# CONTENTION VS. POLLING, A THEORETICAL LOOK AT POPULAR NETWORK PROTOCOLS IN CATV SYSTEMS

Michael Quelly & Eric Frech

E-COM Corporation
A Subsidiary of AM Cable TV Industries, Inc.
Stirling, NJ

## ABSTRACT

This paper addresses a current area of some controversy. Systems employing contention and polling protocols are used over a range of services. The question always arises, "Which is best?" A limited amount of mathematical analysis is employed to demonstrate the strong and weak points of each and to reach some broad conclusions on the optimum applicability of these protocols.

## INTRODUCTION

This paper addresses the relative applicability of each of the two most popular single-channel network communications protocols: polling and contention--in a modern multi-service CATV system.

First, a brief description of each protocol is presented. A discussion of the most common applications of each protocol follows, along with a summary of their relative similarities and differences. Third, a general discussion of the basic network design criteria imposed by typical data services carried on a CATV system is presented, along with a comparative study of the inherent capacity of both polling and contention networks to support such requirements.

## NETWORK PROTOCOLS

Polling and contention are the primary network protocols used in broadband communications networks. Both protocols are well-suited to operation in the conventional tree topology of a CATV system. The decision to advocate one technique over another should be based upon a thorough .comparative evaluation of their strengths and weaknesses relative to the application(s) at hand.

## CONTENTION (RANDOM ACCESS)

This discussion is restricted to the Carrier Sense Multiple Access with Collision Detection protocol, commonly referred to as CSMA/CD. In the CSMA/CD network protocol, a terminal attempts to transmit whenever a message becomes ready. Before transmitting, the terminal checks for the presence of a carrier to determine whether or not the channel is busy. If the channel is unoccupied, the terminal transmits its message immediately. If the channel is occupied, the terminal waits before attempting to transmit the message. There are various methods used to determine how long to wait before attempting to retransmit the message. The method considered here is called 1-persistant CSMA/CD, since the terminal will transmit with a probability of 1 when the channel becomes available.[14] Once the terminal has successfully initiated a transmission, it monitors that transmission on its receive channel. If the data is received correctly as transmitted, the terminal assumes that no collisions have taken place. If incorrect data is received, indicating a collision, the terminal immediately ceases transmission and waits for a random period of time before attempting to retransmit. The throughput of a CSMA/CD system depends greatly on the method of calculation used by the terminal to determine the retransmission delay time. Importantly, the CSMA/CD algorithm is non-deterministic, i.e., a terminal's response time is a random variable, rather than a finite, calculable value.[12]

## POLLING

Although several polling methodologies exist, this discussion is restricted to bus or roll-call polling. In the roll-call polling protocol, each terminal is interrogated in turn by the system controller.[12] As each polled terminal is provided with a response opportunity, it acknowledges by transmitting status information and data to the system controller. Even if the polled terminal does not have a data packet ready or available for transmission, the system controller acquires status information about the terminal. Upon the completion of the transaction, polling of the next terminal is initiated. The order used to poll all of the terminals is determined by the system controller, which can incorporate prioritization or any other method of sequencing required. Importantly, the polling algorithm is deterministic, i.e., in a given situation, a terminal's response time is a finite and calculable value.[12]

## LOCAL AREA NETWORKS

The Local Area Network (LAN) has grown out of the requirements of the business and educational communities for a communications network that can handle a multiplicity of terminal-to-terminal and computer-to-computer connections. Such a network is generally confined to a localized geographical area, with a comparatively small number of users (10 to 1000), and normally transfers large blocks of data at infrequent times. Generally, LANs do not handle real-time applications, so that an occasional increase in network delay is of no real consequence. Carrier Sense Multiple Access with Collision Detection (CSMA/CD) has developed as the logical choice to satisfy these requirements.

