Cross-layer QoE-driven Admission Control and Resource Allocation for Adaptive Multimedia Services in LTE

K. Ivesic*, L. Skorin-Kapov, M. Matijasevic

University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, Zagreb, Croatia

Abstract

This paper proposes novel resource management mechanisms for multimedia services in 3GPP Long Term Evolution (LTE) networks aimed at enhancing session establishment success and network resources management, while maintaining acceptable end-user quality of experience (QoE) levels. We focus on two aspects, namely admission control mechanisms and resource allocation. Our cross-layer approach relies on application-level user- and service-related knowledge exchanged at session initiation time, whereby different feasible service configurations corresponding to different quality levels and resource requirements can be negotiated and passed on to network-level resource management mechanisms. We propose an admission control algorithm which admits sessions by considering multiple feasible configurations of a given service, and compare it with a baseline algorithm that considers only single service configurations, which is further related to other state-of-the-art algorithms. Our results show that admission probability can be increased in light of admitting less resource-demanding configurations in cases where resource restrictions prevent admission of a session at the highest quality level. Additionally, in case of reduced resource availability, we consider resource reallocation mechanisms based on controlled session quality degradation while maintaining user QoE above the acceptable threshold. Simulation results have shown that given a wireless access network with limited resources, our approach leads to increased session establishment success (i.e., fewer sessions are blocked) while maintaining acceptable user-perceived quality levels.

Keywords: Admission control, resource allocation, multimedia services, simulation, LTE

1. Introduction

With the advent of high performance mobile technologies, the capabilities of data transmission on the move have been improved to a great extent. Together with the development of advanced mobile devices, networks like 3GPP Long Term Evolution (LTE) enable the provision of a wide range of multimedia services, like high quality video and online gaming, to mobile users. Ericsson Mobility Report has shown that the volume of mobile data traffic has almost doubled from the second quarter of 2012 to the second quarter of 2013, and further grown 10% between the second and the third quarter of 2013 (Ericsson, 2013). Further increases in mobile traffic volume are expected and capacity shortage is inevitable. This calls for new quality of experience (QoE) driven resource management mechanisms, as the demand and popularity of advanced multimedia services aiming to meet user experience requirements are expected to grow (Agboma and Liotta, 2012; Baraković and Skorin-Kapov, 2013).

Many parameters affect the end user perceived multimedia service quality. At the application layer, examples of such parameters include media type, display resolution, codec type, frame rate, etc. At the network layer, typical parameters to consider are delay, jitter, loss, and throughput. A joint consideration and mapping of these parameters to achievable QoE levels is necessary when making intelligent resource allocation decisions. Resource management mechanisms often focus on lower layer data and Quality of Service (QoS) mechanisms, while neglecting the impact of application layer parameters. For example, typical resource allocation mechanisms may perform high-level traffic classification and prioritization, without considering the knowledge of how given mechanisms actually impact end user QoE, e.g., Nasser and Guizani (2010) classify all services into two categories and distinguish them only by their priorities. Utility functions have been commonly used in the context of formalizing the correlation between network performance and user perceived quality, expressing users degree of satisfaction with respect to corresponding multi-criteria service performance (Reichl et al., 2011).

In our previous work, we have proposed a QoS negotiation and adaptation framework for multimedia sessions (Skorin-Kapov et al., 2007; Skorin-Kapov and Matijasevic, 2009), which has supported the application-level signalling and end-to-end negotiation of a so-called *Media Degradation Path* (MDP) to specify a mapping between session parameters, resource requirements, and corresponding user experience quality levels. The MDP has been defined as

*Corresponding author

Email address: krunoslav.ivesic@fer.hr (K. Ivesic)

an ordered collection of different feasible session *configurations*, where each configuration specifies service operating parameters (e.g., codec, video resolution), corresponding resource requirements of each involved media flow (e.g., required bandwidth), as well as an expected *utility* value. The utility value represents a numerical indicator of user perceived quality with the regarding configuration. For example, consider an audiovisual service that supports different bitrates and encoding types, offering the audio and video flows at different quality levels (e.g., which can be chosen from depending on the type of access network an end user is using, terminal capabilities, operator policy, user subscription type, etc.). In each MDP, there is an optimal configuration (resulting in the highest achievable utility value) and several alternative ones, ordered by decreasing utility value.

During calculation of the MDP, *user* and *service* profiles are taken as input, specifying parameters such as service adaptation capabilities (service profile) and user preferences, e.g., regarding media flow importance and preferences (user profile). In this way, knowledge about the user and the service is embedded in the MDP, and it specifies a "recipe" for controlled degradation of the session to be applied in the case of resource shortage. By *controlled degradation* we refer to the act of switching to an alternative pre-specified session configuration in light of reduced resource availability, rather than experiencing reduced session quality due to uncontrolled degradation of all flows, or relying only on adaptation mechanisms initiated by the application itself. While controlled client-initiated application adaptation has been previously proposed, such as in the context of HTTP Adaptive Streaming (Oyman and Singh, 2012), we focus on network-based mechanisms that consider multiple sessions and domain-wide optimized resource allocation.

In this work we present the application of the MDP to resource management mechanisms, namely (1) admission control, and (2) resource reallocation in case of reduced resource availability, such as that caused by congestion in the network. When a new multimedia session is starting, alternative configurations from the MDP can be used if currently available resources are not sufficient for the optimal configuration, thus increasing admission probability while keeping user satisfaction at an acceptable level, since the alternative configurations reflect user preferences and acceptable quality levels. Additionally, alternative configurations degrade different flows in a different manner, depending on the user preferences. We have introduced this concept in our previous work (Ivesic et al., 2013), which we now extend with more detailed simulation tests used for evaluation of the approach. Additionally, we deal with the problem of resource reallocation in case of resource shortage. Since the multimedia sessions that we consider can, in certain occasions, increase their resource consumption considerably, in comparison to the resources assigned at the admission time, we propose a resource reallocation mechanism in case of reduced resource availability, which conducts graceful degradation by switching active sessions to less resource demanding, but agreed-on configurations. We also build on earlier results (Ivesic et al., 2010, 2011) with extensive validation in a simulated LTE access network. For evaluation purposes, we utilise the simulator tool called ADAPTISE (ADmission control and resource Allocation for adaPtive mulTImedia SErvices) developed by our group at the University of Zagreb (Ivesic et al., 2010, 2011), and the LTE-Sim tool developed by Politecnico di Bari, Italy, for performance evaluation in an LTE network (Piro et al., 2011).

Our tests show that by employing an approach considering MDP calculation, admission probability is increased as compared to a baseline admission control algorithm not considering an MDP. Additionally, we show how our baseline algorithm compares with several works from the literature, thus deriving the general conclusion that the MDP-based algorithm outperforms algorithms that do not consider different service configurations. Furthermore, results show that in cases of reduced resource availability, selected sessions are adapted in a controlled manner while maintaining acceptable quality levels and freeing resources for new incoming sessions. To the best of our knowledge, current literature lacks approaches that perform degradations to alternative configurations that reflect user- and service-related knowledge.

In order to clarify the applicability of the work proposed in this paper, we briefly discuss a possible mapping of our proposed resource management approaches to the standardized LTE Evolved Packet Core (EPC) architecture. The paper is organized as follows. In section 2 we give a survey of related work regarding admission control and resource allocation, as well as a brief overview of resource management mechanisms in LTE networks. Section 3 presents our approaches for resource management based on user- and service-related knowledge. The simulations and analysis of results are given in sections 4 and 5. Section 6 provides a discussion of the proposed approach in the context of standardized LTE QoS management mechanisms, and provides concluding remarks and outlook for future work. We end the paper with a summary of contributions and outlook for ongoing and future work.

2. Related work

2.1. Admission control mechanisms

The basis of many modern admission control algorithms in cellular mobile networks has been established by Hong and Rappaport (1986), who introduced the notion of the *guard channel* as a mechanism for ensuring capacity for handoff calls by reserving a certain number of channels exclusively for them. With the introduction of different user and service categories, other schemes have been introduced, often dividing the available capacity into several zones. For example, in a scheme proposed by Nasser and Guizani (2010) there are two call categories with different priorities and the capacity is divided into four zones.

The first zone is available to all calls, the second one to lower category handoff calls and all higher category calls, the third to all higher category calls, and the last one to higher category handoff calls only. Another approach suggested by Shu'aibu et al. (2011) separates the available capacity based on the bit rate type: there are zones for constant bit rate and zones for variable bit rate services. Additionally, a portion of capacity reserved for handoff sessions can be dynamically adjusted. In either of the cases there is no consideration of multiple media flows nor alternative session configurations. Additionally, there are many admission control algorithms with specific purpose in the literature, e.g., Wang and Qiu (2013) proposed an algorithm specially designed for LTE femtocells.

An approach that enables admission of sessions with different assigned bandwidth values has been proposed by Gudkova and Samouylov (2012). Two call types are assumed, namely voice call and video call, and there are two bit rate values for the video calls: guaranteed, and maximum bit rate. In case of insufficient capacity for an incoming voice call, video calls that have been assigned a maximum bit rate can be degraded to their lower, guaranteed bit rate. Chowdhury et al. (2013) also propose degradation in case of insufficient capacity: non real-time calls can be degraded to admit more real-time ones with separate degradation thresholds for new and real-time handoff calls. The degradation, however, does not consider user preferences in either of the cases. An approach suggested by Posoldova and Oravec (2013) considers session requirements (protocol data unit dropping probability, throughput and average delay) and the available capacity during admission decision. If the requirements can not be fulfilled, a session might still be accepted, but with lowered requirements. However, user preferences are not considered. In recent work, Seppänen et al. (2013) have proposed an admission control and resource management system based on the identification of streams via a two-phased traffic classification algorithm, and taken into account the current and expected state of the network to perform quality estimates and make admission decisions. Their approach relies purely on the network-layer, and does not further consider relevant application-layer information necessary for making quality estimates. Finally, as with other approaches, they do not consider sessions involving multiple interrelated flows, nor the ability to admit one of multiple feasible service configurations based on QoE and resource availability.

Since our MDP-based admission control algorithm cannot be directly compared to the algorithms found in the literature, mainly because they do not consider or include the concept of admitting alternative service configurations according to a service/media adaptation policy (in our case the MDP), we introduce a baseline, non-MDP-based admission control algorithm, named *AC_noMDP*. *AC_noMDP*, as a simple, more generic algorithm, is comparable to approaches found in literature, as summarized in Table 1. Relevant state-of-the-art algorithms are first compared in terms of application, which is general in most cases, but there are also specific purpose algorithms, e.g., those intended for deployment in femtocells, as already mentioned. Further on, we examine the number of considered admission zones in algorithms, since our algorithm relies on a zone-based approach to admission control. Given that we deal with multimedia services, possibly composed of several service media flows, we highlight the number of media flows per session as considered by the algorithms. Since our focus is on adaptive sessions with potentially multiple quality levels, we have also considered the ability of addressed algorithms to adapt session properties and the corresponding resource allocation after a given session has been admitted, as well as whether or not they consider the existence of multiple feasible session configurations. Additionally, we considered several other typical properties of admission control algorithms, namely, network type, number of different service categories considered, session priority basis, handover handling, and admission decision basis. It can be noted that our baseline algorithm *AC_noMDP* is comparable with existing approaches when it comes to zones, service classes, priority values, handover handling and admission decisions. Additionally, it supports sessions consisting of more than a single media flow. Further on, if used in conjunction with our resource management MDP-based algorithm presented in this paper, it enables quality adaptation of active sessions. The MDP-based algorithm is compared to *AC_noMDP* later on in the paper (Section 4), and simulation results have shown that significant improvements (in terms of increased session establishment success) can be achieved by utilising the MDP.

2.2. Resource management

Going beyond admission control, extensive work has addressed resource management in case of network congestion (Meddour et al., 2012). Sharafeddine (2011) focuses on maintaining the quality of voice traffic by proposing softened quality guarantees with controlled tolerance to service degradation in order to substantially reduce required network capacity. A common approach for session degradation in case of congestion is based on maximization of utilities of all the flows. An example given by Shehada et al. (2011) suggests maximization of the utility of video flows and states that such an approach is better than maximization of total bit rate. Brajdic et al. (2011) examine sessions with several media flows of different type and perform degradation by maximizing the sum of utilities of all the flows, with utility functions defined per media flow. A similar approach suggested by Grzech et al. (2010) utilizes penalty functions instead of utility functions and performs degradation by minimizing the total penalty.

