

JOINT NETWORK AND RATE ALLOCATION FOR SIMULTANEOUS WIRELESS APPLICATIONS

D. Jurca¹, W. Kellerer², E. Steinbach³, S. Khan³, S. Thakolsri², P. Frossard¹

¹Signal Processing Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

²Future Networking Laboratory, DoCoMo Communications Laboratories GmbH, Munich, Germany

³Media Technology Group, Technische Universität München, Germany

ABSTRACT

We address the problem of rate allocation and network/path selection for multiple users, running simultaneous applications over multiple parallel access networks. Our joint optimization problem consists of finding the appropriate application rate allocation and network parameters for each individual user, such that an overall quality metric is maximized. We compare our solution to other solutions based on throughput optimization strategies through extensive simulations, and we show the superiority of our approach. Furthermore, our solution proves to be more robust in dynamic systems, when clients can join/leave the access networks.

1. INTRODUCTION

The fast expansion of internet coverage and the increasing availability of wired/wireless network services encourage the development of QoS demanding applications. Wireless users can seamlessly choose from a variety of parallel wireless services (e.g., UMTS/GPRS/WiFi) [1], in order to access these applications; moreover, efforts towards inter-networking different wireless technologies are underway [2], to better meet QoS and cost requirements. In such a context, managing the overall network resources, in the presence of multiple clients accessing simultaneously different applications, becomes of crucial importance for network operators.

Recent wireless technologies allow for parallel access to different applications, like web browsing/downloading, voice conversations and media streaming, each having their distinct QoS requirements and hence, their particular need of network resources. While network resource allocation protocols are standardized for application dedicated networks, e.g., GSM systems for voice applications, or the internet congestion control mechanisms for data traffic, they might fail in a more complex environment, where different types of applications share common network resources.

The recent works in [3, 4, 5, 6] present a new framework for resource allocation and optimization in wireless systems.

This work has been supported by DoCoMo Communications Laboratories Europe GmbH and the Swiss National Science Foundation, under grant PP-002-68737.

They exploit the information available at different layers of the network architecture in order to optimize the overall system performance. The authors of [7] describe a framework for the joint performance optimization of multiple parallel applications sharing the same wireless channels, under a universal quality metric. However, none of these early works address the problem of resource allocation and network selection when multiple users have access to several heterogeneous networks, administered by the same operator.

In this paper we envision a setup where users can access various applications with different Quality-of-Service (QoS) requirements over possibly multiple access networks (Figure 1). We solve a global optimization problem that periodically computes the optimal rate allocation and network selection for each user/application, given a universal quality metric. To this end, we take into account the parameters of the networks available to each user, and the specific characteristics of wireless applications. Then, we define a universal quality metric that models the QoS behavior of all applications as a function of the network parameters. Our final goal is to maximize the overall QoS, under the given network resource constraints.

The rest of this paper is organized as follows: Section 2 presents the considered applications and available access networks. We present our joint optimization problem in Section 3 and we offer a concrete modelling example in Section 4. Extensive simulation results are presented in Section 5, while Section 6 concludes this paper.

2. SYSTEM MODEL

2.1. Multiple Applications

Assume N active users that simultaneously access via a server S any of three different types of applications, namely voice conversation (V), real-time media streaming (M) and FTP download (F). Let user i , $1 \leq i \leq N$ access one of the available applications j , $j \in \{V, M, F\}$, and let $\mathcal{M}_i = r_i$ be the mode of operation of user i , decided by S . It describes the average rate allocated to user i that has chosen application j . We assume that S can scalably adapt the transmission process to the channel conditions of user i . To this end, for each application j , the server can choose the right transmission parameter,

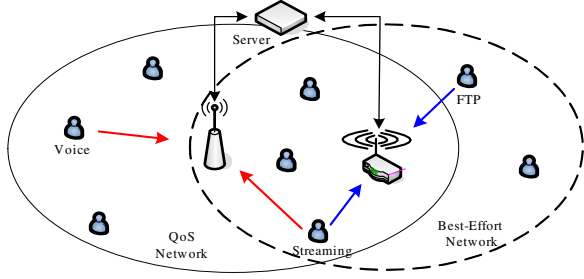


Fig. 1. Multiple wireless networks framework.

from a predefined set of available parameters \mathcal{P}_j .

