MEASURE: Further Information

January 16, 1996

1 Introduction

The MEASURE project is an EC funded 3 year project which began on 1 Jan 1996. The partners are the Computer Laboratory, Cambridge, Telia Research (Telia are the Swedish PTT) and the Theorectical Physics Department of the Dublin Institute of Advanced Studies (DIAS).

The project is investigating a technique of online measurement and estimation in order to perform resource management for multimedia (both in operating systems and in ATM networks).

DIAS provide the theorectical foundation for the work. Telia provide expertise in and access to public ATM networks. Cambridge provide expertise in and access to local ATM networks and operating systems (Nemesis, an output of the ESPRIT Pegasus project — info at http://www.cl.cam.ac.uk:80/Research/SRG/pegasus/).

The research assistant/associate at Cambridge will thus be engaged in the implementation of on line measure monitors (either in ATM switches such as Fairisle — info at http://www.cl.cam.ac.uk/Research/SRG/fairisle.html) or within Nemesis. These monitors will be used to gather information for the development of estimators (done by DIAS) which will have to be implemented in the live systems (by Cambridge for local ATM and operating systems, by Telia for public ATM networks). This will also be the task of the RA.

2 Project Summary

Multiservice networks and multimedia operating systems have a common problem: *resource allocation*. Present resource management schemes in communications networks and operating systems are based on a static model of how the users of the resources behave. Multimedia communication and processing, however, are not static. They are rapidly superseded by better versions, and both user control and the content of the media being processed can drastically change the behaviour of an application during its lifetime. Moreover, such applications require minimal Quality of Service (QoS) guarantees of resource availability over a range of timescales, to function properly.

We propose to investigate the technique of dynamic, on-line measurement and estimation of resource demand, rather than static 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 demonstrate their utility in call admission and network dimensioning in ATM networks, and application resource negotiation in a multimedia operating system. Furthermore, we aim to provide a theoretical unification of resource management across the three domains. If our techniques can be used both in end systems and networks, then there is potential to tackle a serious problem facing distributed multimedia systems, namely the mapping of QoS attributes across the end system/network boundary.

The project will approach the problem of measurement and resource allocation in the three domains, using equipment developed at Telia Research in a public ATM network scenario, equipment developed at Cambridge in a local area ATM network, and a multimedia operating system developed by the Pegasus ESPRIT project.

The theoretical foundation of this work is large deviation theory, which enables us to derive, for a given system, a *rate function* which characterises the probability of rare events occurring. The conventional approach to derive the rate function requires the traffic process to be statically characterised, by fitting model parameters to observed data. Our approach, however, estimates the rate functions directly from on-line measurements, thus bypassing the use of models, by exploiting a close analogy between large deviation rate functions and thermodynamic entropy.

Preliminary work applying this method to predict cell loss and delay on the Fairisle ATM LAN has been performed by the Dublin Institute of Advanced Studies and Cambridge, with good results. The theoretical foundations of the approach are now well established. Experience gained from the preliminary experiments will be used to develop algorithms for resource allocation using entropy as a traffic descriptor, and to test them on both Fairisle and the Swedish Public ATM network. The technique of on-line measurement will also be extended to the operating system in the end-user computing platform, to allocate resources for multimedia applications.

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. This work is thus relevant to network operators, network equipment providers, operating system providers and application providers.

The results of this work will be made available for exploitation in several ways. First, the application of on-line resource allocation in wide area ATM networks is of direct interest to the commercial operation of the Telia AG ATM network. Second, through publication in academic journals and conferences, and through workshops organized by the project, public results of the work will be made available to companies which have expressed their support for our approach. Finally, a World Wide Web site will be set up describing the project and giving access to the results.

2.0.1 Local Area ATM Networks

This domain will be studied by DIAS and Cambridge. Preliminary work on the development of a queue-server paradigm for the Fairisle ATM switch has been completed, and will be refined in the light of further experiments. Measurement software for use on the Fairisle switch will be developed by Cambridge. This will be used to study the resource demands of emerging multimedia applications and traditional LAN traffic using the ATM network.

DIAS will be responsible for the primary objectives of the work in this domain, namely statistical analysis of the traffic measurements and development of estimators for the ATM LAN. The estimators will be implemented by Cambridge directly on the Fairisle switch, where initial evaluation will take place under controlled, repeatable conditions.

Finally, the estimators will be evaluated in a demonstration on the live Cambridge LAN, where they will be exposed to real traffic. The objective is to predict the performance of various traffic types in a demonstration of a prototype Call Admission Control (CAC) system. Measurements of call performance before and after acceptance/rejection will be taken, with the goal of establishing the accuracy of the algorithm.

