# TRAFFIC CHARACTERISTICS OF OPTICAL BURST SWITCH ASSEMBLED AGGREGATED VOIP TRAFFIC

By

MICHAEL ALTON WATERS

Bachelor of Science in Engineering

Oral Roberts University

Tulsa, OK

1996

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 2007

TRAFFIC CHARACTERISTICS OF OPTICAL

BURST SWITCH ASSEMBLED

AGGREGATED VOIP TRAFFIC

Thesis Approved:

Dr. George Scheets

Thesis Adviser

Dr. Martin Hagan

Dr. Qi Cheng

Dr. A. Gordon Emslie

Dean of the Graduate College

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE Page

LIST OF FIGURES

FIGURE Page

# CHAPTER I

# INTRODUCTION

Research related to optimizing networks will continue to be important as packet traffic continues to increase. This increase in traffic can be attributed to common applications that are being redesigned to work over packet based networks. Some examples are multimedia applications such as Voice Over IP (VOIP) and IPTV.

It is forecasted that 44 million US residential lines will be converted to VOIP by 2010[1]. Given the current design of electronic packet switches, it is expected that they will not be able to handle the quantity of traffic that will be required in the future to support this and other real-time applications. There is ongoing research into optical switches that may provide a more efficient solution for carrying very large quantities of packet based traffic. One issue of interest concerns changes to the statistical characteristics of traffic as it traverses typical network hardware.

The goal of this research is to characterize VOIP traffic traversing a link from a VOIP network gateway to a traditional Public Switched Telephone Network (PSTN.) This includes showing the distinctions between traffic from the VOIP gateway and the PSTN network. VOIP traffic will be simulated and passed through an Optical Burst Switched assembly algorithm for comparison of the long range dependence of the input and output VOIP network gateway traffic. The proceeding section will provide some

background on the migration of voice systems from analog to digital to packet based systems.

## VOICE COMMUNICATION

The frequency range for audible voice is approximately from 300 to 3.1 kHz. Older systems would multiplex these signals onto a carrier frequency in order to transport them across the network. Noise is introduced and amplified as the analog signal is transported across the network. This type of noise is referred to as channel noise. Channel noise reduces the distance a signal can be transported as the signal to noise ratio must be above a certain threshold in order for the receiver in the transmission system to recover the signal. Figure 1 shows 3 signals. The 1st waveform is an example of a clean signal. It would be very easy for a receiver to identify this signal. The 2nd waveform has a small amount of noise added to it. A receiver would still be able to identify the signal despite the noise component. The third waveform has a significant amount of noise and the receiver would probably not be able to properly identify the correct signal. A solution to the channel noise problem is to digitize the signal which allows you to regenerate the signal at intermediate points and significantly reduce the impact of the noise component that is added to the signal as it is transported. The benefit to removing the noise is the total span (reach) of a system can be greatly increased if the digital signal is regenerated at the appropriate interval to remove the majority of the effects of channel noise.

