Network performance modeling and control through prediction feedback

Manjunath R
Philips innovation Campus
Bangalore

Shyam vasudev
Philips innovation Campus
Bangalore

Narendranath Udupa
Philips innovation Campus
Bangalore

ABSTRACT
Performance modeling of a network is challenging especially when it involves multimedia traffic. The present day home network makes uses of internet for multimedia content transfer on to the devices in real time. In such a communication system, reliability and in time data transfer is critical. The system has to support the streaming of multimedia and entertainment data from mobile to infrastructure and vice versa. In this paper, a novel modeling method for the network and its traffic shaping is introduced and simulation model is provided. The performance with this model is analyzed.

Categories and Subject Descriptors
I.5.1, G.0

General Terms
Algorithms, Performance

Keywords
Adaptive control, Differential feedback, Multiresolution, Traffic shaping

1. INTRODUCTION
The present day home network predominantly supports multimedia traffic over the wireless medium. In a wireless network, the reflections and multi path result in increased self similarity [1] for the signal at the receiving end. The self similarity of the network refers to the invariance of the shape of the autocorrelation function when observed over multiple time scales. The self similarity imparts long range dependency in to the traffic. As a result of long range dependency, the traffic turns bursty resulting in under or over utilization of the resources, increased cell loss etc.

In order to reduce the loss of multimedia data over the wireless medium, it is required to pump in less data by controlling the degree of compression (and improving channel coding) when the channel is more noisy. To make it possible, a feedback on the channel status in terms of % loss of the data packets over the channel is required to be transferred to the source. Based on this input, a decision has to be taken on the data transmission rate.

For a perfect synergy between the feed forward (FF) and the feedback (FB) path, the properties of forward path that impart aberrations to the signal have to be annulled by generating appropriate signals in the feedback path. Ideally, the controller generating the feedback signal should have the same characteristics of the network. i.e, it should be a network in its own sense.

Meeting stringent constraints on the delay is very important in the networks supporting multimedia traffic for seamless user experience. In addition to the delay and packet losses, the variable delay suffered by the packets i.e. jitter is to be given due weightage. Analysis shows that the jitter degrades the perceptual quality as much as the packet loss.

To support the QoS in the Internet, the IETF has defined two architectures, the Integrated Services (Intsev) and Differentiated Services (Diffserv). They have important differences in both the service definition and the implementation architectures. At the service definition level, the Intserv provides end-to-end guarantees or controlled load service on a per flow basis, while the Diffserv provides a coarser level of service differentiation among a small number of traffic classes.

In an Intserv each router processes per flow signaling messages and maintain per flow transactions including data forwarding and QoS state on the control path, perform per flow classification, scheduling and buffer management on the data path. Performing the per-flow management inside the network affects both the network scalability and the robustness. On the other hand, Diffserv distinguishes between edge and core routers.

The IntServ although provides the agreed QoS guarantees, it results in under utilization of the resources. The scheme poses scalability issues. The Diffserv needs an active resource management mechanism to provide the services with an assurance level of service comparable to those provided by the Intserv. Here a dropper controls the dynamics of the packet forwarding. Thus, the effective design of the dropping mechanism is important in congestion avoidance and rendering a rich class of services

In a home network scenario, there will be limited support to overcome congestion. The best way to fix this issue is to avoid the congestion to happen. So, a proactive control algorithm is required than an active one [2]. A good controller has to foresee the trends in the network traffic variations and provide inputs to the traffic source well in advance. The source would get sufficient time to adjust the traffic rate or provide sufficient redundancies with the appropriate channel coding schemes so that it would not flood the channel when it is disturbed. In the simulation model, the proactive queue management model GREEN [3] together with a neural network is considered.

The simulation study shows that the proposed method can maintain an average queue length and its variations small irrespective of the congestion control methods in end systems. Therefore the proposed method is best suit for real-time multimedia services which have stringent requirements on the delay and jitter. It is verified that the proposed active queue management scheme does not have bias against the bursty flows although it can protect well-behaving flows.

The proposed scheme is scalable because it can provide QoS mechanism without maintaining the per flow states. It is legacy because it can be easily deployed in the current routers by adding predictors and shifters to the existing software.

The neural network model that satisfies the aforementioned requirements of the controller is provided in section 2. The properties of the network are reflected in to this model. Model of the controller provides a mechanism to control the behavior of the network. Performance improvement as a result of this controller model is provided in section 3. Section 4 & 5 provides impact of the shift. Section 6 consolidates the results.

2. THE MODEL

A differentially fed neural network (DANN) [4] has all the characteristics required for a controller mentioned in section I. It is interesting to note that, a neural network exhibits these properties only when differential feedback of different orders is provided from the output to the input of the network. The resulting structure would have its output self similar, long range dependant (on the historical data), autoregressive with its spectrum obeying the power law and the output forming a space of Bayesian estimators for different orders of differential feedback [5] etc.

