Resource Planning and Bandwidth Allocation in Hybrid Fiber-Coax Residential Networks

David Griffith, Kotikalapudi Sriram, Liliya Krivulina, and Nada Golmie
National Institute of Standards and Technology (NIST)
100 Bureau Drive, Stop 8920
Gaithersburg, MD 20899-8920
Email: david.griffith@nist.gov

Abstract

The introduction of new high bandwidth services such as video-on-demand by cable operators will put a strain on existing resources. It is important for cable operators to know how many resources to commit to the network to satisfy customer demands. In this paper, we develop models of voice and video traffic to determine the effect on demand growth on hybrid fiber-coax networks. We obtain a set of guidelines that network operators can use to build out their networks in response to increased demand. We begin with one type of traffic and generalize to an arbitrary number of high-bandwidth CBR-like services to obtain service blocking probabilities. These computations help us to determine how cable networks would function under various conditions (i.e., low, medium, and heavy loads). We also consider how the growth rate of the popularity of such services would change over time, and how this impacts network planning. Our findings will help cable operators estimate how much bandwidth they need to provision for a given traffic growth model and connection blocking requirement.

I. Introduction

As broadband integrated voice and data access finally becomes a reality for millions of consumers worldwide, there will inevitably be steady growth in demand for high-bandwidth services that will make careful deployment planning by cable operators a necessity. It is envisioned that video on demand (VOD), large file downloads (e.g. data, music, and video), and voice over IP (VoIP) will become common over cable networks.

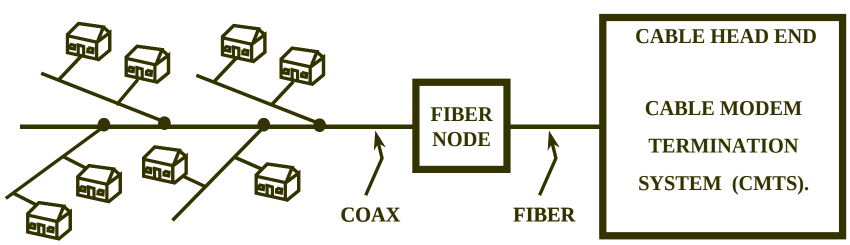

Fig. 1. Example HFC network, showing residences connected to the fiber node by a coaxial tree network.

In a hybrid fiber-coax (HFC) network, illustrated in Fig. 1, the residences are connected to a coax tree and branch network which terminates at a fiber node. At the fiber node, the traffic signals are converted from electrical to optical and transmitted to the Cable Modem Termination System (CMTS) located in cable Head End office. The Data Over Cable System Interface Specification (DOCSIS) protocol is the current industry standard for the physical and medium access control (MAC) layers of upstream/downstream communications between a residential cable modem (CM) and the CMTS over the HFC networks [1]. The spectral allocation for the upstream and downstream cable modem services and the downstream broadcast TV are illustrated in Fig. 2. DOCSIS allows a choice of various frequency channel widths and modulation techniques for the upstream

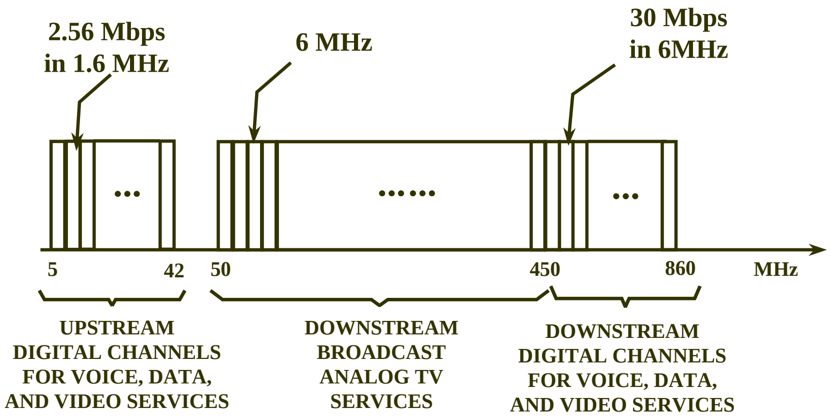

Fig. 2. Spectral allocation for various HFC services.

and downstream communications. So Fig. 2 illustrates a typical implementation wherein the upstream uses 1.6 MHz frequency channels with QPSK modulation and the downstream uses 6 MHz frequency channels with 64-QAM modulation for the DOCSIS cable modem applications. With QPSK, each 1.6 MHz upstream channel provides a data rate of 2.56 Mbps and with 64-QAM, each downstream channel provides a data rate of 30 Mbps [1].

