Tree-Net, A Multi-Level Fiber Optics MAN

Mario Gerla

Computer Science Department
UCLA, Los Angeles, California 90024-1600

Abstract

High bandwidth requirements make the use of fiber optics very attractive for Metropolitan Area Network (MAN) implementations. Several fiber optics MAN architectures have been proposed. A common drawback of such architectures is the presence of active processing components (e.g. ring interfaces, switches, gateways) at intermediate points along the path. Such elements tend to reduce reliability and throughput capacity.

This paper presents a novel fiber optics architecture, Tree-Net, which does not require intermediate processing components. Tree-Net is based on a tree topology, passive station taps, and implicit token protocol. The properties and performance of Tree-Net are evaluated, and possible extensions of the basic scheme are discussed.

1. Introduction

Fiber Optics Local Area Networks (FOLAN's) are becoming increasingly popular for applications involving the exchange of very high data rates [PERS85]. The applications range from the transfer of large files, to the handling of real time control data, to the integration of data, voice and video services. It is to this last application (i.e. integration of services) that the high bandwidth of the FOLAN is ideally suited: therefore, much of the current research on FOLAN's is directed to providing an environment supportive of both real time traffic (voice and video) as well as the more traditional computer traffic (interactive, file transfers etc.) [PERS85].

As the interest in FOLAN's grows, so does the size of the systems to be connected via FOLAN's. One important example is the distribution of integrated services to a metropolitan area via a fiber optics MAN (Metropolitan Area Network).

The design of a MAN poses several challenging problems, including:

- Support of a large number of stations (several hundreds to several thousands)
- Coverage of a large geographical area
- Real-time traffic (voice, video) support
- Very high bandwidth requirements (in the Gbps range)

Additional requirements in a MAN design (which are also common to more traditional LAN designs) are: high availability and fault tolerance, and; growth flexibility.

Several fiber optics networks suitable for MAN applications have been proposed in the literature. An early IEEE 802.6 proposal consisted of a multi-level ring architecture [SZE85]. In contrast, a multi-level bus architecture was proposed in [ALBA87]. Regular mesh topologies (e.g. Manhattan Grid) were described in [MAXE85] and [BORG87]. A two-level architecture with linear busses interconnected by a high level loop via photonic switching was presented in [WONG87]. Finally, Hubnet, a tree architecture featuring random access with capture was described in [LEE83].

A weak point in all the above architectures is the presence of active components at intermediate nodes along the path, which are required to inspect packets and, in some cases, make routing decisions about them (e.g. shift register in the ring interface; switch in the Manhattan Grid; active node in Hubnet; store-and-forward gateway, in the two-level bus architecture). Intermediate processing com-
ponents are not desirable because they tend to reduce network reliability and pose bandwidth limitations.

A further drawback of some of the above schemes (e.g. Hubnet, Manhattan Grid, etc.) is the difficulty in providing guaranteed delay and bandwidth (a necessary requirement for real-time traffic support).

In this paper we propose a novel FOLAN architecture, Tree-Net, which can be used in MAN applications and may overcome some of the above mentioned limitations. Tree-Net has a tree structured topology (to which stations are connected via passive taps), does not include intermediate processing components (although signal amplifiers, i.e. repeaters may be inserted to increase the number of stations supported), and it provides bounded access delay via an implicit token protocol. Because of the lack of intermediate processing components, each packet is broadcast to the entire network.

The basic scheme will permit to support up to a thousand stations, say. Beyond this limit, power budget and network load considerations indicate that it is more cost effective to interconnect several Tree-Nets with gateways in a multi-level configuration. In this regard, we will show that the token protocol can be properly extended so as to alleviate the buffering problems at the gateways. Also, we will discuss several possible enhancements of the basic Tree-Net architecture.

2. The Concept

The Tree-Net architecture here proposed can be viewed as a two level architecture, where the high level is a tree, and the low level is a linear bus (see Fig. 1). Stations are connected to the linear bus via passive taps; the tree itself is built using passive couplers.

In the simplest implementation of Tree-Net, no active components are present on the network path connecting any two stations, thus protecting the system from active component failures. This also implies that the signal transmitted by one station is broadcast to all other stations, with no filtering nor store-and-forward processing at any intermediate node (i.e. gateway). This permits to operate the network at very high aggregate data rate, without suffering of the bandwidth limitations imposed by the gateways.