In summary, a LAN is characterized by small geographic size serving a small number of users handling "bursty" transmissions of data.[14]

## MULTI-SERVICE COMMUNICATIONS NETWORKS

A Multi-Service Communications Network can support a wide range of services such as television (with pay service control, PPV, and IPPV), home security, medical alert, utility meter reading, home banking and shopping, electronic mail, information retrieval, personal computer interfacing, games, videotex, and system status monitoring. The user base of such a system can range from 1000 to 250,000 subscribers. Some of these services are real-time applications (home security and medical alert, for example), that can tolerate a controlled amount of network delay but that must be assured of nearly immediate access to the system. The multi-service network does not exclude those functions normally provided by a LAN.

In summary, a Multi-Service Communications Network is characterized by large geographic size serving a large number of users producing a continuous transmission of data.

## NETWORK DESIGN CRITERIA

Three important network design criteria are imposed by the services provided by a typical Multi-Service Communications Network.

First, the network must be optimized to handle the varying data throughput requirements of the different services carried. For example, a home security terminal requires one bit per second to monitor an alarm and provide a one-second response time. In contrast, a videotex terminal typically requires 2000 to 3000 bytes per active minute. Utility meter-reading services require transmission of approximately 6 bytes per meter per hour to provide hourly readings. The ability to efficiently handle the wide variation in data throughput associated with a diverse variety of services is a primary consideration in the design of an effective Multi-Service Communications Network.

A second design consideration is the wide variation in scheduling required by such an array of services. Some services require continuous or frequent monitoring of the status of remote terminals, while others do not. Citing the above examples, utility meter readings may be required at regular hourly intervals, while security alarms occur at random and must be serviced with a high priority relative to other services. A successful Multi-Service Communications Network must provide convenient and compatible scheduling and prioritization facilities.

Third, as new services are developed for inclusion into a Multi-Service Communications Network, the physical integrity of that network must be assured. Utilities considering the adoption of meter reading services must be assured reliable and continuous operation. In particular, security applications demand that the status of a remote terminal be known at all times. For these and other reasons, network global status monitoring has become an important design consideration.

## COMPARATIVE EVALUATION

The throughput of a polled network is limited by the communications overhead associated with the successive poll and poll-acknowledge messages flowing between system controller and terminals. CSMA/CD's primary strength lies in its lack of overhead. In the absence of a continuous flow of poll and poll-acknowledge messages, extremely fast response times are theoretically possible. In an application that features a low probability of message collisions (as in a system with few terminals or a low message arrival rate per terminal), these fast response times do occur.

The following graphs are the result of simulation programs written in BASIC for a TRS-80 Model III. Many runs of the simulation programs were made with different system parameters to study the sensitivity of the network to these parameters. Details of the models used in the simulation, along with supporting mathematical analysis, may be found in the Appendix.

The simulations consumed many hours of computer time. As a result the data for high loadings is not yet complete. At high loadings a CSMA/CD system experiences a large number of collisions and the individual packet delay varies widely. It appears that some delays will be very large, thereby further reducing the allowable system loading where short delays are mandatory. Further data is being taken to quantify this effect.

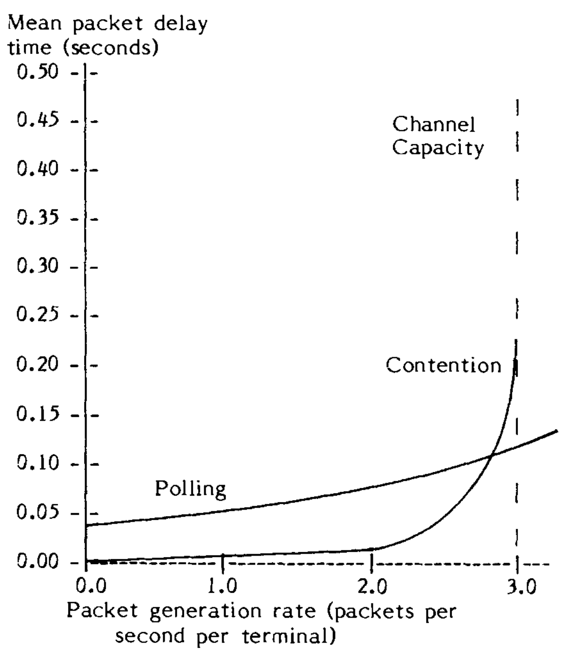
Figure 1 displays the mean packet delay as a function of message rate for a small network (100 terminals) with a long mean message length. The random access network produces lower mean response time than the polled network until it nears its theoretical maximum throughput. Then the polling system provides a lower packet delay. Figure 2 shows the same simulations for a network with 1000 terminals and shorter mean message lengths. As the curves show, there is no doubt that under conditions of low and medium loading, the contention system achieves shorter mean packet delays. However, other performance parameters must be considered along with their relationship to each proposed service.