2.0.2 Multimedia Operating Systems

This domain will be studied by DIAS and Cambridge. Here, the goal is to enable an operating system supporting multimedia applications to allocate resources on demand whilst providing QoS guarantees. In this context the statistical estimators will form part of the scheduler, and will operate, as before, on the basis of the on-line analysis of resource demand. 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 approach will be based on the Nemesis operating system developed in the Pegasus[ref] project. In this domain the initial emphasis will be somewhat different. Measurement of resource demand is straightforward in Nemesis, but the queue-server paradigm for this operating system has not yet been established.

Instrumentation and measurement of Nemesis will be performed by Cambridge. The aim will be to identify key metrics which accurately characterise multimedia application resource demands, and to develop minimally invasive kernel code to measure these. DIAS will undertake first to develop and test a server-queue paradigm for the Nemesis scheduler, using repeatable load conditions. Secondly, DIAS will undertake statistical analysis of measurements made under realistic load conditions, with the aim of developing rate function estimators. It is anticipated that the estimators will be incorporated directly into the Quality of Service Manager within Nemesis, where initial evaluation of the system will again take place under controlled, repeatable conditions. The final objective will be to demonstrate the efficacy of the system under realistic loads.

3 Theoretical Foundations

This note contains a brief introduction to the ideas of large deviations theory, its application to queueing systems, and the connection with thermodynamics.

3.1 Summary

Variable bit-rate traffic, such as that produced by multimedia applications communicating via high-speed communications links, is potentially troublesome to deal with within the network as far as resource allocation is concerned. On the positive side, when a number of traffic streams with widely different characteristics and service requirements are statistically multiplexed, they can be carried on the same link in spite of the fact that the sum of their peak bandwidths may exceed the resource capacity; the *effective bandwidth* of a source describes its effective resource utilisation in the presence of the other traffic, and is typically significantly less than its peak bandwidth. Using the effective bandwidth for admission control permits the network to maximise utilisation of its resources, whilst providing a statistical bound on performance to the applications. In practice, reported methods for computation of the effective bandwidth require a (statistical) characterisation of the behaviour of each of the sources.

What properties are required of a good traffic characterisation scheme? Firstly, traffic descriptors should contain just that information needed to describe the likely effect of the traffic on the network and no more. Secondly, the scheme should work generally, both with forseeable traffic types such as voice, video and data, and also with unforeseeable future applications. Thirdly, it is desirable that the scheme should not be overly complex or make use of descriptors that are difficult to determine in practice.

It is apparent that parametric schemes such as those based on traffic modelling on the basis of precise or statistical descriptions of source behaviour, do not satisfy any of these criteria. Other schemes, based on specifying key parameters such as peak cell rate and cell delay variation tolerance, have more general applicability than those based on a particular model, but do not provide sufficient information for the full statistical multiplexing gain to be acheived. An alternative scheme, based on measuring the large deviation rate-function of a traffic stream, has the potential to meet all three of the above requirements. This scheme, proposed by the DIAS Group (Dublin), is founded on the following observation, now widely accepted: under very general conditions, a single-server queue has a queue-length distribution with asymptotics of the form

$$\frac{1}{q} \ln \mathbb{P}\{Q > q\} \sim -\delta \,, \tag{1}$$

and the asymptotic decay-rate δ can be calculated from the rate-function $I(\cdot)$ of the arrival process:

$$\delta = \min_{c>0} I(c+s)/c \,, \tag{2}$$

where s is the service rate. It follows that, if $I(\cdot)$ can be measured, then δ can be calculated for any value of s. From δ, such important quantities as cell-loss and cell-delay can be estimated.

$I(\cdot)$ is called the 'entropy' of the arriving traffic, in analogy with the thermodynamic entropy of a gas: the entropy function of a gas, which mathematically speaking is nothing other than a large deviation rate-function, contains just the information needed to calculate its bulk properties. Engineers do not bother to construct complex models of a gas in order to calculate its entropy — they measure it. The basis of the LDA approach to traffic modelling is that similar approach may prove fruitful in the study of statistical multiplexing.

A traffic characterisation scheme based on measuring the rate-function $I(\cdot)$ posesses the first and second of the desirable properties listed above: it provides precisely the information needed to describe the effect of the traffic on the network, and it can be applied to traffic of any type. An additional advantage is that, if the rate-functions of several individual traffic streams are known, then the rate-functions of the traffic which results when the streams are multiplexed together can be calculated. Hence, by measuring rate functions at the inputs to a network it may be possible to predict the network's performance at any of its nodes. The following section provides a more thorough comparison of the LD approach and its analogy in statstical thermodynamics, and was written by John Lewis of DIAS, as a technical appendix to the *Measure* project proposal. Thereafter a brief description of the method used to measure the entropy is given. This is drawn from a joint paper presented at the UKTS '95.