A general approach to resource management for sessions with several media flows and predefined configurations has been defined by Lee et al. (1999). In our earlier work (Ivesic et al., 2010, 2011) we already applied such an approach by utilising different configurations from

Table 1: Comparison of approaches from literature to a proposed baseline algorithm (AC_noMDP)

Characteristic	Literature							Proposed approach baseline (AC_noMDP)
	Nasser and Guizani (2010)	Shu'aibu et al. (2011)	Gudkova and Samouylov (2012)	Wang and Qiu (2013)	Chowdhury et al. (2013)	Posoldova and Oravec (2013)	Seppänen et al. (2013)	
Application	General	General	General	Femtocells	General	General	General	General
Zones [#]	4	3	3	3	-	-	-	4
Flows per session	1	1	1	1	1	1	1	1 or more
Network type	Network agnostic	WiMAX	LTE	LTE	Network agnostic	WiMAX	Network agnostic	LTE
Service categories	K service classes	Constant and variable bit rate	Voice, video	Streaming, best effort, conversational	M service classes, distinction between real and non-real time services	Voice, videophone, data, videoconference, IPTV	Interactive, streaming, bulk traffic	5 service classes (see Table 3)
Session priority basis	Service category	Servic category	Voice prioritized	Service category	Real time prioritized	Not specified	2 priority values	3 priority values, defined per session
Handover handling	Prioritized	Prioritized	Not specified	Prioritized for conversational	Prioritized	Not specified	Not specified	Prioritized
Admission decision	Capacity-based	Capacity-based	Capacity-based	Capacity-based	Capacity-based	Capacity- and requirements-based	QoE-based	Capacity-based
Ability to adapt session after having been admitted	No	No	Degrading video to guaranteed bit rate	No	Degrading non-real time calls to admit new and handoff calls	No	QoE-based modifications of streaming and interactive flows	Switching to alternative configurations (if used with Algorithm 2, Section 3.2)
Existence of multiple session configurations	No	No	No	No	Bandwidth degradation thresholds for non real-time calls	Sessions can be admitted with reduced transmission requirements	No	No

the MDP and modelling the resource reallocation problem as a multi-choice multidimensional knapsack problem (MMKP). In this work we take this a step further to evaluate it with simulations and verification in an LTE access network. A comprehensive overview and comparison of user-centric vs. network-centric approaches to utility-based QoE-driven optimization of network resource allocation mechanisms is given in (Skorin-Kapov et al., 2013).

2.3. Resource management and QoS mechanisms in LTE/EPC networks

The LTE network, which is considered as a testbed for our algorithms, enables resource management on several levels. In the following text, knowledge of the LTE architecture is assumed. For interested readers, a more detailed description can be found in (Holma and Toskala, 2011). In the access network, the eNodeB performs scheduling of packets by distributing radio resource blocks to sessions every millisecond, thus being able to prioritize certain services. Scheduling algorithms typically operate in favour of real time services, which are delay sensitive, e.g., the exponential scheduling algorithm EXP (Basukala et al., 2009). In the *Evolved Packet Core* EPC (LTE core network) the key node responsible for QoS management and charging is the Policy and Charging Rules Function (PCRF), which controls setup and modification of radio bearers. The bearer parameters are QoS Class Identifier (QCI), Allocation and Retention Priority (ARP), Guaranteed Bit Rate (GBR) and Maximum Bit Rate (MBR) (3GPP TS 23.203, 2013). In our work we consider the parameters QCI and ARP. QCIs identify nine different traffic classes, each of them specifying different packet forwarding treatment, by specifying resource type, priority, packet delay budget and packet error loss rate, as listed in Table 2. ARP is used for admission control, and also regulates pre-emption by defining relative session priorities for resource reallocation. It consists of the following values:

- priority: 1 to 15, lower value implies higher priority,
- pre-emption capability: possibility of acquiring re-

sources from a lower priority session,

- pre-emption vulnerability: possibility of losing resources in favour of a higher priority session.

At the application layer, QoS and resource management mechanisms are influenced by the IP Multimedia System (IMS) entities, namely Call/Session Control Functions (CSCFs) which are included along the Session Initiation Protocol (SIP) signalling path used to negotiate session parameters. The negotiated flow parameters are then further mapped to their regarding bearers (Poikselkä et al., 2006). In recent editions of LTE Advanced specifications (Release 12), the need for improved QoS and congestion management has been identified (Ali et al., 2013). However, effective congestion management procedures for advanced and dynamic services continue to pose an open research issue.

3. Cross-layer QoE-driven resource management

Our initial approach to quality based optimization was proposed in (Skorin-Kapov and Matijasevic, 2009) whereby a Quality Matching and Optimization Function Application Server (QMO AS) is present along the signalling path, and can participate in the negotiation process at session initiation time. A *service profile* and *user profile* are considered for the purpose of matching profile parameters (e.g., screen resolution and supported network access technologies from the user profile, media types and their parameters from the service profile), resulting with the optimal configuration. Several suboptimal configurations are also calculated by degrading one or more media flows, based on preferences from the user profile and adaptation possibilities from the service profile.

As previously explained, we assume the ordering of suboptimal configurations to be specified in a data structure we refer to as the MDP. We note that the MDP considers user preferences and utility functions for all flows comprising a session when forming a degradation path, rather than calculating optimal adaptation only for a single flow (e.g., as addressed previously in (Wang et al., 2007)). For example, if the service consists of an audio and a video flow, and the user prefers audio, the audio quality will be kept high in alternative configurations and video quality will be degraded. The degradation is conducted based on video adaptation options in the service profile. Following a successful application-level QoS negotiation, a session MDP is signalled to network-level mechanisms responsible for resource management.

Table 2: QoS Class Identifer (QCI) characteristics as standardized according to 3GPP (3GPP TS 23.203, 2013)

QCI	Resource Type	Priority	Packet Delay Budget	Packet Error Loss Rate	Example Service
1	GBR	2	100 ms	10^{-2}	Voice Conversation
2		4	150 ms	10^{-3}	Video Conversation
3		3	50 ms	10^{-3}	Real Time Gaming
4		5	300 ms	10^{-6}	Buffered Video
5	Non-GBR	1	100 ms	10^{-6}	IMS Signalling
6		6	300 ms	10^{-6}	Video, TCP-based
7		7	100 ms	10^{-3}	Voice, Video
8		8	300 ms	10^{-6}	Video, TCP-based
9		9	300 ms	10^{-6}	Video, TCP-based

Since multimedia services may consist of several media flows, the number and properties of which can vary over time, we introduce the notion of the *service state* as a set of media flows that can be active simultaneously at any given time. The service state change happens when a flow is added to or removed from the session. For example, for a complex service such as a collaborative virtual environment (CVE), where users can meet and interact in a 3D virtual world, a video stream or an audio chat can be added to the session consisting of data and 3D world "CVE updates" stream, thus resulting in three different states. These states are termed "state 1", "state 2" and "state 3" (the numbers do not indicate the order, they just serve as labels for distinction). Fig. 1 displays a possible running scenario for such a service: at the beginning of the session, only the CVE is active (i.e., only one flow is active and it corresponds to the exchange of virtual environment and 3D objects related updates) and the session is in state 1. At a certain moment, a video stream is added and the session is in state 2 (i.e., in addition to CVE updates, the session now includes also a video flow). When the video is over, the session goes back to state 1. After a while, an audio chat among multiple users is started and the session moves to state 3 (i.e., environment updates and audio chat traffic is being exchanged in the network), etc. The corresponding MDP of the service is shown in Fig. 2, where configurations are grouped according to the service state they pertain to. If the session needs to be degraded at any time, a less resource demanding configuration from the currently active state is chosen and the session is switched to that configuration.

3.1. MDP-based admission control

The MDP concept is applied to the admission control problem as follows. Similarly to already mentioned approaches in the literature, the available bandwidth is divided into zones. Three different service categories are identified, namely *bronze*, *silver* and *gold*, and the available resources are divided into four different zones accordingly, assuming that higher priority is given to handoff sessions. The first zone is available to all the sessions, the second to bronze handoff sessions and all silver and gold sessions, the third to silver handoff and all gold sessions, and the fourth is reserved for gold handoff sessions, as depicted by Fig. 3a. As the resource consumption is depicted as to increase from left to right, the upper "boundaries" of the four zones are shown as the limit T_l, where l ranges from 0 to 4. Additionally, a limit called *zone critical border* is introduced as a zone-dependent limit for admission of sessions with optimal configurations. The limit is shown in Fig. 3b and denoted B_0. Also in Fig. 3b, for a given amount of

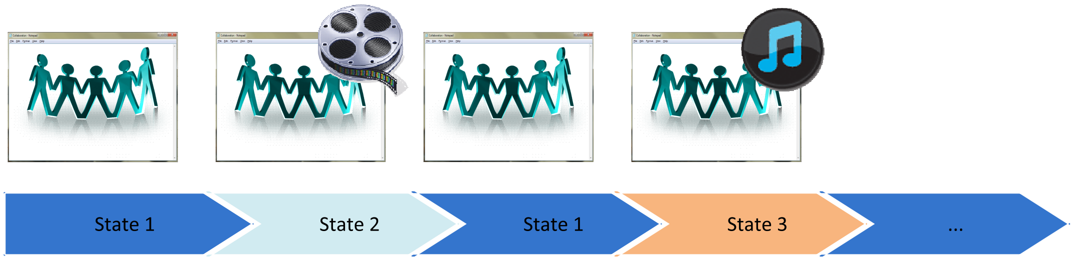
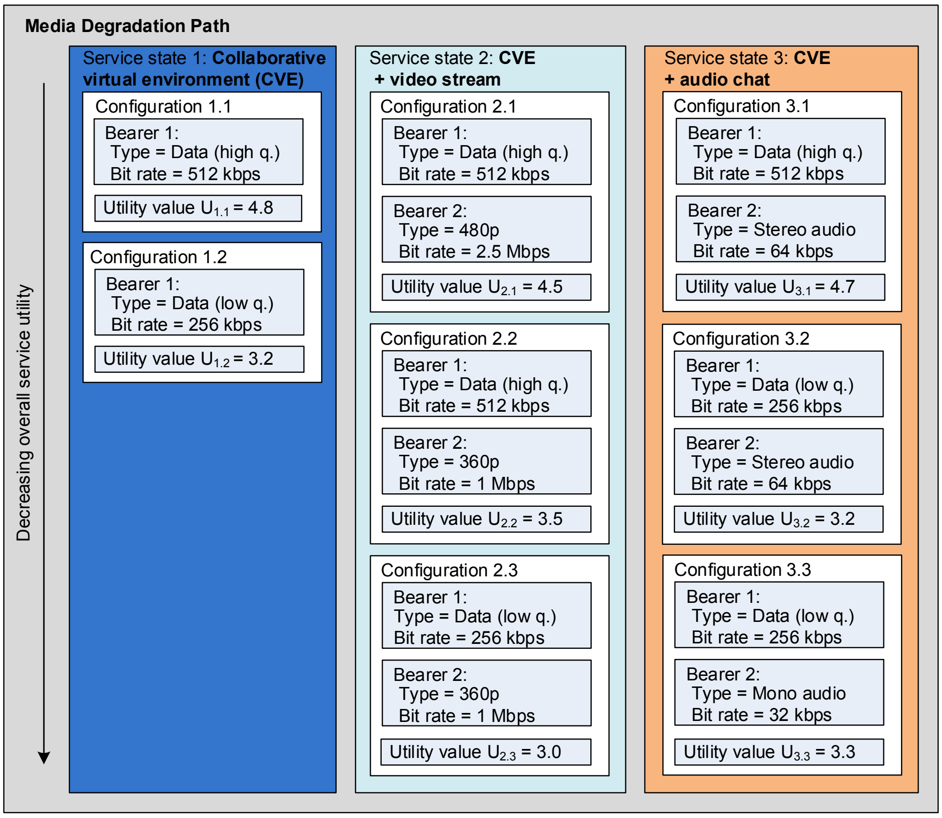