The new services carried over HFC are characterized by a plethora of throughput rates and delay requirements, along with varying degrees of burstiness [2]. An important goal of the system architecture and protocol design for hybrid fiber-coax (HFC) networks has been efficient service integration and real-time statistical multiplexing of various traffic types. With this objective in view, the upstream part of the MAC protocol has been carefully designed to operate with the best efficiency possible, based on contributions from many researchers [3], [4], [5], [6], [7], [8], [9], [10], [11]. The DOCSIS protocol essentially permits statistical packet multiplexing and scheduling with suitable Quality of Service (QOS) for the downstream channels. This is because the downstream channels (unlike the upstream) have no contention due to collisions. However, there are service blocking issues in the downstream due to random call arrivals/departures and inadequate downstream bandwidth. The traffic engineering and bandwidth growth planning for the downstream is becoming increasingly more complex due to the mix of various emerging services that require disparate bandwidth and latency requirements. In this paper, we focus on these downstream issues for a mix of high-bandwidth CBR-like data services of interest, e.g., VOD, large file downloads. Even though these services are transported using TCP/IP or UDP/IP over DOCSIS protocol, they are CBR-like in that they require high instantaneous bandwidth for relatively long durations, they have very high burstiness and low delay tolerance. Essentially, they need to be treated like CBR flows in order to be given adequate QOS treatment as well as for bandwidth allocation and resource planning purposes. The mix of various CBR and VBR services in upstream channels has been analyzed and studied in the literature [2], [4], [5], [9], [10]. The work reported here can also be extended in the future to consider a mix of CBR and VBR (statistically multiplexed) services in the downstream channels. The applicability of our analytical models in this paper is to produce a line card deployment schedule for a cable operator under various applications traffic growth scenarios while serving a mix of heterogeneous high-bandwidth CBR-like downstream services.

The remainder of this paper is organized as follows. In Section II we develop an analytical model of the downstream data channels when the load is provided by multiple types of CBR data services. We use this model to develop expressions for the probability that a customer will experience blocking by being unable to access a desired service or set of services. In Section III we use the model that we developed in Section II to produce a line card deployment schedule for a cable operator under various usage growth scenarios. In Section IV, we summarize our results and discuss how they may be extended to examine other types of data traffic such as compressed video and best effort downloads using TCP.

II. Analytical Model

For the case where the network supports a single CBR service with a bandwidth of B, the analysis is straightforward. The system can be in one of $M+1$ states, where $M = \lfloor B_{\max}/B \rfloor$ is the maximum number of users that can be accommodated using a bandwidth of $B_{\max}$. The blocking probability is given by the Erlang-B loss formula,

$$P_B = \frac{\rho^M/M!}{\sum_{n=0}^{M} \rho^n/n!}, \tag{1}$$

where $\rho = \lambda/\mu$ is the channel utilization, λ is the mean arrival rate of requests for the service, assuming Poisson arrivals, and $1/\mu$ is the average time that the service is used by a customer, where the service time distribution is arbitrary.

It is certain that the bandwidth on HFC networks will be shared by multiple application types. VOD and large file downloads are essentially CBR-like services requiring sustained bandwidth in the downstream channels. Therefore, we next consider the case where there are two CBR services available on the downstream channels. We refer to these two services as Service 1 and Service 2. The state of the downstream data channels is denoted by an ordered pair (n_1, n_2), where n_1 and n_2 are the number of instances of Service 1 and Service 2 that are active, respectively. An example of a state diagram that is generated by two CBR services is shown in Fig. 3. In the figure, the maximum bandwidth available on the downstream digital channels is 30 Mbps, which corresponds to a single 6 MHz channel's being active. The request rates for Service 1 and Service 2 are λ_1 and λ_2 respectively, assuming both services' arrivals can be characterized by Poisson processes; the average time a customer uses Service 1 or Service 2 is $1/\mu_1$ and $1/\mu_2$. The state transition rates along each column and row are the same as those in a system supporting a single CBR service with bandwidth B_1 or B_2, respectively. The bandwidth limitations produce the triangular shape of the state space seen in the figure.

