

A Probabilistic Prediction of Packets on a Diversified IP Network

Christiaan Brand, *Student Member, IEEE*, and Riaan Wolhuter, *Member, IEEE*

Abstract—As more households are given broadband access to the Internet, a huge strain is placed on the underlying IP infrastructure. By being able to predict the packet types that nodes will send, preference can be given to high priority, latency sensitive traffic. Various models characterizing packet flow on a network will be developed and used to predict the packet types to be expected in the immediate future, as accurately as possible. This will later be used as a dynamic QoS criteria for assigning bandwidth to specific hosts.

I. INTRODUCTION

More and more homes are becoming multimedia-equipped. This places a huge strain on the infrastructure of the telecommunications provider to cater for the ever increasing bandwidth demand. This huge demand for broadband computer networks are primarily fueled by the design and deployment of a myriad of online multimedia services.

Internet users no longer only use their systems to check email and browse the web, but rather for real time collaboration and communication using audio and video sources. These applications have different priority classes associated with them and work has been done in the past on reserving a certain amount of resources for a specific application. This is neither a cost effective nor bandwidth optimized approach, and as such, is only offered to institutions that can financially afford these reservations.

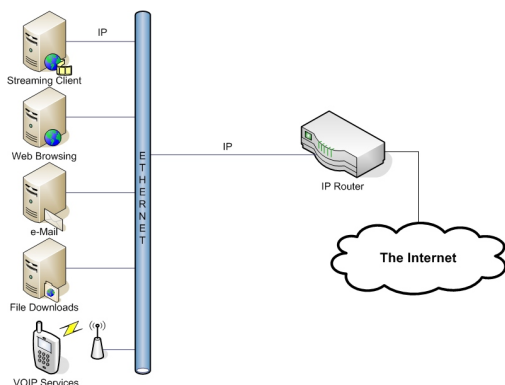


Fig. 1. Typical traffic classes contending for a slot.

Figure 1 shows a typical distribution of packet types on a network at any given time. On most typical IP networks the underlying layer supporting this communication protocol, is Ethernet. Ethernet, being implemented in wired (802.3) or in wireless (802.11) mode, ultimately makes nodes

Christiaan Brand and Dr. Riaan Wolhuter is with the Department of Electrical and Electronic Engineering, Stellenbosch University, South Africa.

contend for a slot in which to transmit. This process is fair, and it is equally likely for any of the nodes with data in their transmit queues to obtain a slot in which to send this data.

Our solution will allow these networks to *favour* a node with more critical realtime data (like VOIP-traffic) above a node that is more latency insensitive (doing a large file download).

We will start our investigation in Section II by developing some models to predict packets types on the network. A brief discussion of the simulation results, using the models developed, will be shown in Section III. This will be followed by Section IV with some conclusions and further research being done in the area.

II. DEVELOPING MODELS

In order to predict the *next* packet that will be sent by a specific node, a model on which to base our decisions is required. In order to obtain such a model, typical traffic flow on a few networks were observed and used to construct the *counting graphs* shown in Figure 2. This graph shows the number of times that a transition occurred from **RTP** to **DNS**. The Y-axis shows the number of times that the transition occurred, and the X-axis shows the number of **RTP** packets received (before the transition to **DNS**).

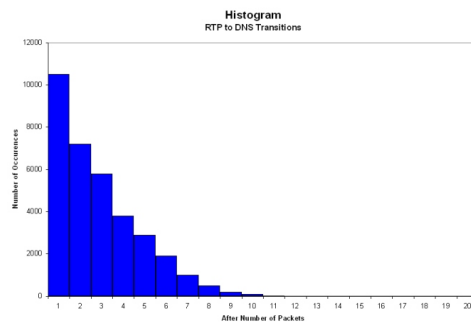


Fig. 2. Counting Graph for RTP to DNS transition.

A. Semi-Markov Model

The first model considered had a specific Semi-Markov property. This simply means that the next state will only depend on the current state, and the amount of time spent in the current state. In our case, the *amount of time* will be the number of packets received while in the current state.

