Supporting Multiple Content Variants in the Multimedia Broadcast/Multicast Service

George Xylomenos and Konstantinos Katsaros
xgeorge@aueb.gr and ntinos@aueb.gr
Mobile Multimedia Laboratory
Department of Informatics
Athens University of Economics and Business
Patision 76, Athens 104 34, Greece

***Abstract*— The Multimedia Broadcast/Multicast Service (MBMS) was recently standardized for use by 3rd Generation cellular networks, aiming to support the economical distribution of multimedia content to large numbers of receivers. This paper proposes an MBMS extension supporting the distribution of multiple variants of the same content to heterogeneous receivers. We first outline the standard MBMS model, along with its state management and signaling procedures, and then describe our extended MBMS model, explaining the modifications it imposes on standard MBMS. We then explain how our approach can be combined either with layered coding or transcoding for the generation of the multiple content variants to be distributed. Finally, we compare our proposal via analysis and simulation with some alternatives based on standard MBMS. Both the analytical and the simulation results indicate that our proposal increases the number of satisfied users without spending excessive resources, thus striking a good balance between the standards based alternatives considered.**

I. INTRODUCTION

Cellular systems are increasingly becoming feasible platforms for multimedia services. While the resource requirements of, say, video distribution, make it too expensive for most users, these costs can be dramatically reduced when they are shared among many users receiving the same service. To this end, the *Universal Mobile Telecommunications System* (UMTS), specified by the *3rd Generation Partnership Project* (3GPP), has introduced the *Multimedia Broadcast/Multicast Service* (MBMS) [1]. MBMS is suitable for services as diverse as media streaming and file downloads [2]; regardless of the application in use, MBMS enables resource sharing throughout the network, including over the air.

The MBMS multicasting mode, similar to the well known IP multicasting, delivers the exact same content to all receivers, using the same *Quality of Service* (QoS) parameters for the entire multicast distribution tree. When the receivers are heterogeneous however, for example, terminals with different screens or users with different budgets, it is difficult to select the proper content variant to transmit. A high quality variant will not be received by low end terminals, while a low quality variant will disappoint users prepared to pay more for better service; in both cases, potential users, and the corresponding revenues, are lost.

As part of our research in the IST B-Bone project, we designed an MBMS extension supporting the distribution of multiple variants of the same content to different receivers; we refer to this extended service as *Multiple Content Variant* (MCV) MBMS. The desired variant is dynamically selected by each receiver, based either on terminal capabilities or user preferences. In the remainder of this paper we describe how MCV MBMS is derived from standard MBMS and how it performs against some standards based alternatives.

In Section II we describe the standard MBMS model, along with its state management and signaling procedures, while in Section III we describe our extended MCV MBMS model. In Section IV we discuss how our approach can be combined with layered coding or transcoding for the generation of proper content variants. In Section V we present an analytical comparison of MCV MBMS against some standards based alternatives, while in Section VI we present a corresponding simulation based comparison; in both cases, MCV MBMS is shown to be clearly superior to the alternatives. We summarize our conclusions and discuss future work in Section VII.

II. THE STANDARD MBMS MODEL

An outline of a UMTS network supporting MBMS is shown in Figure 1. A new functional entity, the *Broadcast/Multicast Service Centre* (BM-SC), controls the provision of MBMS services. The *Gateway GPRS Support Node* (GGSN), the *Serving GPRS Support Node* (SGSN), the *Radio Access Network* (RAN) and the *User Equipment* (UE) are the existing network elements modified for MBMS. Unlike in IP multicasting, where groups are identified by a class D IP address, in MBMS a multicast group is identified both by a class D IP address and an *Access Point Name* (APN); the APN identifies the GGSN serving a UMTS network, therefore MBMS services are defined with respect to a specific network. Also unlike in IP multicasting, where anyone can send to and receive from a group, MBMS uses closed groups: a UE must first subscribe to a group, using some mechanism external to MBMS, in order to be later allowed to join it, and only the GGSN identified by the APN may transmit data to a group. These are ideal properties for commercial services [1].

In order to support MBMS, each network node must maintain additional state. First, packet forwarding state is required so that the node may determine which of its children should receive a packet. Second, user state is required so that the

Fig. 1. Components of MBMS.

network may charge the participating receivers. Each node therefore maintains an *MBMS Bearer Context* (MBC) for each multicast group and an *MBMS UE Context* (MUEC) for each UE that is currently a member of the group [3], as shown in Figure 2 (fields in gray are used by our extended model).