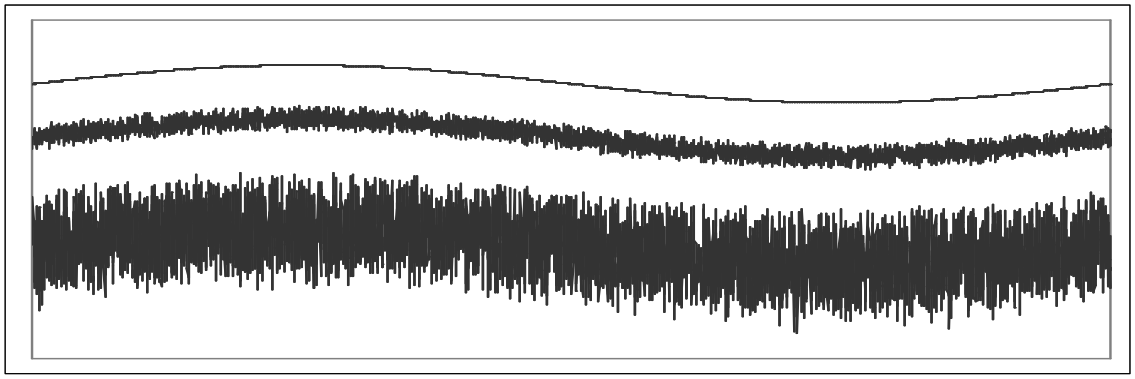

**Figure 1 - Signal and Channel Noise**

Nyquist's theorem can be used to determine the sampling time needed to reproduce a signal. The basic premise is that a signal needs to be sampled at a rate greater than twice the maximum frequency in order to accurately reproduce the signal. In the case of audible voice with a maximum frequency of 3.1 kHz, the signal would need to be sampled more than 6200 times per second in order to reproduce the signal. The sampling rate typically used on the PSTN is 8 kHz.

The conversion of a digital signal to an analog signal also introduces another source of noise. The amplitude of the voice signal is divided up into discrete levels at each sampling point. This process is known as quantization. Figure 2 shows a sine wave and the dashed line indicates how the quantized signal would look.

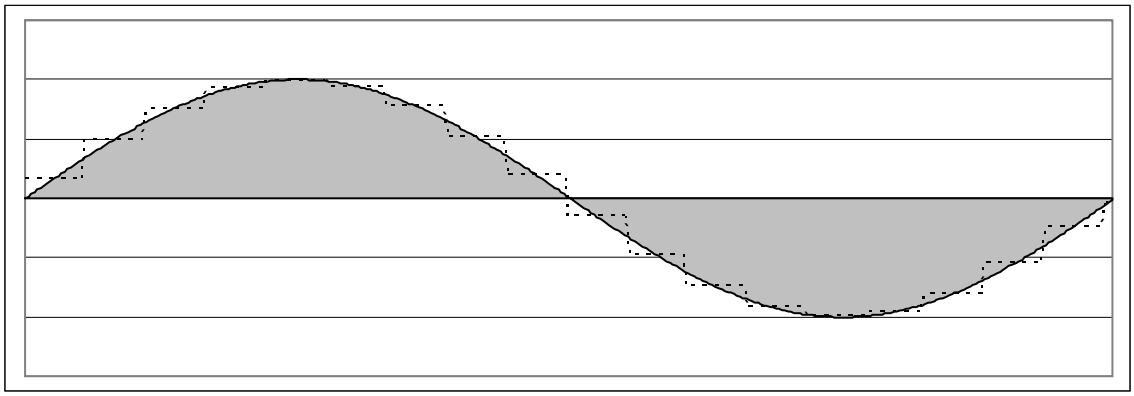

**Figure 2 - Signal Quantization**

Quantization noise occurs as a result of this as the reconstructed signal does not exactly match the original signal. Figure 3 shows the error associated with the quantized signal, where the error is defined to be the difference between the original waveform and its quantized result. Fortunately, the Signal to Noise Ratio (SNR) increases exponentially as the number of bits used to represent the amplitude of the sample is increased[2]. This allows the reduction of noise in the system due to quantization by picking the appropriate number of bits in each sample.