If the network is congested, data packets cannot be transmitted. So FB packets need to be slowed down. However, if the network is free, data packets can be easily transmitted calling for faster FB packet transmission. Here the sources maintain their transmission rate (which otherwise gets slowed down). The constraint is intermediate storage. Therefore, if transmitted, both are transmitted. Else, none are transmitted. Generally the feed back traffic is very small compared to the feed forward traffic. At any point of time, some weightage has to be given for the transmission of FF and FB packets. It depends up on available resources, transmission rate queue and the number of FB packets in the queue etc.

The feedback packets may be intercepted and processed by the intermediate switches to control the rate of transmission.

A small change in the traffic over a small time scale will have a profound impact on the traffic envelop over a longer time. With this idea, a traffic controller has been tried in the feedback path over feedback packets.

The rate of transmission depends up on the resource allotted. But the allotment of resources itself depends up on the transmission rate. This criss-cross dependency is as though two ANNs (or models) connected back to back as shown in figure 1.It reduces to DANN as shown in figure 2.

The dependency is a result of pileup due to previous transmission rates and previous allotments. This architecture implies a differential (delayed) feedback from the output. Thus the proposed architecture better catches this type of dependencies. The feedback on the status of the network resources controls the transmission rate. A shifted feedback has a strong influence on the resources as well as the characteristics of the network.

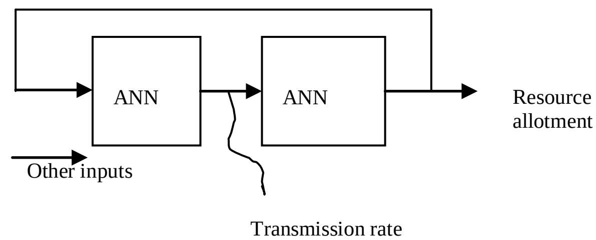

Figure 1. Resource allotment

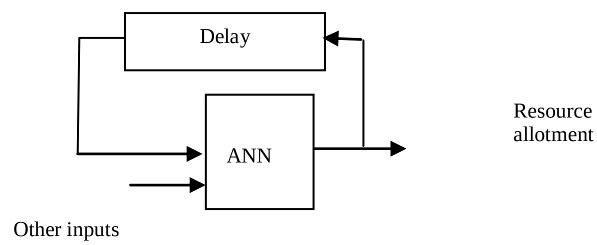

Figure 2. Simplified DANN architecture

3. PERFORMANCE WITH THE MODEL

To generate self-similar input traffic, the superposition of the ON/OFF sources with heavy-tailed distribution is used. In this simulation study, we verify that the proposed active queue management scheme can keep an average queue length and the variation of queue length small when the input traffic is generated by self similar traffic sources. Figure 3 shows the simulation network. Here both the network and the control model may be seen. The DANN gets the training data from the background GREEN algorithm. GREEN is taken as the controller for comparison. For some time, the DANN will be in learning phase. Then it predicts the data k steps in advance. This is given as the feedback to the source [6]. The source then re-computes the transmission rate. It may be seen that the cell loss ratio has been reduced with feedback. In each case, 42 data points computed with GREEN are used for training. The Input consists of 20 sources supporting FTP that exist over the entire simulation time. The maximum buffer size is taken as 8000 with the packet size of 512.

The total cell loss ratio of an ordinary GREEN scheme is found to be 7.4%. With a neural network prediction, it has been reduced to 6% and with first order differentially fed neural network, it is reduced further to 0.55%. A 6 step prediction has been used in the experiment.

Simulation time is set to 40 sec and 180 samples are taken. Matlab version 6 and Simulink have been used to carryout the simulation. Table 1 shows the loss rate reduction with increase in shift.

Table 1 Performance with the proposed method running over GREEN

Sl.No.	Shift	Loss. probability	Queue variance
1	0	7.4%	136.8258
2	1	6%	133.3290
3	2	5%	133.9068
4	4	0.725%	135.5478
5	6	0.55%	136.1364

4. IMPACT OF SHIFT GIVEN TO FEEDBACK SIGNAL ON QOS PARAMATERS

In a general scheduling algorithm the service rate allocations of different classes dynamically get adjusted to meet the proportional guarantees. The service rate allocation is based on the backlog of the classes at the scheduler. If there are two classes with backlogs $B_1(t)$ and $B_2(t)$ at link of capacity C, the service rate of

$$r_1(t) = \frac{B_1(t)}{B_1(t) + \alpha B_2(t)} C \qquad (1)$$

is assigned to the first class. The value $0<\alpha<1$ is the proportional differentiation factor. It represents the ratio of priorities (delay, packet loss) of the first class over the second class. In general the service rate of class i is

$$r_i(t) = \frac{B_i(t)}{\sum_j \frac{s_j}{s_i} B_j(t)} C \qquad (2)$$

i.e. the available capacity is proportionally shared. This implies that the rates are proportional to the queue lengths. Such a scheme results in more absolute delays though the relative delay constraints are satisfied. This calls for a reduced Q length. With this change, the absolute delay constraints are easily met in additional to the relative delay constraints if the queue lengths are reduced. It will be shown here that the shifted or predicted feedback signal actually reduces the queue size.

