Model Based Predictive Rate Controller for Video Streaming

Taner Arsan

Kadir Has University, Engineering Faculty, Computer Engineering Department,
Cibali Kampusu, 34230 Cibali-Istanbul, Turkey
arsan@khas.edu.tr

Abstract

Due to the heterogeneous structure of networks, video streaming over lossy IP networks is very important issues. Infrastructure of the Internet exhibits variable bandwidths, delays, congestions and time-varying packet losses. Video streaming applications should not only have a good end-to-end transport performance but also have a robust rate control, because of variable attributes of the Internet. Furthermore video streaming applications should have a multipath rate allocation mechanism. So for providing the video streaming service quality, some other components such as Bandwidth Estimation and Rate Controller should be taken into consideration. This paper gives an overview of video streaming concept and bandwidth estimation tools and then introduces special architectures for bandwidth adaptive video streaming. A bandwidth estimation algorithm-pathChirp, Optimized Rate Controllers and Multipath Rate Allocation Algorithm are considered as all-in-one solution for video streaming problem. This solution is directed and optimized by a decision center which is designed for obtaining the maximum quality at the receiving side.

1. Introduction

In general, users want maximum speed and minimum delay even using a minimum bandwidth required network applications. In real-time applications such as video streaming, TV broadcasting or video-conferencing, the service quality must be maximized for satisfying the users. Unfortunately, today's technologies cannot provide the best solutions for video streaming users that use best effort delivery networks. Although network devices try to find a best path to the destination, there will be a better path available in the Internet. In 30-80% of the cases, the best paths found by classic routing algorithms are suboptimal from a media perspective [1, 2].

We try to find the most convenient protocol for video streaming. It is not possible to use lower-layer transport layer protocols in video streaming, because TCP has huge amount of delay while retransmitting lost packets, and UDP does not guarantee packet delivery and packet loss recovery. An upper-layer transport protocol will be necessary. In this point, Real-time Transport Protocol-RTP and Real Time Control Protocol-RTCP protocols may be helpful, RTP is designed to provide end-to-end transportation and to support real-time applications, but RTP does not guarantee the best video streaming performance. RTCP, part of the RTP specification, provides QoS feedback and monitors the QoS. RTCP is not intended to support all of an application's control signaling requirements. Still one of the session control protocols or a convenient control mechanism, may therefore be needed. Real Time Streaming Protocol-RTSP is a session control protocol and its aim is to provide video streaming over the Internet. All these algorithms try to find a best solution for routing the packets, and best effort delivery, but none of them guarantees video streaming or quality of service routing. There should be a mechanism to help regulating video streaming packets before or after network devices' routing for considering quality of service routing, and maximizing the video quality at receiving side.

Video streaming system has some constraints such as bandwidth, congestion and delay. The Internet has also a heterogeneous structure and performance measurement is not considered as the primary goal of its architecture. To obtain a better video streaming system, end-to-end transport performance should be optimized. It is possible to understand the network characteristics by using a bandwidth estimator. This makes packet forwarding more effective and evaluates the performance of video streaming via the Internet. Some bandwidth estimation tools are for networking solutions, but we need to consider bandwidth-estimation as video streaming point of view.

Rate controller has an important role in video streaming. If the aim is to maximize the quality at the receiving side, streaming rate should be well regulated and should be optimized. If we consider multipath scenario, it is important to determine which paths should ideally be used for the streaming application, in a given available network topology between a streaming server and a client. This is different from classic routing algorithms that find the best path given some established network metrics. Classic routing algorithms may be optimal in terms of network utilization, but it is certainly suboptimal from the point of view of the media streaming application.

Upper-layer transport protocol usage, bandwidth estimation tools and rate controller are taken into consideration separately as a solution for effective video streaming in the literature, but there is no joint solution on bandwidth estimation and rate control available and only a few control engineering oriented solutions available in literature such as PID or fuzzy controller.

In this paper, an overview for video streaming concept is given and special architectures for bandwidth adaptive video streaming are introduced as a total solution. While Bandwidth Adaptive Streaming Controller uses bandwidth estimation, adaptive rate controllers, and multipath rate allocation algorithm, QoS controller calculates delay and packet loss values. Decision Center optimizes rate-delay-packet loss oriented cost function and determines optimal rate and optimal path. The rest of this paper is organized as follows. In the following section, video streaming overview is provided. We give a brief information on bandwidth estimation tools in Section 3. We introduce complete system architecture of bandwidth adaptive video streaming system in Section 4 and we give the details of software information model of proposed system in Section 5. We conclude the paper with research contributions in Section 6.

