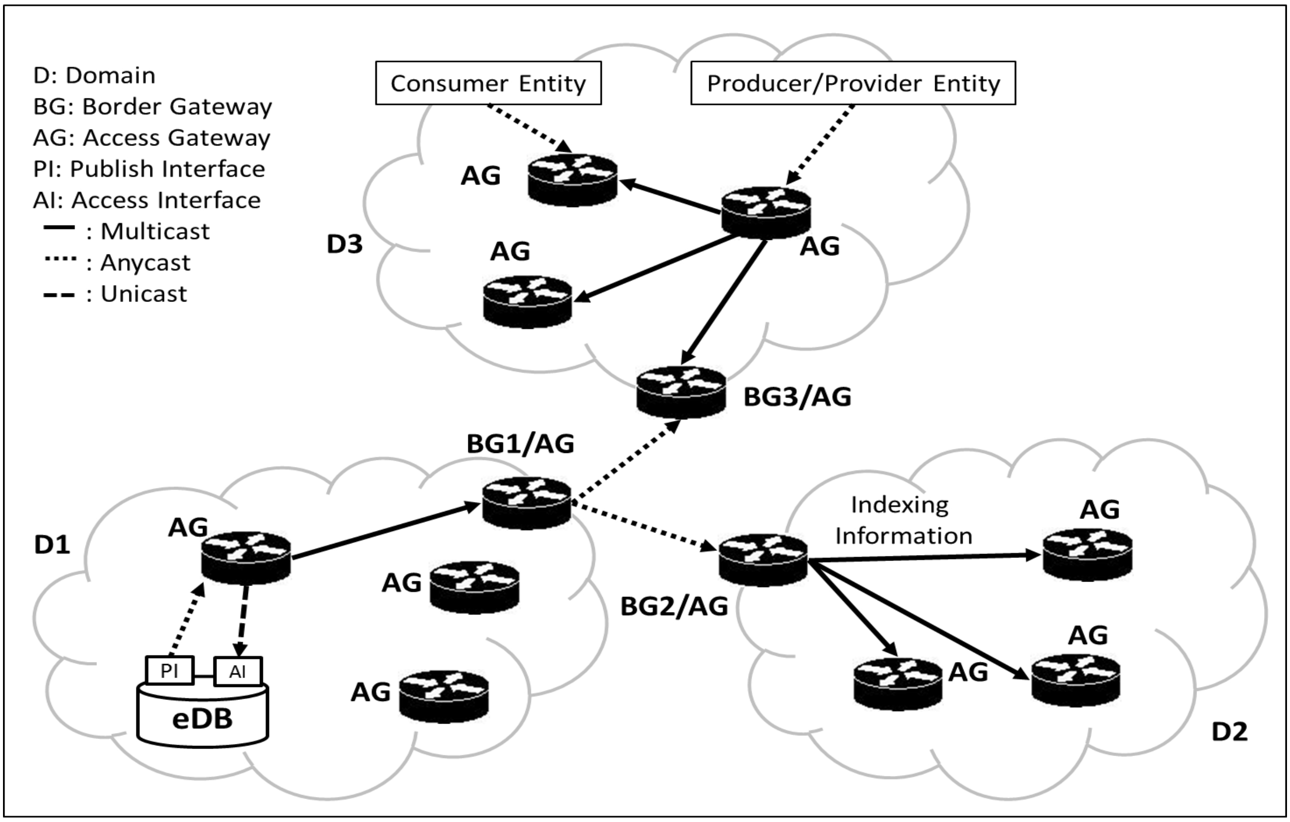

Figure 30: Architecture for cross-layer communication all over the Internet

Among each pool of AGs, a special role is covered by the Border Gateway (BG) that collects indexing information about resources of the remote domains and distributes it to the other AGs of its domain. As already explained, the pool of AGs (i.e. including the associated BG) in a domain belongs to a multicast group for efficient distribution of AG database content (as well as, of indexing information collected by the BG) to the other AGs within that domain. In addition, an anycast address is also assigned to the said pool in order to collect resource information by the AG closest to the producer entity, as well as to provide resource information to the consumer entity by its closest AG. This anycast address is also used by the BGs of the other domains to globally spread the indexing information about the resources in their respective area of competence. The BG of a domain should be advertised to the others as the closest gateway by EGPs for that anycast address. Therefore, each domain has assigned well-known anycast and multicast addresses for its pool of AGs.

An existing database (eDB) can be integrated in the architecture by publishing the indexing information about its content to the closest AG (which will then distribute it within the pool of AGs of the same domain; while the reference BG for that domain is in charge of global spreading). Furthermore, an access interface is associated with the eDB in order to retrieve resource information from it. In the end, both, efficiency, scalability and backward compatibility issues are addressed with the proposed architecture by leveraging on simple interfaces and usage of IPv6 anycast and multicast addresses.

