

Analytic Evaluation of a Hybrid Broadcast-Unicast TV offering

Daniel Catrein, Jörg Huschke, Uwe Horn

Ericsson GmbH

Ericsson Research

Ericsson Allee 1

D-52134 Herzogenrath, Germany

Email: {daniel.catrein,joerg.huschke,uwe.horn}@ericsson.com

Abstract—Today mobile TV services in cellular spectrum are delivered over unicast radio bearers as offered by 3G technologies like WCDMA and HSDPA. Since unicast transmission does not scale very well if mobile TV becomes a true mass service, 3GPP has defined a broadcast extension for UMTS, called Multimedia Broadcast Multicast Service (MBMS). MBMS introduces radio broadcast bearers serving all users in common, thus providing true broadcast capabilities. A not so well known feature of MBMS is its support for hybrid broadcast-unicast transmission. In this mode the more popular channels are broadcasted over point-to-multipoint radio bearers while the less popular channels are delivered over point-to-point bearers only on request. In this paper we will present a theoretical framework which allows an analytic evaluation of the capacity limits when delivering mobile TV services over hybrid broadcast-unicast transmission schemes provided by MBMS.

Index Terms—Mobile TV, IPTV, Unicast Broadcast

I. INTRODUCTION

Mobile networks have emerged from voice telephony networks to multimedia delivery networks. It is expected that mobile data traffic will exceed voice traffic by the year 2010. Mobile TV services have become quite popular data services during the past two years.

The majority of today's mobile TV services are delivered over existing 3G networks since this is the fastest and easiest way to deploy Mobile TV services.

Nevertheless, the increasing popularity of mobile TV and similar services, may lead to situations in which many users want to watch the same content at the same time. Examples are live events of high interest like soccer matches, game shows, etc. For those cases, multicasting, known from the internet, or broadcasting are clearly more appropriate technologies.

Therefore, back in 2002 3GPP created a work item for broadcast/multicast services in GSM/WCDMA called Multimedia Broadcast and Multicast Service (MBMS). The MBMS specifications [1], [2] were functionally frozen in 2005. MBMS introduces only small changes to the existing radio and core network protocols. This reduces the implementation costs both in terminals and in the network.

MBMS supports two basic transmission modes for delivering IP packets: broadcast and multicast.

The new bearer type which was introduced by MBMS is the so-called point-to-multipoint (P-t-M) radio bearer. While

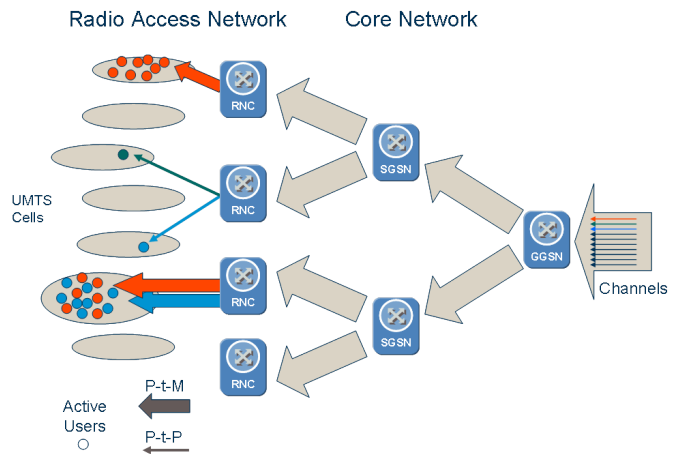


Fig. 1. Hybrid broadcast-unicast transmission in MBMS.

a point-to-point (P-t-P) bearer can only be received by one terminal, the new P-t-M bearer can be received by several terminals in a cell simultaneously.

In MBMS, UTRAN may perform the so-called "counting" or "re-counting" procedures to determine the number of terminals in each cell. Counting is initiated by the RNC as soon as the RNC needs to know the amount of active UEs that want to receive a specific MBMS service. This is used to determine the optimum transmission bearer: Point-to-Multipoint (P-t-M), Point-to-Point (P-t-P) or no transmission at all for a given MBMS service in the considered cell.

