AN OPTIMAL APPROACH TO A FULL-SERVICE BROADBAND COMMUNICATIONS NETWORK

Reem Safadi
General Instrument Corporation
GI Communications Division
2200 Byberry Road
Hatboro, PA 19040

Abstract

There has been a number of past R&D efforts to provide extended-service networks addressing data communications, voice communications, and interactive applications utilizing the cable distribution plant. This paper provides a brief survey of such efforts in terms of special network considerations with respect to a fully integrated system of the aforementioned services. A summary of considerations for an efficient and cost effective approach is provided along with the benefits that such an approach offers to the system operator.

Relevant Work

R&D efforts addressing implementation of non-traditional services on cable systems date as early as late 1970's. Some of these efforts did not come to fruition and others were implemented but lacked consumer demand.

Many such efforts served to be more than mere academic exercises. The larger subset focused on optimization of multiple access protocols (medium access control, MAC) for a cable based metropolitan area network, the other subset described the design of the overall system as in [3] and [7], for interactive video-text based applications. While optimization of medium access control is an important element to enabling non-traditional services, it is not the only area of consideration. Nonetheless a brief account is provided below for illustration. (A summary of the various media access schemes proposed for cable based distribution networks, and referenced herein, is provided in [6] and repeated in [2].)

A common goal to many of these efforts was to achieve performance optimization based on well identified and objective evaluation criteria. The two main figures of merit are network efficiency - amount of network capacity devoted to data (video, audio, and text) transmission, and access efficiency - the time between the queuing and the transmission of a packet. A careful examination of each approach, however, pointed to the following often common drawbacks:

1) Definition and modeling in many cases relied on invalid assumptions. In [4] the approach was optimized specifically for a file transfer application (support of interactive applications was secondary). It pointed out however the practical invalidity of approaches based on memoryless systems.

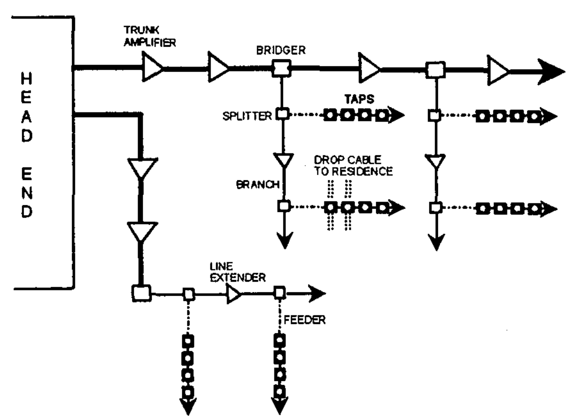

Figure 1. Placing substantial intelligence throughout the distribution plant as suggested by [5], [8], and [9] (at bridgers or splitters) is considered objectionable by Cable Operators.

In [5] the spread spectrum access scheme assumed persistent (next slot) retransmission of unsuccessful transmissions in order to be able to model system performance. However that assumption caused the system to move to a saturation point for throughput values considerably lower than 1. The effects of a random scheduling time were not provided. An entire system was described in [3] and [7] but, again, a common assumption was that traffic was generated by interactive applications only. None of the mentioned references thus far accounted for support of isochronous traffic (for video and voice based applications). An exception to that is found in [8]. The assumption there was that data packets were shorter in length than voice packets. While that facilitated non-blocking access for voice packets, the potential indefinite delay of data packets was not addressed.

2) Imposed certain requirements on the distribution plant that were not favorable from a cable operator's standpoint. In the system described in [8] two tuned notch filters and frequency converters were required at each regional subgroup boundary (neighborhood of less than a mile in diameter). This was an attempt to increase bandwidth reuse on pure coaxial tree & branch distribution networks. A control node was required in [5] consisting of a CPU with memory for buffering at each bridge to aggregate the upstream load from a given neighborhood and forward it to the headend. Similarly [9] proposed placing store and forward switches on trunks (and bridgers as an alternate approach) that performed routing, flow control, and error correction and detection.

3) Each approach grew increasingly complex when an attempt was made to extend it to support and address practical assumptions and requirements. In [5] buffering capacity at the control nodes was assumed to be infinite. Beyond a given threshold (saturation region, i.e., a given number of active users) the system became unstable in terms of successful transmissions. Moreover, spread spectrum processing gain, choice of FEC (imposed to guard against collided transmissions) and packet lengths posed conflicting optimization criteria with respect to one another.