Figure 1: An example session scenario of a service with three states



Service state 1: **Collaborative virtual environment (CVE)**

Configuration 1.1
Bearer 1:
Type = Data (high q.)
Bit rate = 512 kbps
Utility value $U_{1.1}$ = 4.8

Configuration 1.2
Bearer 1:
Type = Data (low q.)
Bit rate = 256 kbps
Utility value $U_{1.2}$ = 3.2

Service state 2: **CVE + video stream**

Configuration 2.1
Bearer 1:
Type = Data (high q.)
Bit rate = 512 kbps
Bearer 2:
Type = 480p
Bit rate = 2.5 Mbps
Utility value $U_{2.1}$ = 4.5

Configuration 2.2
Bearer 1:
Type = Data (high q.)
Bit rate = 512 kbps
Bearer 2:
Type = 360p
Bit rate = 1 Mbps
Utility value $U_{2.2}$ = 3.5

Configuration 2.3
Bearer 1:
Type = Data (low q.)
Bit rate = 256 kbps
Bearer 2:
Type = 360p
Bit rate = 1 Mbps
Utility value $U_{2.3}$ = 3.0

Service state 3: **CVE + audio chat**

Configuration 3.1
Bearer 1:
Type = Data (high q.)
Bit rate = 512 kbps
Bearer 2:
Type = Stereo audio
Bit rate = 64 kbps
Utility value $U_{3.1}$ = 4.7

Configuration 3.2
Bearer 1:
Type = Data (low q.)
Bit rate = 256 kbps
Bearer 2:
Type = Stereo audio
Bit rate = 64 kbps
Utility value $U_{3.2}$ = 3.2

Configuration 3.3
Bearer 1:
Type = Data (low q.)
Bit rate = 256 kbps
Bearer 2:
Type = Mono audio
Bit rate = 32 kbps
Utility value $U_{3.3}$ = 3.3

Figure 2: Example of a Media Degradation Path for a collaborative virtual environment service

"actual" resource allocation (consumption) in any given moment, the amount of remaining available ("free") resources may be seen as spanning from the direction opposite to resource increase. It is represented by the area between F and T_4.

For the purposes of admission control, the following principle is introduced: if the available resources span across the limit B_0, i.e., if the limit F is to the left of the limit B_0, the session is admitted with the optimal, best QoE configuration. Otherwise, the session can be admitted with an alternative, lower QoE configuration.

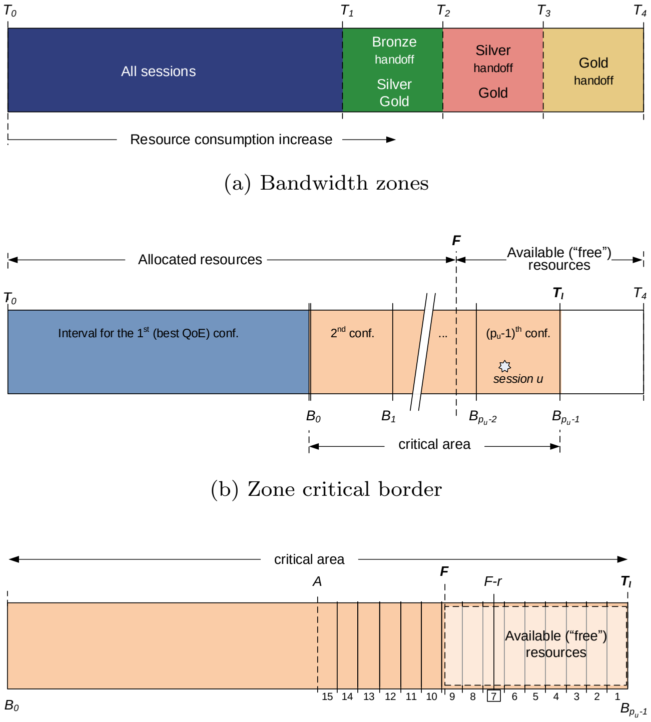
Suppose that the session u pertains to a zone delimited by T_l. Then, the area bounded by the limits B_0 and T_l is called the *critical area*. If the limit F is inside the critical area, the session is admitted with an alternative configuration. For session u with p_u configurations, the critical area is divided into $p_u - 1$ equal intervals, and the choice of the alternative configuration is determined by the relative position of the limit F, i.e., position of F between the limits B_i and B_{i+1} yields the selection of the configuration $i+2$. For example, suppose that the Fig. 3b pertains to a session with 4 different configuration (an optimal one, and 3 alternative ones). Then, the limits B_{p_u-2} and B_{p_u-1} become B_2 and B_3, respectively, and F is between B_1 and B_2, thus denoting the selection of the 3rd configuration. If the free resources do not reach the critical area, i.e., if the limit F is to the right of T_l, there is no space for either of the alternative configurations and the session is blocked.

We consider our approach in the scope of 3GPP Evolved Packet System (EPS). The EPS introduces class based traffic management with specification of different QCIs, as described in section 2 and listed in Table 2. It is assumed that the available bandwidth is divided equally into nine QCIs and the proposed division into zones is conducted within each QCI. As different flows of the same session may be assigned different QCIs, when a new session is about to start, the resource consumption among QCIs is examined (for QCIs required by the flows of the sessions), the QCI with the least amount of available bandwidth is identified and is called the *critical QCI*. Then, the examination of zones and configurations (as explained in the previous paragraph) is conducted for the flow pertaining to the critical QCI. If the appropriate configuration is found, the parameters of the other flows from that configuration are also enforced. As an example, consider a session in, say, state 1, with three flows, namely $f1$, $f2$ and $f3$, that will be assigned QCIs 2, 4 and 9, respectively. Suppose that QCI 4 is critical. In that case, the parameters of the flow $f2$ from the MDP will be considered and if, e.g., the parameters of $f2$ from the configuration 1.3 are found to be appropriate, the parameters of the flows $f1$ and $f3$ from the configuration 1.3 will be enforced as well.

After selecting a configuration, an additional check is conducted by considering the value of the ARP parameter. The limit A is defined as the beginning of the area used for checking the ARP priority value, as shown by Fig. 3c (showing the critical area, enlarged). For the flow f pertaining to the critical QCI, with bandwidth from the selected configuration being equal to r and assigned a zone bounded by T_l, the interval from A to T_l is divided into 15 subintervals (pertaining to 15 ARP priority levels). The session is admitted if its ARP priority is less than or equal to the subinterval indicated by the limit $F - r$, obtained by subtracting the value r from the current value of F. Otherwise, the session is blocked. For example, in Fig. 3c, the limit $F - r$ points to a subinterval with priority equal to 7 and only sessions with this or a higher ARP priority will be admitted (note: lower number means greater priority). The pseudocode is presented in Algorithm 1. Since Algorihtm 1 is invoked each time a new session starts, its execution time depends on the particular sessions properties. Its complexity is $O(n)$ where n stands for the sum of the number of flows and the number of configurations.

Our approach provides an enhanced admission control procedure that considers both the knowledge about the user and the service. The admission probability is increased in comparison to solutions that do not consider alternative service configurations. This is due to the fact that in the case of resource shortage, sessions can still be admitted given they are adapted to an alternative configuration from their MDP that has lower resource requirements. End user QoE corresponding to the delivery of these configurations, expressed in the form of utility functions, is lower than the QoE that would be obtained with the optimal configuration. However, admission of an alternative and agreed on configuration provides a preferable solution as compared to session blocking, both from the perspective of the end user and the service/network provider.

Continuing with the previous example, suppose that the new session, the MDP of which is displayed in Fig. 2, is incoming and that the state 2 is the first state to be active. If the combination of current resource availability and the session priority does not allow the session to be admitted with the optimal configuration (configuration 2.1), that has the highest utility value and the highest resource demands for the CVE and the video stream flows, the session can still be admitted with either of the configurations 2.2 or 2.3. The quality and the resource demands in these alternative configurations are decreased in comparison to configuration 2.1, according to user's preferences. The actual effect on the multimedia service experience varies, based on how the user preferences are set. E.g., in a CVE, the timeliness of virtual world updates is set as a top priority for the user experience, hence in the case of resource shortage, the video flow quality and resource requirements may be decreased (e.g., switch to a lower bitrate codec), while the CVE quality is kept high (e.g., the flow corresponding to CVE updates is kept at a constant rate), which corresponds to the configuration 2.2.



(a) Bandwidth zones

(b) Zone critical border

(c) ARP limit

Figure 3: Bandwidth zones and borders (not drawn to scale)

3.2. MDP-based resource reallocation

After a session has been admitted (with all its constituent media flows), resource consumption can be monitored for triggers on reallocation decisions. While different techniques have been proposed at the traffic management level, such as reassigning flows to different priority queues (Seppänen et al., 2013), we further focus on utilizing available application-level knowledge (i.e., the MDP) in order to provide an efficient and QoE-driven resource management scheme. In case of resource shortage, some sessions

Algorithm 1: MDP-based admission control (AC_MDP)

```
forall the flows do
    determine the zone
    determine the resource consumption of the regarding QCI
$cQCI$ ← critical QCI; $f$ ← flow assigned $cQCI$
if occupied resources of cQCI $< B_0$ then
    select optimal configuration
else if occupied resources of cQCI $< T_l$ then
    //select best feasible subopt. conf.
    $i \leftarrow 0$
    while $B_i < F$ do
        $i \leftarrow i + 1$
    select configuration $i + 1$
else
    reject session
    return;
$r$ ← bandwidth of $f$ in selected configuration
if ARP priority $\leq$ interval pointed by $(F - r)$ then
    admit session
    return;
else
    reject session
```

can be switched to less resource demanding configurations than the currently active ones. Although some studies have shown that quality fluctuations may be perceived negatively by end users, with lower average quality being perceived better than frequent fluctuations (Thakolsri et al., 2011), we argue that controlled quality modifications are justified when the service state changes. Since we introduce service states and assume that state changes occur in an unpredictable manner (typically, due to user input during the session), we are aware that resource consumption of sessions can increase significantly in case of certain state changes, e.g., adding a video flow to the session. If this causes a significant increase in resource consumption, the QoE of existing sessions might be violated and the remaining free resources for the new session might be decreased. For this reason, we propose an approach whereby the existing sessions are degraded if their resource consumption has increased over a certain threshold, due to state changes.

When the resource consumption surpasses a predefined value, the following degradation procedure is performed: the resource consumption of all currently active sessions is inspected and those sessions that increased their resource consumption since admission time due to state changes are extracted. Among these sessions, further selection is made to determine which sessions to degrade to less demanding resource configurations, in a way that maximizes the total utility value of the extracted sessions, subject to resource constraints. The selection of sessions that will be degraded is an optimization problem, addressed initially in our earlier work (Ivesic et al., 2010, 2011). In this work, we provide an evaluation of the proposed approach based on simulation results and with focus on LTE. We investigate the effects of the proposed solution on resource consumption and we inspect the effects of the proposed solution on QoE. The problem class is MMKP. For completeness, we briefly summarize the mathematical formulation of the problem.