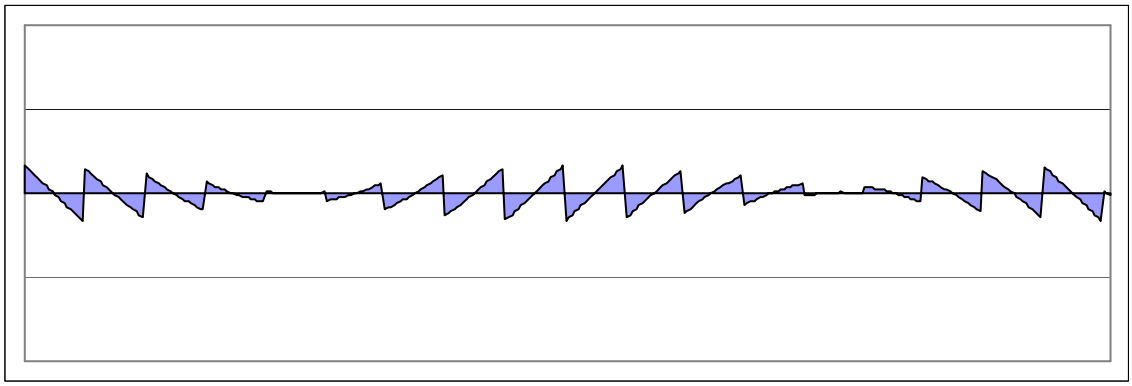

**Figure 3 - Quantization Error of a Signal**

One of the standards for the conversion of voice from analog to digital is G.711. This standard is produced by the ITU (International Telecommunications Union) standards body. This standard is the most adopted for the conversion of analog voice to digital voice. It uses 8 bit codes for the quantization levels and an 8 kHz sampling rate[3].

The North American digital TDM hierarchy is used for transporting these individual voice streams across the PSTN in North America. In this hierarchy lower level signals such as a voice signal are multiplexed with other independent voice signals onto timeslots via a method called time division multiplexing to produce a higher rate signal. A G.711 voice signal has a data rate of 64 kbits/sec. The North American digital

hierarchy takes 24 of these signals and multiplexes them into a 1.544 Mbit/sec signal that can be carried on a physical T1. The standard then specifies that 28 of these DS1s multiplexed together will result in a DS3 with a rate of 44.736 Mbits/sec.

Modern telecommunications systems utilize voice switches, transport systems, cross-connects and multiplexers to provide subscribers with voice communication. These systems are traditional circuit switched systems. Data communication is typically provided also but this is usually on an independent packet based system. Modern business require both voice and data systems and there is a convergence of these two systems occurring.

## VOIP COMMUNCATION

Voice is being provided over packet based networks through a technology referred to as VOIP. From a subscriber perspective, there is definite interest in moving to VOIP. One of the main reasons is from a regulatory perspective; the FCC has decided to limit regulation of VOIP services at this point. This allows VOIP network providers to bypass access charges that legacy telecommunications providers are required to pay. This reduces the cost of providing services and allows VOIP providers to be very competitive from a price standpoint. Also with VOIP technology, subscribers (particularly businesses) would no longer need separate connections to their establishment for voice and data. The introduction of VOIP into the market has also created opportunities for data providers to offer voice services which have increased competition. From a provider perspective, migrating networks to packet based systems reduces complexity and network components from the system.

In a VOIP system, the signal still undergoes an analog to digital conversion. This conversion happens at the customer premise. This digital signal is then inserted into IP packets for transmission across a packet based network. In the VOIP system the signal does not always arrive in a sequential manner as it does in the PSTN network. At the receiver, this requires the use of buffers to hold the packet for a period of time to allow for correct assembly of the signal due to non-sequential packet arrivals. Buffers also play a significant role in removing jitter from the signal. Jitter in VOIP communications results from variability in packet inter-arrival times. The buffer is able to space the output digitized voice signal at constant intervals which reduces the jitter. It is also possible to compress the voice signal because the encoding and decoding of the VOIP signal is done at the endpoints. A signal can be compressed on the POTS network but since switching is performed down to the ds0 level only, considerable work and additional equipment would be needed to perform this type of functionality on a POTS network. A VOIP network has a considerable amount of flexibility that is hard to match in the legacy PSTN network. This includes the ability to change compression of signals either globally or at a line level. Since most data networks have built-in restoration schemes, VOIP can take advantage of this. Much of the currently deployed PSTN networks have no automated restoration scheme. They rely on protection schemes with preplanned protection capacity.