From the form of the state transitions shown in Fig. 3, we can obtain a pair of local balance equations that relate the steady-state probability of being in state (n_1, n_2) to the probabilities of the adjacent states (n_1-1, n_2) and (n_1, n_2-1) and the service utilization levels $\rho_1 = \lambda_1/\mu_1$ and $\rho_2 = \lambda_2/\mu_2$. We have

$$p(n_1, n_2) = \frac{\rho_1}{n_1} p(n_1 - 1, n_2) \tag{2}$$

$$p(n_1, n_2) = \frac{\rho_2}{n_2} p(n_1, n_2 - 1) \tag{3}$$

for $n_1 \geq 1$ and $n_2 \geq 1$. By recursively applying these two equations to an arbitrary state (n_1, n_2), we obtain the following

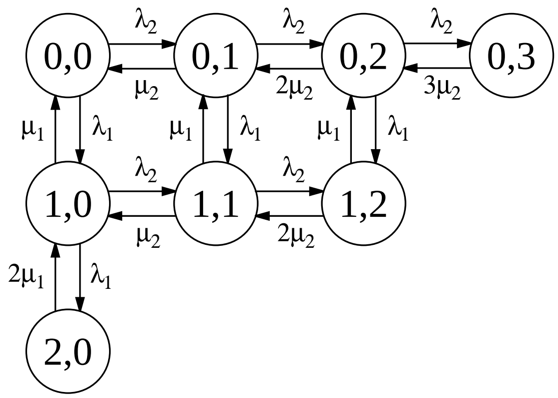

Fig. 3. Two-dimensional state diagram that results when $B_{\max} = 30$ Mbps, $B_1 = 12$ Mbps, and $B_2 = 9$ Mbps.

expression for $p(n_1, n_2)$:

$$p(n_1, n_2) = \frac{\rho_1^{n_1} \rho_2^{n_2}}{n_1! n_2!} p(0,0). \tag{4}$$

To obtain $p(0,0)$, we apply the condition that the state probabilities must sum to one. A state (n_1, n_2) has a non-zero steady-state occupancy probability if $n_1 B_1 + n_2 B_2 \leq B_{\max}$, i.e., if the bandwidth usage associated with being in state (n_1, n_2) does not exceed the maximum available downstream bandwidth. Carrying out the summation and solving for $p(0,0)$ gives

$$p(0,0) = \left(\sum\sum_{\{i,j \mid iB_1 + jB_2 \leq B_{\max}\}} \rho_1^i \rho_2^j / i! j! \right)^{-1}. \tag{5}$$

To quantify the performance of the downstream data delivery system for a given value of $B_{\max}$, we compute the blocking probabilities of the two services, which are

$$P_{B_1} = \sum_{i=0}^{M_1} p(N_1(i), i) \tag{6}$$

and

$$P_{B_2} = \sum_{j=0}^{M_2} p(j, N_2(j)). \tag{7}$$

where $M_1 = \lfloor B_{\max}/B_1 \rfloor$ and $M_2 = \lfloor B_{\max}/B_2 \rfloor$ are respectively the maximum number of instances of Service 1 or Service 2 that can be supported in the absence of any instances of the other service. $N_1(i)$ is the maximum number of instances of Service 1 that can be supported given that i instances of Service 2 are active, and $N_2(j)$ is the maximum number of instances of Service 2 that can be supported given that j instances of Service 1 are active. Thus, $N_1(0) = M_1$ and $N_2(0) = M_2$. Formally, we have

$$N_1(n_2) = \left\lfloor \frac{B_{\max} - n_2 B_2}{B_1} \right\rfloor \tag{8}$$

and

$$N_2(n_1) = \left\lfloor \frac{B_{\max} - n_1 B_1}{B_2} \right\rfloor. \tag{9}$$

In the example shown in Fig. 3, $P_{B_1} = p(2,0) + p(1,1) + p(1,2) + p(0,3)$.

