Benefits and limitations of hybrid (broadcast/mobile) networks

Christoph Heuck

Abstract—**Looking at today's data transmission to mobile terminals, it appears that the capability of existing communication systems is not satisfactory for many future services. Although the mobile network of the third generation (UMTS) is already much more powerful than GSM/GPRS, it still encounters limitations in the transmission of larger files or streams (e.g. audio/video-content) to various users at the same time. By using a hybrid network consisting of a mobile and a broadcast network (e.g. UMTS/DVB-H) those limitations could be overcome. In this paper a queueing model of a hybrid network is derived. The performance and the areas of application of hybrid networks are analyzed in terms of stability and response time. In particular the advantages of a hybrid network in comparison with a classical mobile network are pointed out. In this context two modes of scheduling algorithms for choosing the appropriate branch (mobile or broadcast network) for data transmission are considered.**

Index Terms—**DVB-H, queueing theory, stability, hybrid network**

I. Introduction

FUTURE handheld terminals will be equipped with high resolution displays and improved audio/video capabilities. Extended applications will arise, which take advantage of these new abilities. The applications will require more data to be transmitted than currently known services for handheld terminals. In a mobile environment the required data will be sent via an air interface. Looking at today's solutions for data transmission to mobile terminals, it appears that the capability of existing communication systems is not satisfactory for many new services. Although the mobile network of the third generation (UMTS) is already much more powerful than GSM/GPRS, it still encounters limitations in the transmission of larger files or streams (e.g. audio/video-content) to various users at the same time. By using a hybrid network consisting of both a mobile and a broadcast network (e.g. UMTS/DVB-H [1,2]) those limitations could be overcome. The structure of the hybrid network may encompass a number of mobile cells in the mobile network coinciding with just one broadcast cell in the broadcast network [3]. The performance of a hybrid network depends on both the scheduling algorithm for selecting the appropriate branch (mobile or broadcast network) for data transmission and the mentioned structure of the hybrid network [4].

In this paper a general model of a hybrid network is introduced in section two. In the third section a queueing model of a hybrid network is derived. Subsequently the limitations of the mobile network are discussed. In relation to these results the areas of application of a hybrid network are evaluated in section five. For this purpose the limitations of the hybrid networks are derived analytically in terms of stability and response time. Especially the advantages of a hybrid network in comparison with a mobile network are pointed out. In this context two modes of scheduling algorithms for hybrid networks are introduced and analyzed. Conclusions of this paper are outlined in the final section.

II. General Model of Hybrid Networks

The analysis in this paper is based on the architecture depicted in Fig. 1. Besides the service platform and the mobile terminal, this block diagram shows the network operators (mobile operator, broadcast operator). To establish for instance a connection between the service platform and the mobile terminal the network operators provide access to the different transmission channels. The broadcast network as well as the mobile network can be used for data transmission to the mobile terminal.

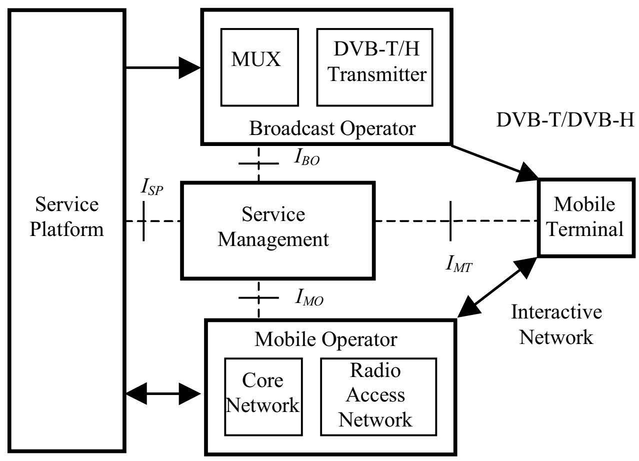

Fig. 1. Architecture of the hybrid network and the location of the service management

This work was supported by the Niccimon project, Lower Saxony, Germany.

C. Heuck is a researcher at Institut fuer Nachrichtentechnik (Institute for Communications Technology) at Technische Universitaet Braunschweig (Braunschweig Technical University), Germany (e-mail: heuck@ifn.ing.tu-bs.de).