The performance of the networks in the overload condition should be considered. In the contention system, when the system is overloaded, very few messages get through. It is difficult for a network to recover once this state is reached. When the polled system is overloaded it simply slows down. Because of the centralized control implicit in a polled system, it is possible to alter the polling sequence so that high priority terminals are still serviced at an adequate rate.

The polled network operates in full-duplex. Different information may be transferred in each direction on the network simultaneously. This two-way transfer of data makes possible 100% utilization of the data channel, i.e., data can be transmitted continuously in both forward and reverse directions. The single-channel CSMA/CD terminal operates in half-duplex. Even though the terminal must monitor the receive channel for collision detection (while transmitting) in what is essentially a full duplex operation, information is transferred in only one direction at a time.

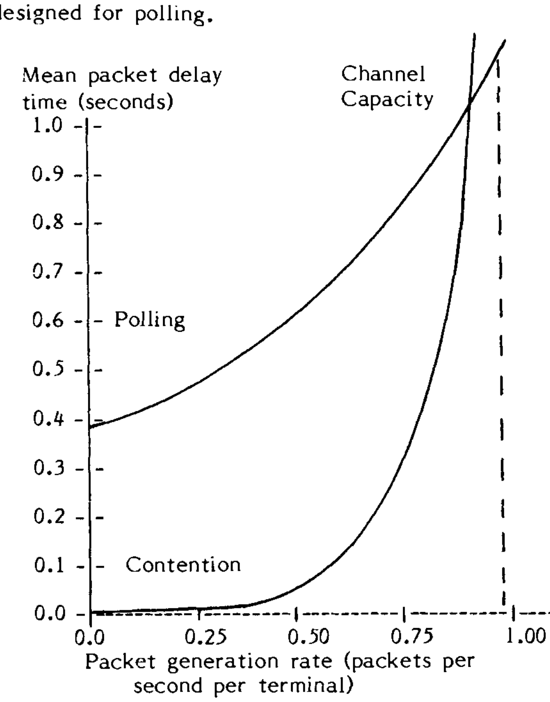
One of polling's strengths is the ease with which complex scheduling and prioritization functions can be implemented. Since the order and frequency of polling are under direct system control, terminals may be polled in accordance with the scheduling and priority requirements of the services which they provide. In addition, adaptive polling strategies may be employed to control network loading. In a CSMA/CD system, the transmission rate of each terminal is individually controlled by that terminal. Scheduling control must be implemented by some external means since it is not inherent in the operation of the contention protocol. A special terminal may serve as a network traffic controller to inform terminals of their scheduled transmission rate.

In a security system the proper operation of each terminal is vital to the security of the installation. Failures, which may be caused by cut cables or tampering, should produce immediate alarms. In the polling protocol, the incorporation of status information into the poll acknowledgement provides automatic global status monitoring. This is quite difficult to implement in a random access network, since there is no automatic notification produced for a terminal failure. A polling-type mechanism may be employed, using one of the terminals as a system monitor to check the operation of each of the other terminals in the network. This type of polling is, however, less efficient than that used in a network designed for polling.