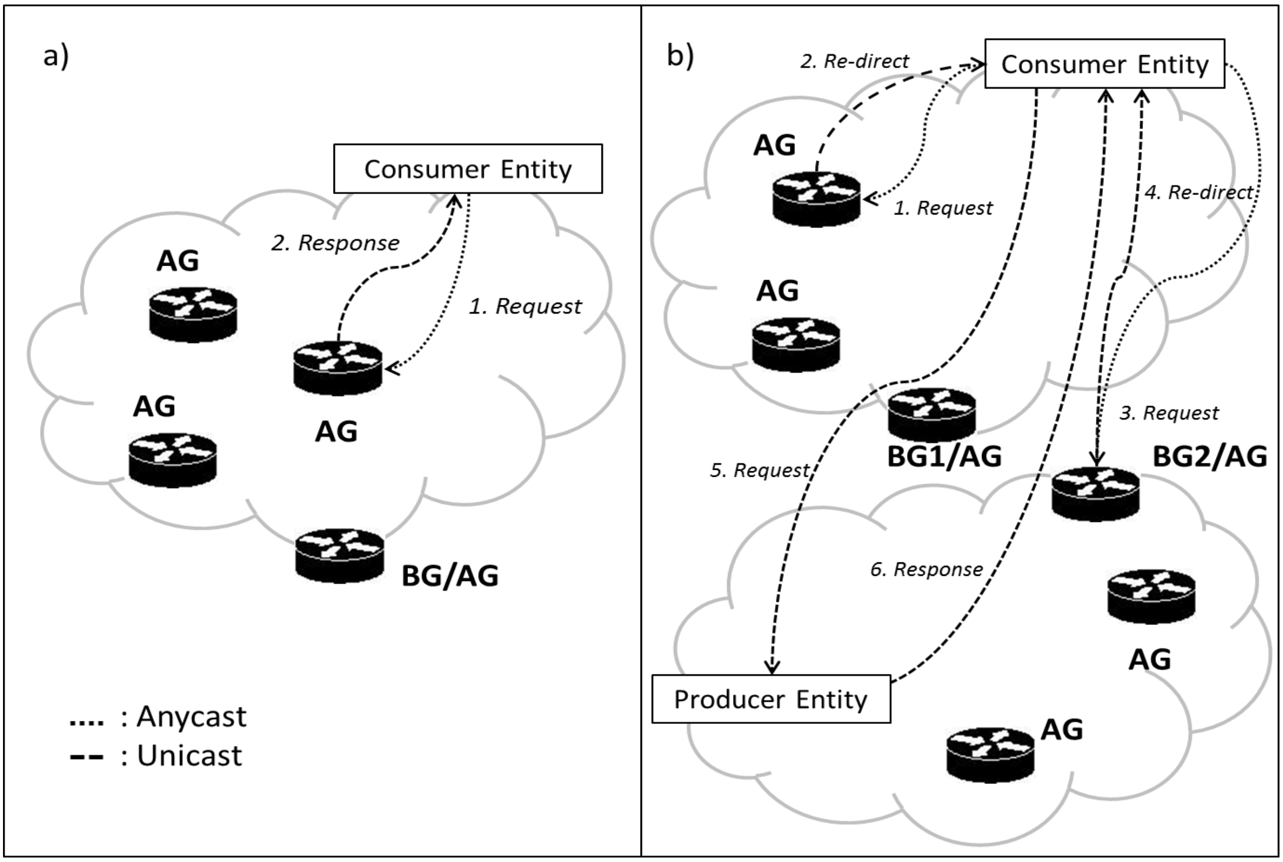

Figure 31: Examples of information retrieval: (a) directly from the closest AG and (b) from the producer entity in a remote domain after re-directions from the local and remote AGs

For the sake of clarity, in Figure 31 the message exchange for the retrieving of resource information is shown for two cases. In (a), the consumer entity obtains the requested information directly from the closest AG in its domain. While in (b), the consumer entity is first re-directed to the closest AG (i.e. the BG) in the (remote) domain where the issued resource is located, and then finally re-directed to the target producer entity. The proposed architecture for collection and retrieval of cross-layer information from a producer entity to a consumer one, all over the Internet has a quite general validity. The specific instance of it to be deployed in a real network and application scenario, in terms of number and location of relevant components and supported functionalities, as well as the protocols, mechanisms and policies to use, results from the analysis of the technical and functional requirements of the use-cases addressed by CONCERTO, and more in detail, by the type and nature of the cross-layer information that should be made available and exploited for joint optimization purposes. For example, the time constraints, the resulting overhead, the updating frequency of such information should be carefully considered in order to fully specify the issued architecture. Therefore, the next step is to evaluate and classify the cross-layer information needed in the CONCERTO's solution against orthogonal axes: timely, variability, size, source and destination endpoint locations and layers, single or multiple use, criticality (for optimization purposes), processing (e.g. filtering or aggregation), QoE impact, backward compatibility (i.e. actual availability) and technology dependency to name a few.

5.2.2 Integration of cross-layer signalling architecture with overall architecture

DDE is a signaling delivery entity, which can reside in the access network and in the core network. In the access network, the DDE (or possibly several DDEs) handles mostly event delivery between MS and BS and between BSs. These events are mostly based on sharing of resource information (mostly between BSs), QoS information from MSs used in traffic adaptation and prioritization in the BSs, and mobility management related information. In the core network, the DDE is defined to disseminate network resource information between access networks and between network domains. The DDE is running in the access gateways and border gateways in this case as described in the previous sections.