A subsystem insuring system control and signalling between the parties involved in the provision of the services on the hybrid network is required. This task is provided by the Service Management and its interfaces. The logical interfaces are: $I_{M(obile)T(erminal)}$, $I_{M(obile)O(perator)}$, $I_{B(roadcast)O(perator)}$ and $I_{S(ervice)P(latform)}$. The indices describe the communication partners of the service management reached via the respective logical interface. In the service management a so called *cost function* should be implemented. This function balances the emerging traffic among both networks by using a suitable scheduling algorithm.

III. Hybrid networks in terms of Queueing Theory

A. Utilization factor, Arrival- and Service rate

In this paper the potential benefits of hybrid networks will be analyzed. Therefore, the utilization factor ρ - well defined in queueing theory - is introduced. In general this parameter depends on the average arrival rate λ and the average service rate μ

$$\rho = \frac{\lambda}{\mu} \tag{1}$$

Thus the utilization factor varies with the average arrival rate of the requests as well as with the varying average amount of data per request and, therefore, the average service rate. This becomes obvious when considering the definition of the service rate as the reciprocal of the average time to serve a request by the network. A system with one service station and a general distribution for both the arrival rate and the service rate (indicated as G/G/1-System) will be stable for $0 \leq \rho < 1$. Sometimes $\rho = 1$ is permitted, too [5]. For the following considerations a unicast network is defined as a network allowing bidirectional point-to-point traffic between two communication partners in general. A broadcast network is defined as a network enabling unidirectional point-to-multi-point communication. In this context, for unicast networks the arrival rate is defined by the number of arriving requests originated by divers users. The utilization factor of such a unicast network depends on the arrival rate of users' requests and the service rate of this network answering these requests. Hence the analysis of a unicast network is related to the number of participating users and, therefore, these queueing parameters can be indicated as *user-based*. Considering a broadcast network the arrival rate is defined by the number of different content elements (e.g. files) fed into the transmission network by the network operator. To determine the utilization factor the average arrival rate is divided by the average service rate this content can be served with. In that case the utilization factor does not depend on the number of involved users. The queueing parameters of the broadcast network therefore are referred to as *content-based*.

Combining a mobile network, which can be considered as a unicast network, and a broadcast network, leads to the question how to evaluate hybrid networks created from the two. If the hybrid network is used to respond clients' requests using both components jointly, the utilization factor of the hybrid network will depend on the users' behavior. In the above sense the queueing parameters describing this hybrid network are *user-based*. However, by taking advantage of the capability of the broadcast network to transmit content only once independently of the number of users requesting this content, the efficiency of the hybrid network does not only depend on the utilization factor of both networks, but also on the popularity of the requested content sent via the broadcast network. To express this advantage gained by the presence of the broadcast network in a hybrid network an efficiency parameter E is introduced. The efficiency E is defined as the ratio of the user-based average service rate of requests μ_B (subscript 'B' for broadcast) that are satisfied via the broadcast network and the content-based average service rate of files μ_{Bc} (additional subscript 'c' for content-based to differentiate this parameter from all the other user-based parameters), which are actually transmitted via the broadcast network:

$$E = \frac{\mu_B}{\mu_{Bc}}. \tag{2}$$

Introducing E additionally allows to calculate the user-based utilization factor ρ_B of the broadcast network defined by the user-based arrival rate λ_B and the content-based service rate μ_{Bc} of different content being actually processed by the broadcast branch:

$$\rho_B = \frac{\lambda_B}{\mu_B} = \frac{\lambda_B}{E \cdot \mu_{Bc}}. \tag{3}$$

With (3) it becomes obvious that the user-based average service rate μ_B can be transformed into the content-based average service rate μ_{Bc} by using the efficiency E.

Relations (2) and (3) can be explained by a simple example: Assume the user-based average arrival rate λ_B of 10 requests per second. All requests refer to the same content. The average content-based service rate μ_{Bc} is defined by the transmission of one file (containing the requested content) per second. In this case 10 requests are satisfied by transmitting the file containing the required content just once. Therefore the user-based service rate μ_B equals 10 requests per second. With (2) the efficiency E is 10. With (3) $\rho_B = 1$ can be found.

