A SIZER ANALYSIS OF IP FLOW START TIMES

J. S. MARRON, FELIX HERNANDEZ-CAMPOS, AND F. D. SMITH

ABSTRACT. The SiZer technique is used to study the homogeneity of a point process of Internet traffic flow start times. It is seen that a homogenous Poisson process is an inappropriate model, because it does not yield observed statistically significant burstiness. Some Weibull waiting processes gives better, but still inadequate performance. A clustered Poisson process gives the best fit.

1. INTRODUCTION

Simulation of Internet traffic is a challenging and interesting problem. It is important for both Internet researchers (who try to improve the performance of the Internet itself), and also for testing of many types of Internet based business applications. There is a strong need for involvement of creative statisticians, both in the development of the simulation methods, and in the assessment of their performance. This paper discusses an example demonstrating the usefulness of the relatively new statistical method, SiZer, and showing that simulation model development and statistical model assessment can and should interact.

Most of Internet traffic is composed of IP (Internet Protocol) flows. These are the transfers of data from one computer to another, as described in RFC 791, see Postel (1981). An IP flow is defined herein as a set of packets carrying IP datagrams that share the same sending and receiving addresses (more specifically, that have identical IP addressing 5-tuples consisting of protocol number, source IP address and port number, and destination IP address and port number). Transport protocols, such as TCP (Transmission Control Protocol) and UDP (User Datagram Protocol), provide higher-level communication services built on top of IP that support the exchange of information between applications. Web browsing, email, telnet and many others use TCP over IP, while audio/video streaming, name resolution, and other applications use UDP over IP. IP is therefore the fundamental building block of the Internet, and its complex behavior is the result of aggregating the communication patterns of very