Measure: Resource Allocation for Multimedia Communication and Processing Based on On-Line Measurement

Brief Project Description

University of Cambridge Computer Laboratory*
Dublin Institute for Advanced Studies
Telia Research

May 1995

1 Summary

Present resource management schemes are based on a model, explicit or implicit, of how customers behave. This works well in two extreme circumstances: where customer behaviour is well understood and relatively static; and where applications are simply forced to adapt to the resources provided.

However, multimedia communication and processing fits neither category. They are not well understood and generating a model for a given application may be pointless as applications are not long lasting – they are superseded rapidly by better versions or products. Nor can the behaviour of multimedia applications be described as static; their behaviour during an invocation can change drastically under user control. Finally they require some minimal levels of resource averaged over a range of timescales.

We propose to examine the technique of on-line measurement and estimation, rather than modelling, to provide information to resource management systems. We will apply this technique to three distinct but related domains: public ATM networks, local area ATM networks and the operating system in end-user computers. The goal is to develop measurement techniques and statistical estimators and to show their utility in each of these domains. Furthermore, we hope to be able to use the technique to provide a unification of resource management across these domains.

Multimedia communication and applications have enormous economic potential. Understanding how to deliver multimedia services in a cost effective way is key to the economic exploitation of these technologies. Cost effective does not simply mean efficient, it means balancing the tradeoff between complexity of algorithm and efficiency of resource usage. This is true whether the resource being consumed is in the network or the workstation. It is the nature of the customer (multimedia application) and thus the consumption pattern that is relevant rather than the nature of the resource being consumed. This work is thus relevant to network operators, network equipment providers, operating system providers and application providers.

The project will approach the problem in the three domains, using equipment developed at Telia Research to investigate measurement in a public network scenario, equipment developed at Cambridge to investigate measurement in a local area ATM network, and the operating system developed by the Pegasus project to investigate on-line measurement in a multimedia operating system.

The theoretical foundation of this work is large deviation theory. Large deviation theory is concerned with the probability of rare events occurring; for a given system the probability of the rare events is characterised

*University of Cambridge Computer Laboratory is the Co-ordinating Organisation.

by a rate function. The key observation is that large deviation rate functions may be estimated directly from measurements thus bypassing the use of models. Preliminary work applying this method has been performed by the Dublin Institute of Advanced Studies. Further mathematical detail can be found in an attached Appendix.

2 Objectives

Multiservice networks and multimedia operating systems have a common problem: *resource allocation*. In order to support quality of service guarantees, network resources must be allocated between users and operating system resources must be allocated between applications. How is this to be done?

- Resources may be allocated on the basis of peak demands, in which case they are likely to be underutilised.
- Statistical multiplexing allows resources to be allocated on the basis of expected aggregate demand, in which case the guarantees which can be made will be probabilistic. When supporting multimedia communication or multimedia applications, a probabilistic guarantee is often sufficient and allows optimal use of resources.

While little work has addressed the problem of resource management for operating systems supporting multimedia applications, the approach in multiservice networks has been to model traffic sources and to determine the expected behaviour of a number of sources sharing a common resource — a transmission link for example.

This strategy has two drawbacks. The first is that the model must be based on accurate source characterisation: as the source characteristics change, the model must be re-calibrated. The second problem is that modelling is wasteful: a model contains more information than is required for the allocation of resources.

We propose to develop techniques which use on-line measurement and estimation to predict the behaviour of resource usage. These techniques are applicable to the problem of call admission in ATM networks and to the problem of application startup and resource negotiation in an operating system designed to support multimedia applications.

If the techniques developed by this project can be used both in the end systems and in the networks, then there is potential to tackle one of the serious problems facing distributed multimedia systems, namely the mapping of quality of service (QoS) attributes across the end system/network boundary.

3 Industrial Relevance

In laboratory experiments [1], as well as in early trials of the European Pilot ATM network, Telia has identified a need for techniques which would allow the network operator/user to allocate bandwidth based on the actual situation in the network rather than on a model of the traffic. It is important to find methods which can accommodate traffic that is not stationary: although traffic may be stationary on one time-scale, shifts in activity will mean that traffic is not stationary on a larger timescale.