We can generalize the preceding discussion to the case where we have S CBR services whose respective bandwidths are $B_1, B_2, \ldots, B_S$. The state of the system is given by the value of the vector $\mathbf{n} = [n_1, n_2, \ldots, n_S]$, where n_k is the number of instances of service k that are occupying the downstream channels. The S-dimensional state diagram associated with the system is a generalization of the two-dimensional state diagram shown in Fig. 3 and leads to a set of S local balance equations for an arbitrary state $\mathbf{n}$, similar to equations (2) and (3); the i^{th} equation is

$$p(\mathbf{n}) = \frac{\rho_i}{n_i} p(n_1, n_2, \ldots, n_{i-1}, n_i - 1, n_{i+1}, \ldots, n_S), \tag{10}$$

where $\rho_i = \lambda_i/\mu_i$ is the utilization level of Service i. The requests for each service are assumed to follow Poisson arrival processes. Recursively solving this system yields the following expression for $p(\mathbf{n})$:

$$p(\mathbf{n}) = p(\mathbf{0}) \prod_{k=1}^{S} \frac{\rho_k^{n_k}}{n_k!}, \tag{11}$$

which is a product form solution with respect to the elements of the vector $\mathbf{n}$. To obtain an expression for $p(\mathbf{0})$, we use the fact that the state probabilities must sum to unity. Since $\mathbf{n}$ is a valid state only if $\sum_{k=1}^{S} n_k B_k \leq B_{\max}$, it follows that

$$p(\mathbf{0}) = \left(\sum \cdots \sum_{\{\mathbf{n} \mid \sum_{k=1}^{S} n_k B_k \leq B_{\max}\}} \prod_{k=1}^{S} \frac{\rho_k^{n_k}}{n_k!} \right)^{-1}. \tag{12}$$

To compute the blocking probability for the i^{th} CBR service, we define the vector $\mathbf{m}_i$ that contains all elements of the vector $\mathbf{n}$ except for n_i:

$$\mathbf{m}_i = [n_1, n_2, \ldots, n_{i-1}, n_{i+1}, \ldots, n_S]. \tag{13}$$

We define the maximum number of instances of Service i that the system can support when the system is supporting the other services in the amounts given by $\mathbf{m}_i$ to be $N_i(\mathbf{m}_i)$. It is found by dividing the residual bandwidth by B_i and rounding down to the nearest integer, giving

$$N_i(\mathbf{m}_i) = \left\lfloor \frac{1}{B_i} \left(B_{\max} - \sum_{\substack{k=1 \\ k \neq i}}^{S} n_k B_k \right) \right\rfloor. \tag{14}$$

The blocking probability for Service i is found by summing over the probabilities of states where the number of instances of Service i, given $\mathbf{m}_i$, is the maximum:

$$P_{B_i} = \sum \cdots \sum_{\{\mathbf{m}_i \mid \sum_{\substack{k=1 \\ k \neq i}}^{S} n_k B_k \leq B_{\max}\}} p(\mathbf{m}_i, N_i(\mathbf{m}_i)). \tag{15}$$

For the numerical results that follow, it is instructive to think of the blocking probability as a function of ρ_1, ρ_2, and c:

$$P_{B_i} = P_{B_i}(c, \rho_1, \rho_2), \tag{16}$$

where, c is the number of downstream RF channels required to provide at least $B_{\max}$ total downstream bandwidth. If we assume that each downstream RF channel can carry 30 Mbps (e.g., using 64-QAM modulation in a 6 MHz RF channel), then $B_{\max}$ and c are related as follows:

$$c = \left\lceil \frac{B_{max}}{30} \right\rceil. \tag{17}$$

For the numerical results related to traffic engineering, we obtain the value of $B_{\max}$ first and then the value of c using Eqs. 15 and 17, respectively. This value of c would be the number of downstream channels that are required to produce a desired set of values for the service blocking probabilities $\{P_{B_i}\}_{i=1}^{S}$.

III. NUMERICAL RESULTS

In this section we use the theoretical models we developed in Section II to develop a hypothetical deployment schedule for a cable operator. We consider a single HFC fiber node serving a community of 100 households. We assume low initial penetration (10% usage of digital HFC services) in the year 2004, and we assume that the usage level versus time follows a sigmoid curve given by the expression

$$\Pi = \left(1 + e^{-4s(t-t_0)}\right)^{-1}, \tag{18}$$

where t_0 is the time where $\Pi = 0.5$ and s is the slope of the curve (i.e. the penetration growth rate) at time $t = t_0$. In Fig.4 we plot curves for three values of s, where $\Pi(2004) = 0.1$. Note that increasing s from 0.05 to 0.1 produces a greater leftward shift and contraction of the curve than does an increase in s from 0.1 to 0.15.