One of the main issues when deploying VOIP solutions is guaranteeing end-to-end delay. This characteristic of voice packet traffic greatly affects network design and the ability of users to communicate. End-to-end delay must be between 0-150ms for most user applications[3]. End-end-delay has many different contributing components including the coder, the queue, the serialization of the message, the transport of the

messages, the network switching delay, the de-jitter queue, and the decoder. The network must be designed such that the total end-to-end delay does not exceed these. Some protocols assist this process by providing set paths for packets to follow. MPLS in particular is used to build virtual circuits across the packet based network. This assists in providing traffic engineering to the packet based networks and VOIP packets end up arriving in order as a result which reduces the size of the jitter buffer.

Since there is considerable growth in VOIP and transport network capacity is increasing there is continued interest in optimizing network resources. Current research and developing technology is working towards using restoration schemes as opposed to protection schemes for providing network restoration. A protection scheme provides network resources for rerouting traffic in the event of a network failure. Protection schemes are typically defined when the circuit or link is established An example would be a secondary circuit that is provisioned at the same time as a primary circuit. A restoration scheme does not use pre-planned capacity like a protection scheme. A restoration scheme will typically reroute circuits using available capacity in the event of a failure on the network. A restoration scheme cannot guarantee service restorability like most protection schemes can. A lot of the current work in this area is focused on optical switching as a solution for providing restoration schemes for networks.

## OPTICAL SWITCHING

There has been considerable growth in the optical switching space in recent years. The primary drivers behind using optical switches are the increased network capacity due to wavelength division multiplexing (WDM) and need for automated restoration and

provisioning. In order to narrow the scope of this paper, the focus will be on wavelength switching and its relationship in optical switching. Since many legacy systems are built around a circuit switched network, the need for circuit switching in optical switch space will continue to be important. There are currently optical switching systems available to help reduce provisioning times and also provide some advanced restoration options for circuit switched networks.

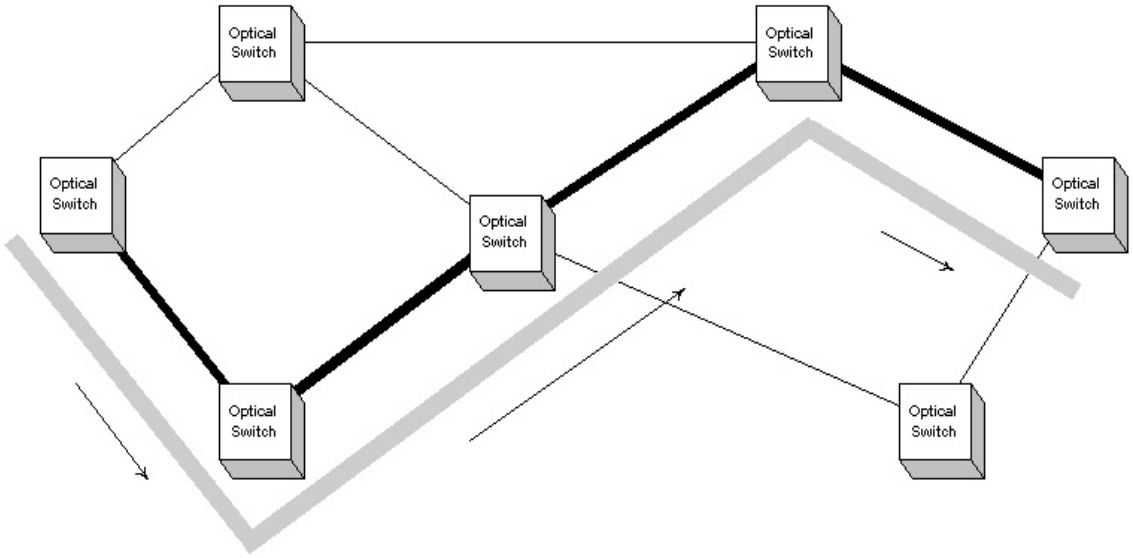


**Figure 4 - Circuit Switched Optical Network**

In a circuit switched optical network, a circuit is provisioned across the network and data is transmitted from the edges of the network. Figure 4 shows an example of a circuit switched optical network. Assume the links between nodes consist of WDM systems on fiber. The path the data is following has been setup prior to any data being transmitted on a certain wavelength on the WDM system. This stays active until the circuit is no loner needed and the user initiates a request to remove the circuit.