Passive taps, full broadcasting and very high aggregate date rates are well advertised advantages of linear bus networks (e.g. Express-Net, U-Net, etc.)[GERL85]. While sharing these advantages, Tree-Net offers additional features which permit to overcome some of the traditional drawbacks of the linear bus architecture. First, Tree-Net extends the number of stations that can be supported by an order of magnitude (from tens to hundreds, say). Secondly, the tree topology is better suited to cover a large geographical area (campus, industrial park, metropolitan area, etc.) than a linear topology. Thirdly, in the tree structure, the problem of transmitter and receiver calibration is simpler than in the linear bus. Finally, the optical couplers used to build the tree are simple 3dB couplers, while the couplers in the linear bus may need to be "tuned" depending on their position on the bus (i.e. the coupling ratio is adjusted to optimize the power budget) [RODR 84].

All the above advantages do not come, of course, for free. Tree-Net has also some drawbacks, most notably, an increased latency delay. In the following sections we review the components and protocols of Tree-Net, evaluate its performance and propose several extensions to the basic scheme.
3. The Bus Component

The basic building block of Tree-Net is a folded bus to which stations are connected via passive, directional taps (see Fig. 2). This architecture has been extensively studied, and several protocols have been reported in the literature. For our application, we assume that the access protocol is a "token" protocol. That is, the "end station" (see Fig. 2) starts a transmission cycle by issuing a token. Upon detecting the token (through the "sensor" port), downstream stations with a backlog of packets to send will append one packet to the token [TSEN83].

If there is more than one backlogged station, a "train" of packets will form after the token. Each station must then be able to locate the end of the train in order to attach its packet to it. This function is accomplished using a "probing" technique: namely, upon sensing the end of a packet (or token) transmission, the backlogged station will start transmitting the preamble of its packet. If, while transmitting, the station hears a packet coming from upstream, it immediately aborts its transmission, "defers" to the incoming packet and tries again at the end of it. The result of this "collision" is damage of a few bits in the preamble. The preamble should be long enough so that proper synch acquisition is not compromised by the collision. More details on probing are found in [GERL85].

The token and the train of packets move from the transmit bus to the receive bus, where each station inspects the address of each packet. If the packet address matches its own address, the station copies the packet into its memory. On the receive bus, the end station looks for the end of the train. Upon sensing it, it issues a new token on the transmit bus, starting a new cycle.

4. The Tree

Two bus segments can be combined in "parallel" using the scheme shown in Fig. 3. Essentially, an "extension" bus is connected to the first bus via two couplers. The original token protocol can be easily extended to handle the two parallel bus configuration. Namely, "leaf" station A in branch A (see Fig. 3) starts the token cycle for branch A (A-cycle). During the A-cycle, all backlogged stations in A can transmit, and both stations in A and B receive. At the end of the A-train, leaf station B starts the B-cycle, picking up the transmissions from the B-branch. Thus, the operation consists of an alternation of A and B cycles.

By applying the "parallel" combination process recursively, we obtain a binary tree structure, where the leaves correspond to linear bus segments (Fig. 1). The token protocol easily extends to the tree structure. The branches (i.e. leaves) take turns in issuing the token, according to a predefined order. A packet issued by a station is broadcast to all other stations. Thus, the tree can be viewed as a repeater-type interconnection of busses.

In our model, stations are connected to the tree only at the leaves. The scheme could be generalized by connecting stations also to internal links. For network access, the "internal" station is associated with the token cycle of one of the leaves in its sub-tree: that is, it is allowed to transmit at the end of the train from that specific leaf. As a special case, the station may be placed at the root of the tree: this is actually the position reserved for the gateway station (see Fig. 1). In this case, to reduce internet delays, the gateway is allowed to attach its packet at the end of a train from any leaf.

In this paper, for the sake of simplicity we will assume that stations can be connected only at the leaves (except for the gateway station). Furthermore, we will assume that the binary tree is a full tree, that is, all the leaves are at the bottom level of the tree, and the bottom level is full.

A simple inspection of Fig. 1 reveals that from the functional standpoint, the tree could be replaced by...
a star. The star would actually provide power savings with respect to the tree and would in fact permit to support more stations. If the star is implemented with a biconical star coupler, however, it would preclude the use of single mode fibers [MARH84]. This may be a serious limitation in very high speed FOLAN's operating at gigabit/sec. speeds. In order to accommodate single mode operations, a modular star based on 2 x 2 couplers may be used [MARH 84]. The modular star, however, requires many more couplers than the tree. A good compromise may be to use a modular star at the root, and to connect several sub-trees to the star. This option will be considered later in the paper.

The actual Tree-Net layout may vary depending on the requirements of the specific application. A typical configuration for metropolitan area coverages is shown in Fig. 4. Here, we see that sets of internal tree nodes (i.e. couplers) may be grouped together in "distribution centers". At the root, a cluster may be created in the "central office" (this cluster may actually be replaced by a modular star, as later discussed). One easily detects the analogy between the Tree-Net layout and the local distribution plant in the telephone network.