2. Video Streaming Overview

Streaming is very popular in today's best effort delivery networks. Delivering audio or video content over the Internet can be achieved by two methods: progressive download and real-time streaming. If the content size is short, the progressive download method is used in general. In this method, media content is directly downloaded from a server into storage units of a client, but in real-time streaming, client software plays media content without storing the content into any storage units. Real-time streaming can be easily explained as delivering media from a server to a client over a network in real-time. The client software is responsible for playing the media as it is delivered. There are two main types of delivery options for real-time streaming: live and on-demand. If the media content contains live events, this type of streaming is called live streaming. On the other hand, if the media contents can be provided on user's demand, it is called on-demand streaming.

2.1. Video Streaming Architectures

There are several common media streaming architectures available in the commercial use. These architectures are Single Sender - Single Receiver Streaming System, Single Sender - Multiple Receivers Streaming System, and Multiple Senders - Single Receivers Streaming System.

In general, streaming technologies support the latest digital media standards such as AAC, H.264, MP3 and MPEG-4. Client software uses a protocol rather than TCP/IP such as RTP or RTSP, because of its long and bandwidth greedy acknowledgement procedure. First, the client sends a RTP or RTSP request to server and server receives the request and calls necessary software modules to satisfy the request and to start the streaming.

2.2. Typical Internal Structure of Video Streaming

Typical video streaming system includes broadcaster, repository for archiving, streaming server and a receiver unit or a client as seen Fig.1.
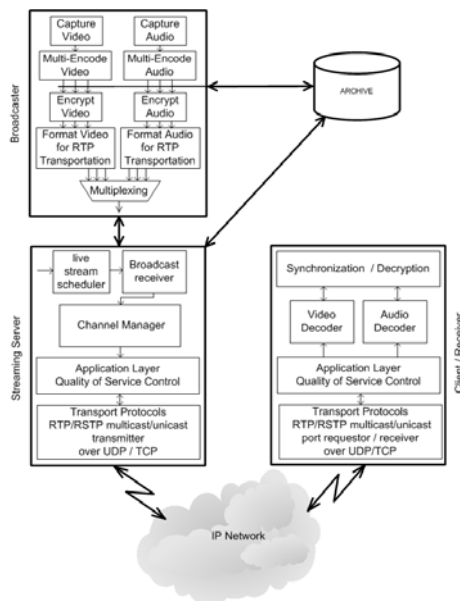

Fig.1. Internal Structure of Video Streaming

The main idea is to capture video and audio signals from the source of the raw video and audio and put these media content into correct form to transmit on IP network. Raw video and audio must be compressed before transmission on IP network. This makes data transmission effective, but it is also noted that transmission of the packets should be under control of the quality of service control. Application layer QoS control tries to achieve this task. It includes congestion control and error control techniques for successful transmission of the packets on IP networks. Congestion control is used for preventing packet loss and reducing the delay. Error control is used for improving video presentation quality, if the packet loss is occurred. Forward Error Correction (FEC), retransmission, error-resilient encoding and error concealment are some examples of error control mechanism. To implement the some methods, continuous media distribution services are necessary. In the Internet, IP protocol can successfully try to achieve this task by means of its best-effort delivery structure.

In video streaming technology, streaming servers have one of the most important roles. Streaming servers offer quality streaming services, and contain transport protocols, operating system and storage system. Although, quality streaming can be successfully achieved by the streaming servers, there will be a media synchronization mechanism to put together the video and audio. Otherwise, video and audio can not be synchronized at receiver's media playing tools. Finally, some protocols are needed to regulate the transmission of the packet by using one of the lower-layer transport protocols such as UDP or TCP, or a higher-level transport protocol such as RTP or RTCP or a session layer protocol such as RTSP.

3. Bandwidth Estimation

Bandwidth is one of the most important factors in network engineering. Streaming video and audio over a network requires huge amounts of bandwidth. The networking professional should find an excellent solution for developing technologies. Network engineers use bandwidth estimation tools for analyzing network performance, optimizing end-to-end transport performance and network devices' abilities. Application developers such as video streaming technology developers are also curious about bandwidth estimation concept and methodology. Different types of bandwidth estimation tools are developed so far. There are two major end-to-end measurement methods available as Active Methods or Passive Methods.

