ITC 11, Minoru Akiyama (Editor)
Elsevier Science Publishers B. V. (North-Holland)

Performance Analysis of a Virtual Circuit Switch Based LAN

A. A. Fredericks

AT&T - Bell Laboratories
Holmdel, New Jersey 07733

In this paper we analyze the performance of a Local Area Network based on a virtual circuit switch. The performance models that are developed have much broader applicability and provide insight into several important problems associated with the performance of computer/communications systems.

For the purpose of describing some of the key performance issues addressed, we may view the switch as consisting of N input and N output ports, each with an available bandwidth B (bits per second). For any point-to-point demand, a given fraction of the available bandwidth can be assigned at both the output and input ports, thus providing a "virtual" circuit communication facility with given bandwidth between the two points.

In order to establish a virtual circuit with a given dedicated bandwidth, say between port i (output) and port j (input) it is necessary that both of these ports have the needed available bandwidth. If we assume for the moment that all requests are for the same bandwidth, b, then we can view each port as having a total of nc = B/b channels available. A point-to-point connection thus requires an available channel at each port.

Perhaps the simplest strategy for managing system capacity is to reject any request for which the needed bandwidth on both ports is not available. (Rejected requests would result in subsequent retries.) This strategy tends to waste capacity since the needed bandwidth might free up and remain idle "waiting" for a retry. Queuing for resources tends to increase their efficient use. Two such strategies we consider are:

1. When a request for a circuit occurs, request a channel at each port and if none is available at either port, a request for a channel is placed in a queue at each of the ports. When the first of these requests is satisfiable, the free channel at that port is held until a channel becomes free at the other port allowing a connection to be established. While this tends to minimize the delay for setup (once the request is made) it does result in wasted channel capacity.

2. When a request for a circuit occurs which cannot be satisfied, wait until the first time that a channel is available on both ports. This strategy tends to eliminate wasted channel capacity, but a given request could be delayed indefinitely, particularly at high loads.

We formulate models for analyzing the performances of these as well as other strategies. An analytic approximation technique is developed which, when applied to these models, is shown to produce quite accurate performance predictions (based on comparisons with simulation). These models provide insight into the basic trade-offs between maximizing overall resource usage and providing fairness of treatment to individual customers.

The above "simultaneous resource possession" problem arises quite often in a variety of computer-communications systems. For example, it is closely related to the problem of obtaining available trunks in a multilink telephone call. Similar problems also arise in connection with shared peripherals in computer systems.