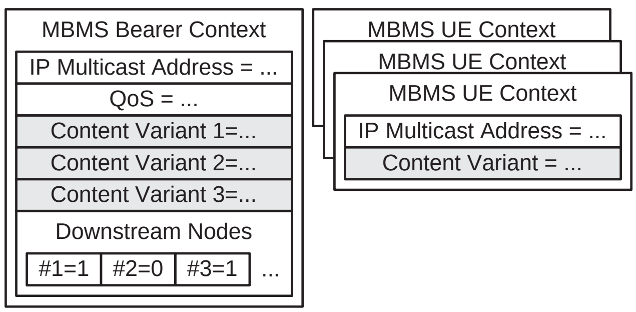

Fig. 2. MBMS UE and Bearer contexts.

The MBC contains information for the entire group, such as its IP multicast address and QoS parameters. A table indicates which downstream nodes should receive packets addressed to that group. For example, in Figure 2 child #1 should receive packets (marked 1) but child #2 should not (marked 0). When a multicast packet arrives, the node examines the MBC for the appropriate group and forwards the packet to all children marked 1. The MUEC on the other hand contains information for a UE belonging to a group; it is linked to an MBC via its address field. When forwarding data to a group, the node uses the MUECs linked to the MBC to charge the UEs. When an MBMS service is to be offered, its attributes are first entered into a new MBC at the BM-SC. Additional MBCs and MUECs are dynamically created at each node based on UE initiated signaling.

In MBMS the *Internet Group Management Protocol* (IGMP) [4] is used for group management, but unlike the standard query/response mode of IGMP used with IP, MBMS uses instead a join/leave mode, which is better suited to UMTS networks [7]. Each UE desiring to join an MBMS group sends an IGMP join message to the GGSN, thus triggering the MBMS multicast activation procedure shown in Figure 3. The GGSN asks the BM-SC if the UE has subscribed to the group and the BM-SC returns the APN of the GGSN that acts as the source. The GGSN asks the SGSN if it can handle the MBMS group. The SGSN responds to the GGSN and notifies the UE to proceed with the activation. At this point (dashed line) the UE knows the real APN of the source, so the signaling continues towards the appropriate GGSN. The UE then requests the SGSN to start sending it data. The SGSN creates the MUEC for the UE and notifies the GGSN corresponding to the APN. The GGSN notifies the BM-SC about the UE, the BM-SC creates the MUEC and responds to the GGSN. Finally, the GGSN creates the MUEC and responds to the SGSN, which responds to the UE.

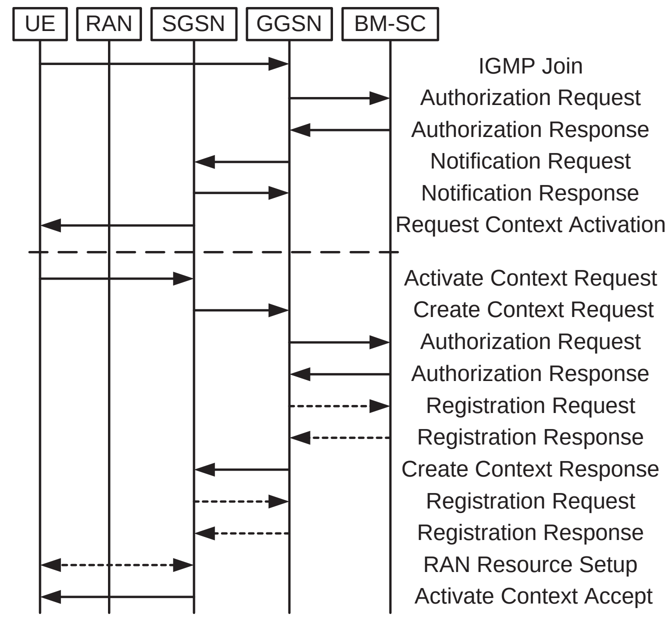

Fig. 3. MBMS multicast activation.