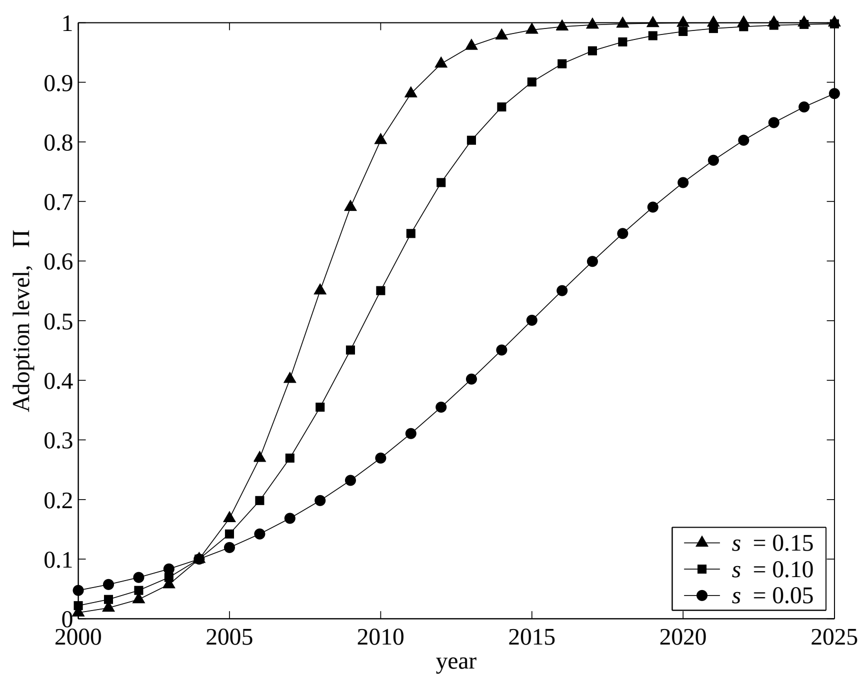

Fig. 4. Adoption curves for data over cable modem using growth rates of 5%, 10%, 15% when $\Pi = 0.5$, assuming 10% penetration in 2004.

For the two services that we are considering, we want the blocking probability for each not to exceed $P_m = 0.001$. In order to determine the number of channels that are required to satisfy this design goal, we need to compute P_{B_1} and P_{B_2} as functions of the service utilization levels ρ_1 and ρ_2 for each value of c (see Eqs. 15, 16 and 17). To compute the utilization levels, we assumed that the average time that a customer spends using the VOD service (Service 1) and the high-speed download service (Service 2) are $1/\mu_1 = 2$ hours and $1/\mu_2 = 15$ minutes, respectively. The maximum value for the arrival rates of requests for the two services was obtained by assuming that each customer requests a service at the same rate that that service completes. In other words, at peak usage, a customer requests the VOD service every two hours and the download service every 15 minutes. Thus, if H households are using the network's digital channels, the maximum utilization of each service is $\rho_1 = \rho_2 = H$. We determine the set of ordered pairs (ρ_1, ρ_2) for which $P_{B_1}(c, \rho_1, \rho_2) \leq P_m$ and $P_{B_2}(c, \rho_1, \rho_2) \leq P_m$. We use the smallest value of c that satisfies the design criteria.

In Fig.5, we show a contour plot of the surface $c_{\min}(\rho_1, \rho_2)$ that is generated when we consider two services with downstream bandwidths $B_1 = 2$ Mbps and $B_2 = 10$ Mbps. The numbers superimposed on each of the boundary lines in the figure apply to the region immediately to the right of the boundary line; the surface in the region that includes the origin has a value of 5. Each of the boundary lines in the figure is approximately linear with a slope of approximately -5, which is the negative of the ratio B_2/B_1.