3.2 Analytical Foundations

To begin with, consider a simple queueing system consisting of a single server with constant service rate s (that is, s customers can be served in each unit of time). Suppose that, on arrival, customers await service in a common buffer. We observe the system at some typical time, $t = 0$ say, and denote the number of customers who have arrived in the previous t units of time by A_t. The queue length at the present instant is related to the queue length at the arrival of the last customer by a simple recursion relation; iterating this relation leads to an equation for the current queue length[**?**],

$$Q = \sup_{t\geq 0}\ (A_t - st). \tag{3}$$

We assume that $\{A_t : t \geq 0\}$ is a stationary process; then $m = \mathbb{E}(A_t/t)$ is independent of t. The queue is stable provided $m < s$. The process $\{A_t : t \geq 0\}$ is said to satisfy a *large deviations principle* with *rate function* I if, for $a > m$,

$$\mathbb{P}\{A_t > at\} \sim e^{-tI(a)}. \tag{4}$$

The range of processes which satisfy a large deviations principle is extremely broad; for examples see[**?**]. In this case, the rate function of the arrivals is related in a simple way to the rate functions associated with the traffic on each line. Glynn and Whitt[**?**] proved that, under general conditions, this distribution has an exponential tail:

$$\mathbb{P}\{Q > q\} \sim e^{-\delta q}, \tag{5}$$

with a decay parameter δ which can be calculated directly from the rate function:

$$\delta = \min_{c>0}\ I(c+s)/c. \tag{6}$$

Work subsequently carried out at DIAS by Nick Duffield and Neil O'Connell[**?**] broadened still further the context in which results similar to (5) are known to hold.

The single-server queue with multiple independent input lines was the first queueing system considered by the DIAS group. In this case, the rate function of the arrivals is simply the sum of the rate functions associated with the traffic on each line. More complex queueing systems are also amenable to a large deviations treatment. Queueing systems generally may be modelled as networks of interconnected single-server queues with stochastic routing schemes. Recent work by Neil O'Connell[**?**] has shown that if the arrivals at a single-server queue satisfy a large deviations principle then so do the departures, with a rate function which can be calculated from the rate function of the arrivals. Therefore the queue length distribution at any node in a network can, in principle, be characterised using the rate functions of the traffic entering the network.

The rate function $I(.)$ is called the *entropy* of the traffic: just as the bulk properties of traffic on a queueing network are described by a large deviations rate function, so the bulk properties of a gas are also described by a rate function, known as the entropy of the gas. We will now briefly explain this connection between large deviations theory and thermodynamics; for a readable account of this approach to thermodynamics (which goes back to Gibbs in 1873), see Callan[**?**].

For simplicity, we consider the thermodynamics of a gas consisting of molecules of a single chemical species. The entropy per unit volume of the gas is a function $s(u, \rho)$ of u, the internal energy per unit volume, and ρ, the mass per unit volume. The function $s(u, \rho)$ is continuously differentiable (expressing the continuity of the intensive variables, such as pressure) and concave (expressing thermodynamic stability). All bulk properties of the gas can be derived from the entropy function, using standard formulae (see[**?**]). For example, the entropy function of an ideal gas is given (up to an additive constant) by

$$s(u, \rho) = k\rho \left\{ \frac{f}{2} \ln u - \left(1 + \frac{f}{2}\right) \ln \rho \right\}, \tag{7}$$

where k is Boltzmann's constant and f is a constant depending on the chemical species. From this function we may derive, for example, the two equations of state:

$$p = kT\rho, \quad u = \frac{f}{2} kT. \tag{8}$$

Boltzmann's remarkable formula

$$S = k \ln W, \tag{9}$$

relating the thermodynamic entropy S of a macroscopic equilibrium state to the number W of microscopic states which correspond to the macroscopic state, gives rise to problems of interpretation. To have any hope of giving the formula a precise meaning, we should pass to the limit in which the volume V of the system becomes infinite. In this limit we might expect a formula:

$$s(u, \rho) = \lim_{V \to \infty} \frac{k}{V} \ln W \left[H_V = uV, N_V = \rho V\right] \tag{10}$$

where $W\left[H_V = uV, N_V = \rho V\right]$ is the phase-space volume of those macroscopic states for which the Hamiltonian H_V takes the value uV and the number of particles N_V takes the value ρV. However, the existence of the limit on the right-hand side of (10) poses obvious difficulties. These were resolved by Ruelle[**?**] in 1965.