Let n be the number of sessions. For each session, let p_u be the number of configurations in the currently active session u. For each configuration i of the session u, let z_{ui} be the number of media flows, such that the flows $1, ..., h_{ui}$ pertain to the downlink direction and the flows $h_{ui} + 1, ..., z_{ui}$ pertain to the uplink direction. Let the vector $\mathbf{b}_{ui} = (\mathbf{b}_{ui1}, ..., \mathbf{b}_{uiz_{ui}})$ contain the bandwidth requirements of configuration i, where the vectors $\mathbf{b}_{uij} = (b_{uij1}, ..., b_{uij9})$ contain bandwidth requirements of the flow j in each QCI. Since each flow is assigned exactly one QCI, the value in the vector $\mathbf{b}_{uij}$ pertaining to the QCI assigned to the flow j contains its bandwidth requirements, while the other values of the vector are equal to zero. It is assumed that the utility value of each configuration is known and denoted by $U(\mathbf{b}_{ui})$. In order to enable fair comparison of utility values of all the sessions, it is required to normalize their values, which is performed by dividing all utility values of the session's current state by the highest value, which pertains to the first (optimal) configuration:

$$U_n(\mathbf{b}_{ui}) = \frac{U(\mathbf{b}_{ui})}{U(\mathbf{b}_{u1})},\ i = 1, ..., p_u \tag{1}$$

Besides the utility value (focusing on the user perspective), we also consider the operator perspective. Operator profit in delivering a given session is denoted by $P(\mathbf{b}_{ui})$ and calculated as the difference between the revenue $R(\mathbf{b}_{ui})$ and the cost $C(\mathbf{b}_{ui})$. The profit value is also normalized, by dividing all profit values of the session's current state by the highest value (in case of profit this does not have to be the profit of the optimal configuration):

$$P_n(\mathbf{b}_{ui}) = \frac{R(\mathbf{b}_{ui}) - C(\mathbf{b}_{ui})}{\max\limits_i \left[R(\mathbf{b}_{ui}) - C(\mathbf{b}_{ui})\right]},\ i = 1, ..., p_u \tag{2}$$

Each session has its own weight factors that define the priority of the regarding user and the service, namely $w_u^{category}$ and $w_u^{service}$, respectively. For example, users may be assigned to different subscription categories, while services on the other hand may also have different priorities (e.g., VoIP services in general may be given a higher priority than streaming services). The session's priority it then determined as a multiplication of these two factors: $w_u = w_u^{category} \cdot w_u^{service}$. The components of the objective function that summarize the utility values and the profit can now be defined as follows:

$$F_{ut} = \sum_{u=1}^{n} \sum_{i=1}^{p_u} w_u x_{ui} U_n(\mathbf{b}_{ui}) \quad (3)$$

$$F_{op} = \sum_{u=1}^{n} \sum_{i=1}^{p_u} w_u x_{ui} P_n(\mathbf{b}_{ui}) \quad (4)$$

where x_{ui} are binary variables that indicate the selected configurations. The objective function that is maximized is the weighted sum of the total users' utility and the total profit defined by equations 3 and 4. For that purpose we define the weight factors w_{ut} and w_{pr}. Hence, the problem is formulated so as to maximize both the operator profit and overall user utility, using weight factors to assign different weights to these multiple objectives. The resource constraints are defined per QCI as maximum total bandwidth for downlink and uplink, represented by $B_{k_{DL}}$ and $B_{k_{UL}}$, respectively, for the QCI k. Than, the optimization problem can be formulated as follows:

$$max(w_{ut} F_{ut} + w_{pr} F_{op}) \quad (5)$$

such that:

$$\sum_{u=1}^{n} \sum_{i=1}^{p_u} \sum_{j=1}^{h_{ui}} x_{ui} b_{uijk} \leq B_{k_{DL}},\ k = 1, ..., 9 \quad (6)$$

$$\sum_{u=1}^{n} \sum_{i=1}^{p_u} \sum_{j=h_{ui}+1}^{z_{ui}} x_{ui} b_{uijk} \leq B_{k_{UL}},\ k = 1, ..., 9 \quad (7)$$

$$\sum_{i=1}^{p_u} x_{ui} = 1,\ x_{ui} \in \{0, 1\},\ u = 1, ..., n \quad (8)$$

The solution to the optimization problem is the vector $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_n)$, where $\mathbf{x}_u = (x_{u1}, ..., x_{u_{p_u}})$, and the value x_{ui} is equal to 1 if the configuration i of the session u has been selected, otherwise it is equal to 0. Having solved the problem, the affected sessions are switched to new configurations if the value determined by the vector $\mathbf{x}_u$ does not match the currently active configuration of the session u. The algorithm is presented in pseudocode in Algorithm 2.

As for the computation complexity, the complexity of the heuristic used to used to solve the optimization problem in the Algorithm 2 is $O(9np_n + np_n log(p_n) + np_n log(n))$ (Akbar et al., 2006), where n stands for number of sessions, p_n for maximum number of configurations per session and number 9 pertains to 9 different QCIs. Two **for** loops before heuristic have complexity $O(n)$ and $O(np_u)$, respectively. Another **for** loop executed after the heuristic has complexity $O(np_n)$. Thus, the whole algorithm has complexity $O(n + 11np_n + np_n log(p_n) + np_n log(n)$ which can be simplified to $O(np_n log(p_n) + np_n log(n))$. One should note that this represents the worst case scenario, i.e., the one where all sessions have to be optimized. In practise, however, only sessions that have increased their resource consumption since admission will be considered by the optimization problem. Since the utility and profit values, and thus the objective function, are calculated based on the configurations' bandwidth requirements, a correlation in the data set is present, thus making the resulting MMKP problem more difficult (Han et al., 2010). However, this is also expected to be the case in the real network.

Algorithm 2: MDP-based Resource reallocation

```
$S \leftarrow \emptyset$
for each session $u$ do
    if $u$ increased its resource requirements then
        $S \leftarrow S \cup u$
for each $u$ in $S$ do
    calculate weight factor $w_u$
    for each configuration $i$ in $u$ do
        calculate $U_n(\mathbf{b}_{ui})$
        calculate $P_n(\mathbf{b}_{ui})$
//solve the optimization problem in $S$
determine the vectors $\mathbf{x}_u$ that maximize the obj.
function
for each $u$ in $S$ do
    /* enforce the new configs. for the
    degraded sessions */
    for $i \leftarrow 1$ to $p_u$ do
        if $x_{ui} = 1$ and config. $i$ in session $u$ inactive
        then
            enforce config. $i$ in session $u$
```

4. Simulations

The simulator tool ADAPTISE (ADmission control and resource Allocation for adaPtive mulTImedia SErvices), demonstrates the complexity of provisioning of adaptive multimedia services. The tool simulates the behaviour of sessions, including arrival, duration, resource allocation, and state changes. The aforementioned algorithms for admission control and resource allocation have also been implemented. Five types of multimedia services have been implemented:

- 3D virtual environment (VE),
- massive multiplayer online role playing game (MMORPG),
- video call,
- voice call,
- streaming video.

For each of these services, up to four different service states (in the context of MDP) have been identified. The number of media flows per state ranges from one to three (depending on the service). ADAPTISE has an abstract view on resource consumption: the resource allocation has been implemented as the assignment of a portion of predefined available bandwidth. For that reason, the simulations conducted using ADAPTISE are further verified in an LTE network simulator, namely LTE-Sim, developed by the Politecnico di Bari in Italy (Piro et al., 2011). LTE-Sim supports four types of flows: constant bit rate, video, VoIP, and infinite buffer. The mapping between ADAPTISE services and flows in LTE-Sim is shown in Table 4. The high level view on the verification methodology is shown in Fig. 4: the simulation is run in ADAPTISE, according to the provided simulation parameters, and it is focused on application layer aspects. In case of the admission control algorithm, the simulation is first run with the proposed AC_MDP algorithm and then repeated with algorithm that does not consider MDP, thus making one pair of simulation traces. In case of the resource reallocation algorithm, the traces are captured before and after the optimization process. Simulation traces are used as input to the LTE-Sim simulator, which focuses on the radio link effects. The simulations are rerun in LTE-Sim, another set of traces is captured from the LTE-Sim and is used for the analysis. The remainder of the section provides the detailed explanation of the whole process.

Fig. 5 depicts the ADAPTISE block diagram (a) and GUI (b). The tool has been written in the Java programming language and implements an event-based simulation. When a new simulation is initialized, a dialog box appears, enabling selection of services to be simulated, along with the parameters of their interarrival and duration times distributions. The following distributions are supported: exponential, normal, lognormal, and Erlang. Having selected services and their parameters, the events of the whole simulation (arrivals, state changes and terminations) are generated and sorted in the queue (Fig. 5a). Then, the events are processed one by one. In the case of a session arrival event, the admission control module is invoked and initial resources are allocated if the session is admitted. The resource consumption is monitored constantly and the optimization process is invoked when necessary. We implemented the heuristic algorithm developed by Akbar et al. (2006) for solving the MMKP problem in terms of milliseconds, as the search for the optimal solution would be too time consuming (Ivesic et al., 2010), given the time critical nature of modifying ongoing sessions, e.g., handover induced service latency is around 1 s (Wu and Tu, 2013). During simulations, the messages are written to the standard output and the most important ones are written in the textbox at the bottom of the simulator window.

The simulator Graphic User Interface (GUI) is shown in Fig. 5b. The menu bar on the top contains simulation control commands (start, stop, pause). When a new session starts, it is added to the bottom of the table labelled "Active sessions", which contains the list of all sessions with their numbers, starting times and unique names. The matrix in the middle of the window, labelled "Active configurations", visualizes the sessions as columns. Each column pertains to a single session and consists of several squares. A square indicates the active configuration, while the white ones indicate the remaining configurations from the currently active state. Grey (shaded) columns indicate terminated sessions. If the session is blocked by the admission control, the corresponding row and column are coloured red. Upon selection of the session, its description is displayed in the panel labelled "Selected session", showing all the states, their configurations and bandwidth requirements of flows along the QCIs. Additionally, the panel displays the values of the ARP parameter and the handoff indication. The panel in the middle labelled "Optimization parameters" contains sliders for adjusting the w_{ut} and w_{pr} weight factors and gauges that indicate bandwidth usage in QCIs. The console at the bottom displays the important messages regarding the simulation.

The evaluation methodology was as follows. First, we ran independent simulation instances in ADAPTISE, with parameters shown in Table 3. Our intention was to demonstrate the behaviour of the proposed algorithms in a "generic" multi service environment, with particular interest in complex multimedia services such as a 3D VE (allowing user interactions and integrated media components in a 3D virtual environment) and MMORPG. We are aware of the existence of session duration models for MMORPGs (Suznjevic et al., 2013; Svoboda et al., 2007) and virtual worlds (Ferreira and Morla, 2010), however, their ratio in current traffic mixes is very low, i.e., according to Cisco Visual Network Index (Cisco, 2014), in 2014 approximately 0.05% of Internet traffic will pertain to online gaming. Thus, we did not intend to generate a traffic mix reflecting the current traffic; instead, we aimed to create a mix (potentially envisioned in the future with advanced and demanding service scenarios) equally representing different traffic categories and with varying behaviour among different categories, i.e., different states and state changes. (In future work we plan to modify this "generic" mix to a more realistic "flavour"). Additionally, we set the weight factors w_{ut} and w_{pr} to 1 and 0.5, respectively, in order to assign higher importance to the objective of maximizing overall utility, as opposed to maximizing profit. The zone limits T_1, T_2 and T_3 have been set to 85%, 90% and 95% of the available resources, respectively. The limit B_0 has been set to 65% of the available resources within the given zone, and the limit A has been set in the middle between the limits B_0 and T_l for session u. These limits have been chosen based on extensive tests that we conducted, which showed these limits to result in the highest number of admitted sessions.

Given the interarrival and duration distribution parameters portrayed in Table 3, it is possible to predict the expected number of sessions, if the interarrival time distribution is exponential with mean λ and the duration distribution's mean μ is known (Adan and Resing, 2002). The system can be modelled as $M/G/\infty$ queue and the expected number of sessions is Poisson distributed with the mean $\lambda\mu$, assuming every session is admitted.