5. Maximum Number of Stations

An important figure of merit of a FOLAN architecture with passive taps is the maximum number of stations that can be connected. In fact, light detectors use much more signal power than their electronic counterparts (there is no such thing as a high impedance light sensor). Thus, the number of passive taps that can be cascaded on a linear bus or on a tree path is fairly limited. As we shall see, the tree structure offers substantial advantages over the linear bus in this regard. Still, even in the tree, the total number of stations is fairly modest unless intermediate amplification is used. Here again, the tree permits much more cost-effective amplification strategies than the bus.

To determine the maximum number of stations supported we must require that the signal transmitted by the light source (laser or LED) is properly detected at the end of the longest network path by the light detector.

First, some definitions and assumptions are in order. Let M be the power margin (in dB), defined as the difference between the laser output power (in dBm) and the minimum power detectable at the light detector (in dBm). Typically, M is on the order of 40 dB. The power reduction at a tree coupler is assumed to be 4 dB (3dB for power splitting, and 1dB for connector excess loss - a fairly conservative estimate). For the bus segment, tap coupling ratios may be individually adjusted to maximize station connectivity. Thus, the ratios vary from tap to tap, and depend on the total number of stations [GERL85]. To simplify the analysis, we assume that each bus tap causes a power reduction of 3dB (2dB for power spilling, and 1dB for excess loss).

Let L be the number of levels in the tree (see Fig. 5); let K be the number of stations on each bus segment. If no intermediate amplification is provided, the following inequality must be satisfied:

\[ 4L + 3K \leq M/2 \]

or

\[ K \leq M/6 - 4/3L \] (1)

Fig. 4 Tree-Net Distribution System

4B.2.4.

0366
Noting that the total number of stations \( N \) is given by:

\[
N = 2^L K
\]  

we have (by substituting (1) into (2)):

\[
N \leq 2^L \left( \frac{M}{6} - \frac{4}{3} L \right)
\]  

For \( M = 40 \), (3) becomes:

\[
N = 2^L \left( 6.66 - 1.33 L \right)
\]  

From (4), the maximum number of stations is plotted in Fig. 6, as a function of number of tree levels. We note that the maximum, \( N = 16 \), is obtained with \( L = 2 \) and \( K = 4 \), or \( L = 3 \) and \( K = 2 \), or \( L = 4 \) and \( K = 1 \). For \( L = 0 \), i.e. single bus structure, the maximum is \( N = 6 \). Thus, the tree structure improves the situation (from 6 to 16), although the maximum number of supported stations is still too low for metropolitan area applications.

Station connectivity can be greatly improved by introducing an amplification stage at the root of the tree. In this case, the following inequality must be satisfied:

\[
4L + 3K \leq M
\]

Following the same steps as before we obtain (for \( M = 40 \) dB):

\[
N \leq 2^L \left( 13.33 - 1.33L \right)
\]

The maximum number of stations for this case is plotted in Fig. 7. The maximum, \( N = 512 \), is obtained for \( L = 9 \) and \( K = 1 \); or \( L = 8 \) and \( K = 2 \) or, for \( L = 7 \) and \( K = 4 \). For \( L = 0 \), we find the maximum achievable with a single bus structure and with intermediate amplification namely, \( N = 13 \). The comparison with the previous result shows that amplification is much more effective (in terms of increasing the maximum number of stations) in Tree-Net than in the single bus structure.

In examining the above results, we note that the optimum is achieved with the pure tree structure (i.e. each leaf of the binary tree supports exactly one station). This result, however, is somewhat misleading because it does not take into account the delay performance. In the next section we address the delay issue and show that it pays to cluster stations on bus segments in order to reduce delays.

6. Access Delay

First, we define walk time \( W \) as the time interval between two successive token visits at the same station, in zero load conditions. Recalling that in Tree-Net each leaf is individually "polled" by the token, and letting:

\[
W = \frac{1}{c_0} + \frac{1}{c_1} + \frac{1}{c_2} + \frac{1}{c_3}
\]
From the above values it follows that $T = (1/2^L) \mu \text{sec.}$ If the application supported by the MAN requires an access delay $\leq 2 \mu \text{sec}$, then we must have:

$$2^L \leq 20$$

or

$$L \leq 4$$

Using the value $L = 4$ in expression (5) of the previous section, we find that the maximum number of stations supported is $N = 128$ (i.e. 16 bus segments, with 8 stations each).