DDE infrastructure can be distributed. The network can possibly be equipped with more than one DDE in order to mitigate the load balancing issues and improve the reliability of the signaling service. For example, some base stations and mobile stations connected to those base stations are using a DDE dedicated for their usage. The DDE implements a possibility to share dissemination events between DDE, which allows wider signaling in case of distributed DDE deployment. In the core network, the communication with and between gateways can be optimized using IPv6 multicasting and anycasting, respectively. When anycasting is used, it is important that each DDE in the core network has a consistent database of the event subscriptions and registrars. For this, the DDE need to be able to exchange their context information between the DDE entities in the core network. Deliverable D2.3 is focusing purely on cross-layer signaling, protocols and message formats used, message flows, message exchange between network domains and integration with the overall system architecture.

5.3 Cross-layer signalling considerations in mobility management

The IPv6 mobility support is one of key parts of the network architecture. This framework manages the network connections and the packet flows of applications. Due to this, a powerful interface is necessary to incorporate, which could interact with other signalling and decision mechanisms in CONCERTO. The following figure depicts the system architecture around the mobility management daemon.

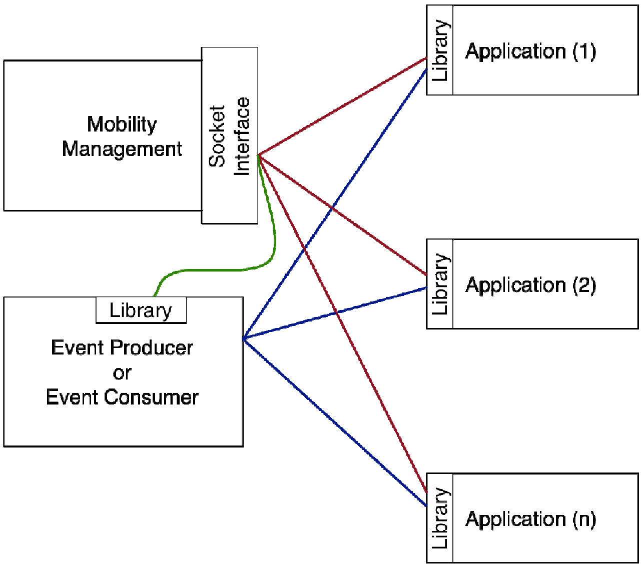

Figure 32: Mobility management system interfaces for cross-layer signalling

The Mobility Management entity manages the packet flows between the existing connections. For this, the application needs to register its parameters. These are: source and destination address, source and destination port, and L4 protocol. To ensure the un-authorized flow management, the Mobility Management module provides and ID and a KEY to the application. These are required for flow management, which changes the output and input interface for the given session in heterogeneous multi-access environments for optimized communication.

The flow redirection could be managed by the Event Producer or Event Consumer entity or by the Application itself. In the first case, the application should provide its ID and KEY (got from Mobility Management module during registration) to the EP/EC, and it will request the redirection of the flow. In this case, the application will get only a notification event about the action. On the other hand, the EP/EC just notifies the application (that could react in case of need), and it precedes the redirection.

The handover initiation could be driven by the Event Producer or Event Consumer entity or by the Mobility Management module. Basically it is a conceptual question, which was not decided yet. In both cases, the Mobility Management entity should be notified about the action to prepare and make the processing of the handover faster. Figure 30 depicts the interaction primitives between the previously introduced components.

First of all, the application should register itself at the Mobility Management entity. During the registration the application provides its source and destination address, port and protocol. These parameters are required to be able to create filter and redirect rules. The Mobility Management module replies with an identifier and a key. The key is necessary for the further modifications, because it will authenticate the application. This method shall avoid the unauthorized redirections. Next, the application should register itself to the Event Producer or Event Consumer as well. The distributed decision engine – through the Event Producer or Event Consumer – will drive the selection of the best communication path and the flow redirection. The application should provide the parameters of its session(s) to the

EP/EC: source and destination address, port, protocol, relevant QoS parameters and requirements, the identifier and the key, which has been replied by the Mobility Management previously. By providing these values to the EP/EC, it will be able to select the best matching interface, and redirect a flow to it. The application does not need to be informed about the selection and the redirection. During the operation phase, the EP/EC could retrieve the list of the available mobility connection from the Mobility Management by the GET_BUL command. The reply will list the available connection, and it will define the mobility level for each, in a per application basis. If the EP/EC would like to redirect the session of the application to another interface or connection, it will send a SET_FLOW command to the Mobility Management. This command has to define the identifier, which identifies the application, the key for authorizing the request and the requested new interface. Next, it could notify the application about the flow relocation, so the application could modify some session parameters based on the information of the new connection and communication path. In that way, an adaptive, flexible and fine grained (i.e. flow level) cross-layer mobility management is achieved, which serves as a basic infrastructure component for emerging wireless/mobile telemedicine applications and services.