We consider N_V available voice transcoders at the server S . Each transcoder v is characterized by its encoding rate ρ_v . We define $\mathcal{P}_V = \{\rho_v : 1 \leq v \leq N_V\}$ as the available parameter set for the voice application. The perceived quality of the voice application at the end client depends on the complexity of the transcoder v , and hence the allocated rate ρ_v , and the error process p that affects the data transmission.

Then, we consider a multimedia streaming application that transmits a scalable encoded stream to the end user. Let L be the number of available encoded media layers available at the server S , where the layer $l \leq L$ is characterized by its average encoding rate ρ_l . Additionally, we assume that the server S can protect each media layer against transmission errors, with one of E forward error correction schemes $FEC(n_e, k_e)$, $e = 1 \dots E$. We define $\mathcal{P}_M = \{\rho_m : 1 \leq m \leq L E\}$ as the set of available streaming modes, where $L E$ represents the total number of feasible combinations between the media encoded layers and FEC schemes, and ρ_m is the total rate imposed by mode m . The final perceived quality at the end user depends on the number of media layers transmitted, and the loss process that affects the media packets after FEC decoding.

Finally we assume $\mathcal{P}_F = \{\rho_f : 1 \leq f \leq N_F\}$ as the available parameter set for the FTP application. ρ_f represents the download rate of the FTP session. The perceived quality of the application will depend on the total download time, hence on the allocated download rate and error process that affects the data transmission.

We define the QoS metric $\Gamma(\mathcal{M}_i) = f(r_i, p_i)$ as a function of the allocated rate r_i and the average loss probability p_i affecting the data transmission of application j , towards user i . Finally, we define $\mathcal{M} = \{\mathcal{M}_i : 1 \leq i \leq N\}$ as the global operation mode of the system, when the server S allocates the rate $r_i = \rho_j \in \mathcal{P}_j$ to each active user i , accessing application j .

2.2. Multiple Networks

Even if the problem formulation proposed here is generic enough, we constrain ourselves to a scenario with two active networks that relay application data between S and user i .

Q.Net is a QoS modelled network, characterized by a guaranteed service to all active users when network loads are inferior to the congestion point (e.g., through spreading codes and transmission time intervals assignment in the case of an HSDPA system), and high blocking probability in saturated regime. Its total resources are characterized by the instantaneous total throughput R^Q , which takes into account the channel conditions of all active users in the network. R^Q is preferentially distributed among active users according to the importance of their accessed application (e.g., HSDPA systems prioritize voice conversations over streaming applications and FTP downloads). R^Q is periodically estimated on time intervals T , possibly with a certain prediction error, which translates into a generally small packet error probability p_i^Q that equally affects the active users.

The second network, BE.Net, is modeled as a Best Effort network that provide services to clients on a first-come-first-serve basis (e.g., a WiFi hotspot). Each active client i in this network can access resources at a maximum data rate R_i^B and is affected by an average loss process p_i^B , over time intervals T . While channel conditions in wireless environments change on very short time scales (e.g., a few to a few tens of ms), we assume that R_i^B and p_i^B represent average values computed, possibly from packet transmission statistics, on larger time scales T (e.g., one to a few seconds), and represent the average channel conditions for user i on the given time interval.