B. Modeling a hybrid network

To analyze the performance of a hybrid network the architectures of both the mobile and the broadcast network and the way of balancing the load among the different networks have to be considered. In Fig. 2 a simplified structure of the cells in a hybrid network is depicted. Without restricting the general validity their shape is idealized for illustration purposes.

Particularly, it can be seen that the service area of the hybrid network is covered by both the mobile network and the broadcast network. In each cell of the mobile communication network requests occur with an average arrival rate $\lambda_1, \ldots, \lambda_n$ corresponding to the number of users and their activity. The total average arrival rate λ_{total} in the hybrid network is defined by the sum of all the individual average arrival rates λ_i from the various mobile cells:

$$\lambda_{total} = \lambda_1 + \ldots + \lambda_n . \quad (4)$$

As already mentioned in section II the cost function implemented in the service management chooses a suitable network for the incoming requests according to a scheduling algorithm. Depending on the decision of the cost function a part λ_{Bi} of the average arrival rate λ_i of each cell of the mobile network is served by the broadcast network. The other part of the arriving traffic (characterized by the average arrival rate λ_{Mi} (subscript 'M' for mobile network)) is served by the mobile network in cell *i*:

$$\lambda_{Bi} = \lambda_i - \lambda_{Mi} . \quad (5)$$

Therefore all requests served by the broadcast network form a subset of the total number of requests. For λ_B the following relation (6) can be found:

$$\lambda_B = \lambda_{B1} + \ldots + \lambda_{Bn} . \quad (6)$$

λ_B represents the average arrival rate of all requests reaching the broadcast network (i.e. assigned to it by the cost function).

The total average service rate μ_{total} from the user's perspective results from the average service rates in the different mobile cells μ_{Mi} and in the broadcast cell μ_B. Both the service rate of the broadcast network μ_B and the service rates of the different mobile cells μ_{M1}, ... , μ_{Mn} are seen in conjunction with the statistical probability distribution of the content size in terms of amount of data that has to be sent to the user per request.

Fig. 3 shows a block diagram of a hybrid network described as a model of the queueing theory. This block diagram considers the structure of the hybrid network as well as the influence of the cost function on the different parameters of average arrival and service rates. In the broadcast branch, the user-based parameter μ_B is transformed into the content-based parameter μ_{Bc} by using the efficiency factor *E* introduced above.

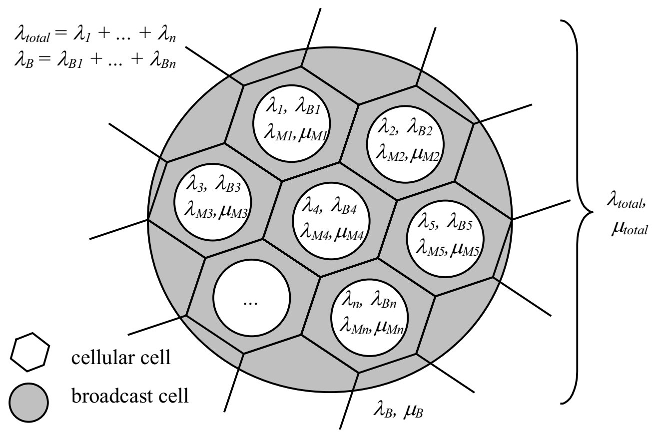

Fig. 2. Simplified structure of a hybrid (broadcast/mobile) network. A certain number of cells of the mobile network coincides with one cell of the broadcast network. An average arrival- and service rate are attached to each cell of the mobile network (λ_n,μ_n) and to the broadcast cell (λ_B,μ_B). Furthermore, each mobile cell is characterized by its share of the arrival rate resulting from requests being served in the broadcast network (λ_{Bn}) and in the individual mobile cell (λ_{Mn}), such that λ_n=λ_{Mn}+λ_{Bn}.