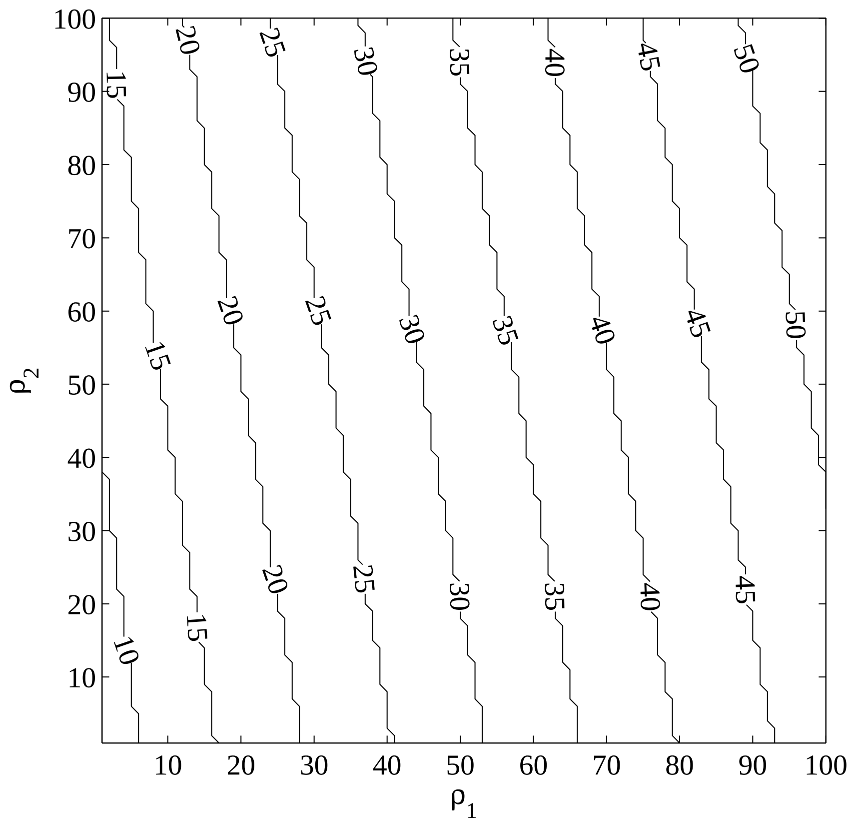

Fig. 5. Number of channels required to achieve blocking probabilities for Service 1 and Service 2 less than $P_m = 0.001$ when $B_1 = 2$ Mbps and $B_2 = 10$ Mbps, plotted vs. service utilization levels ρ_1 and ρ_2.

We carried out a similar set of calculations for the case where $B_1 = 5$ Mbps and $B_2 = 10$ Mbps. The results are shown in

Fig. 6. By examining this figure, we see that each of the boundary lines is again approximately linear, this time with a slope that is close to -2, which is again the negative of the ratio of the two CBR service bandwidths. This observation suggests that for the general case where there are S CBR services, the channel count thresholds in the S-dimensional space associated with the utilization vector $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_S]$ forms a hyperplane whose normal vector is determined by the ratios of the service bandwidths. In this case, there is a region of the (ρ_1, ρ_2) plane where the two service blocking probabilities P_{B_1} and P_{B_2} are both greater than the maximum acceptable blocking probability P_m. This region lies to the right of the line marked, "> 50." Supporting a network load in this region would require adding additional downstream digital channels.

