M.I.T. Laboratory for Computer Science

Request for Comments No. 268
December 19, 1984

Metropolitan Area Network Access Schemes

by David C. Feldmeier

1. Introduction

1.1. The Need for New Access Schemes

A computer network needs an access scheme to allow communication. Access scheme refers to the allocation of bandwidth of a network. Access on a Metropolitan Area Network (MAN) is particularly difficult because of its long propagation delay. The MAN falls between local area networks and satellite networks in size. Unfortunately, existing access systems used for LANs or satellite networks are not suitable for a MAN. A new access system is needed.

Some MANs use polling as an access scheme. A large network (100,000 customers) can take six seconds to poll, causing throughput to be low. [2] The QUBE system uses adaptive polling, in which all stations on the network are polled periodically, the time between polls for a particular station dependent on whether a transmission slot was requested in the previous poll period. All stations are eventually polled within twenty seconds. Another possibility is Carrier Sense Multiple Access with Collision Detection (CSMA/CD), as used on the Ethernet, but CSMA/CD efficiency degrades seriously with network diameter. CSMA/CD has a trade-off of transmission rate, packet length and network propagation delay that may be adequate for small systems, but inefficient with long propagation delays. I propose some new types of access schemes that promise to be more efficient than polling or CSMA/CD.

1.2. Cable Television

I am interested in designing a Metropolitan Area Network on a community cable television system. A cable TV network is a coaxial cable transmission system that is frequency-split into 6 MHz. bands. In the most common, single-cable design used for residential areas, most of these bands are *downstream* channels occupying the spectrum from 54 Mhz. to 300 MHz., 450 MHz., or higher depending on the particular system and converter technology employed. Downstream channels carry signals, typically entertainment video signals, from the cable TV headend to the subscriber. The *upstream* channels occupy the 5.75 - 29.75 MHz. range and carry signals from the subscriber to the cable headend. Frequencies from 29.75 MHz. to 54 MHz. form a *guard band* and no signals are transmitted at these frequencies.

The network consists of stations that transmit data to the headend on an upstream channel. The headend broadcasts the received data from the upstream channel to all stations via the downstream channel. A design goal for data communication over a cable TV network is to change the existing cable system as little as possible. Ideally, the only additions would be a network controller at the headend and modems at the subscriber end. A related goal is complete compatibility with the existing cable equipment, television sets and cable channel allocation. In addition, expensive and complex equipment for data communication should be placed at the headend in exchange for cheaper and simpler equipment at the subscriber nodes. A wideband upstream link is needed in expectation of advanced home computers (comparable perhaps to a LISP Machine) and more complex home applications. To ensure a high upstream bitrate, the upstream link to headend should be via the same cable as the downstream link; other types of upstream channels such as telephone lines are undesirable. The network should support multipoint communication instead of point-to-point communication, because any station may communicate with any other station without prior arrangement. A system must exist to allocate bandwidth on the network to allow communication and to support addressing of the multiple points on the net.

For evaluating various access schemes, a typical CATV MAN might have the following parameters:

1. The data transmission rate, R, is 2 megabits per second
2. The maximum packet length, L, is 1023 bytes or 8184 bits
3. P_{max}, the maximum propagation delay is 100 microseconds

4. P_{ave}, the propagation delay to the average station, is 100/sqrt(2) or 70 microseconds

5. The maximum number of stations, N, is 1023

A data rate of 2 MBPS is a reasonable speed for a 6 MHz video channel on a CATV system. The maximum packet length prevents any station from monopolizing the net. The propagation delay P_{max} is that for a system with 25 km spokes out from a central hub (50 km system diameter). Nodes distributed around the hub of a metropolitan network are likely to be distributed equally over the area that the network covers, rather than linearly with distance from the hub. [3] P_{ave} will be larger than $P_{max}/2$, closer to P_{max}/sqrt(2). The maximum number of stations allows a heavily loaded net to retain good performance.

From traffic analysis done on local area networks, packets seem to be either short (50 byte) packets from interactive sessions or acknowledgments; or long packets (570 bytes) file transfer packets. Most packets are short, but most data is moved in the large packets. [4] [5] Initially the MAN will extend the local area network from the office into the home, connect machines together at various businesses and universities, and form a backbone network for connecting LANs. The traffic on a MAN should be similar to the traffic seen on current local area networks. Therefore, packets on the MAN should be of variable length, with a maximum length. Fixed length packets would waste too much network bandwidth with the expected traffic having mostly packets of less than fifty bytes. The average length of packets on the network is L.