For both algorithms, we ran 10 instances in ADAPTISE with these parameters, and for each instance, we ran 15

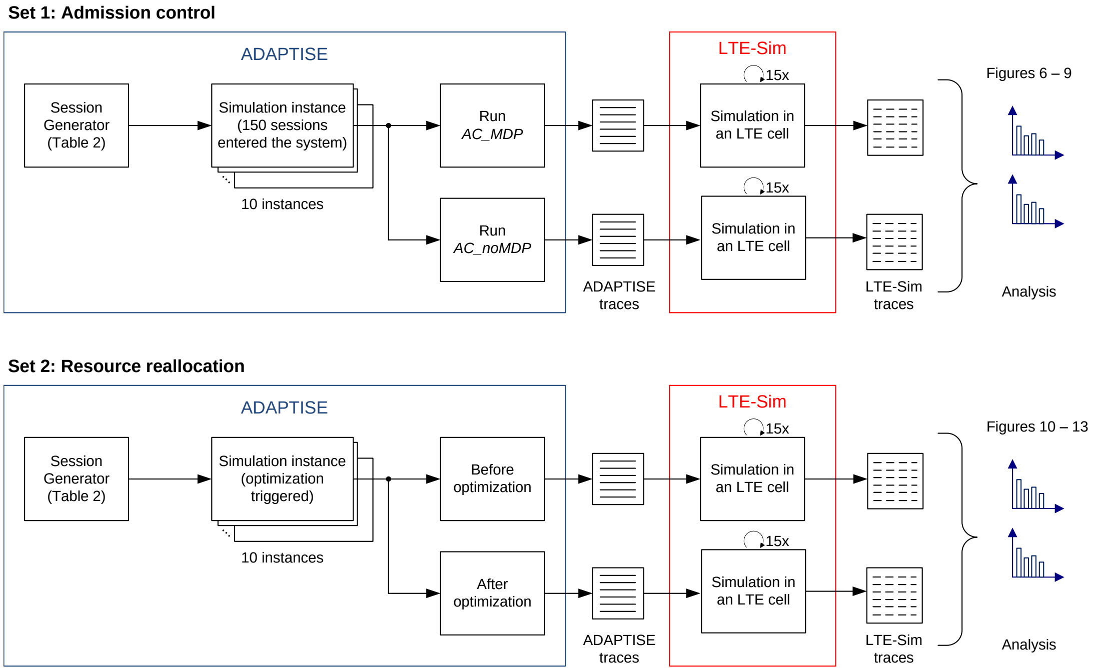

Set 1: Admission control
ADAPTISE
Session Generator (Table 2)
Simulation instance (150 sessions entered the system)
10 instances
Run AC_MDP
Run AC_noMDP
ADAPTISE traces
LTE-Sim
15x
Simulation in an LTE cell
15x
Simulation in an LTE cell
LTE-Sim traces
Figures 6 – 9
Analysis
Set 2: Resource reallocation
ADAPTISE
Session Generator (Table 2)
Simulation instance (optimization triggered)
10 instances
Before optimization
After optimization
ADAPTISE traces
LTE-Sim
15x
Simulation in an LTE cell
15x
Simulation in an LTE cell
LTE-Sim traces
Figures 10 – 13
Analysis

Figure 4: Verification methodology

Table 3: ADAPTISE Session Interarrival Time and Duration Parameters

Service	**Interarrival time**			**Duration**			**Expected sessions [#]**
	Distribution	**Parameters**	**Mean [s]**	**Distribution**	**Parameters**	**Mean [s]**	
VE	Exponential	$\lambda = 0.0001$	10	Exponential	$\lambda = 0.00001$	100	10
MMO	Exponential	$\lambda = 0.0001$	10	Normal	$\mu = 100000, \sigma = 30000$	100	10
Video call	Exponential	$\lambda = 0.0001$	10	Lognormal	$\mu_L = 9.5, \sigma_L = 2$	99	10
Voice call	Exponential	$\lambda = 0.0001$	10	Erlang	$r = 30, \mu = 0.00001$	100	10
Video streaming	Exponential	$\lambda = 0.0001$	10	Exponential	$\lambda = 0.00001$	100	10
Total:							50

instances in LTE-Sim, in order to validate the algorithms in the LTE network. All simulation instances were 10 seconds long and all session activity was logged, including starting times, durations, types, and assigned bit rates of currently active flows. Additionally, any state changes occurring in the observed interval were logged as well. Thus, the whole state of the system in a particular interval was saved. The re-runs in LTE-Sim have been extended by 200 ms, to allow the simulation to complete. Otherwise, since LTE-Sim simulates concrete IP packets, without this time extension, the packets sent at the end of the simulation would be considered lost.

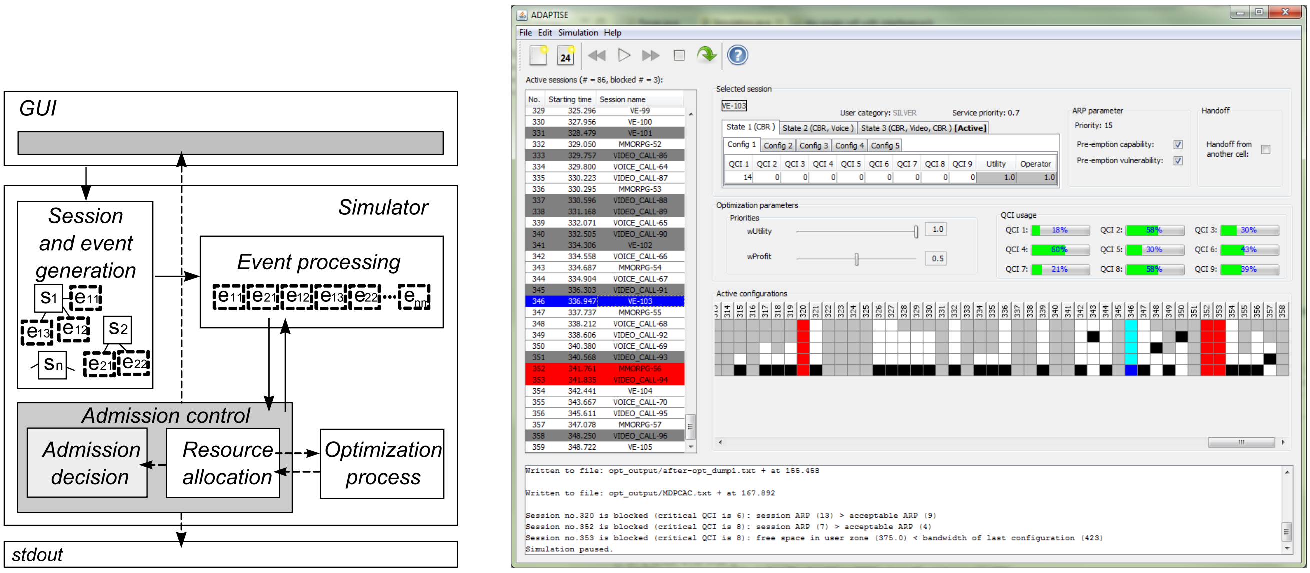
Since LTE-Sim supports only one traffic class at the time of this writing, we set all the flows to use the same QCI in ADAPTISE. The simulation parameters specific to LTE-Sim are listed in Table 5. Among the supported scheduling algorithms in LTE-Sim, we chose the EXP rule algorithm, since our tests showed that it ensures the best treatment of real-time media flows, as they are delay-sensitive. The number of cells has been set to 1 for simplicity. Other parameters have been set as in (Piro et al., 2011). The details of each algorithm are explained in the following paragraphs. Each of the proposed algorithms has been tested independently, as follows.

Table 4: Mapping of ADAPTISE Sessions to LTE-Sim Flows

Service	**States**	**LTE-Sim flows**
3D VE	Virtual world	CBR
	Virtual world + voice chat	CBR, VoIP
	Virtual world + video stream	CBR, Video, CBR
MMORPG	Gaming	CBR
	Gaming + audio chat	CBR, VoIP
	Gaming + download	CBR, CBR
	Gaming + video stream	CBR, Video, CBR
Video call	Video and audio	Video, CBR
	Audio only	CBR
Voice call	Voice	VoIP
Streaming	Video stream	Video, CBR

4.1. Simulations for MDP-based AC algorithm

We ran 10 independent ADAPTISE simulation instances with the MDP-based admission control algorithm (AC_MDP) and the total available bandwidth set to 8000 kbit/s (the value has been chosen to achieve testing conditions, i.e., a certain number of sessions has been re-



Active sessions (# = 86, blocked # = 3):

No.	Starting time	Session name
329	325.296	VE-99
330	327.956	VE-100
331	328.479	VE-101
332	329.050	MMORPG-52
333	329.757	VIDEO_CALL-86
334	329.800	VOICE_CALL-64
335	330.223	VIDEO_CALL-87
336	330.295	MMORPG-53
337	330.596	VIDEO_CALL-88
338	331.168	VIDEO_CALL-89
339	332.071	VOICE_CALL-65
340	332.505	VIDEO_CALL-90
341	334.306	VE-102
342	334.558	VOICE_CALL-66
343	334.687	MMORPG-54
344	334.904	VOICE_CALL-67
345	336.303	VIDEO_CALL-91
346	336.947	VE-103
347	337.737	MMORPG-55
348	338.212	VOICE_CALL-68
349	338.606	VIDEO_CALL-92
350	340.380	VOICE_CALL-69
351	340.568	VIDEO_CALL-93
352	341.761	MMORPG-56
353	341.835	VIDEO_CALL-94
354	342.441	VE-104
355	343.667	VOICE_CALL-70
356	345.611	VIDEO_CALL-95
357	347.078	MMORPG-57
358	348.250	VIDEO_CALL-96
359	348.722	VE-105

User category: SILVER Service priority: 0.7

State 1 (CBR) | State 2 (CBR, Voice) | State 3 (CBR, Video, CBR) **[Active]**

QCI 1	QCI 2	QCI 3	QCI 4	QCI 5	QCI 6	QCI 7	QCI 8	QCI 9	Utility	Operator
14	0	0	0	0	0	0	0	0	1.0	1.0

ARP parameter
Priority: 15
Pre-emption capability: ☑
Pre-emption vulnerability: ☑

Handoff from another cell: □

wUtility 1.0
wProfit 0.5

QCI usage

QCI 1: 18%	QCI 2: 58%	QCI 3: 30%
QCI 4: 60%	QCI 5: 30%	QCI 6: 43%
QCI 7: 21%	QCI 8: 58%	QCI 9: 39%

```
Written to file: opt_output/after-opt_dump1.txt + at 155.458

Written to file: opt_output/MDPCAC.txt + at 167.892

Session no.320 is blocked (critical QCI is 6): session ARP (13) > acceptable ARP (9)
Session no.352 is blocked (critical QCI is 8): session ARP (7) > acceptable ARP (4)
Session no.353 is blocked (critical QCI is 8): free space in user zone (375.0) < bandwidth of last configuration (423)
Simulation paused.
```

(a) Block diagram

(b) Graphical user interface

Figure 5: ADAPTISE simulator

Table 5: LTE-Sim Simulation Parameters

Parameter	Value
Simulation duration	10.2 s
Number of cells	1
Cell radius	1 km
Bandwidth	10 MHz
Scheduler	EXP rule
Frame structure	Frequency-division duplex
User speed	3 km/h

jected). Due to previously mentioned limitations of LTE-SIM, we consider all traffic as belonging to one QCI class. Each of these instances was run from the beginning to the moment when the total number of sessions that arrived into the system (regardless of whether they were admitted or not) reached 150, i.e., three times the expected number of sessions. The goal was to skip the beginning of the simulation, since the starting number of sessions is zero, and the admission control algorithm is not relevant at that point. At the moment when the 150^{th} session arrives, the state of the simulation is saved in a log file (a 10 second time interval, with the middle in the moment of the arrival of the 150^{th} session). The simulation is then repeated from the start, but this time without MDP, i.e., after the session arrives, the admission decision is performed by considering the first configuration only: if there is not enough space for the optimal configuration, the session is dropped, as listed in Algorithm 3 (AC_noMDP). Again, after the arrival of the 150^{th} session, the state is logged. This makes one pair of simulation instances (admission control with and without MDP). Ten such pairs have been created. For each pair, 30 instances in LTE-Sim have been run (15 for each trace), thus making a total of 300 LTE-Sim traces (10 x 2 ADAPTISE traces, each re-run 15 times in LTE-Sim).

The analysis of the impact of the admission control algorithm (AC_MDP) has then been conducted based on the LTE-Sim traces, and it is presented in section 5. A comparison is made with the baseline algorithm (AC_noMDP) in terms of number of admitted sessions and resulting QoS threshold violations, indicating achieved session quality. To the best of our knowledge, there are no previous approaches in the literature that deal with admission control for sessions composed of multiple media flows with different quality configurations. The idea was to show that the number of admitted sessions increases when the proposed MDP-based algorithm is applied, but with little or no negative influence on the overall QoE.