In Active Methods, the idea is to send dummy packets from sender to receiver and try to understanding the network characteristics by the help of transmitted packets [3-5]. Although, Internet Control Message Protocol – ICMP based traditional tools are available such as *ping* and *traceroute*, but it is not possible to obtain satisfactory results by using these tools. There are no details whether the packet was lost or response was lost in ICMP-based tools, because packet loss rates on the forward and reverse path are different from each other.

There is no injected packet in Passive methods, while dummy packets are sent to receiver from sender in active methods. The idea is to observe traffic already present in the network and then estimate the bandwidth of the network. This technique is called non-intrusive measurement technique, and it is used for detecting shared upstream congestion and discovering bottleneck (significant queuing) router link speeds [4].

4. Software Design of an Integrated Architecture

Rate control concept is basically related to application layer Quality of Service control techniques which try to manipulate congestion and error. If we consider video streaming, congestion control takes the form of rate control. It is known that packet loss, delays, and network congestion may affect on video presentation quality. It is necessary to reduce the packet loss and delay by using congestion control techniques, so rate control techniques. The idea is to maximize the quality at receiver side. Especially in lossy IP networks, rate control techniques should not only meet the bandwidth requirements but also assures other quality of service metrics such as delay, delay-jitter and cost.

In this paper, as shown in Fig.2, a special architecture for bandwidth adaptive video streaming is implemented by using two different controllers: Bandwidth Adaptive Streaming (BAS) controller and Quality of Service (QoS) controller. The outputs of these two controllers are collected at a Decision Center. Decision Center is the heart of rate optimization and allocation by considering bandwidth, delay and packet loss metrics.
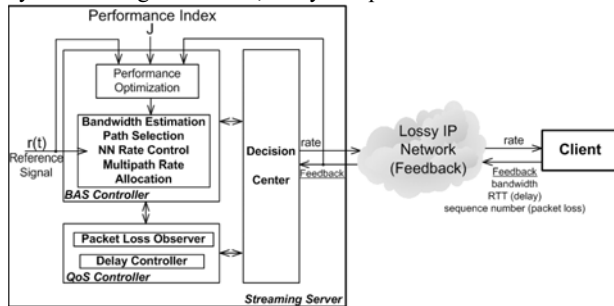

Fig.2. Bandwidth Adaptive Video Streaming Architecture in Video Streaming Topology

Firstly, we set the reference signal for convenient value, and then we run Bandwidth Estimation tools called *pathchirp*. *pathchirp* is the most important agent in this system, because it determines available bandwidth and regulates reference signal. On the other hand, *pathchirp* helps convenient path selection. Nueral Network Rate Controller makes this procedure more intelligent. It has a performance index function to optimize the performance.

Secondly, we try to guarantee the Quality of Service at the receiving side. We use the feedback signal of RTP called RTCP. RTCP includes the values for determining the values of packet loss rate and round-trip propagation delay. We have obtained these two values by the help of Packet Loss Observer and Delay Controller.

4.1. Bandwidth Adaptive Streaming (BAS) Controller

BAS controller is responsible for end-to-end bandwidth estimation, adaptive rate control and multipath rate allocation. This controller also selects the best path to destination by using extended version of bandwidth estimation algorithm for effective video streaming. In order to control the rate and to assure the service quality, feedback of related metrics should be provided. BAS controller needs bandwidth estimation value to determine the path and to regulate and to allocate the rate.

4.1.1. Bandwidth Estimation and Path Selection

In BAS Controller, *pathChirp* algorithm which is one of the best effective publicly available active bandwidth estimation tools is considered as a bandwidth estimator [5]. Original *pathChirp* algorithm estimates the best available network connections bandwidth. It does not however, search any better bandwidth in the entire topology. BAS controller first triggers *pathChirp* algorithm to search and estimate the bandwidth of all available networks. After determining and selecting the best path in the current topology, the bandwidth estimation value is provided to Model-based Predictive Rate Controller which tries to regulate the rate of both live stream scheduler, adaptive encoder and the network.