When the first MUEC for a group is created at the GGSN or SGSN, that node sends a registration message to its parent (BM-SC or GGSN), shown with dotted lines in Figure 3. The parent marks the corresponding entry in its MBC with 1 so as to start forwarding data to that child. Using the information provided in the response, the child creates the MBC for the group. If the group is active, the SGSN also notifies the RAN to establish radio bearers so that data transmission may proceed.

The activation and registration procedures lead to the establishment of a multicast distribution tree from the BM-SC towards all UEs participating in an MBMS service, but they do not reserve any transmission resources. A separate session start procedure is used to establish radio bearers for the actual data, while a session stop procedure is used to release these bearers when they are no longer needed. Finally, a multicast deactivation procedure can be triggered by a UE desiring to leave a group by sending an IGMP leave message to the GGSN. The deactivation procedure essentially reverses the actions performed by the activation procedure described above; when the last MUEC for a group is destroyed at the GGSN or SGSN, the deregistration procedure is used to reverse the actions performed by the registration procedure.

III. The extended MBMS model

Our extended *Multiple Content Variant* (MCV) MBMS model allows a single MBMS service to offer different variants of the same content to different receivers, providing various tradeoffs between bandwidth and quality. We assume that a lower quality variant can always be derived from a higher quality one; in Section IV we discuss two ways to achieve this. The variants are chosen by the content provider to match common terminals and have sufficiently different costs. The number of available variants must be small, to prevent the degeneration of multicast groups to single receivers; in our simulator implementation we support up to three variants, numbered 1 (low quality, LQ), 2 (medium quality, MQ) and 3 (high quality, HQ). A UE specifies the desired variant by including in its IGMP join message a variant number. This number is included in all subsequent MBMS request messages shown in Figure 3. The UE may later modify this request by sending a new IGMP join. For example, the user may request higher quality audio to better hear a passage, or the terminal may request lower quality audio when the bandwidth at its location is limited.

Each node in our MCV model maintains additional information in the MBC and MUEC, shown in Figure 2 with a gray background. The MUEC includes the number of the requested variant, allowing the node to charge the UE accordingly. The downstream nodes table in the MBC is also extended with the number of the variant to forward to each child, i.e. 0 (none) to 3 (high quality). Each node informs its parent which variant it needs to receive, thus allowing the parent to maintain its downstream nodes table. The node asks for the highest quality variant requested by any of its own children; from this variant it can always produce lower quality ones, as we have already assumed. As a result, each node receives and forwards the lowest amount of data possible. In addition, each content variant in MCV MBMS is described by a separate set of UMTS QoS parameters [5]. For example, while all variants of the same service should belong to the same UMTS traffic class, which must be either *Streaming* or *Background* for MBMS services, each variant would normally specify a different bit rate.

In order to determine the highest quality variant requested by any of its children, each node counts the number of MUECs for each variant and stores these counters in the MBC; these counters are updated when MUECs are created or destroyed [8]. When a UE sends an IGMP join, triggering the procedure shown in Figure 3, at the point where in standard MBMS a node would create or destroy a MUEC, in MCV MBMS the node must instead do one of the following: a) if a MUEC was just created (destroyed), the counter for its variant is incremented (decremented), or, b) if a MUEC just modified its variant, the counter for its previous variant is decremented and the counter for its current variant is incremented. In addition, at the point where in standard MBMS a node would create or destroy an MBC, in our model it must instead do one of the following: a) if the first (last) MUEC was created (destroyed), the MBC is created (destroyed) and the parent is informed to start (stop) forwarding data, b) if the counter for a higher quality variant than the current one became nonzero the parent is informed, c) if the counter for the current variant became zero, the next nonzero counter is found and the parent is informed.

IV. Content variant generation and forwarding

As stated in Section III, our MCV MBMS model requires that the content variants are produced in a manner allowing a lower quality variant to be derived from a higher quality one. Two methods that may be used to generate such content variants are *layered coding* and *transcoding* [6]. In the layered coding approach, the source encodes the lowest quality variant as the base layer and then encodes a series of successive enhancement layers. The next higher quality variant consists of the base layer and the first enhancement layer; each successive variant adds another enhancement layer. The source injects all layers to the multicast distribution tree, and each node forwards to each child only the layers required to reconstruct the variant requested by that child. For example, in the MBC shown in Figure 2, if child #3 was marked 2, the node would need to receive at least variant 2 from its parent, that is, the base layer and the first enhancement layer; it would forward the base layer to child #1 and the first enhancement layer to child #3.