A practical consideration such as the effect of round-trip propagation delay on feedback based multiple access protocols often required further refinement (and additional relative complexity) of the proposed schemes as in [2] and [11], where once the distance (from the headend to the most distant communicating node) was changed from the one mile assumption to a more typical value (35 miles), the protocol had to be augmented in order to maintain the same level of desired performance. In this example, impact of round-trip propagation delay is not the only practical factor leading to additional modification of the proposed scheme[11]. The impact of a non-ideal transmission channel (not a major issue in LANs since such networks are environmentally controlled, but is for a cable plant), is yet to be fully characterized. In addition, there has been a general tendency to depend on large upstream transmission rates (in excess of 5-10 Mbps) for performance improvement. Suffice it to state that a signal with such a rate has a considerable wider bandwidth that puts it at a disadvantage when taking into account upstream narrow-band interference.

Optimal-Design Requirements & Considerations

As discussed above, it is imperative to establish a full set of applicable requirements as a first step to defining the overall system. This in turn will have an impact on subsequent definition of system sub components. The system level design of a full service network has to address the following requirements:

1. Enabling a successful deployment of a given service by optimizing system performance, operation (including reliability) / administration, and cost.

2. Enabling incremental investments towards expanded or additional services.

3. Providing an end-to-end solution that is not optimized strictly towards a specific service or application. Such optimization is bound to incur penalties when considering the addition of other services. This is also applicable when considering monitoring and control (operations, administration and maintenance, OA&M) of the overall system.

While the three criteria are not mutually independent, the following addresses each individually.

1. Enabling a successful deployment: The success of a given service is determined by its appeal to the consumer and cost effectiveness of its offering (each governing generated profits). For many services, the appeal to the consumer remains an unknown, thereby potentially invalidating any assumptions technical or otherwise that may be taken into account in the overall design. Consider the previous example of multiple access protocol optimization. In many instances traffic modeling and usage pattern assumptions have been made or even mandated to justify the validity of a given approach. Relying on such assumptions may only serve as a theoretical justification but not a practical one. As illustrated, Medium Access Control is one of many areas that are impacted by the definition of the overall system.

The initial cost effectiveness of a design is mandatory. Relying on economies of scale to justify a more costly initial implementation is undoubtedly risky especially when consumer demand for a given service/application may not be fully characterized. Conducting trials is beneficial and in some instances is required when a given system design mandates a substantial initial investment. The latter is a characteristic of a revolutionary rather than an evolutionary approach (more below). An evolutionary approach allows the system operator to invest incrementally in expanding or adding a given service. As such the initial design has to have a self contained migratory path towards a more capable system. This is applicable to regional and local source equipment design, distribution plant design, and set-top design, etc.

2. Enabling incremental investment: As stated above, the initial design of an evolutionary system has to have a self contained migratory path towards a more capable system. For example, with respect to the distribution network, a gradual increase of return path capacity can be attained initially through extending fiber to a smaller number of homes/node, as in the Hybrid Fiber-Coax (HFC) architecture, illustrated in [12], followed by frequency block up conversion for sub-split systems when it becomes necessary to do so (i.e., when maximum capacity threshold has been reached). With respect to the set-top design, additional capabilities to support enhanced applications can be realized through add-on modules (e.g., increased processing power, memory and graphics capabilities). At a later date secondary modules may be upgraded or integrated into the base unit (when the use of such modules prevail). As for the design of the source equipment, it has to be modular and scalable in order not to dictate an unwanted initial investment while allowing incremental investment as economically deemed desirable. (It is assumed that the necessary switching equipment already meets this criterion). Figure 2 illustrates.

Enabling multiple services and ensuring proper allocation of resources to such services is a key consideration that the system operator has to address. Depending on how the system is designed, system operators may be faced with misplaced choices and options (or lack of). For example, systems that allow dynamic bandwidth allocation per service (and per user) have the flexibility and efficiency of BW reuse that is not present in BW dedicated systems. Additionally, systems that do not allow a transparent transport of information associated with a given service are likely to unnecessarily increase the complexity of the system. This is particularly true for existing services that evolved utilizing different connectivity infrastructures, e.g. telephony (POTS), PCS, data communications, etc. However this approach (transparent transport) if taken to an extreme may also pose unwanted disadvantages: if a given service is to be ported intact onto a Hybrid Fiber-Coax (HFC) infrastructure, such porting while suitable for the associated service may be entirely unsuitable for another, leaving the system operator with significant resource and operational issues. The previous example of multiple access protocols can be extended towards the currently defined MAC schemes such as Ethernet (CSMA/CD), Token Bus, or Token Ring (IEEE802.3, .4, .5 respectively), as an additional illustration. While there are applicable features in some (as is the case for a CDMA scheme), none were intended for a system utilizing a HFC infrastructure with given operational end-node characteristics. The traditional MAC schemes are also unsuitable for applications requiring guaranteed bandwidth. Even when considering DQDB, IEEE802.6, one finds the operational requirements widely varying from those governing the system at hand. All of this however does not imply that isochronous and asynchronous data can not be efficiently transported over the HFC distribution plant. What is important to realize is that there are boundaries that need to be established with respect to transparent information delivery. In the case of telephony, it would not be economically nor technically advantageous to redesign the Host Digital Terminal and switches to recognize new protocols nor reinvent the interfaces, where applicable (consider the effort invested by Bellcore to produce technical requirements specifications for telephony systems). Similarly for data communications, it would not be desirable to impact the design of IP routers, X.25 or ATM switches as an example. Establishing such boundaries becomes a relatively simple task when communication models are applied that separate the various functional layers. Which layers to preserve and which can be replaced in order to optimize overall system design becomes rather self-evident.