Currently proposed resource allocation methods provide only for marginal network utilisation. From the European ATM Pilot there are examples of services which, according to current principles, require allocation of 30–40 Mbit/s even though the average bandwidth is less than 1 Mbit/s. *Without drastic improvements in methods of resource allocation, broadband multimedia communication will lose its commercial interest.*

The situation in operating systems supporting QoS guarantees is even more amenable to on-line analysis. No one has seriously suggested that static models of application behaviour are useful in predicting long term performance. Very little is known about the pattern of demands which will be made by a combination of multimedia applications, particularly those which may be interacting. The ability of users to dramatically alter the demands which an application may make exacerbates the problem. *If general purpose operating systems which allow users to start up any combination of multimedia applications are to be viable, the problem of resource allocation must be solved.*

4 Relevance to ESPRIT Programme

The work relates directly to areas 3.5 (enhance ease of use of multimedia systems), 3.6 (interfaces from multimedia systems to telecommunications services), 6.9 (HPCN application development) and 6.10 (HPCN application execution). Area 3.1 (innovative tools for building multimedia systems) is also related to the work.

The work is related to all these areas in the Programme because it addresses a core problem in both multimedia application platforms and multiservice networks. Multimedia systems will be composed of a number of tools and applications. It is the interaction of these tools and applications that will provide the growth of multimedia. However, without effective resource management these interactions could be destructive rather than beneficial. This applies to the multimedia workstation, the interface between the multimedia workstation and the network and to the network itself.

5 Approach

5.1 Overview

In a system designed to give probabilistic guarantees where the probability of failure (cell loss, missed deadline) is low, the probabilities can often be bounded using large deviation theory. In ATM networks one can use a large deviation approach to predict cell-loss ratios and cell delay. When modelling is used, the large deviation rate function is calculated from the model. However, an alternative approach, which bypasses modelling, estimates the rate function directly from inter-cell arrival times. This exploits the close analogy between large deviation rate functions and thermodynamic entropy. In [6] it was proposed that the entropy of a cell stream be estimated directly at a multiplexer; from this data, predictions can be made rapidly using algorithms which are simple enough to be executed on-the-fly.

Preliminary investigations using trace data suggests that the method is practicable. Using the *Star Wars* movie, reconstructed from a trace of the output of a DCT-based VBR video codec at Bellcore [8], the entropy of the multiplex of several of these cell streams is estimated using the measured cell inter-arrival times. The estimated entropy is used to predict the cell loss ratio for the aggregate traffic and the result is compared with measured values. The original algorithms give good results at 99% load in a virtual buffer. Recent theoretical work has yielded an improved algorithm which gives good results down to 85% load.

Fairisle is an experimental ATM LAN whose hardware design permits a high degree of experimental flexibility, including high resolution clocks for time-stamping and measurement of traffic, and a general purpose CPU on every switch port [3]. Fairisle is equipped with a wide range of ATM traffic sources, including compressed video, audio and LAN data. In addition, traffic can be generated from pre-recorded traces of traffic activity. Further work is already in progress to test the estimation method using real ATM traffic on the Fairisle ATM network.

Telia has many years experience of evaluating network performance and its effect on QoS in ATM networks. For investigations of this sort, Telia has developed an ATM test tool complemented with an analysis software package [2]. It allows the measurement of cell loss, delay and inter-arrival time, with a resolution from hours to cell-times. Telia also have means to create repeatable network loading conditions, allowing systematic evaluation of different options for resource allocation.

The Fairisle and Telia systems provide complementary experimental platforms, giving access to different scenarios; a public ATM network and a local area ATM network are different in many respects. A key difference as far as this work is concerned is the number of traffic streams that one expects to find on a given link. The temporal characteristics of traffic streams may also be different in the two networks. The combined Telia/Fairisle approach ensures that the on-line management techniques will be investigated in both domains.

It is proposed to use the experience gained from the preliminary experiments to develop algorithms for resource allocation using entropy as a traffic descriptor and to test them on both the Fairisle ATM network and the Swedish Public ATM network using the experimental framework developed in Stockholm by Telia

Research. These platforms provide a framework in which we can systematically evaluate the performance of the on-the-fly ATM resource allocation scheme.

The important tasks are to:

- Select appropriate input sources that generate relevant traffic.
- Plan LAN and WAN scenarios that match "average" and "worst-case" conditions.
- Match the switch/measurement system to the experimental needs.
- Collect and analyse data.
- Evaluate quantitative performance measures.
- Design experiments to study how the system responds under perturbations.
- Transient analysis.

The technique of on-line measurement can be extended to the operating system in the end-user computing platform. The approach taken to develop these algorithms for operating system resource allocation, will be based on the Nemesis operating system developed in the Pegasus [11] project. Pegasus is an ESPRIT BRA which provides operating system support for distributed multimedia systems. The Pegasus project has built a completely new operating system with the requirements of multimedia in mind: providing QoS guarantees to applications and accounting the real cost of processing done on behalf of an application are key features of the system. The entire source code of Nemesis is available to the project. The allocation of resources to applications is performed in an extremely flexibly manner. We propose to use Nemesis to monitor the resource usage of applications and to test the algorithms on a live Nemesis system.