With the continued growth of packet based traffic, packet switched optical networks will continue to be an important research area. There are still issues with packet

switching at the optical level. The primary issue at this point is the lack of buffers in an all-optical-network. This requires conversion of signals from optical to electrical prior to buffering or sending signals down delay loops which reduces the total reach of a system due to the increase in the optical path required by delay loops.

In a packet switched optical network, there is not a requirement for physical circuits established through the network. Through the use of signaling protocols virtual circuits may be established. Figure 5 shows an example of a packet switch optical network. In this example, no virtual circuit has been established between nodes and as such each node is able to send the packets down any available link to the destination. Because many packets can arrive at a node simultaneously, buffers would be required.

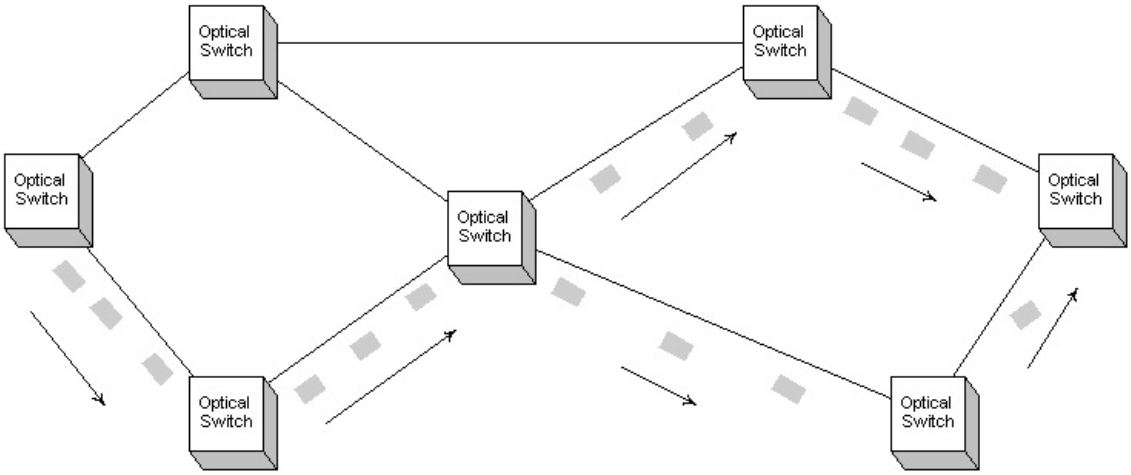


**Figure 5 - Packet Switched Optical Network**

Optical switching is designed to reduce provisioning times as well as provide automatic restoration in the event of failures. One of the basic functions of the optical switch is to build internal connections for connecting external ports together. Optical switching is accomplished within a switch using an electrical or optical switch matrix.

The switch matrix is the part of the switch which has a basic responsibility for building connections between ports on the switch to each other. If the matrix performs a conversion of the signal to electrical prior to building the connection, this is generally referred to as an O-E-O (optical – electrical – optical.) If the signal remains in the optical domain throughout the matrix, the matrix is referred to as O-O-O (optical-optical-optical). A MEMS (Micro-Electro-Mechanical-System) based matrix is an example of a switch fabric which uses tiny optical mirrors to build an optical path between ports on the switch. Switching times in the matrix must be very fast. Research has shown that switching times can be as fast as 4.1ms[5]. It is important to have low switching times as telecommunications carriers are accustomed to sub 50ms protection end to end switching times.