As seen in Figure 2, the distribution of the transitions seems very much exponentially distributed [1]. This is the case for almost all protocols analyzed, and a general equation describing them all can be derived as shown in Equation 1, where β, λ -pairs describe each model approximately.

$$n_{i,l}(x) = \beta_{i,l} e^{-\lambda_{i,l}(x-1)} \quad (1)$$

$i = 1, 2, \dots, I$ and $l = 1, 2, \dots, L$. $I = L$ = the number of protocols analyzed. This means that there is a β, λ -pair for each possible protocol transition.

Evaluating each of these functions for a specific value of x (the number of packets received), the function that yields the highest value is the one most likely to be changed into. It is also possible that a transition to a new protocol **will not** occur which means that we will receive yet another packet of the same type as shown in Figure 3. The chance of no transition occurring, is given by Equation 2.

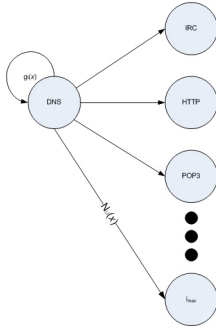


Fig. 3. State diagram of transitions.

$$g_i(x) = \sum_{l=1}^L \sum_{j=x+1}^{\Delta < 0.1} n_{i,l}(j) \quad (2)$$

Δ is given as $n_{i,l}(j)$.

Least Squares Estimation [2] is also used to continually adjust the β, λ -pairs for individual hosts while this model is operating.

B. Highest Average Model

This is a more qualitative approach to a prediction model. Once a certain traffic type is encountered from a node, more of that same type is very likely to follow. Full packets (very close to the MTU) also tends to indicate a continuous stream of data from a node meaning that the next packet from that node will most likely be of the same type as the current one.

Protocols, such as RTP (for Voice over IP communications), also sends a stream of the same sized packets. Attempting to detect and learn this *magic size* for certain protocols is also included in this model.

Incoming TCP data (PUSH) from the Internet to a node will almost certainly invoke (at least) an ACK packet from the specific node. This inbound packet's protocol can then be predicted as the next protocol that node is likely to utilize.

Lastly, averaging over the last 10 received packets from a node, the type occurring the most will be predicted as the most likely type to occur next.

III. SIMULATIONS

Simulations using the Highest Average Model (discussed in Subsection II-B) yields slightly higher results than the Semi-Markov model. A simulation was done examining headers collected on an Internet service provider's router. 13 834 936 packets were examined, 6 630 391 originating from within the network and the remaining 7 204 545 inbound from the Internet. Average prediction accuracy over the whole time period were found to be 89,39%. Figure 4 shows the accuracy of the predictor over the full time period. Initial simulations indicate very promising possibilities for the chosen approach and will be further refined as the strategy is developed.

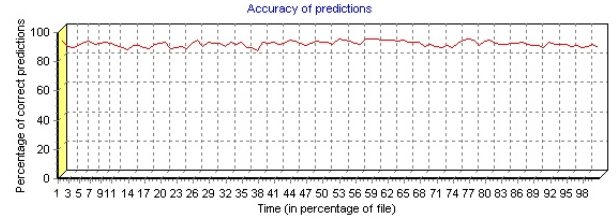


Fig. 4. Prediction Accuracy Results.

IV. CONCLUSIONS AND FUTURE WORK

The global village is a reality, with its highways becoming more and more congested. To counter this surge of users, faster connection methods are being employed. Intelligent channel management - by predicting traffic flows - is a much more economical and elegant alternative to achieve the same ultimate goal: User satisfaction and optimal use of bandwidth.

Porting of the prediction algorithms to take the packets directly from **libpcap** written in **perl** is currently underway. Attempting to combine the Semi-Markov with the Highest Average model to obtain more accurate predictions is the next step, whereafter a software router (Fedora Core Linux with IP Forwarding) will be modified as to increase traffic slots for users with more demanding traffic. Simulations and experiments on this platform should show an increase in user experience, without any additional bandwidth costs.

REFERENCES

- [1] P. Z. Peebles, *Probability, random variables, and random signal principles*, 2nd ed. New York: McGraw-Hill, 1987.
- [2] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, 1st ed. New Jersey: Prentice-Hall, 1974.

Christiaan Brand received his MSc.Eng and B.Eng and is currently working on his PhD at the University of Stellenbosch. His research interests include wired- and wireless digital telecommunication systems and IP networks.