Although counting was originally developed for the multicast mode of MBMS only, it can also be combined with the MBMS broadcast mode. This allows the realization of a hybrid broadcast-unicast transmission scheme. In this transmission scheme, it is for instance possible to broadcast the more popular channels over P-t-M radio bearers while the less popular channels are delivered over P-t-P bearers only if explicitly requested. This situation is depicted in Figure 1.

Hybrid broadcast-unicast transmission was already addressed in [3] using simulations. Here, we will derive a theoretical framework which allows an analytic evaluation of the capacity limits when delivering mobile TV services over a hybrid broadcast-unicast transmission scheme provided by

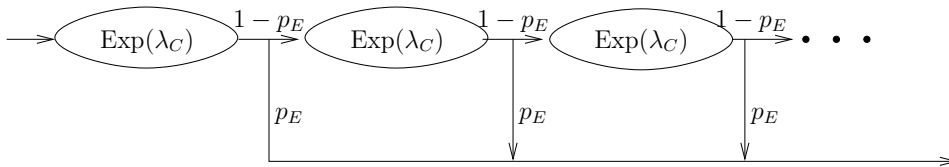


Fig. 2. T_S is Coxian distributed with infinitely many phases.

MBMS. We start with descriptions of the models we used for the overall system, the user behavior, and the P-t-P / P-t-M bearer switching. We will finally present results obtained from the derived analytical model.

II. SYSTEM MODEL

Here, we introduce a model for the behavior of an individual user. This model will be extended towards a model for a multitude of users utilizing the theory of Jackson networks.

III. USER MODEL

Assume each user watches TV in sessions of predefined average length $t_S > 0$. Within one session the user switches from time to time from one channel to another. The length of a sub-sessions, i.e., the time a user stays tuned in some channel, shall be modeled by some exponentially distributed random variable $T_i \sim \text{Exp}(\mu_C)$. Here, $i = 1, 2, 3, \dots$ denotes the number of the sub-session. Let $t_C > 0$ denote the average channel watch time. Thus, we obtain $\mu_C = 1/t_C$. Further, we assume T_i is independent of the selected channel.

At the end of the time period T_i the user quits the session with probability p_E . With probability $1 - p_E$ the user keeps on watching the same or some other channel. We are now interested in the expected value of the overall session length T_S which is a sum over the T_i as depicted in Figure 2.

First assume that a session of a user contains not more than n subsequent sub-session. Given n the session length is in this case modeled by a Coxian distributed random variable

$$T_S^{(n)} \sim \text{Cox}(\underbrace{1 - p_E, \dots, 1 - p_E}_{n-1 \text{ times}}; \underbrace{\mu_C, \dots, \mu_C}_{n \text{ times}})$$

and we obtain for the expectation

$$\text{E}(T_S^{(n)}) = \sum_{i=1}^n p_E (1 - p_E)^{i-1} i t_C.$$

This yields for the expectation of T_S ,

$$\begin{aligned} \text{E}(T_S) &= \text{E}\left(\lim_{n \rightarrow \infty} T_S^{(n)}\right) = \lim_{n \rightarrow \infty} \text{E}\left(T_S^{(n)}\right) \\ &= \sum_{i=1}^{\infty} p_E (1 - p_E)^{i-1} i t_C, \end{aligned} \quad (1)$$

provided all terms exist. The series on the right hand side in (1) converges for $p_E < 1$ and it follows easily $\text{E}(T_S) = t_C/p_E$. This yields $p_E = t_C/t_S$ provided $t_C < t_S$, i.e., the average session lasts longer than the average sub-session, which is a reasonable assumption.

So far, we did not make any assumptions on the selected channel. Assume the system offers $n_C \in \mathbb{N}$ different channels. Each of them may have a different popularity indicated by its selection probability $p_i \geq 0$, $i \in \{1, \dots, n_C\}$ with $\sum_{i=1}^{n_C} p_i = 1$.