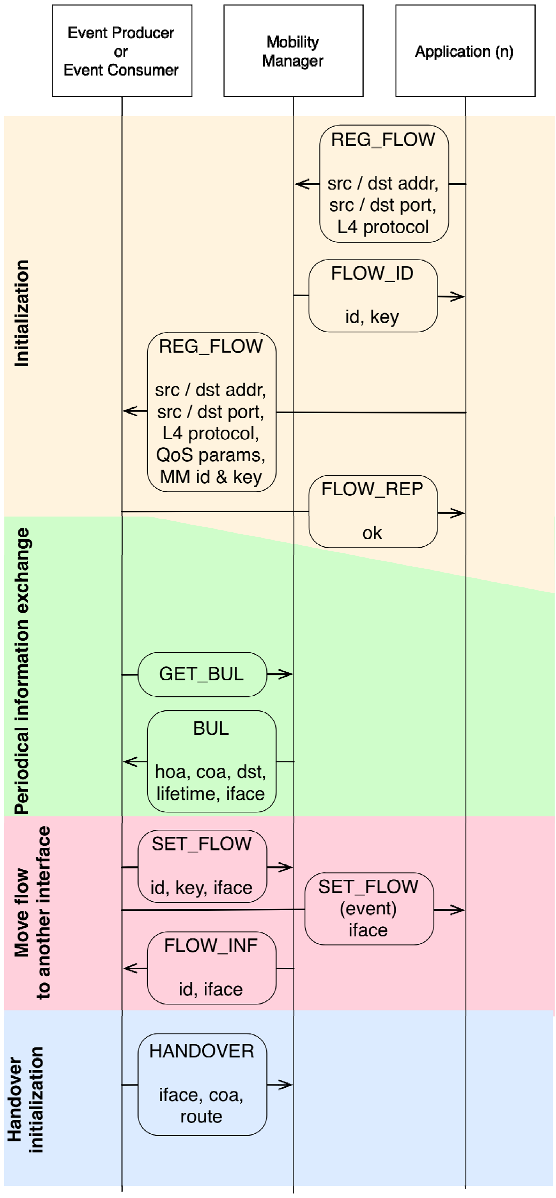

Figure 33: Interaction primitives between functional entities related to mobility management

6 Conclusion

In order to make telemedicine service adaptation according to the current network conditions and network/mobility management peculiarities related to the service properly and with the highest possible QoE, there is a strong need for an appropriate architecture jointly optimized by accurate and timely cross-layer information exchange. This deliverable concerns the issues of the overall architecture setup with all the identified building blocks and their integration. The main requirements, problems, and concepts have been firstly introduced and detailed. Furthermore, an initial specification of the CONCERTO cross-layer signalling and adaptation framework has been provided and discussed, in relation with the considered telemedicine scenarios of the CONCERTO project.

In general, this document focuses on defining the initial version of the cross-layer system architecture using the requirements of the CONCERTO use-cases and scenarios. We first identified the need for cross-layer signalling by defining the functionalities which need or could benefit from the cross-layer information. After identifying the need for cross-layer information, we defined the preliminary requirements for the overall architecture and more specifically for the cross-layer operation. Based on this information together with the state-of-the-art study we proposed the main functional building blocks and a preliminary solution for the cross-layer signalling and adaptation framework aiming at the appropriate integration of the different architectural elements. The proposed initial solution is based on distributed cross-layer information delivery architecture with two main targets: access network or inter-domain solution and broader Internet-wide solution. The main solution for cross-layer signalling is based on a Distributed Decision Engine (DDE) which proposes a producer-consumer –based solution with information database for managing the information delivery. In addition to the main solution, architecture for Internet-wide signalling is proposed for efficient delivery and management of information between network domains.

The documented initial results are fundamental because on the way for the next documents in Task 2.3 they constitute the basis for the future work: further architecture optimizations, extensions and more sophisticated integration of the interworking components will lead to the right and appropriately detailed resulting specifications of an effective as possible cross layer system architecture.

7 References and glossary

7.1 References

[1] G. Tselentis et al., “Towards the Future Internet - Emerging Trends from European Research”. IOS Press, 2010

[2] P. Krishnan, D. Raz and Y. Shavitt, "The cache location problem". IEEE/ACM Trans. on Networking, vol. 8, 2000, pp. 568-582

[3] E. Cronin et al., "Constrained mirror placement on the Internet". IEEE Journal on Selected Areas in Communications, vol. 20, 2002, pp. 1369-1382

[4] L. Qiu, V. Padmanabhan and G. Voelker, "On the placement of Web server replicas". INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE COMSOC. Proc. IEEE, 2001, pp. 1587-1596 vol.3

[5] S. Ratnasamy et al., "Topologically-aware overlay construction and server selection". INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE COMSOC. Proc. IEEE, 2002, pp. 1190-1199 vol.3

[6] Z. Fei, M. Ammar and E. Zegura, "Multicast server selection: problems, complexity, and solutions". IEEE Journal on Selected Areas in Communications, vol. 20, 2002, pp. 1399-1413

[7] P. Quax et al., "Dynamic server allocation in a real-life deployable communications architecture for networked games". Proc. of the 7th ACM SIGCOMM Workshop on Network and System Support for Games, Worcester, Massachusetts: ACM, 2008, pp. 66-71

[8] S. Gargolinski, C.S. Pierre and M. Claypool, "Game server selection for multiple players". Proc. of 4th ACM SIGCOMM workshop on Network and system support for games, Hawthorne, NY: ACM, 2005, pp. 1-6

[9] M. Suznjevic, M. Matijasevic and O. Dobrijevic, "Action specific Massive Multiplayer Online Role Playing Games traffic analysis: case study of World of Warcraft". Proc. of the 7th ACM SIGCOMM Workshop on Network and System Support for Games, Worcester, Massachusetts: ACM, 2008, pp. 106-107