Ruelle's idea can be illustrated simply: let B_a be a disc of radius a centered on the point (u, ρ); one can prove that the limit

$$s[B_a] = \lim_{V \to \infty} \frac{k}{V} \ln W[(H_V/V, N_V/V) \in B_a] \tag{11}$$

exists. Now let B_a shrink to a point, defining

$$s(u, \rho) := \min_{a>0} s[B_a]. \tag{12}$$

The argument used to establish the existence of the limit (11) proves also that $s(u, \rho)$ is a concave function. In the case of an ideal gas, it is easy to verify (using Stirling's Formula) that Ruelle's procedure yields (7).

This simple idea was developed by Ruelle and Lanford to provide a rigorous treatment of statistical thermodynamics, described in detail in Lanford's 1971 Battelle Lectures[**?**]. Ruelle's idea turned out to have a surprising ramification in probability theory: Lanford used it to give a completely new proof of Cramèr's Theorem; this was the first step in an important development in the theory of large deviations.

The modern theory of large deviations began with Cramèr's refinement[**?**] of the weak law of large numbers. Let $A_t = X_1 + \ldots + X_t$, where the X_i are independent and have a common distribution with mean m. Then there exists a function s such that,

$$\mathbb{P}\{A_t > at\} \sim e^{ts(a)}, \text{ for } a > m. \tag{13}$$

Lanford first proved that, for each open interval $B_a := (x - a, x + a)$, the limit

$$\lim_{t\to\infty} \frac{1}{t} \ln P\left\{\frac{A_t}{t} \in B_a\right\} \tag{14}$$

exists. Lanford then defined $s(x)$ by

$$s(x) := \min_{a>0} s[B_a], \tag{15}$$

and proved that (12) holds for each open interval J.

This approach to the theory of large deviations was taken up by Bahadur and Zabell[**?**]; they developed it to prove a powerful generalisation of Cramèr's Theorem. Azencott[**?**] and, later, Deuschel and Stroock [**?**], systematised these developments. A detailed review of thermodynamic aspects of large deviations theory is given in[**?**].

We have illustrated how, from the mathematical point of view, the rate-function of large deviations theory is the same kind of object as the entropy function of thermodynamics. They have this in common: they encapsulate concisely the relevant information about the system; for this reason it makes sense to measure them.

3.3 Measuring the Rate Function

This section describes briefly one method of estimating the rate-function, which has already been applied, with considerable success, to simulated 'bursty' traffic (see [**??**]), and to traffic on the Fairisle ATM network at Cambridge, carrying reconstructed multimedia traffic. Rather than estimating the rate function I itself, it is more convenient to estimate a transform of I called the *scaled* *cumulant generating function* (CGF), defined by

$$\lambda(\theta) = \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\ (\exp(\theta A_n)), \tag{16}$$

where A_n denotes the number of cells (equivalently packets, in a non-ATM system, but subject to simple modifications to incorporate their sizes: i.e. A_n is the amount of *work*) arriving in an interval of length n. The CGF λ is related to I by the Legendre transform:

$$I(x) = \max_\theta\{x\theta - \lambda(\theta)\}, \tag{17}$$

where the maximisation is performed over all values of θ. The decay-rate δ can be calculated directly from λ fom the equation

$$\delta = \max\{\theta > 0 : \lambda(\theta) < s\theta\}. \tag{18}$$

Thus δ is the ordinate of the intercept of the graph of λ and a line of slope s through the origin.

How might we estimate λ? It can be shown that for a very large class of arrivals processes it is possible to find a block-length T such that the aggregated arrivals A_T are approximately independent and identically distributed; in this case the CGF is given approximately by

$$\lambda(\theta) \approx \lambda_T(\theta) = \frac{1}{T} \ln \mathbb{E}\ (\exp(\theta A_T)). \tag{19}$$

The most straightforward approach to estimating λ therefore is to use the CGF of the empirical distribution of the block arrivals:

$$\hat{\lambda}_T(\theta) = \frac{1}{T} \ln \frac{1}{N} \sum_{i=1}^{N} \exp(\theta A_T^{(i)}). \tag{20}$$

The advantages of this estimator are its simplicity and the fact that it makes minimal assumptions about the traffic. To use it in practice, we must make a choice concerning the block length T. This should be sufficiently large so that successive observations of A_T are approximately independent, but not so large that short bursts of activity, which make an important contribution to the queue-length, are effectively smoothed out.