It is important to find a solution for the problem of streaming path allocation in a multi-path network, and it is almost perfect to take media aware metrics into account during the decision process. There were some previous works [4] about path selection methodology based on experimental data. These are some rules that consider network metrics such as available bandwidth, loss rate and hop distance, and media aware metrics such as link connections and video distortion. It will be convenient to consider media-specific metrics to provide a more general framework for the analysis of joint path selection and rate allocation.

4.1.2. Rate Control and Rate Allocation

It is possible to use different type of controllers as rate controller. We consider Model-Based Predictive Control as rate controller which is based on control engineering theory and implementation.

4.1.2.1. Model-based Predictive Rate Controller

Model-based Predictive Control (MBPC) algorithm is a feedback-control methodology suitable to enforce efficently hard constraints on the variables of the controlled system. It is shown that the method hinges upon a constrained open-loop optimal control problem along with the adoption of the so-called receding-horizon control strategy.

MBPC algorithms can be devised with the property of ensuring global feasibility/stability. Considerations on how to deal with disturbances and model uncertainties are also given.

MBPC algorithm may provide a relatively better solution for rate control. NNs become the focus of great attention, because of its capacity to treat complex and nonlinear problems. The Internet has also a heterogeneous structure. It is therefore possible to adapt NN approach to video streaming rate control. Backpropagation algorithm is mostly used for learning about the network topology and it can easily be applied as

1. Initialize weights to small random values.
2. Propagate the input signal forward throughout the network.
3. Compute the sensitivity function for each weight of the network.
4. Update weights.
5. Go to step 2 and repeat the procedure until end criteria is reached.

Predictive control was established as a powerful technique, it became a potential for applications [6]. The idea is based on input and output past information, and then future control actions is predicted along a moving prediction horizon. The algorithm includes the following steps:

1. Predict the system output in period of time via Taylor series,

2. Obtain control signal series by minimizing the error between system output and future system output by considering constraints,

3.   Select the first element in the series as a control signal and repeat same procedure in next future period of time,

Neural Network Predictive controller is a joint solution of backpropagation algorithm and moving horizon predictive control algorithm. The general structure of NN Predictive Controller is given in Fig.3.
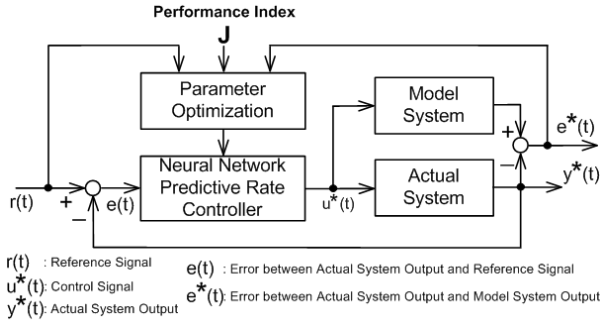

Fig.3. Neural Network Predictive Rate Controller Principle

Finding the optimal rate allocation in multi-path networks is in general a complete problem in generic scenarios. However, we show that a careful analysis of the video distortion evolution allows deriving a linear complexity algorithm for the joint optimal path selection, and flow rate allocation under common network assumptions. In other words, our main objective is to jointly find:

(i)   The optimal encoding or streaming rate of a video stream so that the quality at receiver is maximized,

(ii)   Which network paths should be used for relaying the video stream to the client?

Interestingly enough, our conclusions demonstrate that the answer to these two questions is represented by a careful tradeoff among available network bandwidth (translated into video encoding rate), transmission loss process, and number of utilized paths. And, contrary to the commonly admitted opinion, flooding the network in using all the possible paths rarely provides an efficient strategy.

4.2. QoS Controller: Packet Loss Observer and Delay Controller

QoS controller tries to guarantee the service quality at the receiving side by considering RTT for calculating the delay and sequence number for observing packet loss. QoS controller needs delay and packet loss values to assure QoS at receiving client. Packet loss observation and delay control can be handled within the framework of RTCP protocol.

In the proposed structure, QoS controller uses the feedback values of Round Trip Time-RTT and sequence number which are the feedback metrics of upper-level transport layer protocols *RTP/RTCP*. RTP and RTCP protocols are used for video streaming and QoS controller calculates delay by using RTT and packet loss by using sequence number and tries to make the video streaming much more efficient and makes the video quality reasonable at the receiver side for the available conditions. Both outputs coming from BAS Controller and the QoS Controller are sent to Decision Center. Decision Center runs an optimization algorithm to calculate the best rate to send to adaptive scheduler and adaptive encoder as seen in Fig.4.