IV. CELL MODEL

Each cell shall offer a certain fixed number n_{RES} of resources dedicated either to point to point (P-t-P) or to point to multi-point (P-t-M) connections. Let $n_{\text{P-t-P}}$ and $n_{\text{P-t-M}}$ denote the number of P-t-P and P-t-M bearers, respectively. In general, P-t-M bearers require more resources than P-t-P bearers due additional signaling overhead. Thus, assume a P-t-M bearer requires $\alpha > 1$ times the resources of a P-t-P bearer. This leads to the overall capacity constraint

$$n_{\text{P-t-P}} + \alpha n_{\text{P-t-M}} \leq n_{\text{RES}}. \quad (2)$$

A P-t-M bearer can host arbitrary many users watching the same channel, whereas a dedicated P-t-P bearer is required for each user. We assume a fixed assignment of a sub-set of all channels to P-t-M bearers. In a real system the assignment won't be completely static, however, reconfiguration of bearers takes some time and at least a static assignment for some time period is desirable. Thus, assume without loss of generality that channels $1, \dots, n_{\text{P-t-M}}$ are transmitted via P-t-M bearers and can serve an arbitrary number users.

The remaining $\lfloor n_{\text{RES}} - \alpha n_{\text{P-t-M}} \rfloor = n_{\text{P-t-P}}$ bearers are available for P-t-P connections. Blocking occurs if a user desires to watch a channel that is not broadcasted via a P-t-M bearer while no P-t-P bearer is available.

In case of blocking we assume that the user tries again and again to enter the same channel until he has success or until the program he desires to watch is over, i.e., its sub-session is over.

Model the arrival of new users in the system as a Poisson process with arrival rate λ . Then, the situation described above can be modeled via an open Jackson network. The corresponding network is depicted in Figure 3.

Server 1 shall model the channels broadcasted via P-t-M bearers and sever 2 the channels broadcasted via P-t-P bearers. An user entering the system selects a channel on a P-t-M bearer or a channel on a P-t-P bearer with probability

$$r_1 = \sum_{i \leq n_{\text{P-t-M}}} p_i \quad \text{or} \quad r_2 = \sum_{i > n_{\text{P-t-M}}} p_i, \quad (3)$$

respectively. This results in a splitting of the arrival process in two independent Poisson arrival processes, one on server 1,

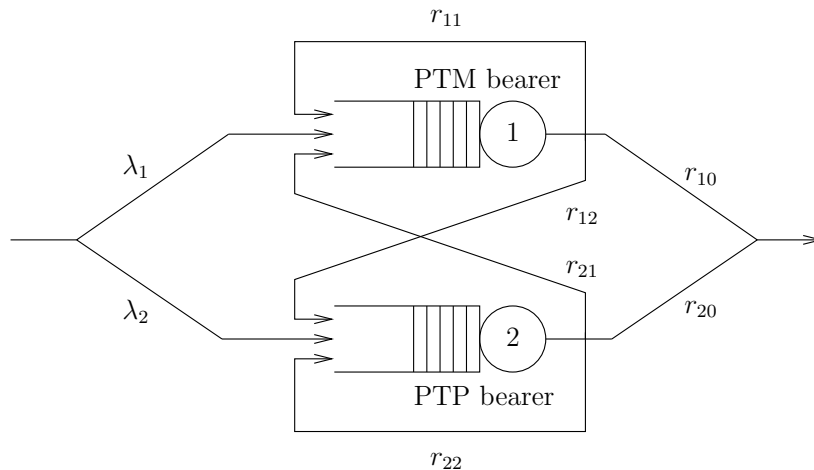


Fig. 3. Jackson network modeling user behavior.

the other on server 2. The arrival rates of this processes are given by $\lambda_1 = r_1 \lambda$ and $\lambda_2 = r_2 \lambda$.