On the other hand, in the transcoding approach the source injects the highest quality variant to the multicast distribution tree and this variant is transcoded, that is, re-encoded, to lower quality variants by the nodes. Each node forwards to each child the requested variant by transcoding the variant received from its parent. For example, in the MBC shown in Figure 2, if child #3 was marked 2, the node would again receive variant 2 from its parent; it would forward it as is to child #3 and transcode it to variant 1 before forwarding it to child #1.

In general, layered coding is not as efficient as transcoding since, among other things, all data must be tagged with the layer that they belong to, in order to allow each node to determine which data to forward to each of its children. This is not needed with transcoding, where all data received by a node are part of a single transcoded variant. On the other hand, layered coding does not require complex computations at each node: each node simply discards some layers. In transcoding, each node may have to transcode the variant received from its parent to produce the variants requested by its children.

In order to reduce the complexity of the UMTS network nodes, in our simulator implementation we used layered coding to produce the content variants. Each packet contains data from a single layer and is tagged with a 2 bit layer identifier. As a result, each node can easily identify and drop any packets belonging to redundant layers. The resulting data stream consists of a base layer and two enhancement layers; users requesting the LQ variant only receive the base layer, users requesting the MQ variant receive the base layer and the first enhancement layer, and users requesting the HQ variant receive the base layer and both enhancement layers.

V. Performance Evaluation: Analysis

We will now compare our MCV MBMS model with two alternative approaches based on standard MBMS. The first alternative is to satisfy all UEs by sending all content variants to everyone; this leads to a waste of transmission bandwidth. The second alternative is to economize on bandwidth by sending only the lowest quality variant to everyone; this leads to unsatisfied users. We refer to the first option as L/M/HQ MBMS, since each UE receives the LQ, MQ and HQ variants, and to the second option as LQ MBMS, since each UE only receives the LQ variant; in contrast, in MCV MBMS each UE receives exactly the variants that it asked for.

In this section we provide an analytical comparison of MCV MBMS against LQ MBMS and L/M/HQ MBMS in terms of their user plane overhead; corresponding simulation results are given in Section VI. Note that the control plane overhead of LQ MBMS and L/M/HQ MBMS is exactly the same, as each UE joins a single service only; their difference with the more complicated MCV MBMS turns out to be negligible however [8].

We assume that in a network comprising a single GGSN, S_n SGSNs and R_n RNCs, N_u users are interested in an MBMS service offered in three variants: LQ, MQ and HQ. The probability that a user will request the LQ, MQ, or HQ variant is p_L, p_M or p_H, respectively. Similarly, the probability that each of the B packets generated by the source are part of the LQ, MQ or HQ layer is q_L, q_M or q_H, respectively. Since we are using separate packets for each layer, an MQ user needs to receive the packets belonging to the LQ and MQ layers and an HQ user needs to receive the packets belonging to the LQ, MQ and HQ layers. Finally, assuming that the UEs are uniformly distributed between the SGSNs and RNCs, we define the average number of users served by each SGSN or RNC as $Sa = N_u/S_n$ or $Ra = N_u/R_n$. The notation used in our analysis is summarized in Table I. In order to simplify the analysis, we also assume that packets are not inflated in MCV MBMS. This means that the 2 bits needed to indicate the layer that a packet belongs to are inserted in the unused parts of the encapsulating headers. Finally, we assume that users do not change their content variant preferences over time.

Variable	**Description**
N_u	Number of users participating in a service
B	Number of packets generated by the source
p_i	Probability that a user requests variant i
q_i	Probability that a packet belongs to variant i
S_n	Number of SGSNs in the network
R_n	Number of RNCs in the network
Sa	Average number of users served by each SGSN (N_u/S_n)
Ra	Average number of users served by each RNC (N_u/R_n)

TABLE I
LIST OF VARIABLES

With respect to user traffic in the RAN, the most important metric is the number of packets each UE should receive over the air, at least if no wireless losses occur. In the L/M/HQ MBMS option, each UE should receive the packets from all variants, therefore the total expected number of received packets is $(q_L + q_M + q_H)BN_u$, or simply BN_u. On the other hand, in the LQ MBMS option each UE should receive only the LQ variant, therefore the expected number of received packets is q_LBN_u. Finally, in MCV MBMS each UE should receive exactly the packets corresponding to its desired variant; the p_LN_u UEs desiring the LQ variant should receive q_LB packets, the p_MN_u UEs desiring the MQ variant should receive $(q_L + q_M)B$ packets, and the p_HN_u UEs desiring the HQ variant should receive $(q_L+q_M+q_H)B$ packets. Summing these up we find that the total user traffic received by all UEs should be $[p_Lq_L + p_M(q_L + q_M) + p_H(q_L + q_M + q_H)]BN_u$ or $[q_L + q_M(p_M + p_H) + q_Hp_H)]BN_u$.