SIMULATED SYSTEM
Mean propagation delay = 182 microseconds
Mean packet length = 100 bytes
Number of terminals = 100

Figure 1




SIMULATED SYSTEM
Mean propagation delay = 182 microseconds
Mean packet length = 10 bytes
Number of terminals = 1000

Figure 2

The importance of proper terminal installation and maintenance cannot be overlooked. The system operator must be able to verify signal levels and error rates of individual terminals in order to verify both correct installation and continuing operation. If a bad terminal is installed, and the error-correction circuitry in the network protocol compensates for the terminal's bad performance, the operator may never know of the failure. The ability of a polling protocol's central controller to determine such an event is difficult to implement in a contention network, without the addition of a polling mechanism.

Due to the centralized nature of polling, the system controller tends to be complex both in equipment and software. The special reliability requirements of a system with highly centralized control contribute to this complexity. In a contention system, the terminals are more complex so that the system cost on a per terminal basis may be considerably greater. While it would appear that CSMA/CD's distributed nature gives it reliability advantages over polling, the need for specialized terminals to control network operation tends to reduce this reliability advantage.

## CONCLUSIONS

In conclusion, it should be clear that the selection of a network protocol that can accomodate the wide variety of data services available to a CATV system must be based on several important requirements:

1. capacity to handle the throughput load requirements of all services carried;
2. capacity to handle timing and scheduling requirements of these sevices;
3. need to insure the physical integrity of the distribution system supporting these services;
4. need for automatic notification of terminal failure;
5. need to have traffic flow monitored and controlled to provide billing information;
6. need to present the lowest individual terminal access time possible and practical for each service.

This paper has attempted to provide the system designer with a basic understanding of the abilities of both polling and contention protocols to handle these design requirements. While the contention protocol has largely evolved from a specific operational system need (such as presented by a Local Area Network), polling has been used to address a number of problems not unique to any one application. It is important not to let certain individual aspects of any one protocol dictate its use in a system that it cannot fully support.

The CSMA/CD architecture is clearly well suited for the multiple host computer, multiple user terminal environment typical in universities and various commercial complexes. Here the probability of one node communicating to any one of the other nodes is usually comparable.

The polled system offers many advantages in situations built around common sources of information, common points of control, services requiring frequent status monitoring and deterministic service requirements as often encountered in CATV residential services. Accepting the intrinsic delays, a polled system can transfer more data in a given channel and does not suffer paralysis in overloaded conditions.

In closing, the following thought: the system should not be selected to fit a protocol, but rather the protocol should be selected to fit a system.

## APPENDIX

### MODEL SYSTEM

In order to compare polling and random access communications networks, it is necessary to select a model. The model chosen is used in modeling both the polling and the random access protocols to provide an accurate comparison.

Messages are generated at the terminals using a poisson distribution (exponentially distributed inter-arrival times). Geometrically distributed message lengths are assumed, with maximum message length constrained to fall below 251 bytes, the maximum data field length in the communications protocol we are considering.

We make the assumption that each terminal is independent of the others. Since we are comparing two systems that perform the same services, the processing time for the packets is neglected. This is reasonable, since in practice the systems run in a multiprocessor environment where the communications processor is separated from the processor handling the actual service.

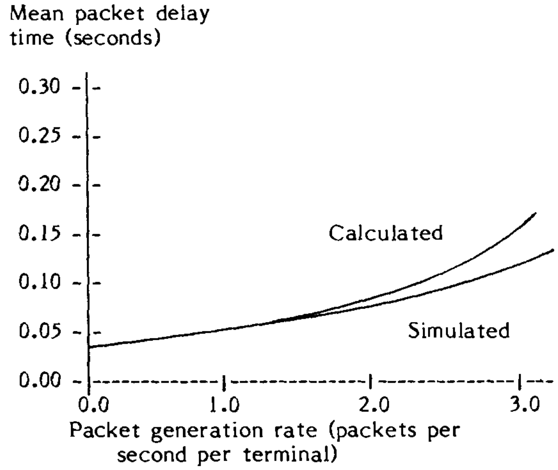
The propagation delay through the network is considered to be a constant. The average delay between two points in the network is considered to be one-half the maximum network delay. This is a reasonable assumption for a network where the terminals are uniformly distributed throughout the network. For a fifty-mile CATV system using .7 as the velocity constant in the cable, we arrive at 384 microseconds for the delay to the farthest point in the system. One half of this number, or 7 byte periods, is used for the one-way propagation delay through the network.