Algorithm 3: Admission control without MDP (AC_noMDP)

```
forall the flows do
    determine the zone
    determine occupied resources in the regarding QCI
    if occupied resources of QCI >= $B_0$ then
        reject session
        return;
select optimal configuration
admit session
```

4.2. Simulations for MDP-based resource reallocation algorithm

In a separate set of traces, we evaluate the performance of our MDP-based resource reallocation algorithm, targeted towards optimally reallocating resources for previously admitted sessions in light of resource shortage. The threshold for invocation of the resource reallocation algorithm has been set to 85% of the available resources, and the total available bandwidth has been set to 10000 kbit/s (a larger value was required, in comparison to admission control simulations, since the number of session has been increased because every session has been admitted). For the affected sessions (i.e., those sessions subject to modification and reallocation of resources), the new maximum resource consumption constraint is set to 75% of the resources previously occupied by these sessions. This value is used as input to the optimization algorithm. Thus, the total bandwidth consumption of the affected sessions will be decreased by at least 25% after the optimization (not all sessions have to be degraded, but their total bandwidth consumption will be at least 25% lower).

We logged 10 different instances of the optimization process in ADAPTISE. For each of them, we saved two 10 seconds long traces from ADAPTISE, one before the optimization process was about to run, and the other one upon running it, which makes ten pairs of traces, as it was the case with the tests of the admission control algorithm. Again, after 15 re-runs of each trace in LTE-Sim, we obtained another 300 LTE-Sim traces. We aimed to show that the sessions will benefit from the resource reallocation algorithm, by comparing the resulting network conditions. The analysis of the traces is presented next.

5. Results analysis

The resulting LTE-Sim traces have been processed and the following results have been obtained: for each algorithm, the plots display the number of active sessions, number of active bearers, total cell throughput, and the ratio of violated real-time flows regarding loss. We consider *violation* to refer to the situation in which a given QoS parameter (in our case, loss) was found to be above a specified threshold used to indicate acceptable user perceived quality. Consequently, we consider these violations as proxy measures for QoE. We analysed violations of real-time flows, as specified by ITU-T G.1010 (2001): a VoIP flow was considered violated if its average packet loss had been $\geq$ 4%, or, if its average delay had been $\geq$ 150 ms, while a video flow was considered violated if its average packet loss had been $\geq$ 1%, or, if its average delay had been $\geq$ 150 ms. Additionally, the packets pertaining to real-time flows were dropped from the queue if they had waited longer than the preconfigured delay, which was set to 200 ms in our simulations. Thus, a portion of all lost packets originates from the scheduler, as this behaviour is specific to the schedulers in LTE-Sim that prioritize the *real-time* traffic and the EXP rule that we used is among them. The results are shown in two sets of plots: Figures 6-9 show the results of verification of the admission control algorithm, and the Figures 10-13 show the results of verification of the resource reallocation algorithm.

5.1. Evaluation of MDP-based AC algorithm

The plots pertaining to the admission control algorithm are shown in Figures 6-9. The number of admitted sessions is depicted in Fig. 6. In each of the ten instances, the number of admitted sessions increased and the increase ranges from 12% (instance 7) to 74% (instance 9), with an average of 32%. Without MDP, the average number of active sessions in the traces was 38.9 and with MDP, it increased to 51.2, which is close to the expected number of sessions from Table 3, meaning that almost all sessions were admitted in case of AC_MDP algorithm. The number of bearers also increased in all instances, ranging from 7% to 80% increase (instances 7 and 9, respectively), with an average increase of 29%. The high variability in the increase in the number of sessions and bearers is caused by differences in alternative configurations regarding the number and type of flows: our analysis shows that the increase in number of flows is not uniform among different flow types and the increased number of admitted sessions usually yields a higher percentage of non real-time flows. In other words, for the services containing video flows, the number of admitted sessions pertaining to these services can be increased at the expense of enforcing configurations containing lower quality video flows or no video flows at all.

The throughput increased in all instances in case of the AC_MDP algorithm as compared to the AC_noMDP algorithm (except for instance 7 where it decreased slightly, less than 1%). Real-time bearers have not been violated regarding delay in either instance, both in case of AC_MDP and AC_noMDP algorithm. Considering packet loss, a minor portion of real-time bearers has been violated, as depicted by Fig. 9. It can be noted that the ratio of violated bearers is around 5% in most instances and that the increase in the case of the AC_MDP algorithm is minor. The highest individual increase occurred in instance 8: from 9% to 12%. However, the number of real-time bearers in AC_noMDP and AC_MDP instances was 59 and 65, respectively, meaning that the number of violated bearers increased from 5.33 to 8, in average. Thus, we can note that the number of sessions has increased considerably with the AC_MDP algorithm, at very little or no "expense" regarding real-time service violations. The size of the confidence intervals can be explained by varying conditions across the 15 re-runs of each iteration in LTE-Sim. Namely, while the number of users is constant in each of the 15 reruns in LTE-Sim, as is their speed (3 km/h, according to Table 5), their starting locations and motion patters are random and differ in each re-run. This affects the results because the distance from the cell center has a great impact on signal strength and transmission properties and hence the user mobility leads to variable packet loss patterns.

In addition to our previous work (Ivesic et al., 2013), where we evaluated the AC_MDP algorithm in LTE network simulator, in this paper we verified the previous results with more comprehensive simulations (number of

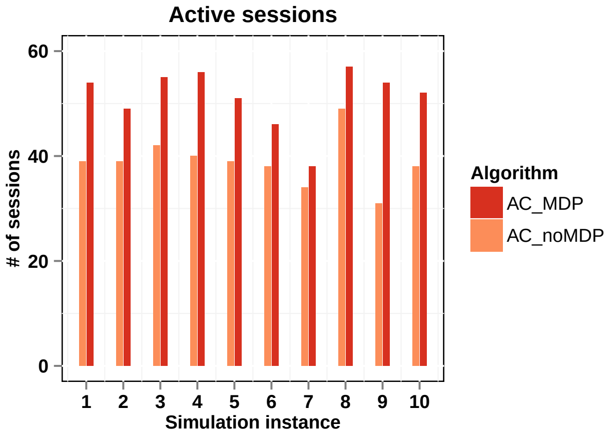


Figure 6: Number of active sessions per simulation instance

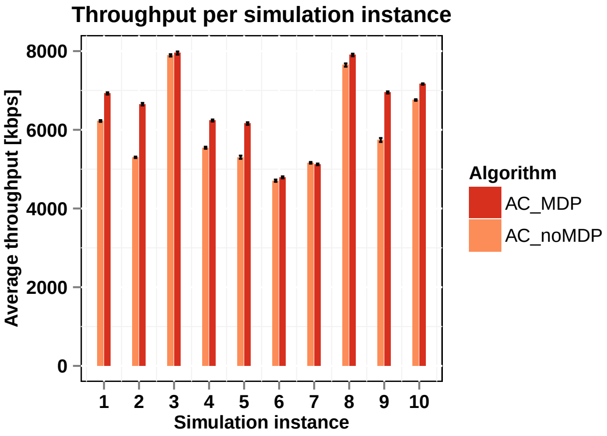


Figure 8: Total throughput with 95% confidence intervals shown

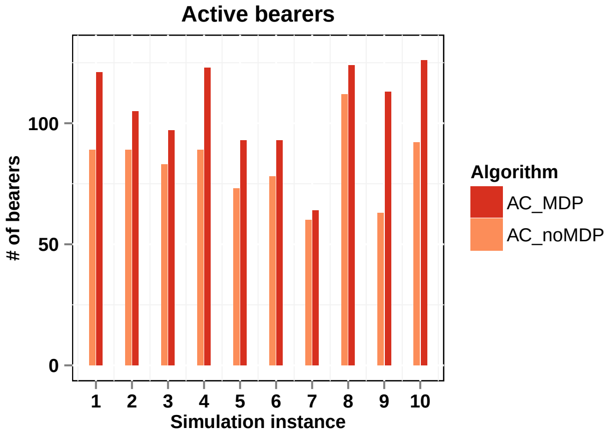


Figure 7: Number of active bearers per simulation instance

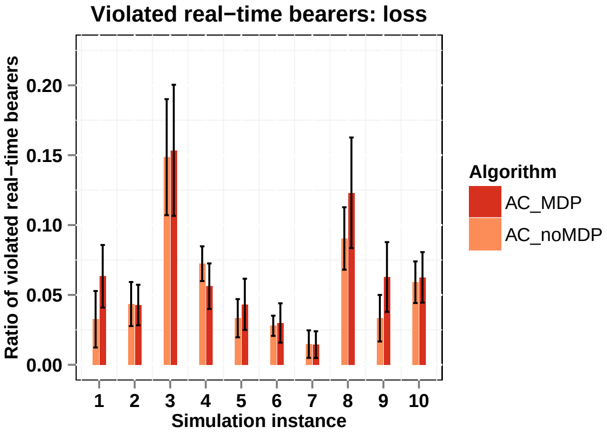


Figure 9: Real-time flows with violated loss threshold, 95% confidence intervals shown

traces and simulation duration have been doubled). In Section 2 we showed that our AC_MDP algorithm cannot be directly compared to the literature so we compared the literature to the baseline algorithm AC_noMDP (listed in Table 1). We assumed that introducing MDP to the process of admission control will increase session admission rate and therefore users' satisfaction. Having implemented MDP-based admission control algorithm in ADAPTISE, we showed that admission probability increased. By re-running the simulations in LTE-Sim the approach has been verified: increase of admission probability has no negative effect on users' perceived quality. We can therefore assume that other algorithms from literature would benefit by applying the concept of different session configurations.

5.2. Evaluation of MDP-based resource reallocation algorithm

Figures 10-13 depict results of the optimization algorithm tests. Each run represents the state of the network before and after invocation of the optimization algorithm. The number of active sessions across different simulation instances is shown in Fig. 10 and, since the algorithm only switches some sessions to lower quality configurations, the overall number of sessions is not affected by it. The number of sessions in individual instances ranges from 48 to 59, and averages at 54.8. Most of these values are greater than the expected number of sessions from Table 3, which can be explained by the fact that the optimization algorithm is invoked when the network is running out of resources due to increased traffic load. It should also be noted that the resource consumption depends not only on the number of sessions, but also on the traffic mix. This is taken into account through the selection of the service type.

The number of bearers is shown in Fig. 11. A slight decrease in the number of bearers is noticeable, 3.3 bearers in average, representing 3.4% of active bearers, which means that very few sessions have been degraded to configurations that discard some bearers. Throughput across the simulation instances is depicted by Fig. 12 and decrease after the optimization ranges from 3.2% to 11.6% with the average of 6.4%. The throughput in the instances is noticeably smaller than the total available throughput (set to 10000 kbit/s) because the optimization threshold was set to 85% of available resources, which was required since the admission control was not used while testing the optimization algorithm. However, with the admission control coupled with the optimization algorithm, this threshold could be increased.

The violation of loss threshold for real-time bearers is shown in Fig. 13. It is evident that the loss decreased after the optimization in most cases. The improvement ranges from 1.1% (instance 2) to 10.9% (instance 4). There are, however, two exceptions: in instance 3 the loss rate remained unchanged, and in instance 7 it increased, but the increase in average value after optimization is inside the confidence interval that describes the state before the optimization. We can thus conclude that the algorithm enhances the network conditions in most cases, and otherwise if it does not contribute to better QoE, it does not degrade the current QoE level. Additionally, by penalizing the "greedy" sessions that increased their resource demands since admission, a portion of resources is freed for the new incoming sessions. The examination of the proposed algorithm in our previous work (Ivesic et al., 2010, 2011) has now been verified by simulations in LTE network. In comparison to congestion handling approaches from literature, where degradation is performed on a per-flow basis (Gudkova and Samouylov, 2012; Chowdhury et al., 2013; Seppänen et al., 2013), we showed that session composed of several flows can be considered as a whole and degraded according to user preferences and predefined quality levels. According to 3GPP (2012), congestion handling can be performed in either a preventive or a reactive way. While preventive way means that actions are taken before congestion occurs, reactive way pertains to action taken after the congestion occurrence, either by the entity experiencing the congestion, or by an external entity. In the latter case congestion information is signalled to the external entity where it needs to be processed rather quickly. Hence, our proposed algorithm is a good candidate since it performs optimal resource reallocation in terms of milliseconds.

The benefits of a cross layer approach stem from taking into account application level knowledge to make intelligent resource allocation decisions that will provide an end user with optimal service quality under given network conditions. Both proposed algorithms build on user- and service-related knowledge and utilise it to offer improved resource management.