[10] B. Krishnamurthy, C. Wills and Y. Zhang, "On the use and performance of content distribution networks". Proc. of the 1st ACM SIGCOMM Workshop on Internet Measurement, San Francisco, California, USA: ACM, 2001, pp. 169-182

[11] J. Agogbua, M. O'Dell and J. McManus, “Requirements for Traffic Engineering Over MPLS”. IETF RFC 2702, September 1999

[12] K. Shiomoto and D. Brungard, ”Generalized MPLS (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)”. IETF RFC 6001, October 2010

[13] A. Reibman and M.-T. Sun, “Compressed Video Over Networks”.New York: Marcel Dekker, 2000

[14] M. van der Schaar and S. Shankar, “Cross-layer wireless multimedia transmission: Challenges, principles, and new paradigms”. IEEE Wireless Commun., vol. 12, no. 4, Aug. 2005, pp. 50–58

[15] X. Wang, Q. Liu and G. B. Giannakis, “Analyzing and optimizing adaptive modulation coding jointly with ARQ for QoS-guaranteed traffic”. IEEE Trans. Veh. Technol., vol. 56, no. 2, Mar. 2007, pp. 710–720

[16] T. Holliday, A. Goldsmith and P. Glynn, “Optimal power control and source-channel coding for delay constrained traffic over wireless channels”. In Proc. IEEE Int. Conf. Commun., May 2002, vol. 2, pp. 831–835

[17] F. Fu and M. van der Schaar, “A New Systematic Framework for Autonomous Cross-Layer Optimization”. IEEE Trans. On Vehicular Technology vol. 58, no. 4, May 2009

[18] IETF ALTO WG charter, http://datatracker.ietf.org/wg/alto/charter/

[19] P4P DICIA WG home page, http://cs-www.cs.yale.edu/homes/yong/p4p.html

[20] FP7 ICT ALICANTE home page, http://www.ict-alicante.eu/

[21] FP7 ICT ENVISION home page, http://www.envision-project.org/index.html

[22] FP7 ICT ADAMANTIUM home page, http://www.ict-adamantium.eu/

[23] 3GPP IMS home page, http://www.3gpp.org/article/ims

[24] D. Harrington et al., "An Architecture for Describing SNMP Management Frameworks". IETF RFC 2261, January 1998

[25] B. Wijnen et al., "View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)". IETF RFC 2265, January 1998

[26] R. Enns, "NETCONF Configuration Protocol". IETF RFC 4741, December 2006

[27] T. Nadeau, "A Framework for Multi-Protocol Label Switching (MPLS) Operations and Management (OAM)". IETF RFC 4378, Februry 2006

[28] G. Swallow, H. Ishimatsu and Y. Rekhter, "Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support". IETF RFC 4208, October 2005

[29] ITU-T Y.2011, "General principles and general reference model for Next Generation Networks". October 2004

[30] ITU-T Y.2012, "Functional Requirements and architecture of the NGN". April 2010

[31] 3GPP TS 23.401 General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN), V12.0.0 (2013-03)

[32] 3GPP TS 23.402 Architecture enhancements for non-3GPP accesses V12.0.0 (2013-03)

[33] Sutinen T., Vehkaperä J., Piri E., Uitto M, "Towards Ubiquitous Video Services through Scalable Video Coding and Cross-layer Optimization", EURASIP Journal on Wireless Communications and Networking, 2012.

[34] K. Lindqvist, "Operation of Anycast Services". IETF RFC 4786, December 2006

[35] T. Pusateri, "Protocol Independent Multicast - Sparse Mode (PIM-SM)". IETF RFC 4602, August 2006

[36] J. Postel, "Transmission Control Protocol". IETF RFC 761, January 1980

[37] J. Postel, "User Datagram Protocol". IETF RFC 768, August 1980

[38] T. Dierks, "The Transport Layer Security (TLS) Protocol Version 1.2". IETF RFC 5246, August 2008

[39] E. Rescorla and N. Modadugu, "Datagram Transport Layer Security". IETF RFC 4347, April 2006

[40] D. Harrington et al., "An Architecture for Describing SNMP Management Frameworks". IETF RFC 2261, January 1998

[41] B. Wijnen et al., "View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)". IETF RFC 2265, January 1998

[42] Wuttipong Kumwilaisak, Wenwu Zhu, A Cross-Layer Quality-of-Service Mapping Architecture for Video Delivery in Wireless Networks, IEEE J. Select. Areas, vol. 21, pp 1685 – 1698, Dec 2003

[43] Piri, E.; Uitto, M.; Vehkapera□, J.; Sutinen, T.; , "Dynamic Cross-Layer Adaptation of Scalable Video in Wireless Networking," IEEE Global Telecommunications Conference (GLOBECOM 2010)

[44] G. Panza and S. Grilli. "A scalable solution to provide Quality of Service in Next-Generation Networks". Future Networks and Mobile Summit 2011

[45] Yuan-Cheng Lai, Wei-Hsi Li, "A novel scheduler for proportional delay differentiation by considering packet transmission time," IEEE Communications Letters, vol. 7, No. 4, pp. 189-181, April 2003.

[46] Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An architecture for Differentiated Services – RFC2475," IETF DiffServ WG, 1998.

