

STOCHASTIC MODELS OF INTERNET TRAFFIC DATA FOR DIFFERENT TYPES OF INTERNET USERS: THE CASE OF EGERTON UNIVERSITY, KENYA.

Abstract

The phenomenal growth of internet usage and size has brought about a unique set of challenges in maintaining the Quality of Service (QoS) as measured by the speed of access to information which is determined by the amount of bandwidth made available to the users. In the past, network administrators have attempted to improve QoS by supplying more bandwidth but internet links have remained perennially slow. Previous studies have focused on determining the optimum bandwidth supply using a variety of stochastic models but this has not helped to reduce the delays experienced by users. Other studies have focused on bandwidth demand as opposed to bandwidth supply to explain the behavior of internet users. However, bandwidth demand has not been studied when taking into account the type of internet user demanding the bandwidth. The purpose of this study was to model internet traffic data while taking into account the different types of internet users generating the data. The Internet traffic data was collected from a total of 133 users comprising of students, lecturers and other university staff with ages varying between 19 to 48 years. The Internet Traffic Data obtained in the study was found to be skewed, heavy tailed and non-stationary. In addition, the data exhibited strong self-similarity and long range dependence (LRD). The TAR, ARMA-GARCH, ARFIMA models captured the non-stationary, nonlinear, self-similar and LRD characteristics of the data. The ARFIMA model exhibited the least variability among the fitted models. The MMPP model did not capture any of the key characteristics of the data. No models were fitted to the data sets from the different types of internet users since there were no significant differences observed in their marginal distributions ($p > 0.05$) after carrying out the Z-test for means comparison. The study thus established that the ARFIMA model exhibited the least variability among the fitted models.