Stations can only monitor the downstream channel to determine the usage of the upstream channel because amplifiers isolate cable segments, making upstream detection impossible. Stations know their distance from the headend only after sending a packet. Measuring the round-trip time of a packet is a simple way to determine distance from the headend, but this distance is unknown when the first packet is sent. Allowing stations to automatically calculate their own distance from the headend assures that the system will work correctly without having to adjust the network electronics manually.

2. Access Scheme

The access scheme for a MAN is an important and difficult problem. Because the network serves the home or small business, the expected packet distribution is bursty. Bursty traffic is difficult to schedule on a network with a large propagation delay; static allocation of time or frequency slots would make network utilization unacceptably low. A more dynamic allocation of bandwidth is needed efficient network utilization.

2.1. CSMA/CD

The Carrier-Sense Multiple-Access with Collision Detection (CSMA/CD) systems, such as the Ethernet, trade-off minimum packet transmission time and maximum propagation delay. CSMA/CD systems follow two simple rules. If the channel is busy, all stations defer to the active station. When the channel clears, if two or more stations begin transmitting and collide, then all stations stop transmitting and try again later. The system depends on short propagation delays on the network relative to packet transmission times so that collisions are short and waste little of the network bandwidth. The packet transmission time is the packet length divided by the bitrate of the network. As transmission speed or propagation delay increases, or packet length decreases, network performance degrades. RF collision detection is difficult; because the system is broadband, detection of conflicting signals on the upstream channel is uncertain, so the only way to know that the network is in use is to watch the downstream channel. Use of the downstream channel only increases the delay of collision detection.

A 10 megabit Ethernet has a minimum packet size of 72 bytes and a maximum end-to-end length of 2.5 km; with repeaters in line, the maximum round-trip propagation delay is 51.2 microseconds. [6] Round-trip propagation delay for the CATV MAN is 400 microseconds - eight times more; but transmission rate is 2 megabits, one fifth the rate. Slightly lengthening the minimum packet size brings the numbers near the maximum.

A simple efficiency analysis uses the Ethernet efficiency formula [7] which is

$$\frac{1}{1+5.4\tau}$$

where 1 is the normalized packet transmission time and τ is the maximum propagation delay in scaled by the packet transmission time L/R.

$$\tau = \frac{2RP_{max}}{L}$$

This formula gives an efficiency of 79% using maximum length packets and 42% using 72 byte packets. Collisions cannot be detected until the signal has traveled to the headend and back again to the originating station. On the average, this is $2P_{ave}$ or 140 microseconds. The 2 megabit data rate means that on the average 280 bits be transmitted (really more because the average farthest station out in a collision will be farther out than the average station). 79% is probably the highest efficiency that such a net would have, and because of these problems the actual figure will be somewhat lower.

2.2. Store and Forward

If the headend were intelligent, all station on the network could transmit to the headend, which would verify correct packet format and retransmit correct packets to the appropriate station on the downstream link. An Aloha type system of totally random transmissions could be used, but the efficiency would be low. Spread-spectrum allows an alternative. If all stations transmit at the same power level, spread-spectrum trades channel capacity for bit error rate. Multiple station can transmit at once, the only problem is that bit error rate could become unacceptably low.

The headend could monitor bit error rate and transmit "suggestions" downstream to allow stations to compute the probability that their additional transmission would increase the bit error rate above an acceptable amount.

2.3. Polling

The network should be controlled by a system at the headend of the cable system to lower the cost of the network nodes and to increase the maintainability of the system. A centralized system can support access control if it is desirable to prevent certain users from transmitting packets. A cable TV system already has a point of centralization, the headend, which transmits upstream signals onto downstream channels, so a centralized access assignment scheme adds no disadvantages. In addition, polling from the centralized point that is the upstream-to-downstream converter is the fastest way to poll.