[47] J Heinanen, F Baker, W Weiss, J Wroclawski, "Assured forwarding PHB - RFC 2597," IETF DiffServ WG, 1998.

[48] Sutinen, T.; Huusko, J.; , "MAC layer QoS architecture for optimized and fair transmission of scalable video," IEEE Consumer Communications and Networking Conference (CCNC 2011)

[49] CONCERTO deliverable D2.1: "Use cases, usage models and related requirements"

[50] IEEE Standard for Local and Metropolitan Area Networks. Part 21: Media Independent Handover Services, IEEE Std. 802.21, 2009.

[51] 3GPP System Architecture Evolution (SAE); Security Aspects of non-3GPP Accesses (Release 8), 3GPP Std. TS 33.402, Rev. 8.8.0, 2012.

[52] R. Fracchia, G. Jeney, M. Mazzotti, and E. Piri, "Optimisation of Multimedia over wireless IP links via X-layer design", in Proc. NEM Summit, Barcelona, Spain, October 2010.

[53] 3GPP Technical Specification Group Services and System Aspects; Feasibility study for Proximity Services (ProSe) (Release 12), 3GPP TR 22.803, Rev. 12.1.0, 2013.

[54] Lea Skorin-Kapov and Maja Matijasevic, "Analysis of QoS Requirements for e-Health Services and Mapping to Evolved Packet System QoS Classes," International Journal of Telemedicine and Applications, vol. 2010, 18 pages, 2010.

[55] R.Wakikawa et al.: Global HA to HA Protocol Specification. IETF Draft draft-wakikawa-mext-global-haha-spec-02 (2011).

[56] Soliman, H. et al.: Hierarchical Mobile IPv6 (HMIPv6) Mobility Management. IETF RFC 5380 (2008).

[57] Arkko, J. et al.: Enhanced Route Optimization for Mobile IPv6. IETF RFC 4866 (2007).

[58] Bertin, P. et al.: A Distributed Dynamic Mobility Management Scheme Designed for Flat IP Architectures. New Technologies, Mobility and Security, 2008. NTMS '08. pp. 1 –5 (2008).

[59] Wakikawa, R. et al.: Multiple Care-of Addresses Registration. IETF RFC 5648 (2009).

[60] Tsirtsis, G. et al.: Flow Bindings in Mobile IPv6 and Network Mobility (NEMO) Basic Support. IETF RFC 6089 (2011).

[61] Nordmark, E. et al.: IPv6 Socket API for Source Address Selection. IETF RFC 5014 (2007).

[62] Global internet phenomena report. Technical report, Sandvine, 2011.

[63] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012–2017, white paper, February 6, 2013.

[64] Powers, R.A., "Batteries for low power electronics," Proceedings of the IEEE, vol.83, no.4, pp.687-693, Apr 1995.

[65] Phillips, C.; Singh, S., "An Empirical Activity Model for WLAN Users," INFOCOM 2008. The 27th Conference on Computer Communications. IEEE, vol., no., pp.2065-2073, 13-18 April 2008.

[66] The DICOM Standard, http://medical.nema.org/standard.html

[67] Taubman, D. S., and Marcellin, M. W., "JPEG2000: image compression fundamentals, standards, and practice", Kluwer Academic Publishers, ISBN 978-0-792-37519-7 (2001).

[68] ISO/IEC 15444-1 and ITU-T Rec. T.800, "Information technology – JPEG 2000 image coding system: Core coding system" (2002).

[69] Wavelet-based Coding And its application in JPEG2000, Monia Ghobadi, CSC561 final project, http://www.docstoc.com/docs/52129672/Wavelet-based-Coding-And-its-application-in-JPEG2000

[70] Scalable Wavelet Compression of Higher Dimension Medical Images, William A. Pearlman, IEEE COMSOC MMTC E-Letter

[71] ISO/IEC 15444-10 and ITU-T Rec. T.809, "Information technology – JPEG 2000 image coding system– Part 10: Extensions for three-dimensional data" (2008).

[72] Standardized Technologies for Volumetric Image Coding, Tim Bruylants, Joeri Barbarien Adrian Munteanu, Jan Cornelis and Peter Schelkens, EEE COMSOC MMTC E-Letter 2011.

[73] DICOM. "Digital Imaging and Communications in Medicine (DICOM): Part 5 – Data Structures and Extensions", 2008.

[74] What is Different About Medical Image Compression?, David A. Clunie, IEEE COMSOC MMTC E-Letter Vol.6, No.7, July 2011

[75] Partial encryption of images for real-time applications, Marc Van Droogenbroeck, Fourth IEEE Benelux Signal Processing, Hilvarenbeek, The Netherlands, pages 11-15, April 2004.

[76] Efficient selective encryption for JPEG 2000 images using private initial table, Jiang-Lung Liu, Pattern Recognition, Volume 39, Issue 8, August 2006, Pages 1509–1517

[77] Concerto MS3, Preliminary definition of use cases and requirements.

[78] Fenner B. et al., Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification, RFC 4601, August 2006.

[79] A. Adams, J. Nicholas, W. Siadak, Protocol Independent Multicast - Dense Mode (PIM-DM), RFC 3973, January 2005.