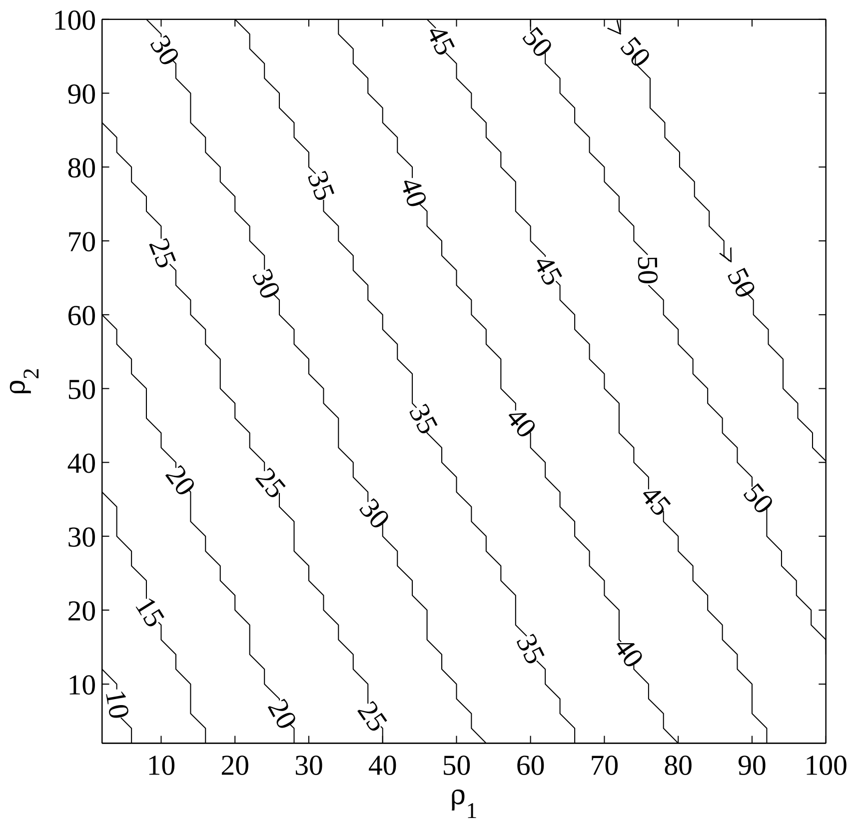

Fig. 6. Number of channels required to achieve blocking probabilities for Service 1 and Service 2 less than $P_m = 0.001$ when $B_1 = 5$ Mbps and $B_2 = 10$ Mbps, plotted vs. service utilization levels ρ_1 and ρ_2.

Once we know how many downstream digital channels are required to achieve a given level of service accessibility, we can use the projected growth curves in Fig. 4 to determine a set of deployment schedules for the downstream line cards. These will allow the network operator to provide an acceptable level of service to his customers while avoiding the excessive costs associated with rolling out unnecessary equipment. We expect that, in practice, the downstream line cards cannot be procured with the granularity of a single downstream RF channel. Therefore, in illustration of our results, we assume that line cards come with a granularity of 5 downstream RF channels per unit. So each time a line card is deployed by the service provider, an addition of 5×30 Mbps = 150 Mbps of downstream bandwidth occurs.

For each year from 2005 to 2025, inclusive, we computed the number of households in the set of 100, attached to the fiber node, that are using digital services. This number, H, corresponds to the maximum utilization for each of the two CBR services. Plotting this point on the graph in Fig. 5 yields the minimum number of channels needed to support peak load while maintaining a service blocking probability of at most 0.001. Plotting the required number of channels vs. time for each of the three growth rates that we considered (in Fig. 4) produces the graph shown in Fig. 7. The number of downstream channels grows in steps of five because of the above-stated assumption regarding deployment granularity (i.e., five RF channels per line card). This graph is similar to the adoption growth curves in that a peak growth rate of 5% produces a rollout schedule that is considerably less aggressive than the ones that result from a growth rate of 10% or 15%.

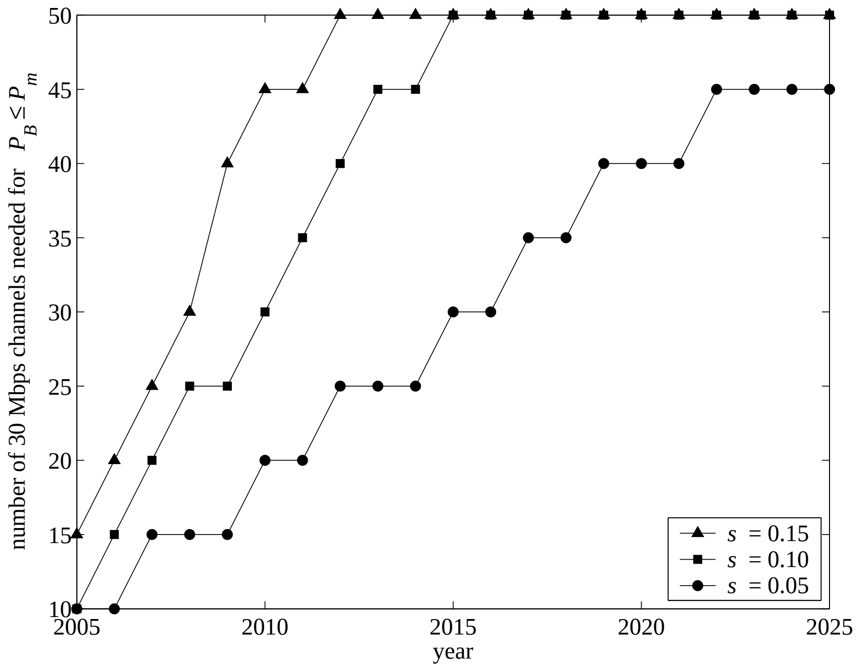

Fig. 7. Number of 30 Mbps channels required to achieve a maximum blocking probability of 0.001 for Service 1 and Service 2 plotted versus time using the adoption curves in Fig. 4.