7. Gateway Interconnections

A Tree-Net can be connected to other Tree-Nets or to a wide area network (e.g. ISDN backbone) via a gateway. The gateway can be conveniently installed at the root of the tree (see Fig. 1). Its function is to inspect every packet coming up from the leaves and to extract and forward to other gateways the packets with a "foreign" address. Likewise, the gateway injects in Tree-Net the packets received from other networks.

It is important to note that the gateway, because of its privileged position in the tree, can append packets to the end of any "train" coming from any of the leaves. Thus, external packets can be transmitted on Tree-Net with very little latency delay, without waiting for an entire "polling" cycle to be completed. This feature clearly alleviates the buffer management requirements at the gateway.

The coverage of a typical metropolitan area may require several Tree-Nets. The Tree-Net gateways can be interconnected to each other and to WAN (Wide Area Network) gateways in various ways. The choice will depend on the geographical proximity of the gateways (in the limit they may all be located in the same central office), the internet traffic requirements, and the reliability and fault tolerance requirements.
A common cause of congestion at gateways is the fact that the destination network cannot accept packets as fast as the origin networks generate them. Fortunately, in Tree-Net "foreign" packets have higher priority over local packets, so that queues will not form at the gateway-to-Tree-Net interface; rather, packets will queue in the origin station, waiting for the token. Once transmitted, the packets are essentially granted a congestion free path all the way to destination.

8. Transmit Power Calibration

In a fiber optics broadcast network with passive taps the paths between different transmitter/receiver pairs may have different attenuation characteristics. Thus, if the transmit power level is the same for all transmitters, a receiver will receive packets from different sources at different power levels. This situation (often referred to as "dynamic range") makes proper reception difficult because the receiver cannot adjust to rapid fluctuations in input power. It is then necessary to "calibrate" the transmit power at each station, so that each detector receives the same power level from any sending station.

In Tree-Net, calibration can be accomplished with the following procedure. A station is chosen as a reference station (possibly, the leaf station on the longest path from the root). This station periodically issues a token (as part of the access protocol). Each station will compare the power level of the token received from the reference station, with the power of the echo of its own transmission. It will dynamically adjust its transmit level until its echo power equals the reference power. A mechanism to implement this dynamic calibration was described in [GERL85].

9. Fault Recovery Procedures

The ability to recover from faults is an important requirement for a network supporting many diverse applications. In Tree-Net, proper operation must be preserved even after failures of stations, gateway and amplifier. We review the various failure modes and present recovery schemes.

First, we consider the failure of the leaf station on a bus. Recall that the leaf station is responsible for issuing the token after detecting the end of train from the preceding bus (in the pre-established order). If the leaf station fails, no token is issued and the entire network fails. One way to avoid this problem is to modify the bus topology as shown in Fig. 8. In this modified topology, each station acts as if it were the leaf station; that is, upon detecting the end of train, it issues the token. However, a downstream station (e.g., station 8 in Figure 8) upon starting transmission of its token will hear the token from the last station and, therefore, it will immediately abort its transmission. This solution is inspired to the scheme used in Express-Net [TOBA83].

Another possible solution (which will work with the unmodified topology) consists of implementing staggered time-outs on the stations along the bus. Each station starts a time-out upon detecting the end-of-train; and, it issues a token if the time-out expires before a token has been heard on the bus. Since the time-outs are properly staggered, only one station will transmit the token. When the leaf station comes back up, it is automatically reinserted in the cycle.

Next, we consider the case in which all the stations on a bus segment (say, bus n in the polling sequence) have failed. Let T be the round trip delay from a leaf station to the root of Tree-Net and back. Bus n+1, i.e. the bus following bus n in the polling sequence upon noticing a silence period larger than T during which no token is heard, will assume that bus n has failed and issue a token, thus bypassing bus n.

Gateway and amplifier failures are handled by making these resources redundant. When the main unit fails, the backup is switched in. Monitoring and maintenance of these units is made easier by the fact that they are generally located in the Central Office (in the case of a Metropolitan Area Network) rather than on customer premises.
10. Extensions of the Basic Concept

In Section 5 we have shown that the maximum number of stations supported by Tree-Net is about 500, using off-the-shelf technology and assuming amplification of the optical signal at the root. Here, we explore extensions to the basic Tree-Net structure which permit us to increase the number of stations.

The first modification consists of replacing the bus segments at the bottom of Tree-Net with "tree segments" (see Fig. 9). The stations, instead of being aligned along the bus as in Fig. 9a are now placed at the leaves of a tree segment as in Fig. 9b. The stations are also connected by a "control wire", a low bandwidth, unidirectional coaxial cable providing access control signaling [NASS85].

