Forecasting Models and Adaptive Quantized Bandwidth Provisioning for Nonstationary Network Traffic

Balaji Krithikaivasan (University of Missouri-Kansas City)

Abstract

With the emergence of multitude of loss and delay-sensitive applications along with increasing processing power of routers and switches, it is the *link bandwidth* that can become the bottleneck resource for an Internet service provider influencing observed data loss, delay, jitter, etc. Additionally, it is equally important from the provider's perspective to maintain the utilization of provisioned bandwidth envelope in acceptable bounds at all times. Thus, an effective bandwidth management framework that has the ability to adapt to inevitable traffic fluctuations becomes a critical requirement for a service provider. As a first step, it is essential to characterize the traffic that flows through a network. In this context, it has been found in recent years that a non-stationary trend of second order (both the mean and the variance are time-dependent) is increasingly prevalent in the volume of traffic traversing any IP backbone network on a 24-hour window. Consequently, it is quite difficult and at times, not quite possible to develop an objective traffic model without making simplified assumptions on the stochastic nature of the traffic. Furthermore, the applicability of such objective models strongly depends on the validity of their assumptions on an actual traffic trace. In our work, we overcome this issue by adopting a measurement-based approach to characterize the traffic dynamics *statistically* (through time series models) rather than objectively. In addition, statistically derived model has an advantage that it can be tuned in an incremental manner in order to take into account both the short term and long term traffic fluctuations that may arise in a network. Essentially, we propose a *two-stage* adaptive bandwidth management framework: In the first stage, we use statistical time series to characterize the network traffic collected at periodic intervals and forecast the traffic throughput at future instant; In the second stage, we map the obtained forecast into quantized bandwidth requirement and provision it.

The traffic trace used in this work is collected on the Internet access link connecting University of Missouri-Kansas City (UMKC) to MORENET, Internet service provider of UMKC. Initially, we have explored the applicability of various time series models to effectively capture the dynamics of real traffic trace. The inherent time-of-the-day effect as well as the non-stationary variance evident in the traffic trace motivated us toward deriving a seasonal ARIMA model with *conditionally heteroscedastic* disturbances (i.e., ARIMA model with ARCH extension). Furthermore, as against the conventional modeling of Gaussian innovation in time series models, we have found the innovation process to be *heavy-tailed* which we capture using Student-t distribution. It can be noted that Gaussian distribution is a limiting case of Student-t distribution as the degrees of freedom $\nu \to \infty$. The utility of deriving a time series model is to be able to extrapolate the future. It has been in practice to use the Minimum Mean Square Error (MMSE) forecast as the desirable predictor for systems modeled using time series. However, in a system such as bandwidth provisioning and management system, short term data loss (arising from under-forecast) is more stringent than the short-term under-utilization (arising from over forecast). Thus, we introduce a generalized forecast cost function that takes into account the relative importance of short term data loss over under-utilization of provisioned bandwidth envelope. We achieve this by incorporating a penalty ratio factor in the forecast cost function that enable us to associate different weights to the deviation of forecast from the *unknown true series value*. In this way, the service provider of a network can tune the point forecast to meet the agreed-upon guarantee in terms of data loss as well as desirable utilization bounds. We have studied various class of penalty functions and proved that an optimal forecast always exist and it is unique for a given penalty ratio. Furthermore, we have shown that Minimum Mean Square Error (MMSE) forecast can be obtained as the special case for these class of penalty functions. Once a point-forecast is obtained, we smoothen it through discretized bandwidth units. Since bandwidth is always available in discrete set of modules, such a smoothing operation is rather necessary. In addition to managing data loss and utilization, a secondary issue that inevitably arises in a bandwidth management framework is to control the signaling overhead due to frequent bandwidth updates. We propose a simple approach to address this issue as well in the provisioning stage. In summary, through three different traffic traces, we have found that our approach indeed provides a reasonable bandwidth envelope while honoring the user-level as well as administrative requirements.