## OPTICAL BURST SWITCHING

Circuit switching is an established product for optical switches. Packet switching still requires optical buffering which is being researched. Optical Burst Switching (OBS) has emerged as a fit to bridge the gap between circuit switching and packet switching in optical networks[6]. The basic concept behind OBS is to send bursts of traffic to the optical network. This burst is assembled on the edge of the network through the use of electrical buffers. The optical connections are built across the network via burst switching protocols which only hold each link open long enough for the burst to pass. This allows reuse of network resources for other traffic. For very small bursts of data, there may never be an end-to-end circuit established as the initial links have already passed the data and are being switched for an independent burst to a new location. Figure 6 shows an example of a burst that has traversed an OBS network. The last link is shown as partially

activated below indicating that the switch has not completely finished establishing the cross connect in the switch matrix. Also notice that the initial links the burst traveled down are no longer provisioned.

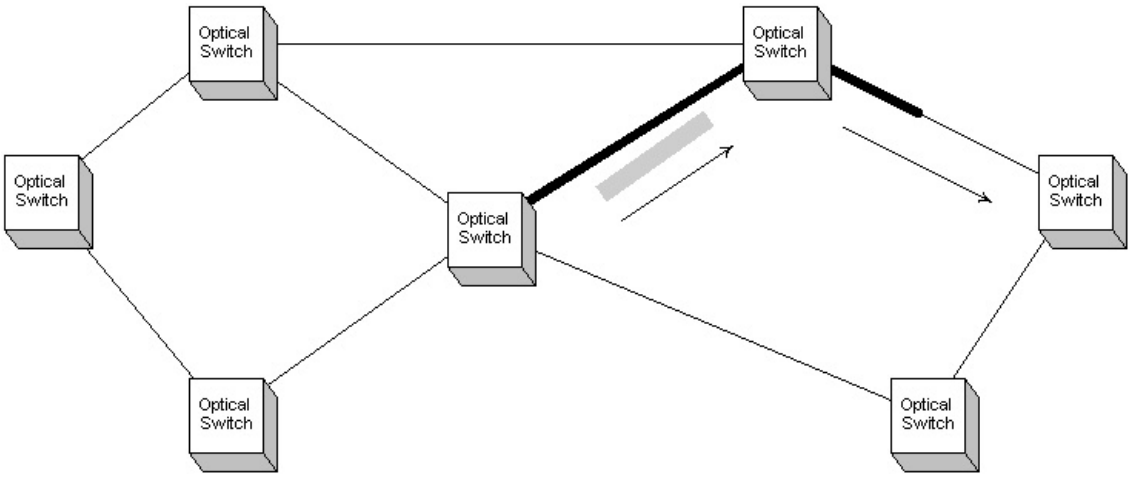


**Figure 6 - Optical Burst Switch Network**

In an OBS network control channels are required for the nodes to communicate via the OBS protocols. Control channels can be in-band or out-of-band. The nodes are required to process the information in the control channels in order to establish the appropriate connections. The separation of control channels from data channels is referred to as the control plane. The control plane is a very important resource in OBS due to the necessity of sending messages to nodes prior to the actual data arriving at that specified node. These messages contain the necessary information to allow the switch to build the connections in the switch matrix prior to receiving data so there is no interruption in the transfer. Since there is a small amount of time required to build the connections in the matrix, the control messages in OBS must arrive in advance with at least enough time for the last switch in the path to build the connection before the data arrives.

Two of the most important components of optical burst switching that distinguish it from optical circuit switching are the burst assembly algorithms and the control plane protocol used to transport the burst across the network. In the OBS network burst assembly occurs on the edges of the network through the use of electrical buffers. The packets being routed across the network are held in these buffers until an appropriate condition is met. These conditions can be based on a burst size limit, buffering time limit or a combination of both. This shapes the input traffic and since packet traffic is typically bursty in nature, the buffers smooth the traffic at the input to the OBS network.