We have used RTP and RTCP protocols that are described by RFC 3550 and 4586 for transmitting the video streaming signals. *RTP provides end-to-end delivery services for data with real-time characteristics, such as interactive audio and video. RTP header contains timing information and a sequence number that allow the receivers to reconstruct the timing produced by the source. The sequence number can also be used by the receiver to estimate how many packets are being lost.* it is possible to calculate both packet loss ratio and round-trip propagation delay according to RFC 2550.
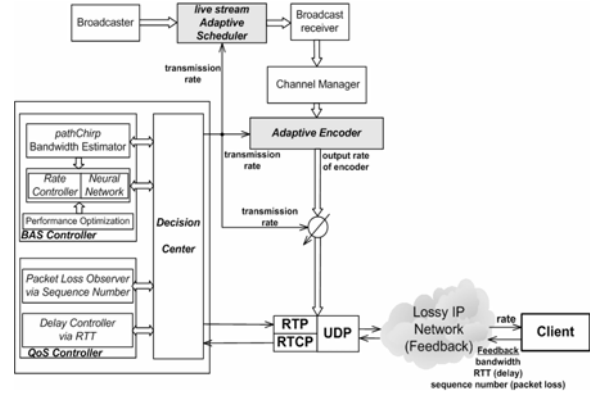

Fig.4. BAS Controller, QoS Controller and Decision Center with Encoder and Transmitter

4.3. Decision Center and Optimal Rate Determination

Decision Center is the heart of this architecture, because all controller information is gathering here. It should therefore have a mechanism to determine optimal rate by considering Bandwidth Adaptive Streaming Controller output (rate ratio) and QoS Controller outputs (delay and packet loss). Decision center also considers a performance index function. The constraints are also rate ratio, delay and packet loss. If other metrics are necessary such as jitter and congestion, it is also possible to calculate such these values by RTP/RTCP sender-receivers reports.

In this work, the aim is to minimize transmission cost while limiting delay and packet loss. The transmission packet loss (L), delay (D) and rate (R) are constraints for transmission cost function. We have to consider encoder (E) parameters and network (N) parameters. If performance index is characterized by one of these three costs,

$$J = \int_0^\infty |e(t)| dt \quad , \quad J = \int_0^\infty e^2(t) dt \quad or \quad J = \int_0^\infty te(t) dt$$

We try to obtain the minimized cost function as

$$\min J(E, N)$$

We have the constraints as

$$L(E, N) \le L_0 \quad Packet\ Loss$$
$$D(E, N) \le D_0 \quad Delay$$

considering Lagrange multipliers as $J(E, N) + aL(E, N) + bD(E, N)$ where a and b weighting parameters, and try to find a solution for this performance index by Lagrange Optimization Method.

5. A Software Information Model of Bandwidth Adaptive Video Streaming

A software information model of bandwidth adaptive video streaming system is given in Fig.5. This model includes Broadcaster Component, Streaming Server Component and Clients Component. All the methods regarding to component functionalities are shown inside each component. All components run a middleware bus software which enables transparent distributed communication between them. Now it is time to explain these components one by one.
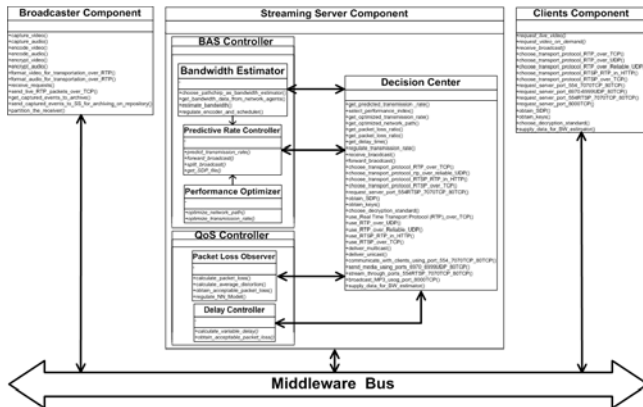

Fig.5. Software Information Model of a Bandwidth Adaptive Video Streaming

Broadcaster Component captures the video and audio signals separately, and then encodes and encrypts these signals in convenient forms. At this point, data is ready to transmit to a streaming server over convenient transport algorithm such as RTP over UDP or TCP. If needed, this data can be stored at the repository for archiving. This repository can be jointly used by a broadcaster and/or streaming server components.