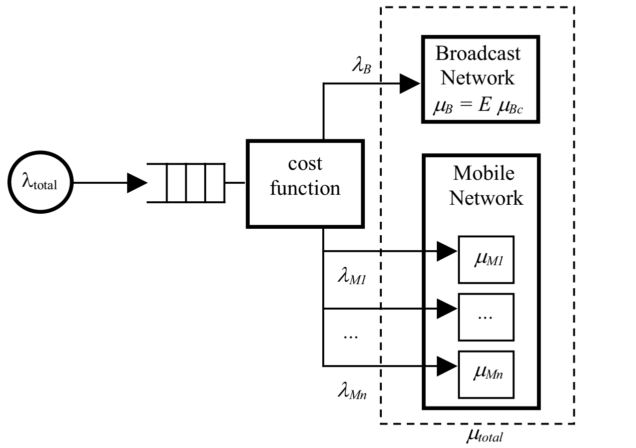

Fig. 3. A hybrid network incorporating the cost function described as a model of the queueing theory

IV. Limitations of Mobile Networks

In a mobile network each user is served by a specific cell, since the person is located in its coverage area. The number of users being served at the same time, the available data rates etc. of each cell are dimensioned according to the expected network load (e.g. considering users' activity) at the time of network planning.

In the near future, applications like video and audio download and streaming are expected to become widely used. Such applications cause heavy traffic on mobile networks, since the requested amount of data might be very high. In general a network becomes instable for an average arrival rate higher than the average service rate. Before reaching this point the system performance decreases rapidly.

To prevent such a situation in a mobile network three mechanisms can be chosen: admission control (AC), packet scheduling (PS) and load control (LC) as described in [6] for a UMTS network. Furthermore the multimedia broadcast/multicast service (MBMS) can be used for a more efficient transmission of popular data in mobile networks, which is not yet considered here.

The AC handles all the new incoming service requests. It checks whether a new circuit or packet switched connection can be established by the mobile network. According to the decision a user's request may be rejected in the case of high network load. Hence a blocking rate can be calculated to specify the probability of the rejection of a connection attempt.

The PS decides how to handle non-real-time traffic by considering e.g. users' priority or the quality of service requirements.

The LC works closely with the AC and the PS. If the load originated by the admitted users exceeds a threshold the LC will take some countermeasures to get the system back to a manageable load situation. For example, the LC may reduce the available data rate for suitable user terminals.

All those mechanisms lead to an improved utilization of the mobile network and assure that the system works in a suitable operation point. However they do not increase its service rate. Therefore a value of the average arrival rate exists up to which the requests can be appropriately served by the mobile network. If this value is exceeded, the network does not work satisfactorily (in terms of an acceptable service from the users point of view). Since there are so many parameters which can be set by the operator, it is hard to calculate an appropriate value of the arrival rate for a given service rate analytically. Therefore this value may be found by practical experience or may be indicated by simulation. For all *n* mobile cells a necessary condition to be stable is that the utilization factor ρ_{Mi} defined in (7) remains in the range between zero and one.

$$\rho_{Mi} = \frac{\lambda_{Mi}}{\mu_{Mi}} \quad (7)$$

V. BENEFITS OF HYBRID NETWORKS

As shown in the previous section the mobile network can be kept in a stable state because of AC, PS and LC. However exploitation of the methods of PS and LC results in the blocking of call attempts. With a rising utilization factor (e.g. caused by new emerging applications or changing user activity) the blocking probability rises, too. To counteract this process an additional broadcast network can optimize the service quality of the network perceived by the user.

When using a hybrid (broadcast/mobile) network the additional broadcast network needs to be utilized in an efficient way. Hence, a capable scheduling algorithm assigning requests to one of the two networks has to be implemented in the cost function. In general, the scheduling algorithm can be utilized in two modes: The first mode only considers the additional data rate made available by the broadcast network. In this case the broadcast network is only used for additional unicast transmissions of requested data to each single user. For the following discussion, this mode is referred to as *unicast mode*. In contrast, a second mode takes the broadcast ability of the broadcast network into account. In this case the broadcast network is used to send popular data to multiple users by sending the data only once. For the following discussion this mode is referred to as *broadcast mode*.