With respect to user traffic in the *Core Network* (CN), i.e. the traffic sent from the GGSN to the SGSNs and from the SGSNs to the RNCs, we will again consider the number of packets received by each SGSN and RNC, respectively. For simplicity, we will assume that at least one UE controlled by each SGSN or RNC has joined the service. In this case, in the L/M/HQ MBMS option each such node will receive B packets, that is, all variants, therefore the total user traffic received by all SGSNs and RNCs should be $(S_n + R_n)B$. On the other hand, in the LQ MBMS option each such node will receive q_LB packets, that is, only the LQ variant, therefore the total user traffic received by all SGSNs and RNCs should be $q_L(S_n + R_n)B$.

The analysis for MCV MBMS is slightly more involved. Since we have assumed that at least one UE is served by each SGSN or RNC, each such node will definitely receive at least the LQ variant, which consists of q_LB packets. The MQ variant, which consists of q_MB packets, will only be received by a node if at least one of the UEs that it controls has asked for either the MQ or the HQ variants; this probability is the complement of the probability that all the UEs that the node controls have asked for the LQ variant, or p_L^{Sa} for an SGSN, thus the expected number of MQ packets received by each SGSN is $(1 - p_L^{Sa})q_MB$. Finally, the HQ variant, which consists of q_HB packets, will only be received by a node if at least one of the UEs that it controls has asked for the HQ variant; this probability is the complement of the probability that none of the UEs that the node controls have asked for the HQ variant, or $(1-p_H)^{Sa}$for an SGSN, thus the expected number of MQ packets received by each SGSN is $[1 - (1 - p_H)^{Sa}]q_HB$. We can treat the RNCs identically, replacing Sa with Ra. Summing these up we find that the total user traffic received by all SGSNs should be $\{q_L + (1 - p_L^{Sa})q_M + [1 - (1-p_H)^{Sa}]q_H\}S_nB$, while the total user traffic received by all RNCs should be $\{q_L+(1-p_L^{Ra})q_M+[1-(1-p_H)^{Ra}]q_H\}R_nB$.

VI. PERFORMANCE EVALUATION: SIMULATION

As part of the IST B-Bone project we have created a MBMS simulator based on the 3GPP Release 6 specifications. The simulator uses the Opnet Modeler 11.0 platform and, among other extensions to standard MBMS, it fully supports the MBMS MCV model described above. In this section we will use this simulator to compare MCV MBMS with the LQ MBMS and L/M/HQ MBMS approaches discussed above in terms of their user traffic overhead in a specific scenario; we will also apply the analysis presented in Section V to this scenario, in order to compare the analytical predictions with the simulation results.

We simulated the topology shown in Figure 4, consisting of a single GGSN, two SGSNs (S_n = 2), four RNCs (R_n = 4) and six Node-Bs, or cell controllers, where two of the Node-Bs control cells with 4 UEs and the other four Node-Bs control cells with 9 UEs. We varied the number of UEs joining

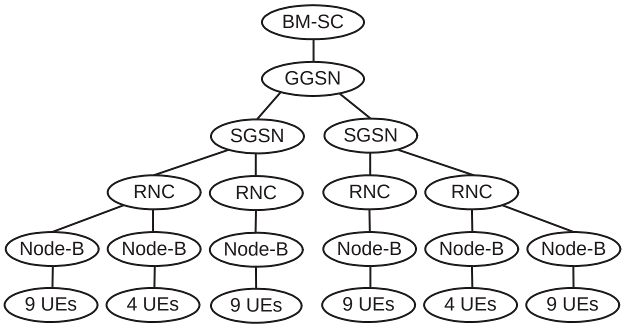

Fig. 4. Simulated topology.

the group in each experiment (N_u) from 1 to 40 and repeated each experiment 30 times. In each experiment N_u UEs were randomly chosen to join a multicast group in random order. Each UE randomly selected a content variant with probabilities $p_L = 0.7$, $p_M = 0.2$ and $p_H = 0.1$. The source generated $B = 1000$ packets distributed to the three content variants with probabilities of $q_L = 0.5$, $q_M = 0.25$ and $q_H = 0.25$.