In our model, RF transmitters are assumed to turn off in zero time and collisions are detected instantaneously. This allowance reduces the period of vulnerability of the contention message, which tends to reduce the number of collisions.

The data rate used for our simulation is 307.2 kilobits per second, yielding a byte time of 26 microseconds. The communications protocol is similar to X.25. Data is synchronous with 8-bit bytes. The message formats are shown in Figures 3 and 4.

## MATHEMATICAL ANALYSIS

While the mathematical analysis of a polling system is inherently difficult due to the interdependence of the status of messages waiting at message sources, a similar analysis of contention is much more difficult due to its non-deterministic nature.[7] Konheim and Meister, however, have carried out the analysis of a polling model similar to ours, i.e., poisson arrival processes at each message source, geometric message lengths, symmetric equilibrium statistics at all sources, etc.[2,4,12] Calculations of mean inbound packet delay, employing formulas developed in this analysis, tend to support our simulation results. Figure 5 compares simulated and calculated delays for a small network (100 terminals) with long mean message length (see also Figure 1).




POLLING
Mean propagation delay = 182 microseconds
Mean packet length = 100 bytes
Number of terminals = 100

Figure 5

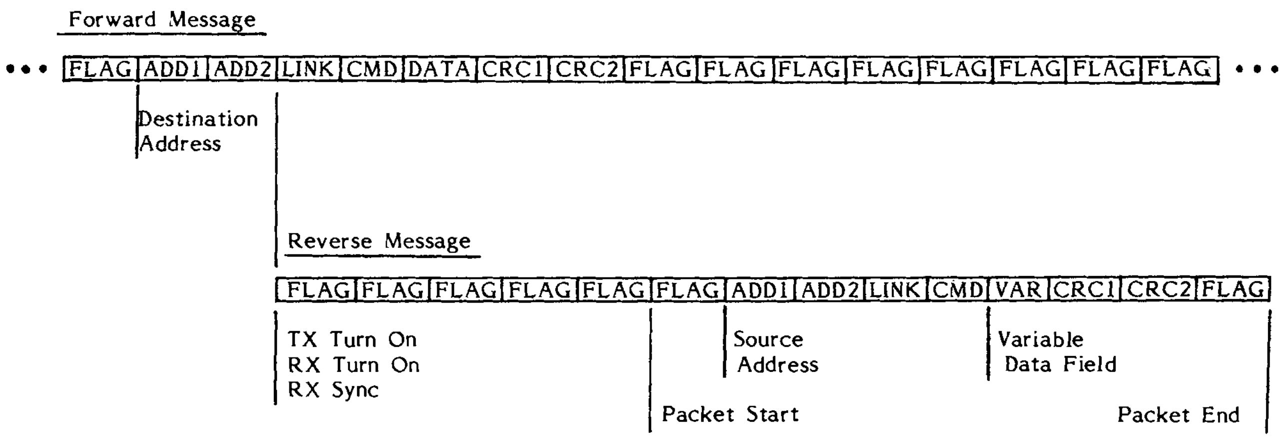



Figure 3 Polling Packet Format

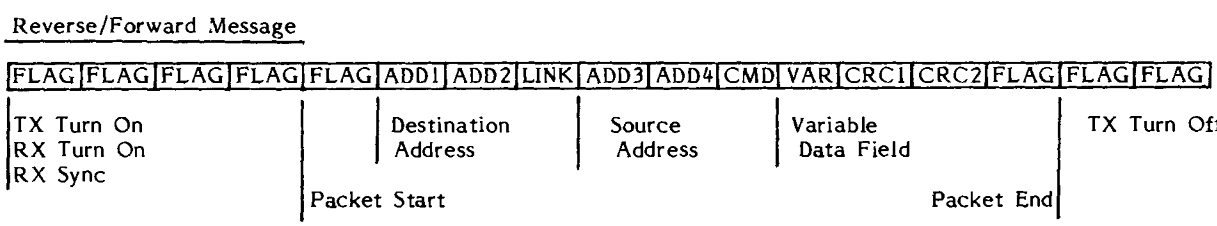



Figure 4 Contention Packet Format

Expected value of the packet delay:

$$E(D) = \frac{t}{2}(1-r) + \frac{NVk}{2}\,\frac{1}{(1-Nr)} + \frac{(1-r)k}{2} \qquad (1)$$

E(D) = expected value of the packet delay (seconds)

Variance of the number of data units arriving in k seconds:

$$V = \frac{lk}{p^2}(2-p) \qquad (2)$$

V = variance ($bytes^2$)

Mean Scan Time (the mean time required to poll all the units):

$$t = \frac{L}{1-Nr} \qquad (3)$$

t = mean scan time (byte periods)
N = number of terminals
r = traffic density (unitless)
L = walk-time (byte periods)

Traffic Density (the average rate of data flow in the network):

$$r = \frac{lk}{p} \qquad (4)$$

r = traffic density (unitless)
l = packet rate (packets/sec)