Both the mobile network and the broadcast network have to remain stable. Therefore there are two conditions that have to hold. They take the utilization factor of both networks defined in (3) and (7) into account. Both the utilization factor of the mobile network cells ρ_{Mi} and the utilization factor of the broadcast network ρ_B have to be smaller than one. It follows:

$$\begin{aligned} &a)\quad \lambda_B < \mu_B \\ &b)\quad \lambda_{Mi} < \mu_{Mi} \end{aligned} \quad (8)$$

With (2), (4), (5) and (6) equation (9) follows:

$$\begin{aligned} &a)\quad \lambda_{total} < E \cdot \mu_{Bc} + \sum_{i=1}^{n} \lambda_{Mi} \\ &b)\quad \lambda_{Mi} < \mu_{Mi} \end{aligned} \quad (9)$$

Fig. 4 illustrates the results. Here the areas of application of the hybrid network using the unicast and the broadcast mode are plotted as a function of λ_{total}. Four different areas labeled A1...A4 can be identified: If the limitation of the mobile network is not yet reached there is no need for a hybrid network, since all users' requests can be satisfied by the mobile network alone. In Fig. 4 area A1 visualizes that situation. If the average arrival rate of requests λ_i will be higher than the average service rate μ_{Mi} in some or all cells of the mobile network coinciding with a single broadcast cell, then the mobile network is not able to service all arriving requests and the *mobile limit* is reached. This is the point at which the additional broadcast branch will be used in order to enhance the network capacity. The portion of the arrival rate of requests reaching the broadcast network is defined by λ_B. Its value depends on λ_{total} and the decision of the cost function.

The different areas A2, A3 and A4 reflect the different operational modes of the hybrid network. Note that the position and the boundaries of the different areas depend on several parameters e.g. the average service rate of the broadcast network μ_B and the average service rates of the mobile cells μ_{Mi}. Therefore Fig. 4 should be understood just as a rough indication of the arrangement of the different areas.

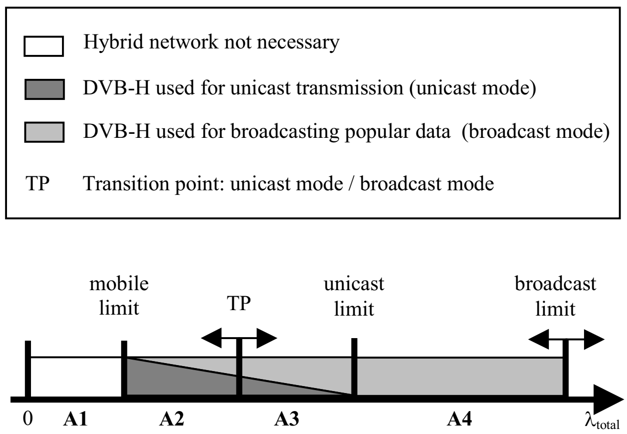

Fig. 4. Visualization of the areas A1 ... A4 of application of hybrid networks using the unicast and the broadcast mode.

Both in A2 and in A3 the unicast mode can be used in the hybrid network. For unicast, the popularity of the requested content is not taken into account. With *E* set to one equation (9) can be simplified to:

$$\begin{aligned} &a)\quad \lambda_{total} < \mu_{Bc} + \sum_{i=1}^{n} \lambda_{Mi} \\ &b)\quad \lambda_{Mi} < \mu_{Mi} \end{aligned} \quad (10)$$

Following (10a) λ_{total} has to be smaller than the sum of the average service rate of the broadcast branch and the cumulated average arrival rates λ_{Mi} served in the n cells of the mobile network to comply with the stability condition. For a properly working hybrid network the average arrival rate λ_{Mi} in the individual cells of the mobile network has to be limited by the cost function such that no overload situation occurs. (10b) takes care of this requirement. At the *unicast limit* the hybrid network using the unicast mode exclusively leaves the stable state.

If the broadcast mode will be used in the hybrid network, the stability constraint for the mobile network stays the same. However, for the broadcast branch the efficiency E might be larger than one (depending on the popularity distribution of the requested content). Therefore the hybrid network using the broadcast mode is able to serve a higher total average arrival rate of requests λ_{total} instead of a hybrid network using just the unicast mode. This effect also can be seen by comparing equations (9) for E>1 and (10). A hybrid network operated in the broadcast mode remains stable in areas A2, A3 and A4. At the *broadcast limit* the hybrid network using the broadcast mode leaves the stable state. The difference between the positions of the unicast limit and the broadcast limit on the λ_{total}-axis depends on the efficiency E. Its value varies with the popularity distribution of the requested files. If only some files in the content set are quite popular, this results in a very high efficiency E. In that case, the position of the broadcast and the unicast limit on the λ_{total}-axis strongly differentiates (see Fig. 4). But the broadcast limit can approach the unicast limit, if many users request different content and therefore the efficiency E is close to one.