The time required for each station is $2P_{ave}$ for the poll request to get from the headend to the station and back, and 1/R for a single bit of data from the polled station. If the station has a packet, it should be immediately appended to the poll response for an additional L/R. If the number of stations is N and the number of stations with traffic is S, then the total polling cycle time is

$$2NP_{ave} + (N+SL)/R.$$

The system efficiency is

$$\frac{SL/R}{2NP_{ave}+(N+SL)/R}$$

Using the values for the typical MAN, efficiency for a single station transmitting full length packets is only 10%, but efficiency for all stations sending the maximum length packet is 99%. It would be desirable if any number of nodes transmitting any length packets could use nearly all of the network bandwidth, rather than a small fraction of it.

2.3.1. Multiple Polling

The network controller at the headend of the cable TV system polls subscriber stations and assigns transmission order in the following way: the controller broadcasts a poll request to all the stations on the network. Each station that has data to transmit asserts an unmodulated carrier on a unique frequency permanently assigned to that station. The headend controller receives the carriers of all stations wishing to send, calculates the frequency of each and makes a list of stations with traffic. The headend broadcasts the transmission order to all stations as a serial bit string (one bit per station) on the downstream channel. All stations transmit in order and when the last station is finished, the headend polls again. Polling is continuous until someone has a packet to send. Note that the larger the number of stations, the closer together the carriers must be and the longer the headend controller must observe a carrier to determine its frequency to sufficient accuracy.

If N is the maximum number of stations and R is the bitrate, then in a perfectly coordinated system, the polling time would be N/R because each station N requires one bit 1/R using either time division multiplexing (TDM) or frequency division multiplexing (FDM) on the polling channel. But if a station may not know its distance from the headend without first transmitting a packet, then the distance from the headend must be unknown during the first time that a station polls. To allow for this, each slot in a TDM system would have to be $1/R + P_{max}$ for a total of $N/R + NP_{max}$. In a FDM system, since all stations transmit simultaneously, the total polling time is $N/R + P_{max}$, for a total savings of $(N-1)P_{max}$. Now to calculate the total cycle time for each polling cycle.

1. Time for the headend's poll request to reach the farthest station: P_{max}
2. Time for the farthest station to respond to the poll request: P_{max}

3. Time to transmit one bit from each of N stations: N/R
4. Time to receive entire bit string of stations to send: N/R
5. Time for results of poll to reach the first station: P_{ave}
6. Time from first station to headend: P_{ave}
7. Transmission time of packet is L/R

Assuming that S stations are ready to send, the total is $2P_{max} + 2N/R + S(2P_{ave} + L/R)$. If S is large, the factor of $2SP_{ave}$ becomes large. The waste is because packets are not of a fixed length and it is impossible to determine when transmission can begin without destroying the packet in progress.

If packets contain a length field near their beginning, then the next station in line with knowledge of its distance from the headend can start to transmit packet before the downstream rebroadcast of the previous station stops, and yet avoid collision. The position of the end of the packet currently being transmitted can be calculated because both the network transmission speed and the packet length are known. As soon as the tail of the packet is closer to the headend than the next station to transmit, that next station may begin transmission. This technique is advantageous if several stations know the transmission order in advance and can proceed without waiting for the network controller at the headend to issue permission on an individual basis. Calculation of a station's distance from the headend is simple and can be done every time that station sends a packet by clocking the packet round-trip time. The station assumes the worst case if the distance to the headend is unknown by allowing the end of the previous packet to reach the headend before its transmission begins; by that time the end of the previous packet must be past all stations. The length field is simply prefixed to the packet by the network hardware.

The new packet length becomes $L' = L + \log_2 L_{max}$, L_{max} being the maximum packet length. The slight overhead more than makes up for the $2SP_{ave}$ that is saved. The only problem is that a station must know its distance from the headend in order to determine when transmission can begin. The first time that a packet is sent, the distance is unknown. This is not a problem - the station simply assumes that it is a close as possible to the headend. On the average this leaves a gap of P_{ave} between the two packets, but the gap occurs only the first time that a station transmits a packet. If a station has to recalculate its distance from the headend whenever it is turned on, this waste is likely to be spread throughout the day and negligible. If S stations are ready to send, the total is $2P_{max} + 2N/R + P_{ave} + SL'/R$.

The P_{ave} can be eliminated if the first station begins transmission immediately after the bit string transmission. The length of the bit string is known, so it is no problem to use the same procedure after the bit string. If all station begin to transmit their poll response immediately after the last packet (everyone knows the last packet because the entire bit string is received by everyone), then a factor of P_{max} is also saved. The new cycle time is $P_{max} + 2N/R + SL'/R$.