l/p = mean packet size (bytes/packet)
k = time interval (sec/byte)

Walk-time (the irreducible portion of the time required to poll all the terminals):

$$L = N(C_F + C_R + 2d + D_{min}) \qquad (5)$$

L = walk-time
$C_F$ = constant portion of the forward message
$C_R$ = constant portion of the reverse message
$D_{min}$ = minimum length of the data field

## BIBLIOGRAPHY


1. Bond, Ralph C. "Collision Avoidance Improves Broadband Bus Networks." Systems & Software. November 1982, pp. 67-71.

2. Chu, Wesley W., Alan G. Konheim. "On the Analysis and Modeling of a Class of Computer Communications Systems." IEEE Transactions on Communications. Vol. Com-2, No. 3, June 1982, pp. 645-660.

3. Hillier, Frederick S., Gerald T. Lieberman. Operations Research. San Francisco: Holden-Day, Inc., 1974.

4. Konheim, Alan G., Bernard Meister. "Waiting Lines and Times in a System with Polling." Journal of the Association for Computing Machinery. Vol. 21, No. 3, July 1974, pp.470-490.

5. Kuo, Franklin F. Protocols and Techniques for Data Communications Networks. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1981.

6. Martin, James. Systems Analysis for Data Transmission. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1981.

7. Meisner, Norman B. "Estimating the Throughput of Broadband LANs." Systems & Software. November 1982, pp. 75-77.

8. Meyer, Paul L. Introduction to Probability and Statistical Applications. Reading, Mass: Addison-Wesley Publishing Co., 1970.

9. Puzman, Josef, Rodoslav Porizek. Communication Control in Computer Networks. New York: John Wiley and Sons, 1980.

10. Riley, Wallace B. "Local Area Networks Move Beyond the Planning Stage." Systems & Software. November 1982, pp.50-60.

11. Sasieni, Maurice, Arthur Yaspan, Lawrence Friedman. Operations Research Methods and Problems. New York: John Wiley and Sons, 1959.

12. Schwartz, Mischa. Computer Communication Network Design and Analysis. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1977.

13. Sharma, Roshan Lai, Paula J. T. de Sousa, Ashok D. Ingle. Network Systems. New York: Jan Nostrand Reinhold Co., Inc., 1982.

14. Tanenbaum, Andrew S. Computer Networks. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1981.