Any number of users watching a channel on a P-t-M bearer can be served simultaneously. Further, we assumed that a users stays on some channel for a time interval of exponentially distributed length. Thus, server 1 can be modeled as $M/M/\infty$ queuing system with a serving rate $\mu_1(n_1) = n_1 \mu_C$ for a given number of users at server 1, $n_1 \in \mathbb{N}$.

A user leaving a P-t-M bearer, i.e., server 1, may leave the system with probability $r_{10} = p_E$ or will stay and choose again a channel with probability $1 - p_E$. If a user stays, she will choose a channel transmitted on a P-t-M bearer with probability r_1 and a channel that requires a P-t-P bearer with probability r_2 . Thus, $r_{11} = (1 - p_E)r_1$ and $r_{12} = (1 - p_E)r_2$.

Only a limited number of P-t-P bearers is available for channels not transmitted over P-t-M bearers. Let the random variable N_{PTP} denote the number of users that want to chose a channel on a P-t-P bearer, i.e., the number of users at server 2. Blocking occurs whenever more than n_{PTP} users have to be served by server 2. Thus, the blocking probability p_B is given by

$$p_B = P(N_{\text{PTP}} > n_{\text{PTP}}).$$

Our assumption for the channel switching behavior in case of blocking leads to a serving rate of $\mu_2(n_2) = n_2 \mu_C$ for server 2 for any given number of users $n_2 \in \mathbb{N}$. Using the same arguments as above we obtain the transition probabilities $r_{20} = p_E$, $r_{21} = (1 - p_E)r_1$ and $r_{22} = (1 - p_E)r_2$.

V. EVALUATION OF THE MODEL

A. Blocking Probability

The following theorem describes the distribution of users watching a channel distributed over a P-t-P bearer, i.e., the number of users at server 2.

Theorem 1. The stationary distribution of the number of users aiming at watching a channel transmitted over a P-t-M bearer

is given by

$$N_{\text{PTM}} \sim \text{Poi}(\rho_1), \quad \text{with} \quad \rho_1 = t_S \lambda \sum_{i \leq N_{\text{PTM}}} p_i.$$

The stationary distribution of the number of users aiming at watching a channel transmitted over a P-t-P bearer is given by

$$N_{\text{PTP}} \sim \text{Poi}(\rho_2), \quad \text{with} \quad \rho_2 = t_S \lambda \sum_{i > N_{\text{PTM}}} p_i. \quad (4)$$

N_{PTM} and N_{PTP} are stochastically independent random variables.

Proof: Let

$$\Lambda_i = \lambda_i + \sum_{j=1}^2 \Lambda_j r_{ji}, \quad i \in \{1, 2\} \quad (5)$$

denote the rate equation of the Jackson network depicted in Figure 3. Solving (5) yields

$$\Lambda_i = \frac{r_i \lambda}{p_E} \quad \text{for } i \in \{1, 2\}.$$

The stationary distribution of the whole system has a product form solution. The stationary distribution at server i is obviously the same as the stationary distribution of a $M/M/\infty$ queuing system with utilization $\rho_i = \Lambda_i / \mu_C$, $i \in \{1, 2\}$. Thus, the number of users at server 1 and at server 2 is Poisson distributed with parameter ρ_1 and ρ_2 , respectively. Furthermore,

$$\rho_i = \frac{r_i \lambda}{p_E \mu_C} = \frac{r_i \lambda}{\frac{t_C}{t_S} \frac{1}{t_C}} = t_S \lambda r_i,$$

for $i \in \{1, 2\}$.

Independence follows from the fact that the stationary distribution of a Jackson network has product form. \blacksquare

It is an interesting observation that the distribution of the number of users does no longer depend on the average sub-session length.