A slight improvement is to send station addresses of all stations responding rather than the entire bit string. There is a savings if the number of station addresses plus the prefixed length is less than the bit string size:

$$S\log_2 N + \log_2(N\log_2 N) < N$$

A better idea explained below that also reduces the size of the bit string.

2.3.2. Probabilistic Multiple Polling

Ideally, the polling period should be proportional to the number of stations responding to the poll. The frequency-determination time is proportional to the minimum bandwidth, Δ, between two adjacent carriers. Static frequency slots mean that in the worst case, the bandwidth Δ is B/N, where N is the number of stations. No matter how many stations wish to send, the carrier frequency must be determined to the same accuracy; therefore the poll always takes the same amount of time. The minimum possible distance between two carriers determine how long the polling cycle takes. If the number of stations active is much smaller than the total number of stations on the network , then time is being wasted by having a large number of slots. Time could be saved by making Δ larger to reduce the time needed to determine which stations are ready to transmit.

The system can be made faster by reducing the number of frequency slots because headend carrier frequency determination time is proportional to the number of frequency slots that share the channel. Suppose that S stations are queued to transmit. Divide the bandwidth into M slots, where $M \geq S$, and have each station randomly transmit its carrier in one of these M slots during polling. Since this is a probabilistic system, there is some chance that no one will ever be able to transmit successfully. M can always be adjusted so that the expected polling time is less than or equal to the deterministic polling time.

In general, only a fraction of the stations will respond to any poll. If the number of polling slots and therefore the bit string were M instead of N, the savings would be 2(N - M)/R. One way to make M smaller than N is to use M slots in a contention mode. Let the expected number of stations expected to respond to a poll be S. Then the headend should select some M that maximizes expected throughput or minimizes expected delay. The problem comes when multiple stations use the same frequency slot notify the headend. The headend assigns as many transmission slots as there are filled polling slots. If several stations fill the same request slot, the returning bit string will have no indication of this. Any stations that are using the same polling slot will collide and none will get a packet through. But if M is large enough, perhaps the probability of collision will be low enough to allow a faster cycle time than if N request slots were used.

The first thing to determine is the probability that S stations each selecting one of M slots with a probability of 1/M will leave exactly *e* slots empty and *f* slots with exactly one station. We denote this P(M, S, e, f). Let $P_m(a, b)$ be the probability that *a* balls placed in *b* cells randomly and with equal probability will leave *m* empty slots. This is a standard combinatorial problem [1] whose solution is

$$P_m(a,b) = \binom{b}{m} \sum_{\nu=0}^{b-m} (-1)^{\nu} \binom{b-m}{\nu} \left(1 - \frac{m+\nu}{b}\right)^{a}$$

The sum over all *f* of P(M, S, e, f) is just $P_e(S, M)$ from above. Now given M, S, and e; how many of the slots contain just a single object? This is almost exactly the same problem again. The number of slots that contain any objects is (M-e). Placing a single object in each of the (M-e) slots known to contain at least one object reduces the number of objects left to (S-M+e). The number of original slots with a single object is equal to the number of slots that now contain one object (M-e) and will receive no more of the (S-M+e) objects, but this is just $P_f([S-M+e], [M-e])$. The product

$$P_e(S, M)*P_f([S-M+e], [M-e]) = P(M, S, e, f):$$

$$\left[\left[\binom{M}{e}\sum_{i=0}^{M-e}(-1)^{i}\binom{M-e}{i}\left(1-\frac{e+i}{M}\right)^{S}\right]\cdot\left[\binom{M-e}{f}\sum_{j=0}^{M-e-f}(-1)^{j}\binom{M-e-f}{j}\left(1-\frac{f+j}{M-e}\right)\right]^{[S-M+e]}\right]$$

The efficiency of the system is

$$\frac{\frac{fL'}{R}}{P_{max}+\frac{2M}{R}+\frac{fL'}{R}+(M-e-f)\,C}$$

where C is the collision cost. The collision cost is the amount of time lost to the network due to a collision. An estimate of this is $2P_{max}$. It is assumed that only the transmitting stations can detect a collision. A station at the maximum distance will begin to hear his original message after P_{max}. If he stops transmitting immediately, then another time of P_{max} must pass before the channel is clear. Because the length field packet could be corrupted, this also implies that the minimum length for a packet should be $2RP_{max}$ bits (50 bytes), so that a collision does not draw in even more stations.