Using the analysis presented in Section V for the RAN traffic, we find that the expected number of received packets with the L/M/HQ MBMS option is $BN_u = 1000N_u$, with the LQ MBMS options it is $q_L BN_u = 500N_u$, and with MCV MBMS the number is $[q_L + q_M(p_M + p_H) + q_H p_H)]BN_u = [0.5{+}0.250.3{+}0.250.1]1000N_u$ or $600N_u$. These functions of N_u are plotted in Figure 5 against the actual simulation results; each point in the simulation curve represents the average value from 30 experiments along with the 99% confidence intervals. From the figure it is clear that the agreement between the analytical predictions and the simulation results is nearly perfect. In addition, it can be seen that while the L/M/HQ MBMS option requires double the traffic of the LQ MBMS option in order to fully satisfy the 30% of the MQ and HQ users, the MCV MBMS option achieves the same goal by only inflating the LQ MBMS traffic by 20%.

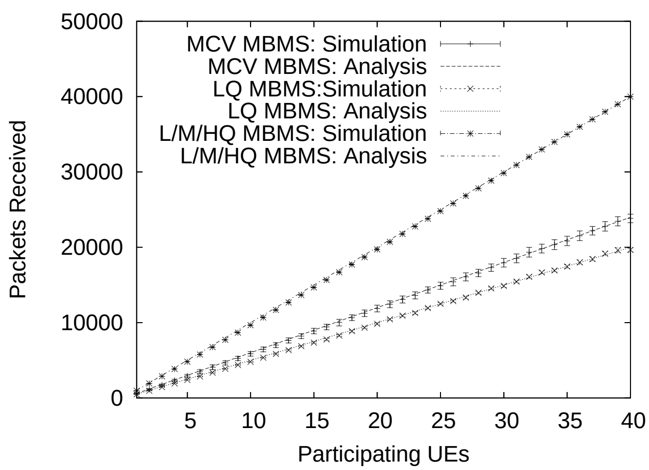

Fig. 5. Number of packets received in the RAN.

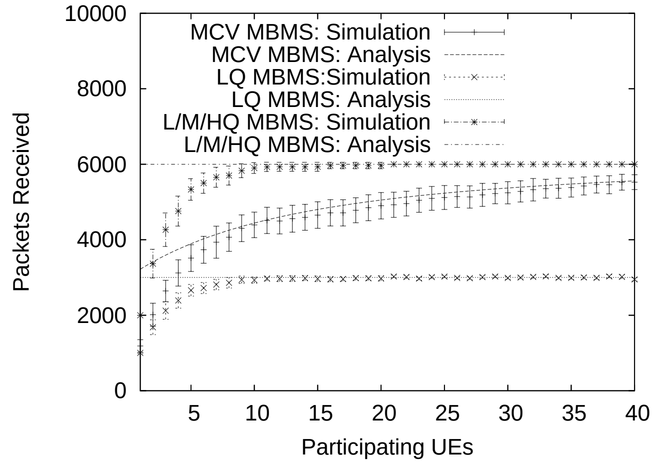

Fig. 6. Number of packets received in the CN.