Architecturally, there is a common denominator to enabling additional services on the cable infrastructure. PCS, POTS and video telephony serve as an example. Each may have a suitable interface from an origination/destination point and a corresponding interface to the controlling (switching) point. As long as the interfaces to equipment residing at the demarcation points is kept intact, a single architecture can be supported.. Preservation of a given service platform outside the demarcation points is desirable since it provides the system operator the freedom to select individual platforms with the most competitive offerings. The same applies architecturally to interactive or Video on Demand applications. The later, however, is an example application where end-to-end system definition (of all applicable functional layers) has to be provided since it is a new application.

3. Providing an end-to-end solution: A fully integrated system implies a fully integrated Operations, Maintenance, Administration, and Provisioning subsystem (OAM&P) covering: billing, network, resource and service management, access control, remote monitoring and control, etc. for the different types of services: broadcast video, NVOD, VOD, POTS, video telephony, interactive applications, telecomputing, etc. This however is a long term objective especially when integrating in an overlay manner two or more existing systems of distinct service types (to achieve near term objectives).

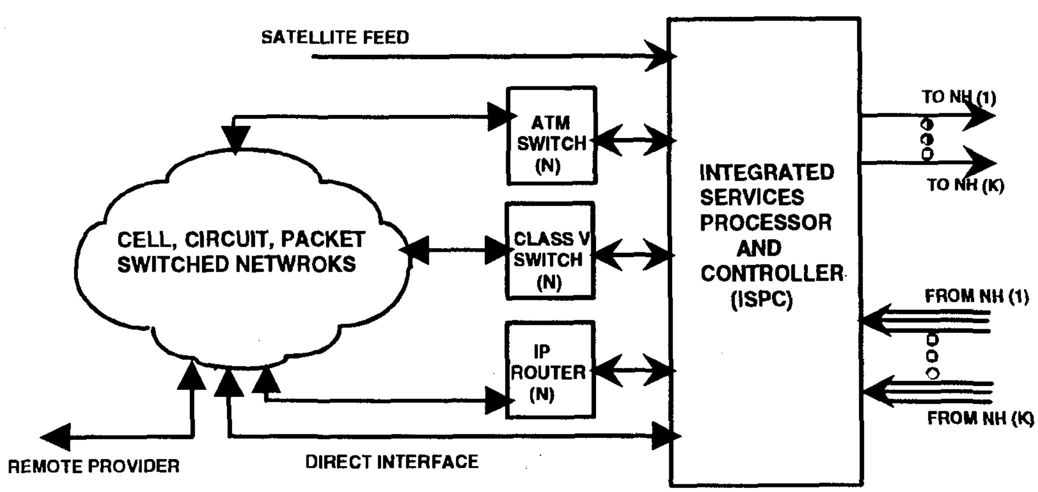

Figure 2a. Example of Scalable Network Stages External to the ISPC. (Some of the elements in the intermediate stage, to the left of ISPC, may be combined via ATM).