Among the observables in an operating system (which should be measured on a per-process basis) are:

- processor utilisation;
- deadline misses;
- network bandwidth;

These, including cache hit rate on DEC Alpha platforms, are all measurable by suitable software modifications within Nemesis. These measurements can then be fed into resource allocation (scheduling) algorithms.

5.2 Methodology

The following methodology has been followed in preliminary studies, and is proposed for the project. Its application in the domains of ATM networks and the Nemesis operating systems is detailed in the work-package descriptions (not included in this document). Summarising, the steps are:

- Collecting trace data from the system (i.e., network or operating system)
- Designing and testing a server-queue model of system operation
- Preliminary statistical analysis of data
- Construction of data simulators on basis of statistical analysis
- Designing, testing and developing of rate function estimators under:
  - (i) controlled exposure to traffic controlled by application or statistical models
  - (ii) exposure to real data
- Designing performance indices and evaluating the performance of the estimator in a real system
- Implementing the rate function-based traffic descriptor in an on-line demonstration.

Particular importance is attached to the testing and refinement of estimators at each stage of evaluation: this is necessarily an iterative process.

Collection of traffic data.

Data will be collected from a wide class of sources and applications under which the rate function measurement scheme is intended to operate, and within each class under varying conditions.

Establishment of server-queue paradigm.

The server-queue paradigm provides the particular setting for the application of large deviation methods. Its specification provides a particular relation between the measured rate function and probabilities of failure. Hence, the paradigm for a particular system (multiplexer, operating system) must be established and verified. Since (low) failure probabilities can be calculated directly for a given choice of paradigm and model data, the paradigm is verified by comparing its predictions against measured failure probabilities for traffic generated on the basis of various statistical data models.

Preliminary statistical analysis of data.

Measuring the rate function of data is a problem of statistical estimation, and a variety of methods are available or can be designed for this purpose. In order to choose an estimation method some prior knowledge is required of the data's broad statistical properties. Successful estimation techniques must take account of features such as the stationarity time of the data (that is, the length of time over which its statistical properties remain the same), its correlation time, the presence of 'outliers', etc.

Development of rate function estimators.

Whereas we have spoken of the rate function of a data stream, in practice we are presented, in any given application, with a representative of that stream: a finite set of data. Measuring the rate function amounts, in practice, to estimating it from the data set. Thus the measured rate function is a statistical quantity and the problem becomes to find the best method of estimation appropriate to the data. This is done by testing and refining the estimator under progressively more rigorous conditions. Testing involves comparing the estimator's predictions concerning observable quantities against those realized in the a given test configuration. The statistical properties of estimators will be examined.

Testing estimators using controlled traffic. Initial testing of estimators will be done by testing their predictions on simulated traffic, derived from a traffic model arising from the statistical analysis of data.

The advantage of simulation over the use of real data, initially, is the degree of control that it allows over the data used in the development of rate function estimators. Specifically:

- **Repeatability.** Multiple, non-identical data sets drawn from the same statistical model can be used for repeated testing. This is particularly important as the estimators we use are statistical objects: a significant component of their accuracy and reliability is their statistical variance.
- **Adjustability.** During the development of an estimator, there will be particular strengths and weaknesses which we will need to identify and explore. These can be explored over a range of source behaviour by varying the parameters of the simulator model in a more controlled manner than is possible with real sources. This applies also to fundamental statistical properties, such as time-scales of stationarity, which will be established for real data only by measurement.
- **Analytic tractability.** Analytic large deviation results can often be obtained for simple models, allowing rapid insight to be gained into the origins of experimental observations.

Testing estimators against real traffic. Once an estimator with favourable properties against simulated data has been developed, it will remain to verify that the estimates work well with real data. If not, the task will be to identify which features of real data, not present in the simulation models, cause the poor estimation results. These features can then be incorporated in the models, allowing systematic testing of estimation schemes against them.

PERFORMANCE EVALUATION.

DIAS algorithms for the rate function estimators will be encoded in a form suitable for on-line measurement on the experimental platforms in Telia and Cambridge. Performance indices will be devised to encapsulate the utility of the algorithms in:

- accounting for measured values of the key observables;
- predicting future demand from present measurements.

Both types will be based on the methods of quantitative evaluation of the estimators during their development. Amongst the latter type will be those based on a comparison of the following time-scales:

- *Sampling Length:* the time over which data must be collected to provide a reliable one-time estimate of the rate function.
- *Prediction Length:* the time over which a given estimate is sufficiently reliable. This determines the frequency at which sampling for on-line estimation must be carried out.