To maximize the expected throughput with respect to M, we must maximize the following equation over integer M

$$\sum_{e=M-S}^{M-1}\sum_{f=0}^{M-e}\left\{\frac{P[M,S,e,f]\ \frac{fL'}{R}}{P_{max}+\frac{2M}{R}+\frac{fL'}{R}+(M-e-f)\,C}\right\}$$

Using the assumed network parameters, a computer generated the throughput for all M given S. For S = 2, the optimal M is 92, the expected efficiency over 96.6%. For the polling system with N deterministic slots, the efficiency is 87.9%. The probabilistic system does better for M between 12 and 1018. For S = 10, the probabilistic system has an efficiency of 99.18% at M = 115 versus 97.3% for the deterministic system. Here M can be between 18 and 999 and outperform the deterministic system. For S = 20 (the highest number that the computer could process in a reasonable amount of time), the efficiency is maximized for M = 218, the efficiency being 99.3%.

For the deterministic system, the efficiency is 98.5%. This is exceeded for M = 51. The upper bound is unknown because the computation exceeded the precision of the program.

To optimize M, S must be known. S is not known, but an estimate of S can be made. The estimation of S and the calculation of M can be distributed, but this would increase the cost and the complexity of the network nodes. The headend could estimate S, determine M and transmit this value to all stations after the bit string of a polling cycle for the next polling cycle. A simple estimate of S is the number of stations that responded last time. The headend could also have a traffic monitor that a statistical history over time and even examines what type of packets are going by to give a better estimate of S.

2.3.3. Implementation

All stations need a distinct channel to reply to the headend simultaneously. Division of the network into frequency slots assures that the poll responses are orthogonal and transmitted simultaneously. The headend needs a bank of filters to determine which carriers are being received and thus the transmission order. A method of implementing a bank of identically spaced, identical characteristic filters is digitally using a Fast Fourier Transform (FFT) algorithm.

Frequency division of the upstream bandwidth is one possible way, but because of the interference from ingress noise, some of the frequency bands may be unusable and these frequencies may shift with time. If spread-spectrum modulation is used on the upstream channel, then CDMA (Code Division Multiple Access) could be used. Code division multiplexing multiplies the data output of a station with a pseudo-noise sequence that is different for each station. The receiver multiplies the the received signal with the appropriate pseudo-noise sequence to extract the original data. If all stations have a fixed power level, then CDMA has a trade-off of transmission rate versus bit error rate. The headend polls using CDMA at some slower speed based on estimates of number of replies to keep bit error rate low and then allows stations to transmit one at a time at full speed.

2.4. Intelligent Bridger Amplifier Polling

Replacing the upstream amplifiers has some advantages. In addition to forwarding one of N signals, upstream amplifiers can digitally regenerate the upstream signal for lower noise and fewer bit errors. Current linear amplifiers amplify noise as well as the signal and all the noise over the entire cable system is funneled into the headend. Disadvantages include the cost and complexity of this type of upstream amplifier and installation and service of systems in difficult to access areas (on a telephone pole).

2.4.1. New Upstream Amplifiers

The previous ideas assumed that the upstream amplifiers are simple linear amplifiers and that it is impractical to change them. Dropping this assumption leads to an interesting idea. Assume that the largest number of separate stations in the network that can be accommodated by any single section of cable served by any one amplifier is N. During polling, each station may respond by transmitting its address in an assigned frequency slot at 1/N of the network speed. The upstream amplifier decodes any responses and then chooses to forward only one of the N possible station replies. The signal is retransmitted forward on the amplifier's slot (not necessarily the same frequency as the received slot) to the next upstream amplifier. This process continues, each upstream amp selecting one of N slots and forwarding it until the address of a single station reaches the headend. The network controller broadcasts the received address and the victorious station continues its transmission at full speed. The bridger-amplifiers used here are intelligent and autonomous, each decides which of the N signals to forward.