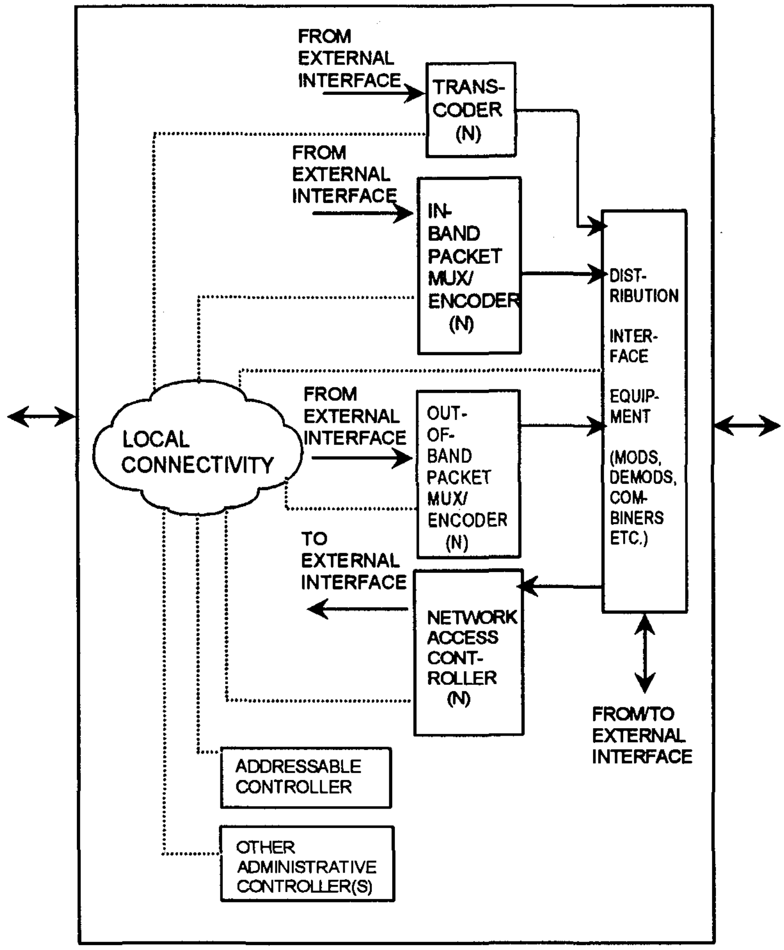

Figure 2b. Example of ISPC Functional Blocks.

Notes: 1. Some functional blocks may be remotely located.

2. Future (conditional) ubiquitous use of ATM may facilitate combining certain functional elements.

3. See [12] for example illustration of subscriber side. (There ISPC is ISHDT.)

A homogenous operational system offers advantages associated with a single set of skills required to administer the system as well as understanding the underlying OAM&P processes. However fully-developed enabling technologies for a homogeneous system (not just the OAM&P subsystem) are a few years away and thus it is best to rely on proven approaches rather than risk offering a system that can be considered experimental at best. What is learned from experimental systems however can be used in further advancing existing systems and as such should not be disregarded.

Acknowledgment

The writer would like to thank the countless individuals especially those who prefered anonymity for their strong interest and ongoing support of this effort and all other related activities.

References

[1] IEEE 802.6-1990," Local and Metropolitan Area Networks, Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network," IEEE Inc., July 1991.

[2] G. Campbell, R. Khasawneh, "DQRAP: A Proposed Standard For the 'Last Mile' of An IEEE802.6 MAN, " Dept. of Comp. Science, Illinois Institute of Technology, Submittal to IEEE802.6 Committee (IEEE P802.6-93/13).

[3] M.L. Ellis et al., "INDAX: An Operational Interactive Cabletext System," IEEE J. on Selected Areas in Communications, vol. SAC-1, no. 2, February 1983.

[4] D.C. Feldmier, "An Efficient Access Scheme for A CATV-Based High-Speed Packet-Switching Metropolitan Area Network", Lab. for Computer Science, Massachusetts Institute of Technology, IEEE INFOCOM '88, March 1988, New Orleans, LA.

[5] H.M.Hafez et al, "Spread Spectrum Access in Two Hop CATV Data Networks," IEEE J. on Selected Areas in Communications, vol. SAC-3, no.2, March 1985.

[6] A.I. Krashmer, J. N. Thomas, "Computer Networking on Cable TV Plants," IEEE Network, November 1992.

[7] R. McNamara et al., "CATV Communication System", General Instrument, U.S. patent 4533948, August 1985.

[8] N.F. Maxemchuck, A. Netravali, "Voice and Data on a CATV Network," IEEE J. on Slected Areas in Communications, vol. SAC-3, no. 2, March 1985.

[9] T.N. Saadawi, M. Schwartz, "Distributed Switching for Data Transmission Over Two-Way CATV," IEEE J. on Selected Areas in Communications, vol. SAC-3, no. 2, March 1985.

[10] A. S. Taylor, "Characterization of Cable TV Networks as the Transmission Media for Data," IEEE J. on Selected Areas in Communications, vol. SAC-3, March 1985.

[11] W. Xu, G. Campbell, "A Distributed Queuing Random Access Protocol for a Broadcast Channel," Dept. of Comp. Science, Illinois Institute of Technology, Submittal IEEE 802.6 Committee (IEEE P802.6-93/13).

[12] C. Eldering et al, "Engineering Requirements for Hybrid Fiber-Coax Telephony Systems," Proceedings of the 1994 NCTA Convention & Exposition, (New Orleans, May 22-25, 1994).