[80] Mzomuhle T. Nkosi, Fisseha Mekuria, Samson H Gejibo, Challenges in Mobile Bio-Sensor Based mHealth Development, In proc. of IEEE 13th International Conference on e-Health Networking, Applications and Services, pp. 21 - 27, Columbia, MO, June 2011.

[81] A. Kollmann, M. Riedl, P. Kastner, G. Schreier, B. Ludvik, “Feasibility of a mobile phone-based data service for functional insulin treatment of type 1 diabetes mellitus patients”, Journal of Medical Internet Research, vol. 9, Mars 2007.

[82] Eva Lai, Karl E. Friedl, Digital Soldiers: Transforming Personalized Health in Challenging and Changing Environments, In proc. 6th International Workshop of Wearable Micro and Nano Technologies for Personalized Health (pHealth), pp.5-8, Oslo, June 2009.

[83] T.Gao, T.Massey, L.Selavo et al., “The Advanced Health and Disaster Aid Network: A Light-Weight Wireless Medical System for Triage,” IEEE Trans. Inf. Tech. Biom., vol.1, no.3,pp. 203-216, 2007.

[84] Elizabeth Fife, Francis Pereira, Digital home health and mHealth: Prospects and challenges for adoption in the U.S, In proc of 50th FITCE Congress (FITCE), pp.1-11, Palermo, Italy, Sept. 2011.

[85] JunHua Li, Nathan Moore, Shahriar Akter, Steven Bleisten, Pradeep Ray: mHealth for Influenza Pandemic Surveillance in Developing Countries, In proc. of 43rd Hawaii International Conference on System Sciences (HICSS), pp. 1 - 9, Honolulu, HI, Jan. 2010.

[86] C. Delgorge, F. Courreges, L. A. Bassit, C. Novales, C. Rosenberger, N. Smith-Guerin, C. Bru, R. Gilabert, M. Vannoni, G. Poisson, P. Vieyres, “A Tele-Operated Mobile Ultrasound Scanner Using a Lightweight Robot,” IEEE Trans. Info. Tech. Biomed., vol. 53, no. 4, pp. 50–58. Mar. 2005.

[87] Bruno M. Silva, Ivo M. Lopes, Joel J. P. C. Rodrigues, Pradeep Ray, SapoFitness: A Mobile Health Application for Dietary Evaluation, In proc. of 13th IEEE International Conference on e-Health Networking Applications and Services (Healthcom), pp.375 - 380, Columbia, MO, June 2011.

[88] Murakami, H., Shimizu, K., Yamamoto, K., Mikami, T., Hoshimiya, N., and Kondo, K., Telemedicine using mobile satellite communication. IEEE Trans. Biomed. Eng. 41:488–497, 1994.

[89] Hwang, S. C., and Lee, M. H., A WEB-based TelePACS using an asymmetric satellite system. IEEE Trans. Inf. Technol. Biomed. 212-215, 2000.

[90] Anogianakis, G., Maglavera, S., and Pomportsis, A., Relief for maritime medical emergencies through telematics. IEEE Trans. Inf. Technol. Biomed. 254–260, 1998.

[91] Lin, C. F., Hung, S. I., and Chiang, I. H., 802.11n WLAN transmission scheme for wireless telemedicine applications. Proc. Inst. Mech. Eng., H J. Eng. Med., 224(10):1201-8., Oct. 2010.

[92] Lin, C. F., and Li, C. Y., A DS UWB transmission system for wireless telemedicine. WSEAS Trans. Syst. 578–588, 2008.

[93] Rasid, M. F. A., and Woodward, B., Bluetooth telemedicine processor for multichannel biomedical signal transmission via mobile cellular networks. IEEE Trans. Inf. Technol. Biomed. 9 (1):35–43, 2005.

[94] Mundt, C. W., Montgomery, K. N., Udoh, U. E., Barker, V. N., Thonier, G. C., Tellier, A. M., Ricks, R. D., Darling, R. B., Cagle, Y. D., Cabrol, N. A., Ruoss, S. J., Swain, J. L., Hines, J. W., and Kovacs, G. T. A. A., Multiparameter wearable physiologic monitoring system for space and terrestrial application. IEEE Trans. Inf. Technol. Biomed. 9(3):382–391, 2005.

[95] Pierucci, L., and DelRe E. E., An interactive multimedia satellite telemedicine service. IEEE Multimed. 76–83, 2000.

[96] Lin, C. F., Chen, J. Y., Shiu, R. H., and Chang, S. H., A Ka band WCDMA-based LEO transport architecture in mobile telemedicine. In: Martinez, L., and Gomez, C. (Eds.), Telemedicine in the 21st Century (pp. 187–201). USA: Nova Science Publishers, 2008.

[97] Hung, K., and Zhang, Y. T., Implementation of a WAP-based telemedicine system for patient monitoring. IEEE Trans. Inf. Technol. Biomed. 7(2):101–107, 2003.

[98] Salvador, C. H., Carrasco, M. P., González, M. A., Carrero, A. M., Montes, J. M., Martín, L. S., Cavero, M. A., Lozano, I. F., and Monteagudo, J. L., Airmed-cardio: A GSM and internet services based system for out-of-hospital follow-up of cardiac patients. IEEE Trans. Inf. Technol. Biomed. 9(1):73–85, 2005.