Fig. 9 Tree Segment Scheme

Access control in the tree segment works as follows. A station with packet (or token) to transmit, turns on its signal on the control wire. The leftmost station (station A in Fig. 6b) is responsible for issuing the token, so it always turns on the control signal. Upon detecting the end of the train generated by the previous segment, station A transmits the token followed by a packet (if any), and it drops the signal on the control wire. The next station along the control wire is now allowed to transmit its packet. After transmitting its packet (if any), it will drop the signal on the control wire, enabling the following station to go, and so on until all the backlogged stations have transmitted their packets.

To protect the tree segment from leftmost station failure, a station will automatically take the role of leftmost station if it does not hear a control signal on the wire. Namely, it turns on its control signal and generates a token when required.

Note that the tree-segment can modularly replace the bus-segment in the network. In fact, the two types may actually co-exist in the same network.

The advantage of the tree-segment is that of reducing tap insertion loss with respect to the bus segment. Referring to the assumptions in Section 5, we find that the maximum number of stations grows from 16 to 32 (without amplification at the root); and, from 512 to 1024, with amplification.

Although this improvement may not seem large enough to justify considering the tree segment solution, there are other features that make this solution quite attractive. Referring to Section 5, one recalls that \( N = 512 \) was achieved with a bus segment with only two stations on it \( (K = 2) \). This in turns led to very high latency (as discussed in Section 6). To keep delay within an acceptable range, one is forced then to reduce \( N \) well below 512.

In the tree-segment solution, the choice of cluster size (i.e. the number of stations per segment) does not affect the maximum number of stations that can be connected to the network. Thus, the size can be properly chosen to suit the delay constraints.

Another very effective way of improving the total number of stations consists of interconnecting several Tree-Nets with a passive, modular star [MARRH84], as shown in Fig. 10. Recall that the number of stages in the modular star is \( \log P \), where \( P \) is the number of ports [MARRH84]. Each stage introduces a 4db attenuation. Using the assumptions in Section 5, a system with a 4 stage star and with 3 level Tree-Nets can support 128 stations without requiring any amplification. This is a major improvement with respect to the 16 stations in a pure Tree-Net!

4B.2.8.
If amplification is used, the modular star permits a dramatic increase in the total number of stations. Assuming that amplifiers are used both at the inputs and the outputs of the star, a configuration with a 10 stage star and 10-level Tree-Nets becomes feasible (from the power budget standpoint). This configuration supports up to 1,000,000 stations. Probably, such a large system is not practical for other reasons (the delay latency may be excessive; 5,000 couplers are required in the star; and, even the gigabit/sec bandwidth may not be sufficient to support such a large population). However, a scaled-down version of hybrid star/tree architecture can be properly engineered and provide a very cost effective solution to the metropolitan coverage problem.

One may note that calibration can be carried out in the star/tree network in the same way as in Tree-Net, using a reference station. As for the interconnection of the star/tree network with other networks, this can be accomplished via a gateway connected across the star (see Fig. 10).

11. Conclusions

The proposed Tree-Net architecture satisfies the basic requirements for MAN implementation. Throughput efficiency is obtained through the use of an implicit token protocol within Tree-Net, and of fast packet switching and mesh topology at the higher level. Delay guarantee derives from the use of virtual circuits for real time traffic; bandwidth preallocation to real time connections at call set up time; and, built in priority for internet traffic. Fault tolerance follows from the fact that all network components in Tree-Net are passive, except for repeaters and gateways, which are redundant and are installed in a Central Office where extensive monitoring and maintenance facilities are presumed. Expandability within the tree topology is easily obtained by replacing a leaf with a subtree. At a higher level, new Tree-Nets can be deployed and connected to the internet mesh via gateways.

Work is currently in progress on refining and extending several aspects of Tree-Net. A hybrid random access/token mode scheme is being designed to overcome token latency. The trade-offs of star/tree combinations are investigated. Flow control and bandwidth allocation protocols for the internet are defined, with special attention to fairness issues. Various internet connection alternatives are investigated including the integration of fast packet switching and time division slot switching. Finally, the suitability of the Tree-Net concept for large FOLAN systems other than the MAN (e.g. automated factory; interconnection of existing FOLANs etc.) is investigated.

Acknowledgements

This research was supported by the State of California and Pacific Bell (through a MICRO grant); by NSF (INT 85-16798 and ECS 80-20300); by the US ARO (DAAG 29-85-0101); and, by DARPA (MDA 903-87-0063).

References


[BORG87] Borgonovo, F., E. Cadarin, "Routing in the Bidirectional Manhattan Net-


[YEA87] Yea Li-Sun, "Interconnection Topologies for FOLANs", UCLA Internal Report, August 1987.