6 Conclusions

In this dissertation we have investigated the variability of packet-arrival processes produced by individual engineering workstations on an Ethernet local-area network. Chapter 2 presented the measurement methodology, the measurement error analysis, and a simple description of the network's aggregate traffic in terms of network load, packet length distribution, and interarrival time distribution. Chapter 3 dealt with the stochastic analysis of interarrival time series. We focused on the second-order properties as they provide the basis of our definition of variability. In Chapter 4 we defined variability in terms of indices of dispersion for intervals and counts. We estimated these indices for a number of workstations, and discussed their properties. In addition, with an example based on the rather artificial Markov-modulated Poisson process, we illustrated a way to introduce variability into analytical models. Finally, in Chapter 5 we derived a model of variability whose structure follows the structure of the data: short and long packets produce disjoint sets of interarrival times; lengths of sequences of short and long packets form a discrete-time Markov chain, a generalized semi-Markov process provides a good approximation to the autocorrelation coefficients of interarrival times for short time scales. The generalized semi-Markov process approximation requires only estimates of the first- and second-order moments of interarrival times. To complete the model, we also provided simple characterizations of short and long packet lengths. Because of the recursive nature of the model's equations, the model is suited for in simulation studies.

The relevance of the dissertation lies in the prominence of the type of computing system that we have analyzed: personal engineering workstations represent one of the most successful products developed by the computer industry over the last decade. Because of this prominence, researchers and computer analysts alike are always looking for data about system usage and modeling examples. Our analysis is quite specific, as we developed the notion of variability and tailored traffic models to reproduce it. Computer communication research is proceeding apace toward future Gigabit networks. These networks will require knowledge of the statistical behavior of data streams if they are to switch data efficiently. Analyzing the variability of arrival processes with indices of dispersion can provide some of that knowledge. Models of variability can become useful as input generators to gigabit network simulation studies.

II. Future Work

In spite of the wide range of material presented in the previous chapters, we have just scratched the surface of the possible research topics that can be identified around the issues of stochastic analysis and modeling of packet arrival processes. Here, we address briefly, in the form of research questions, some of the work that remains to be done.

By far the most interesting and pressing question is the relationship between our models of variability and queuing behavior. A key notion is whether the two-moment approximation of the generalized semi-Markov process is sufficient to guarantee an accurate queuing performance. What is the relationship between index of dispersion curves and queuing? Are there simple rules that relate the initial slope, curve knee, and asymptote of the indices to, say, waiting times? The work of Whitt et al. [26,84] provides some initial answers, but we feel that more research is needed to understand these questions completely.

Many researchers are working on future packet-switching gigabit-speed networks. These networks might provide performance guarantees. Can indices of dispersion characterize the traffic sources that can be used by algorithms that guarantee the performance of communication channels in terms of delay, jitter, and/or throughput [27]?

We discussed the difficulties associated with removing nonstationary components from packet arrival processes. Can a nonstationary version of the generalized semi-Markov process be derived that models nonstationary packet arrival processes?

How general are the models of variability that we have developed? Can they be adapted to handle, say, compressed audio and compressed video data?

Cluster models are important in applications since clusters can be treated as units by the operating system for the purpose of scheduling or processing. What is the relationship between cluster models and models of variability? Can one type of models be expressed in terms of the other?

We have expressed file server queues, composed of the superposition of several client workstation queues, in terms of the same models. Can the generalized semi-Markov model be used to represent the entire aggregate traffic?

III. Epilogue

The analysis of a massive amount of data such as that described in this work has required time and resources out of the ordinary. Real data is almost never linear and immediately suitable for simple models. We found no easy recipes for giving structure to what at first (and for quite some time) looked like an orderless aggregate of arrival times. Trial and error, experimentation, and exploratory-data-analysis techniques have slowly shaped our final solution. Unfortunately, modeling is still an art, one in which the form of the final product depends as much on the raw material as it depends on the inclination of the artist. But as artistic invention relies on techniques and tools, so our modeling of packet arrival processes has benefited from the framework and the methodology of point processes. We hope that the analyses and the models that we have developed can be the base upon which others can build.