The settling time of delay is reduced with feedback i.e. the time taken to stabilize the ratios of delays get reduced. This is because the Queue flush time reduces as the Queue size reduces with shift. The average delay also gets reduced. With higher orders of the feedback or the larger shifts, the constraint on the absolute delay is more easily satisfied as the instantaneous delay falls off with shifts.

The shift also helps in reducing the absolute delay of the packets. The absolute delay constraint (ADC) is the sum of service delay and the stranding time in the queue. With shift given to feedback signal, the queue length gets reduced while the input rate increases. Hence, the ADC is met easily with shifted feedback signal.

5. ACTIVE QUEUE MANAGEMENT METHOD FOR QOS PROVISIONING

In the proposed scheme, an average input traffic rate for the next interval is predicted at a large time scale based on the history of

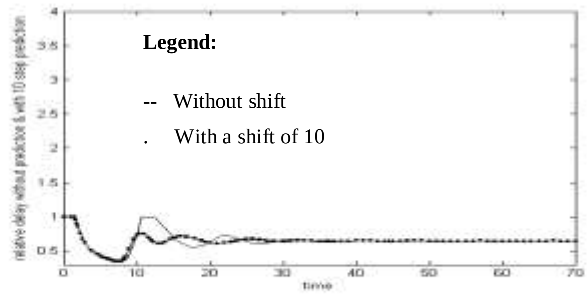

Figure 3. Relative delay for shift =10

the traffic activity. The expected average input traffic rate is used to detect the congestion. If the congestion is detected, the primary packet marking probability is calculated considering the predicted average input rate and the link capacity. By doing so, the proposed scheme can react fast to the large variation in the input traffic level.

Since the congestion is detected at a time scale longer than the round trip time (RTT) of a flow, sources can react to the congestion signal before the queue size exceeds a certain threshold value. Therefore, compared to the other schemes of the congestion control, the propose scheme does not need a large buffer space. This means that average delay and the delay jitter can be reduced when the REP is used.

Because the proposed scheme detects the congestion at a large time scale, senders can react to the network conditions more rapidly. They can receive the congestion signal and reduce their sending rates before the network goes into severe congestion. With this, the aggregated input rate at a congested router will be below the service rate. So, the amount of buffer space required accommodating an applied load greater than the link capacity gets reduced. It also reduces the end-to-end delay and the delay jitter.

The proposed active queue management scheme is superior to RED in that it can maintain the average queue length and the variation of the average queue length small. In general, as the RTT decreases, the response time of the sources to the network congestion becomes short. Therefore, the amplitude of the queue length variation gets reduced.

If the gateway uses the active queue management scheme proposed here, the average queue length and the variance of the average queue length increases as the RTT increases. But the amount of increment is negligible. This is because the congestion is detected by using the estimated input rate at a larger timescale.

The sources can detect the network congestion before it really happens. Because sources reduce their sending rate before network goes into congestion, the average queue length fluctuations are minimized. Therefore, the amount of fluctuation of the variance of the average queue length is much smaller when the gateway uses the proposed scheme than when the gateway uses the RED.

The longer the RTT of a flow or the smaller the window size of a flow is, the burstier the flow becomes. When a gateway uses random drop queue, packet drop occurs due to the packet overflow. Therefore, the packets of flows with small RTT and large window size are more likely to occupy the buffer space. So, the probability of packet drop for a bursty flow becomes higher compared to a non-bursty flow.

Figure 3 shows the effect of shifts on the relative delay. From the simulation, it is evident that the shifts help in attaining the relative delay ratio faster. The simulation has been carried out on MATLAB version 6. For the simulation, the capacity C is set to 1Gbps, Buffer size of 8000 packets is considered

6. CONCLUSION

Modeling and controlling the multimedia traffic is challenging. In this paper, the characteristics of a typical home network are listed and related to the properties of a neural network. The neural network is then inserted in to the network in the feedback path to indicate the congestion status and adjust the transmission rate of the content sources. Presence of the controller in the feedback path ultimately leaves a well behaved network.

7. ACKNOWLEDGMENTS
Our thanks to Philips Innovation Campus, Bangalore for the support rendered to publish this work..

8. REFERENCES

[1] W.Willinger, M.S.Taqqu, R.Sherman and D.V.Wilson, 1997."Self-similarity through high Variability: Statistical analysis of Ethernet LAN traffic at the source level", IEEE/ACM transaction on networking, Vol 5, pp71-86.

[2] C. Hollot, V. Misra, D. Towsley and W. Gong, 2001. A Control Theoretic Analysis of RED, INFOCOMM.

[3] Wu-chun Feng; Kapadia, A.; Thulasidasan, S. GREEN: proactive queue management over a best-effort network, IEEE GLOBECOM, pp 1774 - 1778

[4] Manjunath.R, Gurumurthy.K.S.,2002. Information geometry of differentially fed artificial neural networks, TENCON'02, Vol 3, PP 1521 – 1525

[5] Manjunath.R, K.S.Gurumurthy,2003. Bayesian decision on differentially fed hyper planes, ITSim2003

[6] Manjunath.R, K.S.Gurumurthy, 2004. Maintaining Long-range dependency of traffic in a network, CODEC'04