Let $X \sim \text{Poi}(\rho)$ with parameter ρ . The cumulative distribution function of X is given by

$$F_X(k) = P(X \leq k) = \frac{\Gamma(\lfloor k+1 \rfloor, \rho)}{\lfloor k \rfloor!} = \sum_{i=0}^{\lfloor k \rfloor} e^{-\rho} \frac{\rho^i}{i!},$$

for $k \geq 0$. Here,

$$\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt$$

denotes the incomplete gamma function.

The blocking probability follows immediately.

Theorem 2. The blocking probability p_B is given by

$$p_B = 1 - P(N_{\text{PtP}} \leq n_{\text{PtP}}) = 1 - \sum_{k=0}^{n_{\text{PtP}}} e^{-\rho_2} \frac{\rho_2^k}{k!},$$

with ρ_2 according to (4) and $n_{\text{PtP}} = \lfloor n_{\text{RES}} - \alpha n_{\text{PtM}} \rfloor$.

B. Number of Users

The number of users watching TV in the system, i.e., in a cell, is controlled via the arrival rate λ . However, the arrival rate is not a very intuitive measure. An alternative measure is the average number of active users $E(N_{\text{active}})$. It holds that $N_{\text{active}} = N_{\text{PtM}} + N_{\text{PtP}}$. Thus, $N_{\text{active}} \sim \text{Poi}(t_S \lambda)$ and $E(N_{\text{active}}) = t_S \lambda$

VI. OPTIMIZING THE NUMBER OF BROADCAST BEARERS

One can easily proof that the blocking probability is minimal if and only if the channels with highest selection probability are transmitted of over P-t-M bearers.

Theorem 3. Assume a fixed number of P-t-M bearers n_{PtM} and a fixed maximal number of P-t-P bearers n_{PtP} . The blocking probability is minimal if and only if $p_i \geq p_j$ for all $i \leq n_{\text{PtM}}$ and $j > n_{\text{PtM}}$.

Proof: Let $\mathbf{p} = (p_1, \dots, p_{n_C})$ with $p_i < p_j$ for some $i \leq n_{\text{PtM}}$ and $j > n_{\text{PtM}}$ and assume the corresponding blocking probability $p_B(\mathbf{p})$ is minimal. Let \mathbf{p}' denote the corresponding vector \mathbf{p} with elements i and j swapped.

It follows immediately from the definition that $\Gamma(k, \rho)$ is strictly monotonous decreasing in ρ . Thus, $1 - \Gamma(k, \rho)$ is strictly monotonous increasing in ρ . Furthermore, $\sum_{i>N_{\text{PtM}}} p_i > \sum_{i>N_{\text{PtM}}} p'_i$, such that

$$\begin{aligned} p_B(\mathbf{p}) &= 1 - \frac{\Gamma(n_{\text{PtP}} + 1, t_S \lambda \sum_{i>N_{\text{PtM}}} p_i)}{n_{\text{PtP}}!} \\ &> 1 - \frac{\Gamma(n_{\text{PtP}} + 1, t_S \lambda \sum_{i>N_{\text{PtM}}} p'_i)}{n_{\text{PtP}}!} = p_B(\mathbf{p}'), \end{aligned}$$

which contradicts the assumption that \mathbf{p} is minimal and concludes the proof. ■

An important consequence of Theorem 3 is that we can easily calculate the optimal number of P-t-M bearers. Sorting the channel watch probabilities and calculating the blocking probability for any number of P-t-M bearers yields immediately the optimal number of P-t-M bearers. Algorithm 1 illustrates this procedure.

Algorithm 1 $p_B(\mathbf{p})$

Require: Channel probabilities $\mathbf{p} = (p_1, \dots, p_{n_C})$.