Let $[r_i^Q, r_i^B]$ be the rate allocation of user i over the two networks, with $r_i = r_i^Q + r_i^B$. Please observe that application rates $r_i^Q = 0$ or $r_i^B = 0$ imply that user i is inactive in the given network. Finally, let the tuple $\tau_i = [r_i^Q, p_i^Q, r_i^B, p_i^B]$ characterize the application rates and channel conditions for each user i in the two networks. The following resource constraints apply:

$$\sum_{i=1}^N r_i^Q \leq R^Q, \quad \sum_{i=1}^N \frac{r_i^B}{R_i^B} \leq 1 \quad (1)$$

for Q.Net and BE.Net respectively. Finally, under these conditions, the total error probability that affects the transmission to user i , reads: $p_i = \frac{r_i^Q \cdot p_i^Q + r_i^B \cdot p_i^B}{r_i^Q + r_i^B}$.

3. NETWORK SELECTION AND RATE ALLOCATION PROBLEM

We assume that the server S periodically solves the optimization problem, in full knowledge of the connection parameter tuple $\tau_i, \forall i : 1 \leq i \leq N$, and of the application parameter sets $\mathcal{P}_j, \forall j \in \{V, M, F\}$. Within each time interval T , we optimize the allocation of network resources among the N clients, with the final goal of maximizing the overall quality of the system. In other words, we are looking for the optimal global operation mode $\mathcal{M}^* = \{\mathcal{M}_i^* : 1 \leq i \leq N\}$ containing the optimal application mode for each client i , where $\mathcal{M}_i^* = r_i^* \in \mathcal{P}_j$, j being the application accessed by client i :

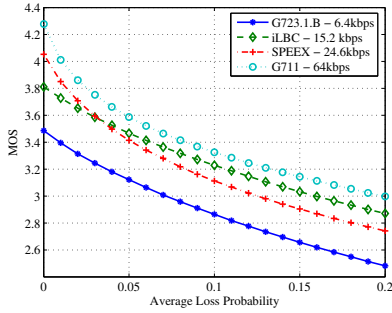


Fig. 2. Voice Application MOS

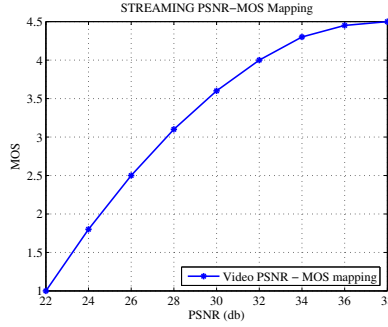


Fig. 3. Streaming Application MOS

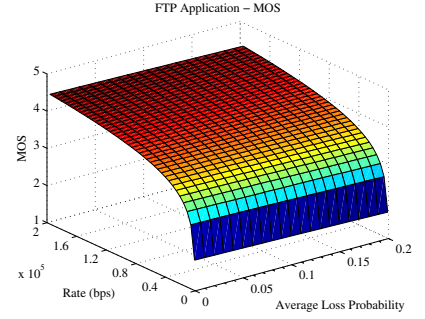


Fig. 4. FTP Application - MOS

$$\mathcal{M}^* = \arg \max_{\forall \mathcal{M}} \sum_{i=1}^N \Gamma(\mathcal{M}_i) \quad (2)$$

under the constraints provided by Eq. (1). A discrete search through all operation modes leads to the solution \mathcal{M}^* with optimal overall QoS. Alternatively, in the case of a large number of clients N , heuristic search algorithms based on stepwise iterative local optimizations can be employed in order to obtain a close to optimum resource allocation [8].

4. MOS QUALITY METRIC

In this section we exemplify on a concrete quality metric Γ based on the *MOS* (Mean Opinion Score) value. *MOS* reflects the average user satisfaction on a scale of 1 to 4.5. The minimum value reflects an unacceptable application quality, and the maximum value refers to an excellent QoS. The perceived quality of each of the three applications is converted into an equivalent *MOS* value, which is later used in the optimization problem.

The performance of different voice transcoders as a function of network losses is mapped to *MOS* values using the *PESQ* algorithm on a representative set of voice samples [7] in Figure 2. We observe that, while good network conditions lead to increased user experience, high packet error rates will degrade the perceived quality of the voice communication.