PROTOTYPE DEMONSTRATION.

Once the development of estimators is complete, the aim of the project is to demonstrate how resource allocation using rate function-based source descriptors can maintain users' perceived QoS at agreed levels while making maximal use of resources. The demonstrations will operate within the same environments as were used to collect data initially.

The rate function of the data stream will form the basis of the source descriptor. Full specification of the descriptor will include a definition of the type of data required, algorithms to determine the quality of the data available, estimates of the accuracy of the descriptor obtained, and methods of deducing QoS information from the descriptor.

The functions of data collection and data processing (the on-line calculation of the descriptor) will be logically separated. While it is reasonable to assume that a workstation running Nemesis would have the resources required to implement the descriptor, we do not expect the port controllers of an ATM switch to have the spare processing capacity or the resources available to implement a full traffic descriptor itself. We envisage the port controllers simply collecting data, and sending it elsewhere for processing in return for the estimated descriptor or derived QoS parameters.

The prototype descriptor will be implemented as a server to which the network manager or the multimedia operating system can connect as clients. The interface will be defined by the type of data the implementation needs to form the descriptor, and the kind of information the client needs for resource allocation. The client will connect to the server and start offering data, which the server will start to process. When the client needs the descriptor, or QoS information derived from it, it will request it from the server, which will return it along with estimates of the accuracy of the information. Also possible is the development of a feedback mechanism, where the server may suggest to the client changes in the volume of data collected, or in the data collection method.

References

[1] N. Bjœrkman, Goldstein, Hedman, A. Latour-Henner, Tholin, and Gil. Network performance and its relationship with quality of service in an experimental broadband network. In *Proceedings 1st IFIP TC6 Workshop on Broadband Communications*, pages 157–168, Estoril, 1992.

[2] N. Bjœrkman and A. Latour-Henner. High time resolution cell loss measurements of mixed cbr and vbr traffic. In *Proceedings 2nd IFIP TC6 Workshop on Broadband Communications*, Paris, 1992.

[3] R. J. Black, I. M. Leslie, and D. R. McAuley. Experiences of building an ATM switch for the local area. In *Proceedings, 1994 SIGCOMM Conference*, London, UK, August–September 1994.

[4] D. D. Botvich and N. G. Duffield. Large deviations, the shape of the loss curve, and economies of scale in large multiplexers. Technical report number dapg-9412, Dublin Institute for Advanced Studies, 1994. To appear in *Queueing Systems*.

[5] Simon Crosby, Ian Leslie, John T. Lewis, Neil O'Connell, Raymond Russell, and Fergal Toomey. Bypassing modelling: an investigation of entropy as a traffic descriptor in the fairisle atm network. In *Proceedings of the 12th UK Teletraffic Symposium*, Old Windsor, 1995.

[6] N. G. Duffield, J. T. Lewis, Neil O'Connell, Raymond Russell, and Fergal Toomey. Entropy of atm traffic streams: a tool for estimating qos parameters. Technical report number dapg-9430, Dublin Institute for Advanced Studies, 1994. To appear in *IEEE Journal of Selected Areas in Communications: Special Issue on the Fundamentals of Networking*.

[7] N. G. Duffield, J. T. Lewis, Neil O'Connell, Raymond Russell, and Fergal Toomey. The entropy of an arrivals process: a tool for estimating qos parameters in atm traffic. In *Proceedings of the 11th UK Teletraffic Symposium*, Cambridge, 1994.

[8] M. W. Garrett and W. Willinger. Analysis, modeling and generation of self-similar vbr video traffic. In *Proceedings of ACM SIGCOMM 94*, pages 269–280, London, UK, August 1994.

[9] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson. On the self-similar nature of Ethernet traffic. In *Proceedings of ACM SIGCOMM 93*, 1993.

[10] N. Bjœrkman M. Goldstein and A. Latour-Henner. Cell loss effects on perceived audio/video quality and quality of service (qos). In *Sixth International Workshop on Packet Video*, Portland, Oregon, 1994.

[11] S. J. Mullender, I. M. Leslie, and D. R. McAuley. Operating-system support for distributed multimedia. In *Proceedings of Summer 1994 Usenix Conference*, pages 209–220, Boston, Massachusetts, USA, June 1994. Also available as Pegasus Paper 94–6.

[12] Neil O'Connell. Large deviations in queueing networks. Technical report number dapg-9413, Dublin Institute for Advanced Studies, 1995.