Ensure: The minimal number of P-t-M bearers minimizing the blocking probability

```

1:  $\mathbf{p}' \leftarrow \text{sort}(\mathbf{p}, \text{"descending"})$ 
2:  $n_{\text{PtM}, \text{min}} \leftarrow 0$ 
3:  $p_{B, \text{min}} \leftarrow 1$ 
4: for  $n_{\text{PtM}} = 0$  to  $\lfloor n_{\text{RES}} / \alpha \rfloor$  do
5:    $\rho = t_S \lambda \sum_{i>N_{\text{PtM}}} p'_i$ 
6:    $n_{\text{PtP}} = \lfloor n_{\text{RES}} - \alpha n_{\text{PtM}} \rfloor$ 
7:    $p_B \leftarrow 1 - \sum_{k=0}^{n_{\text{PtP}}} e^{-\rho} \frac{\rho^k}{k!}$ 
8:   if  $p_B < p_{B, \text{min}}$  then
9:      $p_{B, \text{min}} = p_B$ 
10:     $n_{\text{PtM}, \text{min}} = n_{\text{PtM}}$ 
11:   end if
12: end for
13: return  $n_{\text{PtM}, \text{min}}$ 

```

Channel	i=1	i=2	i=3	i=4	i=4,...,20
sel. prob. 1	0.35	0.25	0.15	0.10	0.009
sel. prob. 2	0.20	0.20	0.15	0.10	0.022
sel. prob. 3	1/2	1/4	1/8	1/16	1/2 ⁱ
sel. prob. 4	0.095	0.090	0.086	0.081	(21 - i)/210

TABLE I
FOUR DIFFERENT CHANNEL SELECTION PROBABILITY SCENARIOS.

VII. NUMERICAL ANALYSIS

Analogously to [3], we derive the parameters for the numerical analysis from [4] and assume a system with 20 channels and resources for 20 P-t-P bearers. Table I depicts channel selection probabilities for four different scenarios being analyzed. For comparison purposes scenario 1 and scenario 2 are taken from [3]. It is noted that the analytical model developed here generally follows the principles used in [3]. However, the assumed user behavior and the definition of blocking is not exactly the same. Therefore, the results cannot directly be compared.

Figure 4 depicts the blocking probability as a function of the average number of active users with channel probabilities according to scenario 1. Here, we extend the results of [3] to a much higher cell load. We see that increasing the number of P-t-M bearers decreases the blocking probability only up to a certain minimum. Here, the blocking probability is minimal with four P-t-M bearers in the observed load range.

This is further illustrated by Figure 5, which depicts the blocking probability as a function of the number of P-t-M bearers. Observe that Figure 5 applies a logarithmic scale at the y -axes. We see that for an average load of $t_S \lambda = 20$ users the blocking probability is heavily affected by the number of P-t-M bearers. Furthermore, the optimal number of P-t-M bearers depends also on the channel selection probabilities. The higher the differences in the probabilities are, the higher is the affect on the blocking probability when changing the number of P-

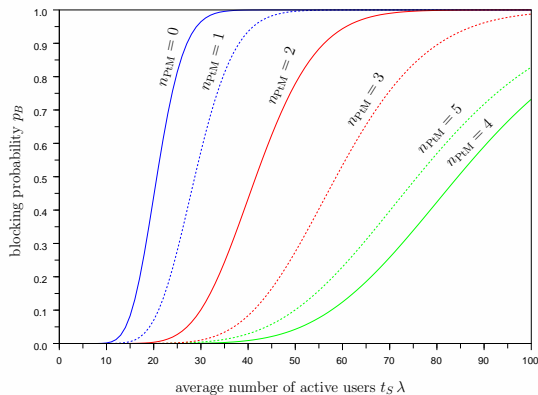


Fig. 4. Blocking probability as a function of the average number of users. Selection probabilities according to scenario 1, resources for $n_{RES} = 20$ P-t-P bearers, each P-t-M requires $\alpha = 2$ times the resources of a P-t-P bearer.