The perceived media streaming quality is initially mapped into an *MSE* (mean square error) distortion measure: $D(MSE) = D_S + D_L$, which takes into account the video source distortion D_S (which is a function of the encoding rate) and the loss distortion D_L (which is a function of the transmission loss rate). In general, the source distortion is an exponentially decreasing function of the encoding rate, while the loss distortion is linearly dependent on the quantity of media information lost in the transmission process¹. Later on, a nonlinear mapping between *MSE* and *MOS* values is used, as illustrated in Figure 3.

¹The parameters for these functions can be computed off-line for each individual media encoder and for each media sequence [9].

Finally, the perceived quality of the FTP application is mapped to *MOS* values according to a logarithmic function of the achieved throughput: $MOS = a \cdot \log(b \cdot r(1 - p))$. Values a and b are system dependent parameters, and can be set by the network operator (Figure 4).

5. SIMULATION RESULTS

We test the performance of our proposed rate allocation and path selection method, and we compare its performance against a classic optimization solution based on application throughput as a quality metric.

We use 4 voice transcoders, namely G.723.1B, iLBC, SPEEX and G.711 with average encoding rates of 6.4, 15.2, 24.6 and 64kbps respectively. To simulate the media streaming application, we encode the *foreman_qcif* sequence (300 frames) with the H.264/SVC codec. We encode one base layer and one enhancement layer, each of 70kbps. Additionally, we use one forward error correction mode *FEC*(20, 17) which can correct up to 3 packet errors in a block of 20 packets. For FTP downloads, we set 4 available download rates of 50, 100, 150 and 200kbps respectively. A total of 6 clients are placed in the coverage area of both considered networks (3 voice, 2 FTP, and one streaming user). Server S performs the rate allocation optimization periodically, every $T = 1s$. The average throughput R^Q of Q_Net varies in the interval [100, 150]kbps and the prediction error p_i^Q is kept around 1%. The connection data rate R_i^B of the users in the BE_Net is set in the interval [220, 310]kbps, and the individual average loss probabilities p_i^B are randomly chosen in the interval [1, 15]%. We average our results over 100 simulation runs of 10 seconds each.

We first compare the average performance of the overall system, when the optimization is performed according to the *MOS* and throughput quality metrics. We start by identifying the traffic distribution obtained by each optimization metric over the two networks. Table 1 presents the fraction of traffic that passes through both networks, for each application. We observe that the *MOS* optimization rightfully uses the Q_Net resources for the voice and streaming applications, while the FTP traffic is forwarded through BE_Net. On the

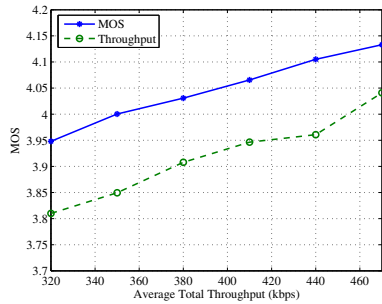


Fig. 5. Average system *MOS* values: *MOS* vs. Throughput Optimization

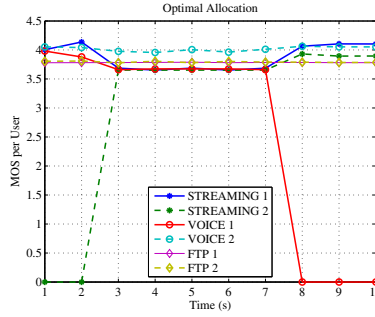


Fig. 6. Client performance when users are added/removed to/from the system: *MOS* optimization.