Using the analysis presented in Section V for the CN traffic, we find that the expected number of received packets with the L/M/HQ MBMS option is $(S_n + R_n)B = 6000$, with the LQ MBMS option it is $q_L(S_n + R_n)B = 3000$, and with MCV MBMS it is $\{q_L + (1 - p_L^{Sa})q_M + [1 - (1 - p_H)^{Sa}]q_H\}S_n B = 2000 - 500(0.7^{Nu/2} + 0.9^{Nu/2})$ for the SGSNs and $\{q_L + (1 - p_L^{Ra})q_M + [1 - (1 - p_H)^{Ra}]q_H\}R_n B = 4000 - 1000(0.7^{Nu/4} + 0.9^{Nu/4})$ for the RNCs, or $6000 - 500(0.7^{Nu/2} + 0.9^{Nu/2}) - 1000(0.7^{Nu/4} + 0.9^{Nu/4})$ in total. These functions of N_u are plotted in Figure 6 against the actual simulation results. The agreement is nearly perfect for 10 UEs or more; with fewer UEs the assumption that at least one UE is served by each SGSN and RNC is not satisfied, therefore the analysis overestimates the, more realistic, simulation results. Again, the L/M/HQ MBMS option requires double the traffic of the LQ MBMS option in order to fully satisfy the MQ and HQ users, since it always forwards the MQ and HQ layers to all nodes. On the other hand, the MCV MBMS option tends to reach L/M/HQ MBMS, since as the number of UEs grows, so does the number of UEs per SGSN and RNC. As a result, the probability that at least one UE served by each such node will request the HQ variant tends to one, meaning that the node will have to receive all content variants. Fortunately, the CN is a wired network, thus the user traffic overhead incurred by each option is of secondary importance compared to the overhead incurred over the air in the RAN.

VII. Conclusions and Future Work

We have presented an extended MBMS model that transmits different variants of the same content to each UE, aiming to maximize the number of UEs participating in an MBMS service while minimizing the amount of user traffic transmitted, especially over the air. We explained how our extended MBMS model can be derived from the standard MBMS model by describing its state management and signaling procedures, and evaluated our extensions against two alternatives based on standard MBMS via analysis and simulation. Our results indicate that our Multiple Content Variant MBMS model can satisfy all users with a small increase in transmission overhead over the air, thus striking a good balance between the standards based MBMS alternatives.

Regarding future work, we are currently focused on obtaining more detailed results of the impact of MCV MBMS to the RAN. Nodes in the CN communicate via dedicated wired links, therefore the number of packets received is an adequate CN performance metric. In contrast, for an interference limited UMTS network the most important performance metric is not the number of packets received by the UEs, but the amount of transmission power spent in order to send them; this shows how much of the available power in a cell was consumed for these transmissions, or, equivalently, how much power remains available for other transmissions. For cells where UEs are served by dedicated radio bearers, the average amount of transmission power spent per multicast packet is proportional to the number of UEs receiving it, therefore the number of packets received is again an adequate metric. However, as the number of UEs receiving a service in a cell increases, it is more economical to employ a (more expensive) common radio bearer, thus making the amount of transmission power spent per multicast packet fixed, regardless of the number of UEs receiving it. Indeed, with MCV MBMS some layers may be sent via common radio bearers and some layers via dedicated radio bearers, since many UEs request the lowest quality layers but only a few request the highest quality ones. Our research is therefore currently focused on extending the analysis and the simulations in order to compare our approach against the standards based alternatives in terms of their actual transmission power requirements, thus taking into account the performance of each alternative with both dedicated and common radio bearers.

Acknowledgements

The work reported in this paper was supported by the IST B-Bone project under contract IST-2003-507607. The simulation platform was provided by the OPNET University Program.

References

[1] 3GPP: “Multimedia Broadcast / Multicast Service (MBMS); Stage 1”, *TS 22.146, V6.6.0*, (2004).
[2] 3GPP: “Multimedia Broadcast / Multicast Service (MBMS) user services; Stage 1”, *TS 22.246, V6.2.0*, (2004).
[3] 3GPP: “Multimedia Broadcast / Multicast Service (MBMS) user services; Architecture and functional description”, *TS 23.246, V6.4.0*, (2004).
[4] W. Fenner: “Internet Group Management Protocol, Version 2”, *RFC 2236*, (1997).
[5] R. Koodli, M. Puuskari, “Supporting packet-data QoS in next-generation cellular networks”, *IEEE Communications*, vol. 39, pp. 180–188, (2001).
[6] B. Li, J. Liu: “Multirate Video Multicast over the Internet: An Overview”, *IEEE Network*, volume 17, pp. 24–29, (2003).
[7] G. Xylomenos: “Group Management for the Multimedia Broadcast / Multicast Service”, *Proc. of the IST Mobile Summit* (2005).
[8] G. Xylomenos, K. Katsaros, G.C. Polyzos: “Analysis of a Multiple Content Variant Extension of the Multimedia Broadcast / Multicast Service”, *Proc. of the European Wireless Conference* (2006).