6. Applicability of the proposed approach in the context of the LTE/EPC

The relevance of the proposed approach is related to the QoS management mechanisms in the LTE/EPC architecture. While further details are out of scope for this paper, the overall framework is presented.

A mapping of our proposed resource management approaches to the LTE/EPC architecture is depicted in Fig. 14. Our admission control algorithm is designated to be run by the eNodeB, while resource reallocation decisions in case of congestion map to the functionality of the PCRF node. According to 3GPP standards, if a mobile station (MS) is connected to the network and is in idle state, it is assigned a default bearer. Given that multiple applications (with differing QoS requirements) can be running simultaneously from a user device, different bearers supporting different QoS classes may need to be established. When a new bearer (or the modification of the existing one) is needed, an option is for the MS to send a request to the application server (AS) on top of the default bearer (Holma and Toskala, 2011). The AS then sends the request for setup of a session to the PCRF, which creates corresponding Policy and Charging Control (PCC) rules to determine how a certain data flow will be treated (including mapping to QCIs and bit rates). PCC rules are then passed to an underlying Packet Data Network Gateway (P-GW) for QoS enforcement.

Considering the notion of a negotiated and signalled MDP, we assume the MDP to be included in the PCC

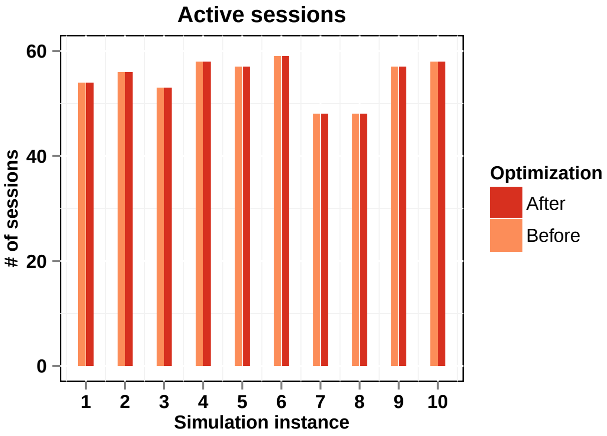


Figure 10: Number of active sessions per simulation instance

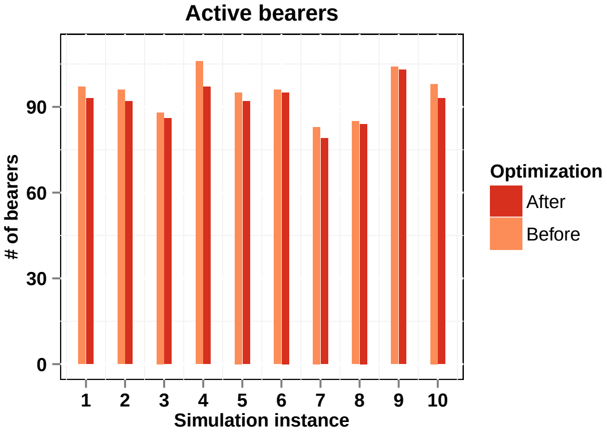


Figure 11: Number of active bearers per simulation instance

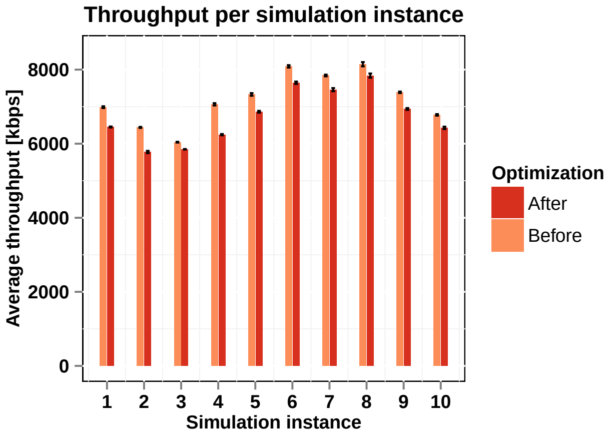


Figure 12: Total throughput with 95% confidence intervals shown

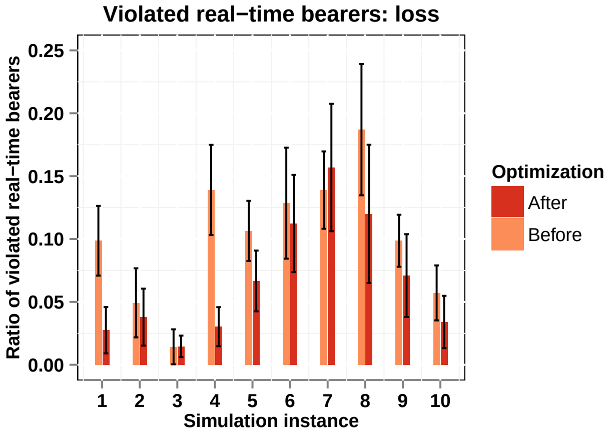


Figure 13: Real-time flows with violated loss, 95% confidence intervals shown

rules (for details regarding the mapping of the MDP to standardized PCC rules for the purposes of charging an interested reader is referred to (Grgic et al., 2009)). The PCRF may request resources (bearer establishment/modification) for the optimal (highest quality) configuration derived from the MDP by sending a request to the P-GW, where it is further passed on to the eNodeB (via the Serving Gateway S-GW) and Mobility Management Entity (MME)). If multiple requested flows are assigned the same QCI, they can be mapped to the same bearer, otherwise, several bearers are required. The eNodeB performs the bearer admission control and passes back a response via the aforementioned nodes to the PCRF. In case of bearer establishment failure, the PCRF now has the knowledge regarding alternative feasible session configurations to attempt establishment of bearer(s) for another configuration from the MDP. The bearer establishment attempt is iterated until there is a successful outcome, otherwise the session is blocked.

The overhead of the proposed solution due to inclusion

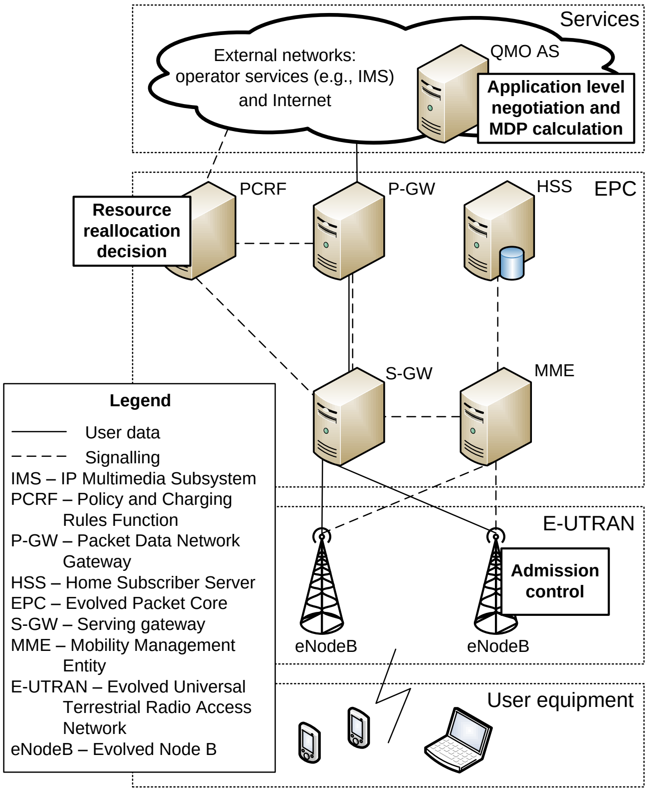

Services
External networks:
operator services (e.g., IMS)
and Internet
QMO AS
Application level negotiation and MDP calculation
EPC
PCRF
P-GW
HSS
Resource reallocation decision
S-GW
MME
Legend
User data
Signalling
IMS – IP Multimedia Subsystem
PCRF – Policy and Charging Rules Function
P-GW – Packet Data Network Gateway
HSS – Home Subscriber Server
EPC – Evolved Packet Core
S-GW – Serving gateway
MME – Mobility Management Entity
E-UTRAN – Evolved Universal Terrestrial Radio Access Network
eNodeB – Evolved Node B
E-UTRAN
Admission control
eNodeB
eNodeB
User equipment

Figure 14: Mapping of the proposed resource management schemes to the LTE architecture (Holma and Toskala, 2011)

of the QMO AS is induced by the calculation of the MDP, discussed in detail in our previous work (Skorin-Kapov and Matijasevic, 2009). We note again that the QMO AS is included along the session establishment signalling path and is responsible for matching client, service, and network parameters in order to derive an optimal session adaptation path (i.e. the MDP). With regards to signalling, the ability to share the MDP information among the service and network layers might obviate the need for a trial-and-error approach to network resource reservation specified in existing 3GPP standards, hence leading to reduced signalling.

Following the calculation of the MDP, we consider signalling overhead during the admission control procedure in the case that more than one session configuration has to be tried: for each configuration (starting from the optimal one), a message is passed from the PCRF to the eNodeB (via P-GW, S-GW and MME), and a response has to be sent back indicating whether or not requested resources are available. This procedure is repeated until a feasible resource allocation can be made. While this procedure may incur additional delay for session establishment, it will allow the network to try multiple session configurations without the need for additional end-to-end signalling exchanges (between session endpoints) in the case that an initial session configuration cannot be established. Consequently, as we have shown, such an approach can lead to an increased number of admitted sessions.

In case of reduced resource availability (e.g., due to

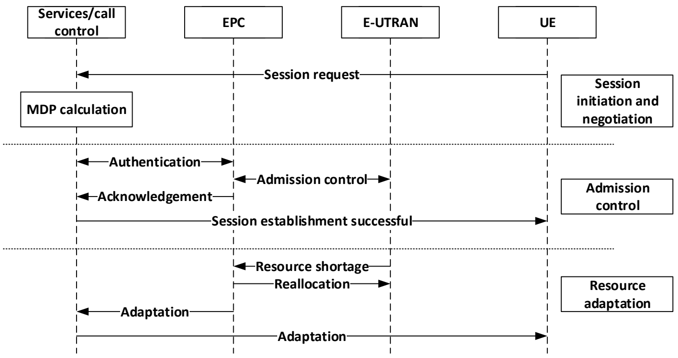


Figure 15: High level overview of signalling procedures

congestion), eNodeB can notify the EPC about resource shortage, and the message is forwarded to the PCRF, which performs degradation of chosen sessions to less resource-demanding configurations derived from stored session MDPs. This decision then propagates to the eNodeB in the form of bearer modification/removal requests, and to the ASs which change the regarding sessions parameters and notify the MSs. The decision on sessions to be degraded is made by using the heuristic, in terms of milliseconds. The only overhead left is signalling of changes to the affected sessions. Since only a small number of sessions (those which increased their resource consumption after admission time) can be affected, the number of signalling messages would not be significant. In terms of performance, these modifications should be quick, e.g., shorter than 1.5 s, as in (Wu and Tu, 2013). The high level overview of signalling procedures required for calculation of MDP and proposed admission control and resource reallocation mechanisms is displayed in Fig. 15.

7. Summary of contributions and future work

In this paper we have presented two approaches for resource management based on incorporation of user- and service-related knowledge in the decision-making process in form of MDP. As a first step we have proposed an admission control algorithm utilizing MDP knowledge. Our results have shown that by employing our AC_MDP algorithm, the average increase (as compared to the AC_noMDP algorithm) in admitted sessions was 32%, while the average increase in the number of bearers was 29%. Furthermore, while the number of sessions has increased considerably, this came at very little or no expense regarding real-time service violations. Employing such an AC algorithm would hence result in the increased session admission rate.

In the second step, we proposed and evaluated an MDP-based resource reallocation and optimization procedure, invoked in light of resource shortage in order to degrade sessions to less resource demanding configurations. Simulation results have shown decreased loss for real-time bearers following the optimization procedure, as compared to before running the procedure, with the average improvement of 3.5%. In this way, additional resources are freed for new sessions and the QoE of the existing sessions is improved.