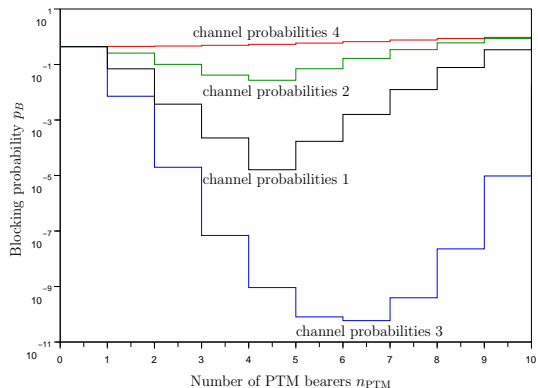


Fig. 5. Blocking probabilities for a scenario with resources for up to $n_{RES} = 20$ P-t-P bearers. Every P-t-M bearer requires $\alpha = 2$ times the resources of a P-t-P bearer and in average $t_s \lambda = 20$ users are using the system.

t-M bearers.

Finally, Figure 6 depicts that the optimal number of P-t-M bearers depends heavily on the resources occupied by a single P-t-M bearers. In contrast to P-t-P bearers a P-t-M bearer does not use (hybrid) ARQ nor link adaptation or channel dependent scheduling. A robust transmission mode with sufficient power needs to be configured that ensures the desired coverage probability, typically 95%. Therefore, on average a P-t-M bearer requires more radio resources than a P-t-P bearer. The resource ratio is in the range of about 2, for the case that adjacent cells transmit the same P-t-M bearer so that it can be soft combined in the receiver, to about 10, for the case that adjacent cells do not transmit the P-t-M bearer. As expected, we see that the more resources a single P-t-M bearer requires, the less P-t-M bearers are optimal.

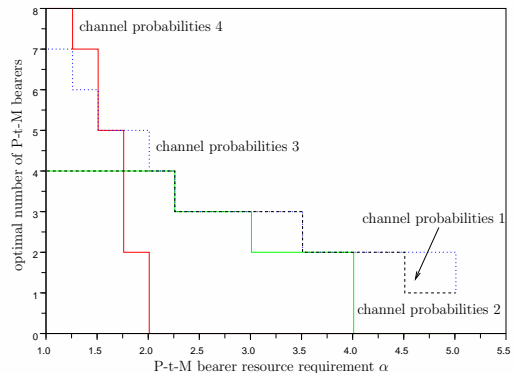


Fig. 6. Optimal number of P-t-M bearers as a function of the resource requirement α . In average $t_s \lambda = 50$ users are watching TV simultaneously, resources for up to $n_{RES} = 20$ P-t-P bearers are available.

VIII. CONCLUSIONS

Based on a theoretical model for the user behavior we showed in this paper that there is a need for only a small number of Point-to-Multipoint radio bearers in a mobile TV service offering. Only the most attractive channels need to be transmitted via P-t-M bearer. The less requested channels can be more efficiently transmitted via Point-to-Point bearers using e.g. HSDPA. Furthermore, we derived a method to compute the optimal number of P-t-M bearers given the channel selection probabilities, the overall available resources and the amount of resources occupied by a single P-t-M bearer.

In reality the exact channel probabilities might not be known. Hence, future work will include an analysis of the impact of optimizing the number of P-t-M bearers in case of unreliable channel probability information. Furthermore, extending the model to remove the possibility to zap immediately from one channel to the same channel again might be of practical interest.

REFERENCES

- [1] 3GPP TS 23.246 v6.10.0, "Multimedia broadcast/multicast service (MBMS); architecture and functional description," 3rd Generation Partnership Project (3GPP), Tech. Rep., 2006. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.246/
- [2] 3GPP TS 26.346 v6.5.0, "Multimedia broadcast/multicast service (MBMS); protocols and codecs," 3rd Generation Partnership Project (3GPP), Tech. Rep., 2006. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/26_series/26.346/
- [3] T. Lohmar and U. Horn, "Hybrid broadcast-unicast distribution of mobile tv over 3g networks," in *Proceedings: 31st IEEE Conference on Local Computer Networks*, Florida, USA, 2006, pp. 850–851.
- [4] C. Södergård, "Mobile Television – Technology and user experiences. Report on the mobile-TV project." VTT Information Technology, Espoo, Sweden, Tech. Rep., 2003.