It is possible to apply both the unicast mode and the broadcast mode in the areas A2 and A3. When considering the response times for both concepts, however, it becomes obvious that A2 is the area where unicast mode has to be favored, while for the area A3 the broadcast mode should be preferred:

To clarify that proposition the average response time, defined as the time between the request and the reception of the requested content by the clients is introduced. In this context, R_U is the response time in a hybrid network using the unicast mode. In contrast R_B is the response time in a hybrid network using the broadcast mode.

After the mobile limit is reached the response time R_U is smaller than R_B, since a hybrid network using the unicast mode can serve the arriving requests immediately. When the scheduling algorithm has chosen a branch the data is transmitted instantaneously. In contrast a hybrid network applying the broadcast mode separates popular and non popular data. To serve multiple users by broadcasting popular content, requests considering this content have to be collected over some time. This time can also be used to determine the popularity distribution among the offered content by analyzing the number of requests related to the different files. However, the transmission of the requested files is suspended and the response time R_B increases. With an increasing λ_{total} the broadcast mode regains this lost time by broadcasting popular files to more and more users. By this concept, load is removed from the hybrid network expressed quantitatively with the factor E. A hybrid network applying the unicast mode exclusively can not exploit this advantage. Instead it has to allocate the same amount of resources for each user in a unicast transmission Hence it is more loaded. A rising network utilization leads to an increasing average waiting time in the queues of the different service stations. Thus the unicast mode shows worse performance with a rising λ_{total} in comparison to the broadcast mode. A transition point separating A2 from A3 can be found where the additional collecting time of the broadcast mode is compensated by the longer waiting time in the queues of the unicast mode.

The position of this transition point also varies with the popularity of the requested files and the chosen additional collecting time selected when using the broadcast mode. For a concentrated popularity in only some of the requested files the transition point approaches the mobile limit of Fig. 4. If the popularity is distributed over a great number of files the transition point is close to the unicast limit. In the extreme case that all users request different content the unicast limit and the transition point coincide with the broadcast limit.

VI. CONCLUSION

This paper presents a queueing model of a hybrid (broadcast/mobile) network. In particular, a new parameter E measuring the efficiency of the additional broadcast branch in a hybrid network has been introduced. By using the queueing model conditions for stability of hybrid networks are derived. Two modes for a scheduling algorithm balancing the load between the mobile and the broadcast branch are presented. Their areas of application are analyzed.

The results of this paper are a basis for evaluating a hybrid network. In particular, a network operator can estimate the stability of a hybrid network in consideration of the network structure and the behavior of the users in its service area. Furthermore, it is shown that the performance of the hybrid network depends on the scheduling algorithm balancing the load among the broadcast and the mobile communication network.

REFERENCES

[1] U. Reimers, "Digital Video Broadcasting",2nd edition, *Springer Berlin Heidelberg New York*, 2004.

[2] M. Kornfeld, "DVB-H - The emerging standard for mobile data communication", IEEE *Proc. of IEEE ISCE'04*, Reading/U.K., 1-3 September 2004.

[3] P. Unger, T. Kuerner, "Scenarios for Radio Planning of DVB-H / UMTS Hybrid Networks", *European Cooperation in the Field of Scientific and Technical Research (COST 273)*, Duisburg, 20.–22. September, 2004.

[4] C. Heuck, "Analysis of the performance of hybrid (broadcast/mobile) networks", *Proc. of IEEE ISCE'04*, Reading/U.K., 1-3 September 2004.

[5] L. Kleinrock, "Queueing Systems – Volume I: Theory", *Wiley New York*, 1975.

[6] J. Laiho, A. Wacker, T. Novosad, "Radio Network Planning and Optimisation for UMTS", *Wiley New York*, 2002.