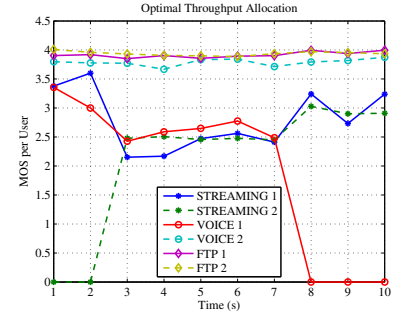


Fig. 7. Client performance when users are added/removed to/from the system: Throughput optimization.

other hand, the throughput optimization favors the FTP application, as it forwards part of its traffic over Q_Net (hence increasing the offered rate for the application), at the expense of lower available resources for the voice and streaming applications that share the same network. This explains the lower overall system performance obtained for the throughput metric, compared to *MOS* (Figure 5). For total average system throughput varying from 320 to 460kbps, the *MOS* optimization outperforms the throughput optimization in most cases by as much as 0.15 *MOS* points.

Table 1. Traffic distribution over the two networks (in %).

Application	<i>MOS</i> Optimization		Throughput Optimization	
	Q_Net	BE_Net	Q_Net	BE_Net
Voice	100	0	100	0
Streaming	88.5	11.5	94	6
FTP	1	99	12	88

Finally, we test the two optimization metrics in dynamic systems where users are allowed to join/leave the networks. We start with 5 clients (2 voice, 1 streaming and 2 FTP users). At time $t = 3s$ we add a streaming user, and at time $t = 8s$ we remove one voice user. Figure 6 and Figure 7 present the average application performance for each user. We observe that in the case of *MOS* optimization, the system is able to cope with the extra user at the expense of a small quality degradation for the existing users. On the other hand, the throughput optimization is unfair, as some of the clients are penalized more than the others, and the overall performance is worse.

6. CONCLUSIONS

We introduce a new optimization framework for the rate allocation and network selection for clients accessing multiple applications over parallel networks. In the optimization process we take into account the available network resources and the connection parameters of each client, along with the specific quality requirements of each application. We unify the performance of all applications under a single *MOS* quality metric, which is later used in the optimization process. Com-

pared to traditional optimization metrics based on throughput, the *MOS* approach achieves a more fair resource allocation among active clients, and proves to be more scalable in dynamic systems. In our future work we will further investigate this optimization approach, and provide concrete heuristic algorithms for solving the described problem.

7. REFERENCES

- [1] Swisscom Mobile Unlimited UMTS/GPRS/WLAN, 2006. http://www.swisscom-mobile.ch/scm/gek_mobile-unlimited-en.aspx.
- [2] K. Ahmavaara, H. Haverinen, and R. Pichna. Interworking architecture between 3GPP and WLAN systems. *IEEE Communications Magazine*, pages 74–81, November 2003.
- [3] M. van der Schaar and S. Shankar. Cross-layer wireless multimedia transmission: Challenges, principles, and new paradigms. *IEEE Wireless Communications*, 12(4):50–58, Aug. 2005.
- [4] W. Kellerer, L. U. Choi, and E. Steinbach. Cross-Layer Adaptation for Optimized B3G Service Provisioning. In *Proceedings of WPMC*, Japan, Oct. 2003.
- [5] S. Shakkottai, T. Rappaport, P. Karlsson. Cross-layer design for wireless networks. In *IEEE Communications Magazine*, 41(10):74–80, Oct. 2003.
- [6] M. Ivrlac, and J. Nossek. Cross Layer Design - An Equivalence Class Approach. In *Proc. IEEE International Symposium on Signals, Systems, and Electronics*, 2004.
- [7] S. Khan, S. Duhovnikov, E. Steinbach, M. Sgroi, and W. Kellerer. Application-driven cross-layer optimization for mobile multimedia communication using a common application layer quality metric. In *Proceedings of Second International Symposium on Multimedia over Wireless, ISMW*, Canada, July 2006.
- [8] Dan Jurca. Algorithms and metrics for cross-layer optimization of multiple applications over parallel access networks. Technical report, DoCoMo Communications Laboratories Europe GmbH, 2006.
- [9] D. Jurca and P. Frossard. Media-specific rate allocation in multipath networks. *IEEE Transactions on Multimedia*. accepted for publication.