The Streaming Server Component has the most complex, and at the same time an integrated structure. This component includes BAS-Bandwidth Adaptive Streaming Controller, QoS-Quality of Service Controller and a Decision Center subcomponents. BAS includes Bandwidth Estimator, Predictive Rate Controller and Performance Optimizer components. Bandwidth Estimator object is used for determining available bandwidth by means of a bandwidth estimation algorithm called *pathchirp* [5]. It also regulates encoder and scheduler for the received video information. Predictive Rate Controller object predicts transmission rate and forwards broadcasts. Performance Optimizer object tries to find the best network path to destination. It also maximizes transmission rate to Client Components. QoS Controller subcomponent contains two important objects, Packet Loss Observer and Delay Controller. Using an RTP feedback protocol RTCP with its frame fields, Packet Loss Observer object calculates packet loss ratio and average distortion. Delay Controller object calculates variable packet delay and acceptable packet loss. Decision Center of Streaming Server Component is the heart of this software information model which has many tasks to control and regulate the streaming. Decision Center receives BAS Controller and Qos Controller information and processes them to generate the following important functionalities: get predicted transmission rate, select performance index, get optimized transmission rate, get optimized network path, get packet loss ratio, get delay time,

regulate transmission rate, receive broadcast, forward broadcast, and choose convenient transport protocol and then send it to destination client components.

Clients Component on the client side is the main component which generates client video requests, sends them to Streaming Server, and receives processed streaming video from Streaming Server Component. It requests live or video on demand streaming, receives their broadcasts, chooses transport protocol, requests streaming server port, chooses decryption standard, as well as supplying rate feedback to bandwidth estimator. All communication between these components are seamlessly accomplished through the middleware bus.

6. Research Contributions and Conclusions

In this work, first, first, an overview of both video streaming methodology and bandwidth estimation tools are given, and bandwidth adaptive video streaming architecture is introduced. This streaming server architecture includes Bandwidth Adaptive Streaming (BAS) Controller, QoS Controller and Decision Center components and their related objects. While Bandwidth Adaptive Streaming Controller tries to obtain the best path, an acceptable rate and optimized rate allocation by making use of *pathChirp* bandwidth estimation tool, the QoS controller calculates delay and packet loss values. A novel and carefully designed Decision Center determines the path, and optimizes the rate under the delay and packet loss constraints, by means of client produced RTCP feedback values. A global optimization method is used for optimizing parameters. A Model-Based Predictive Controller is used for video rate control. We have developed a complete distributed software information model with Broadcaster Component, Streaming Server Component and Client Components communicating over a middleware bus. Model-Based Predictive controller and in an all-in-one solution within our Integrated Software Architecture is now being implemented by using *ns-2* Network Simulator [7].

7. References

[1] S. Savage, A. Collins, E. Hoffman, "The end-to-end effects of internet path selection," *in Proceedings of ACM SIGCOMM*, 1999, pp. 289–299.

[2] D. Jurca and P. Frossard, "Media-Specific Rate Allocation In Multipath Networks", *Signal Processing Institute Technical Report* TR-ITS-2005.032, 2005.

[3] R. S. Prasad, M. Murray, C. Dovrolis and K. Claffy, ``Bandwidth estimation: metrics, measurement techniques and tools'', *IEEE Network* , Nov 2003, Vol. 17, No. 6, pages 27-35.

[4] R. L. Carter and M. E. Crovella. Measuring Bottleneck Link Speed in Packet-Switched Networks. In Performance Evaluation, 27,28:297-3 18, 1996.

[5] V. Riberio, R. Riedi, R. G. Baraniuk, J. Navratil, and L. Cottrell, "pathChirp: Efficient Available Bandwidth Estimation for Network Paths ." *Passive Active Measurement Workshop -- PAM2003* (Apr. 2003).

[6] E. Ronco, T. Arsan and P.J. Gawthrop, Open-loop Intermittent Feedback Control: Practial Continuous-time GPC, *IEE Proceedings of Control, Theory and Applications*, Vol.146, Issue 5,p.426-434, September 1999.

[7] K. Fall, K. Varadhan, The ns Manual, *The VINT Project A collaboration between researchers at UC Berkeley*, LBL, USC/ISI, and Xerox PARC.