Evaluating Computer Network Packet Inter-Arrival Distributions

Dennis Guster
St. Cloud State University, USA

David Robinson
St. Cloud State University, USA

Richard Sundheim
St. Cloud State University, USA

INTRODUCTION

The past decade could be classified as the "decade of connectivity"; in fact, it is commonplace for computers to be connected to an LAN, which in turn is connected to a WAN, which provides an Internet connection. On an application level this connectivity allows access to data that even five years earlier were unavailable to the general population. This growth has not occurred without problems, however. The number of users and the complexity/size of their applications continue to mushroom. Many networks are over-subscribed in terms of bandwidth, especially during peak usage periods. Often network growth was not planned for, and these networks suffer from poor design. Also, the explosive growth has often necessitated that crisis management be employed just to keep basic applications running. Whatever the source of the problem, it is clear that proactive design and management strategies need to be employed to optimize available networking resources (Fortier & Desrochers, 1990). This is especially true in today's world of massive Internet usage (Zhu, Yu, & Doyle, 2001).

BACKGROUND

Obviously, one way to increase network bandwidth is to increase the speed of the links. However, this may not always be practical due to cost or implementation time. Furthermore, this solution needs to be carefully thought out because increasing speed in one part of a network could adversely effect response time in another part of that network. Another solution would be to optimize the currently available bandwidth through programming logic. Quality of service (QOS) and reservation bandwidth (RB) are two popular methods currently being utilized. Implementation of these optimization methods is rarely simple and often requires a high degree of experimentation if they are to be effectively configured (Walker, 2000). This experimentation can have detrimental effects on a live network, often taking away resources from mission critical applications. The client/server model so popular in Internet communication is a prime example of a system that can benefit from an analytical modeling strategy (Postigo-Boix, Garcia-Haro, & Melus-Moreno, 2005).

THE BENEFITS OF SIMULATION

Therefore, the most efficient way to ascertain the potential benefit and derive baseline configuration parameters for these optimization methods is through simulation or mathematical modeling. Simulation can be very effective in planning a network design. For example, what if network link number three was increased to 10Gbs? Would workstations on that link experience an improvement in response time? What would happen to workstations on the other part of the total network? Another approach to ascertain if a given network will exceed its capacity is based on network calculus (Cruz, 1991; Le Boudec, 1998). In this method the characteristics (such as speed, maximum packet size, and peak rate) of the network architecture are analyzed, and performance bounds are defined. The goal then is to devise control/management programs (such as QOS and RB) that will keep the workload within those defined bounds. There are numerous applications of this control/management logic, such as Cruz (1995), Cruz and Tsai (1996), Firoiu, Le Boudec, Towsley, and Zhang (2002), and Vojnovic and Le Boudec (2002). Recent work has focused on integrated service on WANs across service providers (Cruz & Santhanam, 2000). These control/management programs have proved very effective under a variety of circumstances, but are influenced by the packet inter-arrival rate as well. Therefore, if a network designer is contemplating invoking one of these options, simulation could be used to test how the option in question would improve performance on his or her system, provided an adequate method could be found to describe the distribution within that network. Simulation has been used for many years in network design; however, the time and cost of its use have often been prohibitive. In recent years, new windows-based point and click products such as Comnet III (and its successors Network & Simscript) have eliminated the drudgery and the cost of writing simulations via a command line interface (CACI, 1998). Under Comnet III the appropriate devices are selected, connected together, and their characteristics defined. There is still a limiting factor in this process: the definition of the distribution of the packet inter-arrival rates. The theoretical model often used to describe computer networking is the Poisson. This model may have been adequate for some of the first single tier, single protocol networks. However, it lacks validity in describing the total stream in today's hierarchically complex multi-protocol networks. In the classical Poisson process model (such as M/M/1), when the number of arrivals follows a Poisson probability distribution, then the time between arrivals (inter-arrival time) follows a decaying exponential probability distribution. A number of studies confirm that the actual inter-arrival distribution of packets is not totally exponential as would be expected in the classical model (Guster, Robinson, & Richardson, 1999; Krzenski, 1998; Partridge, 1993; Vandolore, Babic, & Jain, 1999). However, in light of recent changes in networking regarding line speed and number of connected hosts that in effect have tripled in magnitude, the Poisson model is being reevaluated. Specifically, Karagiannis, Molle, and Faloutsos (2004) found that the stream as a whole may not be exactly Poisson, but some of it components might fit quite well. Those cases involved sub-second streams and large multi-second streams.

The inter-arrival distribution selected can have a major impact on the results of the simulation (Guster, Safonov, Hall, & Sundheim, 2003; Guster, Sohn, Robinson, & Safonov, 2003). In a study by Krzenski (1999) that analyzed the simulated performance on a shared Ethernet network, 12 different inter-arrival distributions were tried within the same simulation problem. These included the gamma distribution, which is a generalization of the exponential distribution, allowing for a modal inter-arrival time (the most commonly occurring time between arrivals) to be moved out away from the very short, nearly instantaneous time occurring with the exponential distribution. Another distribution among the 12 was an integer distribution, whereby equal probabilities are assigned to different values that are equally spaced throughout the possible inter-arrival times. Among the 12 distributions, there were vast discrepancies in the results. For example, the number of collision episodes varied from 310 with a gamma distribution to 741 with an integer distribution. These results further support the need to have the correct distribution in simulations designed to provide design and management feedback about computer networks. The frustration of the past work and the need for additional research is best summarized by Partridge (1993, p. 3):

> *... We still do not understand how data communication traffic behaves. After nearly a quarter of a century of data communication, researchers are still struggling to develop adequate traffic models. Yet daily we make decisions about how to configure networks and configure network devices based on inadequate models of data traffic. There is a serious need for more research work on non-Poisson queuing models.*

A number of different strategies have been employed in the development of models used to describe packet inter-arrival rates (Guster, Litvinov, Richardson, & Robinson, 2002). Perhaps the most valid is to record all of the packet arrival times for the time period desired and use that to generate the distribution. The advantage of this strategy is accuracy, but it often requires massive amounts of data to be recorded and processed. To lessen this burden, often a representative sample from the time period is used. However, validating the sample period is often difficult, especially if the file size is not large. Known distributions have been used with limited success (Guster & Robinson, 1994, 2000; Guster, Robinson, & Juckel, 2000). For simple networks, exponential distributions provide some promise; however, they fail to deal with the intricacies of complex multi-protocol networks. Tabular distributions, in which one column describes the interval and a second column describes the probability of a value from that interval occurring, offer a moderate degree of accuracy, but they take time to derive, and their sophistication is related to the number of rows included. Regression and ANOVA have been used in some cases but lack the ability to describe the peaks and valleys associated with packet arrival data. Time series deals with these variations better but still lacks the sophistication needed and requires relatively complex models to even come close (Guster & Robinson, 1993). Packet trains are very effective in describing packet traffic from a single session such as telnet (Vandolore et al., 1999) but lack the complexity to deal with multiple concurrent sessions on the same network. However, the basic idea of breaking the total stream into subparts has been improved using a multilevel traffic approach. Karagiannis, Papagiannaki, and Faloutsos (2005) have been able to devise a classification schema based on the applications that generate them and have been able to classify 80-90% of traffic with more than 95% accuracy.

FUTURE TRENDS

Two non-Poisson queuing models have offered a degree of promise. One method involves viewing the observation interval as containing several independent Poisson processes rather than as a single exponential distribution. In a study by Guster et al. (1999), actual data were analyzed and shown to contain three Poisson processes of differing characteristics. During the first phase activity was increasing. During the