2.4.2. Time Budget

Assume an A bit station address, a maximum of S stations per cable segment, P_{ave} second propagation delay to the average station, and a bitrate of R bits per second. Transmission speed per frequency slot is R/S and the time to transmit A address bits at this speed is AS/R seconds. Assuming D bits delay in each of a maximum of U upstream amplifiers there is DSU/R seconds of delay. Once the headend receives the address there is a P_{ave} seconds of propagation delay to notify the victorious station and an upstream delay of P_{ave} seconds, for an average delay of $(UD+A)S/R + 2P_{ave}$. On the Newton cable system, the number of upstream amplifiers is limited to $U = 25$. Let $P_{ave} = 70$ microseconds, bits of delay $D = 1$, the station address have $A = 12$ bits, the maximum number of stations per cable segment $S = 20$. The total delay then becomes 273 microseconds. The efficiency is 42% with 50 byte packets and 93% with maximum length packets.

2.4.3. Distributed Access

A better scheme is available if the upstream amplifiers can directly communicate with their downstream nodes. Intelligent upstream amplifiers can have less feedback delay if they could also transmit in the downstream direction. Transmission in the downstream direction would be in the upstream frequency bands to avoid interference with other CATV downstream transmissions.

If the upstream channel is clear, a station may begin transmitting on the cable. The upstream node both stores the incoming packet in a buffer and also immediately retransmits the regenerated packet to the next upstream node if the node's outgoing channel is not busy. Since RF collision detection is difficult, the upstream amplifier can assert a "busy signal" to show that the channel is in use. If there is a collision, then the stored packet is retransmitted later. If the transmission is successful, then the stored packet is discarded. By buffering and transmitting simultaneously, packets are delivered with the smallest propagation delay and the least wasted network bandwidth. This is similar to a series of Ethernets connected end-to-end with repeaters, but better because packet retransmission may begin from where the collision occurred instead of from the packet source.

3. Conclusion

Metropolitan Area Networks have a propagation delay that is three orders of magnitude less than satellite networks, allowing more interaction. Dynamically assigned time or frequency slots used in some satellite systems is unattractive for MANs because the loads on the computer network change too quickly for these slotted allocation schemes. The slots on satellite networks are long which leads to inefficient use of bandwidth under highly-variable loads. The access schemes proposed in this paper may be better for MANs or any type of network that has bursty traffic and long propagation times.

Local area networks have one or two orders of magnitude less propagation delay than MANs, so LAN access schemes may be too feedback dependent to use on a MAN. CSMA/CD throughput degrades seriously when round trip propagation time becomes a significant fraction of the packet transmission time. A centralized form of polling is always faster than decentralized polling (token) systems because each token must pass through the headend. Of the centralized polling methods, multiple polling is faster than either conventional or adaptive polling because little time is wasted polling inactive stations.

Replacing the upstream amplifiers on the CATV MAN makes upstream transmission easier and allows simpler access schemes at the expense of more complex components in less accessible locations. Replacing the upstream amplifiers is undesirable; but if replacement is necessary to reduce the upstream bit error rate to an acceptable level, then there is no additional disadvantage to increasing the complexity to handle a more distributed access scheme.

References

[1] Choudhury, Gagan L. and Rappoport, Stephen S.
Diversity Aloha - A Random Access Scheme for Satellite Communications.
IEEE Transactions on Communications COM-31(3):450-57, (March, 1983).

[2] Estrin, D. L.
Data Communications Via Cable Television Networks: Technical and Policy Considerations.
Technical Report MIT/LCS/TR-273, MIT Laboratory for Computer Science, May, 1982.

[3] Deborah Estrin.
Some Technical Considerations in Using a 400-MHZ catv System for Data.
Local Network Note No. 29.

[4] Feldmeier, David C.
Empirical Analysis of a Token Ring Network.
Undergraduate Thesis, Electrical Engineering and Computer Science Dept., Massachusetts Institute of Technology, May, 1984.

[5] Shoch, John F. and Hupp, Jon A.
Measured Performance of an Ethernet Local Network.
Communications of the ACM 23(12):711-721, (December, 1980).

[6] Shoch, John F. et. al.
Evolution of the Ethernet Local Computer Network.
Computer :10-27, (August, 1982).

[7] Tanenbaum, Andrew S.
Computer Networks.
Prentice-Hall, Inc., 1981.