Predicting Flow Rate Densities

Tarem Ahmed and Mark Coates
McGill University
{tahmed, coates}@tsp.ece.mcgill.ca

Introduction

- Optical time-division multiplexing (OTDM) allows a wide-area AAPN to effectively share bandwidth. Time-slot reservation at the core switch is an essential component.
- Propagation delays from an edge node to a core node in a wide-area network can be very large, and reservations have to be made several (potentially up to 10) frames in advance of actual traffic arrival times.
- We thus need accurate estimates of demand ahead of time. An approach to estimating demand is to predict Flow Rate Densities.

Definitions

- An IP flow is defined to be a set of packets that share a common key.
- We define the key to be the following 4-tuple:

  key = {source IP address, source port number, destination IP address, destination port number}
- A Flow Rate Density indicates the expected fraction of the total number of flows of every rate.

Motivation

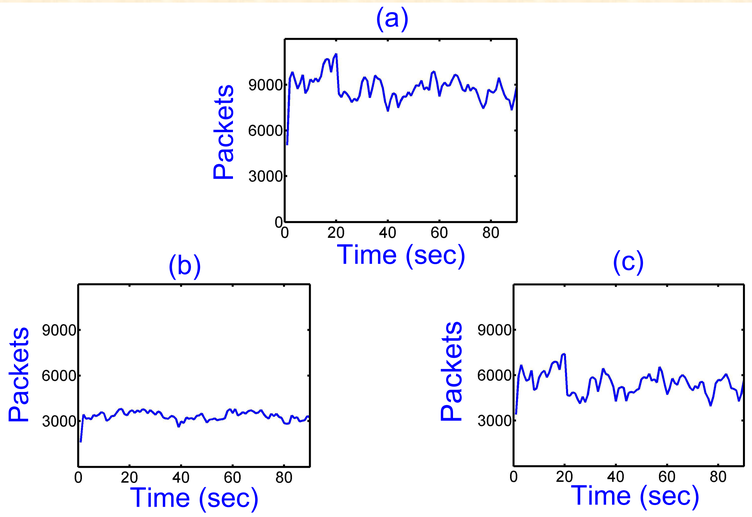

Fig. 1: Decomposition of (a) the total number of packets per second in the BWY packet trace into (b) low-rate flow packets and (c) high-rate flow packets. The BWY trace is collected at an OC3c link at Columbia University. The trace is publicly available at the NLANR repository: http://pma.nlanr.net/.

- We classify a flow as high-rate if it generates more than 50 packets in a one-second time interval. Figure 1 depicts a decomposition of an example data trace into low-rate and high-rate flows.
- Figure 1 clearly indicates that the fluctuation in the total number of packets per second is due primarily to the variation in the high-rate flows. This result (observed in many traces) suggests that the number of packets generated by low-rate flows is reasonably stable and predictable.
- We have used flow rate densities to predict the packets contained in the low-rate flows.

The Model for the Low-Rate Flows

- To represent the low-rate flows, we suggest a mixture of M-1 geometric distributions and a Dirac delta function centered at x=1:

$$y = \alpha_0 .\delta(x-1) + \sum_{m=1}^{M-1} \alpha_m (1-p_m) p_m^x$$

such that

$$\sum_{m=0}^{M-1} \alpha_m = 1$$

where y = Fraction of flows, x = Flow rate

The Prediction Algorithm

The Expectation-Maximization (EM) algorithm:

- The EM algorithm is an iterative algorithm for obtaining the Maximum-Likelihood (ML) estimates of the parameters of an underlying distribution.

The Estimation Step:

- We use the EM algorithm to generate ML estimates of the parameter set for the current 1-second interval:

$$\Theta = \{\alpha_0, \ldots, \alpha_{M-1}, p_1, \ldots, p_{M-1}\}$$

- We use previous interval's parameter values to initialize the algorithm, in order to speed up convergence.
- Convergence is generally seen to occur with 5 iterations.

The Prediction step:

- Then use current interval's estimates for the parameter values, to predict flow rate densities for the next 1-second interval.

Predicting Low-Rate Flows

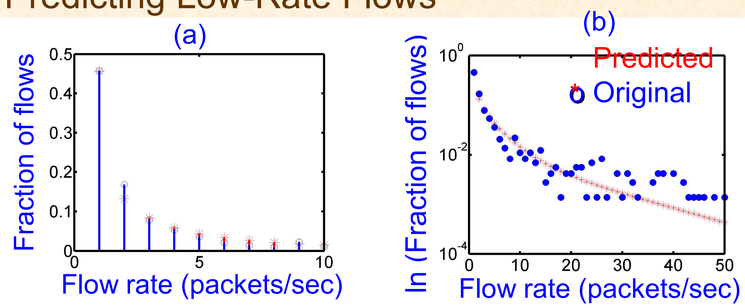

Fig. 2: Overlaid predicted and original flow rate densities for an example 1-second interval in the BWY trace, in (a) linear and (b) logarithmic scales. Here M=4.

- Mean relative error in predicting packets in the low-rate flows over 90 1-second intervals, is only 0.06. Here

  relative error = (*number of predicted packets* - *number of original packets*) / *number of original packets*

Anomalous Behaviour of High-Rate Flows

- Figure 3 shows the number of packets that arrive in consecutive 1-second intervals, in an example high-rate flow. Individual high-rate flows exhibit such substantial fluctuations.
- Only about 10% to 20% of the high-rate flows are persistent for more than 5 to 10 frames.
- Hence it is difficult to predict the packets in the high-rate flows.

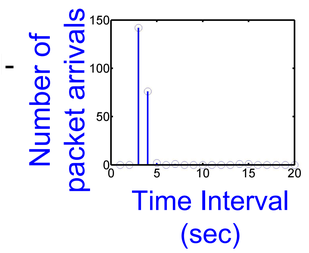

Fig. 3: Variation in the number of packets in an example high-rate flow, over consecutive 1-second intervals.

Future Work

- Predict flow characteristics from sampled statistics, as it is infeasible to record every packet.
- Develop routing algorithm that balances load across the available core nodes.

Acknowledgements

Garrick Ing also contributed to the project.