More extensive simulation traces: By running all of our simulations in LTE-Sim, we were able to assess the performance of the proposed algorithms in a simulated LTE radio access network. In our ongoing work, we are currently running longer simulation traces to further test algorithm performance over an extended period of time. Further, while we have conducted tests with a future envisioned traffic mix incorporating complex multimedia services, we also intend to verify our algorithms with a traffic mix that is simpler in terms of session dynamics and resembling the current traffic patterns in mobile networks.

QoE Assessment: While we have considered the impact on end user QoE from the perspective of performing utility-driven adaptation decisions, improving session establishment success, and meeting QoS requirements (i.e. loss thresholds), we will further aim to make QoE estimations using existing objective QoE prediction models. Such estimations will enable us to quantify overall improvements in terms of estimated Mean Opinion Score (MOS), thus obtaining better insight into the influence of the algorithms on QoE.

Mapping to LTE/EPC: Considering applicability of the proposed approach, while we have briefly discussed a mapping of our mechanisms to standardized LTE/EPC QoS and bearer management mechanisms, we are further investigating a more detailed mapping addressing concrete interfaces and signalling protocols in line with 3GPP standards. In this way we will be able to precisely assess the signalling overhead induced by implementation of the proposed algorithms.

QCIs in LTE-SIM: We note that a current limitation of the LTE-Sim tool is the lack of support for mapping service flows to multiple QCIs. Given that our approach incorporates the notion of mapping different service flows to corresponding QCIs (depending on flow type and resource requirements), as is in line with the standard 3GPP QoS architecture, it would be beneficial to run tests in a simulated LTE environment whereby the eNodeB scheduler accounts for different QoS classes. Hence, future efforts will be targeted towards incorporating this functionality into LTE-Sim, or conducting tests with other potentially available tools.

Acknowledgements

The authors acknowledge the support of the European Community's Seventh Framework Programme under grant agreement no. 285939 (ACROSS), and of the project "Policy Based Resource Control" funded by Ericsson Nikola Tesla, Croatia. The authors would also like to acknowledge the help of Mario Kusek, Iva Bojic and Ivan Dinjar in development of simulation scenarios, and Mirko Suzn-

jevic for internal review and suggestions regarding future work.

References

3GPP . Feasibility study on user plane congestion management. 3GPP TR 22.805, Release 12; 3GPP; 2012.

3GPP TS 23.203 . Policy and charging control architecture. Technical Report 23.203, Release 12; 3GPP; 2013.

Adan I, Resing J. Queueing theory. Department of Mathematics and Computing Science, Eindhoven University of Technology; 2002.

Agboma F, Liotta A. Quality of experience management in mobile content delivery systems. Telecommunication Systems 2012;49(1):85–98.

Akbar MM, Rahman MS, Kaykobad M, Manning E, Shoja G. Solving the multidimensional multiple-choice knapsack problem by constructing convex hulls. Computers & Operations Research 2006;33(5):1259–73. doi:DOI: 10.1016/j.cor.2004.09.016.

Ali N, Taha AE, Hassanein H. Quality of service in 3GPP R12 LTE-Advanced. Communications Magazine, IEEE 2013;51(8):103–9. doi:10.1109/MCOM.2013.6576346.

Baraković S, Skorin-Kapov L. Survey and challenges of QoE management issues in wireless networks. Journal of Computer Networks and Communications 2013;2013:1–28.

Basukala R, Mohd Ramli HA, Sandrasegaran K. Performance analysis of EXP/PF and M-LWDF in downlink 3GPP LTE system. In: Internet, 2009. AH-ICI 2009. First Asian Himalayas International Conference on. 2009. p. 1–5. doi:10.1109/AHICI.2009.5340336.

Brajdic A, Kassler A, Matijasevic M. Quality of Experience based Optimization of Heterogeneous Multimedia Sessions in IMS. In: Baltic Congress on Future Internet Communications. Riga, Latvia; 2011. p. 25–32.

Chowdhury MZ, Jang YM, Haas ZJ. Call admission control based on adaptive bandwidth allocation for wireless networks. Communications and Networks, Journal of 2013;15(1):15–24. doi:10.1109/JCN.2013.000005.

Cisco . Cisco Visual Networking Index. http://www.ciscovni.com/; 2014. Accessed 14th May, 2014.

Ericsson . Ericsson Mobility Report, November 2013. Technical Report; Ericsson AB; 2013. http://www.ericsson.com/res/docs/2013/ericsson-mobility-report-november-2013.pdf, Accessed 10th January 2014.

Ferreira M, Morla R. Second Life In-world Action Traffic Modeling. In: Proceedings of the 20th International Workshop on Network and Operating Systems Support for Digital Audio and Video. New York, NY, USA: ACM; NOSSDAV '10; 2010. p. 3–8. URL: http://doi.acm.org/10.1145/1806565.1806569. doi:10.1145/1806565.1806569.

Grgic T, Ivesic K, Grbac M, Matijasevic M. Policy-based charging in IMS for multimedia services with negotiable QoS requirements. In: Telecommunications, 10th International Conference on. 2009. p. 257–64.

Grzech A, Świtek P, Rygielski P. Dynamic resources allocation for delivery of personalized services. In: Cellary W, Estevez E, editors. Software Services for e-World. Springer Berlin Heidelberg; volume 341 of *IFIP Advances in Information and Communication Technology*; 2010. p. 17–28.

Gudkova IA, Samouylov KE. Modelling a radio admission control scheme for video telephony service in wireless networks. In: Andreev S, Balandin S, Koucheryavy Y, editors. Internet of Things, Smart Spaces, and Next Generation Networking. Springer Berlin Heidelberg; volume 7469 of *Lecture Notes in Computer Science*; 2012. p. 208–15.

Han B, Leblet J, Simon G. Hard multidimensional multiple choice knapsack problems, an empirical study. Comput Oper Res 2010;37:172–81. URL: http://portal.acm.org/citation.cfm?id=1595074.1595252. doi:10.1016/j.cor.2009.04.006.

Holma H, Toskala A, editors. LTE for UMTS: Evolution to LTE-Advanced, Second Edition. John Wiley & Sons, Ltd, 2011. doi:10.1002/9781119992943.ch16.

Hong D, Rappaport SS. Traffic model and performance analysis for cellular mobile radio telphone systems with prioritized and non-prioritized handoff procedures. In: ICC. 1986. p. 1146–50.

ITU-T G.1010 . End-user multimedia QoS categories. ITU-T Recommendation G.1010; 2001. URL: http://www.itu.int/rec/T-REC-G.1010-200111-I/en.

Ivesic K, Matijasevic M, Skorin-Kapov L. Utility based model for optimized resource allocation for adaptive multimedia services. In: IEEE 21st Int. Symp. on Personal Indoor and Mobile Radio Communications (PIMRC). Istanbul, Turkey; 2010. p. 2638–43. doi:10.1109/PIMRC.2010.5671784.

Ivesic K, Matijasevic M, Skorin-Kapov L. Simulation based evaluation of dynamic resource allocation for adaptive multimedia services. In: CNSM 2011. Paris, France; 2011. p. 1–4.

Ivesic K, Skorin-Kapov L, Matijasevic M. Admission control for adaptive multimedia services based on user and service related knowledge. In: EUROCON, 2013 IEEE. 2013. p. 437–45. doi:10.1109/EUROCON.2013.6625019.

Lee C, Lehoczky J, Rajkumar RR, Siewiorek D. On quality of service optimization with discrete QoS options. In: RTAS '99: Proc. of the Fifth IEEE Real-Time Technology and Applications Symp. Washington, DC, USA: IEEE Computer Society; 1999. p. 276.

Meddour DE, Abdallah A, Ahmed T, Boutaba R. A cross layer architecture for multicast and unicast video transmission in mobile broadband networks. Journal of Network and Computer Applications 2012;35(5):1377–91.

Nasser N, Guizani S. Performance analysis of a cell-based call admission control scheme for QoS support in multimedia wireless networks. Int J Commun Syst 2010;23(6–7):884–900. URL: http://dx.doi.org/10.1002/dac.v23:6/7. doi:10.1002/dac.v23:6/7.

Oyman O, Singh S. Quality of Experience for HTTP adaptive streaming services. Communications Magazine, IEEE 2012;50(4):20–7.

Piro G, Grieco L, Boggia G, Capozzi F, Camarda P. Simulating LTE Cellular Systems: An Open-Source Framework. Vehicular Technology, IEEE Transactions on 2011;60(2):498–513. doi:10.1109/TVT.2010.2091660.

Poikselkä M, Mayer G, Khartabil H, Niemi A. The IMS: IP Multimedia Concepts and Services in the Mobile Domain, Second Edition. J. Wiley & Sons, 2006.

Posoldova A, Oravec M. Fuzzy logic based admission control method in wireless networks. In: EUROCON, 2013 IEEE. 2013. p. 431–6. doi:10.1109/EUROCON.2013.6625018.

Reichl P, Tuffin B, Schatz R. Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience. Telecommunication Systems 2011;:1–14.

Seppänen J, Varela M, Sgora A. An autonomous QoE-driven network management framework. Journal of Visual Communication and Image Representation 2013;doi:http://dx.doi.org/10.1016/j.jvcir.2013.11.010.

Sharafeddine S. Capacity assignment in multiservice packet networks with soft maximum waiting time guarantees. Journal of Network and Computer Applications 2011;34(1):62–72.

Shehada M, Thakolsri S, Despotovic Z, Kellerer W. QoE-based cross-layer optimization for video delivery in long term evolution mobile networks. In: Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on. 2011. p. 1–5.

Shu'aibu DS, Yusof SKS, Fisal N, Ariffin SHS, Rashid RA, Latiff NMA, Baguda YS. Fuzzy logic partition-based call admission control for mobile WiMAX. ommunications and Networking 2011;2011. URL: http://dx.doi.org/10.5402/2011/171760. doi:10.5402/2011/171760.

Skorin-Kapov L, Ivesic K, Aristomenopoulos G, Papavassiliou S. Approaches for utility-based QoE-driven optimization of network resource allocation for multimedia services. In: Biersack E, Callegari C, Matijasevic M, editors. Data Traffic Monitoring and Analysis. Springer Berlin Heidelberg; volume 7754 of *Lecture Notes in Computer Science*; 2013. p. 337–58.

Skorin-Kapov L, Matijasevic M. A QoS negotiation and adaptation framework for multimedia services in NGN. In: 10th Intl. Conf. on Telecommunications, ConTEL. Zagreb, Croatia; 2009. p. 249–56.

Skorin-Kapov L, Mosmondor M, Dobrijevic O, Matijasevic M. Application-Level QoS Negotiation and Signaling for Advanced Multimedia Services in the IMS. Communications Magazine, IEEE 2007;45(7):108–16. doi:10.1109/MCOM.2007.382669.

Suznjevic M, Stupar I, Matijasevic M. A model and software architecture for MMORPG traffic generation based on player behavior. Multimedia Systems 2013;19(3):231–53. URL: http://dx.doi.org/10.1007/s00530-012-0269-x. doi:10.1007/s00530-012-0269-x.

Svoboda P, Karner W, Rupp M. Traffic Analysis and Modeling for World of Warcraft. In: Communications, 2007. ICC '07. IEEE International Conference on. 2007. p. 1612–7. doi:10.1109/ICC.2007.270.

Thakolsri S, Kellerer W, Steinbach E. QoE-Based cross-layer optimization of wireless video with unperceivable temporal video quality fluctuation. In: Communications (ICC), 2011 IEEE International Conference on. 2011. p. 1–6. doi:10.1109/icc.2011.5963296.

Wang J, Qiu Y. A New Call Admission Control Strategy for LTE Femtocell Networks. In: 2nd Intl. Conf. on Advances in Computer Science and Engineering. 2013. p. 334–8.

Wang Y, Kim JG, Chang SF, Kim HM. Utility-based video adaptation for universal multimedia access (UMA) and content-based utility function prediction for real-time video transcoding. Multimedia, IEEE Transactions on 2007;9(2):213–20.

Wu JS, Tu JY. A Fast IMS Service Recovery Mechanism for the Handover Over MIH-Capable Heterogeneous Networks. Wireless Personal Communications 2013;68(4):1761–87. URL: http://dx.doi.org/10.1007/s11277-012-0549-y. doi:10.1007/s11277-012-0549-y.