[99] Pavlopoulos, S., Kyriacou, E., Berler, A., Dembeyiotis, S., and Koutsouris, D., A novel emergency telemedicine system based on wireless communication technology—AMBULANCE. IEEE Trans. Inf. Technol. Biomed. 2(4):261–267, 1998.

[100] Chu, Y., and Ganz, A., A mobile teletrauma system using 3G networks. IEEE Trans. Inf. Technol. Biomed. 452–456, 2004.

[101] Gállego, J. R., Ángela, H. S., Canales, M., Javier, J., Antonio, V., and Julián, F. N., Performance analysis of multiplexed medical data transmission for mobile emergency care over the UMTS channel. IEEE Trans. Inf. Technol. Biomed. 13-22, 2005.

[102] Ghini, V. Ferretti S., and Panzieri F.: M-Hippocrates: Enabling Reliable and Interactive Mobile Health Services. IT Professional. 14, 29 –35, 2012.

[103] National Emergency Number Association (NENA) VoIP/Packet Technical Committee Long Term Definition Working Group, "Next Generation 112 - Long Term Definition". EENA, Revision Date: 06-03-2013

[104] J. Rosenberg et. al., "Session Initiation Protocol". Internet Engineering Task Force, RFC 3261

[105] R. Droms, "Dynamic Host Configuration Protocol". Internet Engineering Task Force, RFC 2131

[106] Sung M. and Pentland A., Livenet: Health and Lifestyle Networking through Distributed Mobile Devices, Tech. report 575, MIT, 2004.

[107] 3GPP TS 23.401 General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access, Rel-10, 2011.

[108] 3GPP TS 23.402 Architecture enhancements for non-3GPP accesses, Rel-10, 2011.

[109] CONCERTO Deliverable D3.1, Description of Preliminary Algorithms for Image/Video Coding, Security, and QoE 2013.Mobile Broadband for Critical Communications Users : A review of options for delivering Mission Critical solutions,, December 2013

[110] https://www.thalesgroup.com/en/worldwide/security/press-release/thales-launches-eiji-suite-secure-mobile-telecommunications

7.2 Glossary

3GPP	3rd Generation Partnership Project
AES	Advanced Encryption Standard
AG	Access Gateway
ALTO	Application Layer Traffic Optimization
ANDSF	Access Network Discovery and Selection Function
AVC	Advanced Video Coding
AWTP	Advanced Waiting time Priority
BG	Border Gateway
CBR	Constant bitrate
CDN	Content Delivery Networks
CN	Correspondent Node
CP	Collection Point
CT	Computer Tomography
DAPSK	Differential Amplitude and Phase Modulation
DDE	Distributed Decision Engine
DICOM	Digital Imaging and Communications in Medicine
DLC	Data Link Control
DLL	Data Link Layer
DMM	Distributed and Dynamic Mobility Management
EDCA	Enhanced Distributed Channel Access
eHealth	Electronic Healthcare
EP/EC	Event Producer or Event Consumer
EPS	Evolved Packet System
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
FEC	Forward Error Correction
FTV	Free Viewpoint Television
FVV	Free Viewpoint Video
GBR	Guaranteed bit rate
GCR	GroupCast with Retries
GHAHA	Global Home Agent to Home Agent protocol
GMPLS	Generalized Multi-Protocol Label Switching
HA	Home Agent
HTTP	Hypertext Transfer Protocol
HTTPS	HTTP Secure
I2AEX	Infrastructure to Application Information Exposure
ICB	Internal Communication Bus
ICT	Information and Communication Technology
IETF	Internet Engineering Task Force
JPEG	Joint Photographic Experts Group
LDPC	Low-density parity-check code
LTE	Long Term Evolution
LTE-A	LTE-Advanced
MAC	Medium Access Control
MAP	Mobility Anchor Point
MCT	Multiple Component Transformations
mHealth	Mobile Healthcare
MIMO	Multiple-input and multiple-output
MIPv6	Mobile IPv6
ML	Maximum Likelihood
MMOG	Massively Multi-player Online Games
MN	Mobile Node
MPLS-TE	MPLS Traffic Engineering
NAT	Network Address Translation
NIST	National Institute of Standards and Technology

OAM	Operation Administration and Management
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PC	Personal Computer
PCRF	Policy charging and rules function
PDM	Proportional Differentiation Model
PDN GW	Packet Data Network Gateway
PHY	Physical layer
PIM-DM	Protocol Independent Multicast – Dense Mode
PIM-SM	Protocol Independent Multicast – Sparse Mode
PL-FEC	Packet-level Forward Error Correction
PSK	Phase-shift keying
QAM	Quadrature Amplitude Modulation
QCI	QoS class identifier
QoS	Quality of Service
QoE	Quality of Experience
RO	Route Optimization
RP	Rendezvous Point
RRC	Radio Resource Control
RTP	Real-time Transport Protocol
RTSP	Real Time Streaming Protocol
SC-FDMA	Single-Carrier Frequency Division Multiple Access
SNMP	Simple Network Management Protocol
SVC	Scalable Video Coding
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TRG	Triggering
TTCA	TETRA + Critical Communications Association
TTCHM	Turbo Trellis Coded Hierarchical Modulation
TTCSPM	Turbo Trellis Coded Superposition Modulation
UDP	User Datagram Protocol
WDS	Wireless Distribution System
WiFi	Wireless-Fidelity
WIMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network