IV. Summary

In this paper, we examined some of the issues associated with deploying resources in HFC networks to support emerging downstream data services. We developed a theoretical model that allows us to obtain the blocking probability of an arbitrary CBR-like service out of a set of such services that are simultaneously supported over the HFC. We primarily focused on VOD and large file download applications that are essentially CBR-like services requiring sustained bandwidth in the downstream channels. For these types of applications, we presented suitable connection blocking and traffic forecasting models. Using these models, we obtained a set of downstream bandwidth deployment and line-card rollout schedules for a variety of usage growth scenarios. These results can be particularly useful for HFC-based ISPs to minimize cost and connection blocking probability while planning and deploying the HFC resources to accommodate higher take rates and traffic growth over time. These results can also be extended to the case where there is a mixture of CBR and VBR (statistically multiplexed) services.

References

[1] "Data-Over-Cable Service Interface Specifications DOCSIS 1.1," http://www.cablemodem.com/downloads/specs/SP-CMCI-I09-030730.pdf, July 2003.

[2] J.E. Dail, M.A. Dajer, C.-C. Li, P.D. Magill, C.A. Siller, Jr., K. Sriram, and N.A. Whitaker, "Adaptive Digital Access Protocol: A MAC Protocol for Multiservice Broadband Access Networks," *IEEE Communications Magazine*, pp. 104-112, March 1996.

[3] D. Sala and J. Limb, "A protocol for efficient transfer of data over fiber/cable systems", *Proc. of IEEE INFOCOM*, 1996.

[4] N. Golmie, Y. Saintillan, and D. Su, "A Review of Contention Resolution Algorithms for IEEE 802.14 Networks," *IEEE Communications Surveys*, 1Q99. Also appears in a book entitled, **Cable Modems: Current Technologies and Applications**, IEEE Press, ISBN 0-7803-5395-X, pp. 193-231 (1999).

[5] K. Sriram, "Performance of MAC Protocols for Broadband HFC and Wireless Access Networks," *Advances in Performance Analysis*, vol. 1, no. 1, pp. 1-37, March 1998. Also appears in a book entitled, **Cable Modems: Current Technologies and Applications**, IEEE Press, ISBN 0-7803-5395-X, pp. 193-231 (1999).

[6] N. Golmie, S. Masson, G. Pieris, and D.H. Su, "A MAC protocol for HFC networks: Design issues and performance evaluation," *Computer Communications*, vol. 20, no. 12, pp. 1042-1050, November 1997.

[7] B.T. Doshi, S. Dravida, P.D. Magill, and C. A. Siller, Jr., and K. Sriram, "A Broadband Multiple Access Protocol for STM, ATM, and Variable Length Data Services on Hybrid Fiber-Coax Networks," *Bell Labs Technical Journal*, Volume 1, Number 1, Summer 1996, pp. 36-65.

[8] C. Bisdikian, B. McNeil, R. Norman, "ms-START: A random access algorithm for the IEEE 802.14 HFC network," *Computer Communications*, (special issue on "Recent advances in networking technology"), pp. 876-887, Sept. 1996.

[9] K. Sriram and P.D. Magill, "Enhanced Throughput Efficiency by Use of Dynamically Variable Request Mini-Slots in MAC Protocols for HFC and Wireless Access Networks," *Telecommunications Systems* Vol. 9, Nos. 3,4, pp. 315-333 (1998).

[10] K. Sriram, "Performance of ATM and Variable Length Packet Access in Broadband HFC and Wireless Networks," *Proc. of the IEEE ICUPC*, Vol. 1, pp. 495-501, Florence, Italy, October 5-9, 1998.

[11] M. D. Corner, J. Liebeherr, Nada Golmie, Chatschik Bisdikian, and David H. Su, "A Priority Scheme for the IEEE 802.14 MAC Protocol for Hybrid Fiber-Coax Networks," *IEEE/ACM Transactions on Networking*, Vol. 8, No. 2, pp. 200-211, April, 2000