Once the burst is assembled the OBS initiates control plane signaling to forward the burst to the appropriate endpoint of the network. In circuit switched networks, the circuit is established between all links on the network prior to transmission of the signal. In OBS, bursts are typically small and as such they do not require the entire path to be established prior to transmission. This partially establishment of a path is accomplished by sending control messages to nodes shortly before arrival of the burst to configure the switch matrix for transmission of the burst to the next node in the path. Another control message is forwarded to that node also to configure the switch matrix prior to the burst arrival there. This process continues until the burst arrives at the destination.

Sometimes there are no available links to the next node in the path. Typically a blocking condition would exist. There has been research into deflection routing which would send the data signal down delay loops to allow the signaling protocol enough time to configure the switches in an alternate path. Also bursts can be broken up to allow them to fit in between already scheduled larger bursts. This is referred to as Burst Segmentation.

Optical Burst Switching is a bridge between a circuit switched optical network and a packet switched optical network. One of the main benefits of using OBS over optical circuit switching is there is a gain in utilization through statistical multiplexing of links.

This section has provided an overview of analog and digital voice. It also compared the differences between optical circuit switching, optical packet switching and optical burst switching. The next chapter will provide background research on packet voice models and optical burst switching. It will also identify research related to the estimation of long range dependence.

# CHAPTER II

# REVIEW OF LITERATURE

Traffic characterization is very important from a network modeling perspective. Early work on traffic characterization led to network optimization through the research and implementation of statistically multiplexed traffic on network trunks in the voice network. The identification of the statistical nature of inter-arrival times and call holding times for customer generated traffic was particularly important. This research led to the application of erlang and poisson traffic distributions for the sizing of PSTN voice trunks.

## MODELS FOR VOIP TRAFFIC

Dang, Sonkoly and Molnar[7] indicate in their research that IP traffic call interarrival times are exponentially distributed. They further indicate that call holding times are not exponentially distributed as commonly assumed. They model call holding times using a pareto distribution.

Fractional Gaussian Noise has also been suggested as a model for IP traffic traces[7]. Recent research by Hassan, Garcia and Bockstal[8] provides guidelines for packet interarrival times of VOIP network models. Their research indicates that for light traffic intensities, the exponential model can be used. For heavy traffic intensities they claim that there is a long term covariance in the arrival time that appears at larger time scales which causes poor performance. They approximate the heavy traffic intensity with

a Markov Modulated Poisson Process (MMPP). As will be shown, traces from a carrier packet network do not have an exponentially distributed call interarrival time and call holdings times do not follow the exponential distribution as referenced above.

## OPTICAL BURST SWITCHING

There is considerable research activity into the use of OBS for the transport of internet traffic. OBS was first proposed as a solution as a bridge between optical circuit switching and optical packet switching[6]. In an OBS switched network links between nodes are only active long enough for a burst of data to pass. This is a deviation from Optical Circuit Switching in that the entire path is not necessarily active at the same time. In an OBS network, control packets are sent to adjacent nodes with enough lead time to allow the adjacent node to configure the switch matrix shortly prior to the optical burst arriving. Research into the protocols associated with the timing of sending bursts is not the focus of this thesis (Just In Time[9] and Just Enough Time[6] signaling). The size of the bursts can affect the blocking probability associated with these two signaling protocols in the core of the network and as such has to be considered in burst assembly analysis. This research has shown that large bursts increase the blocking probability in the core of the network. The design of the control plane also has an effect on the size of bursts in the OBS network so as not to overrun the capacity of the control plane network by the quantity of the control packets sent[10] . Figure 7 shows the minimum burst duration required to keep utilization below 100% on the control channels for two different control channel configurations. This graph assumes the control channel has the same bitrate as the data channels and the control packet size is 64 bytes. The formula used to calculate the minimum burst duration is $L_{\min} = \frac{(K-k)}{k} \cdot \frac{B_{length}}{R}$ where K is the

number of available channels, k is the number of control channels, $B_{length}$ is the burst control packet length, and R is the link bit rate. The minimum burst duration plays a key factor in the design of the burst assembly algorithm which is another primary element of an OBS network.

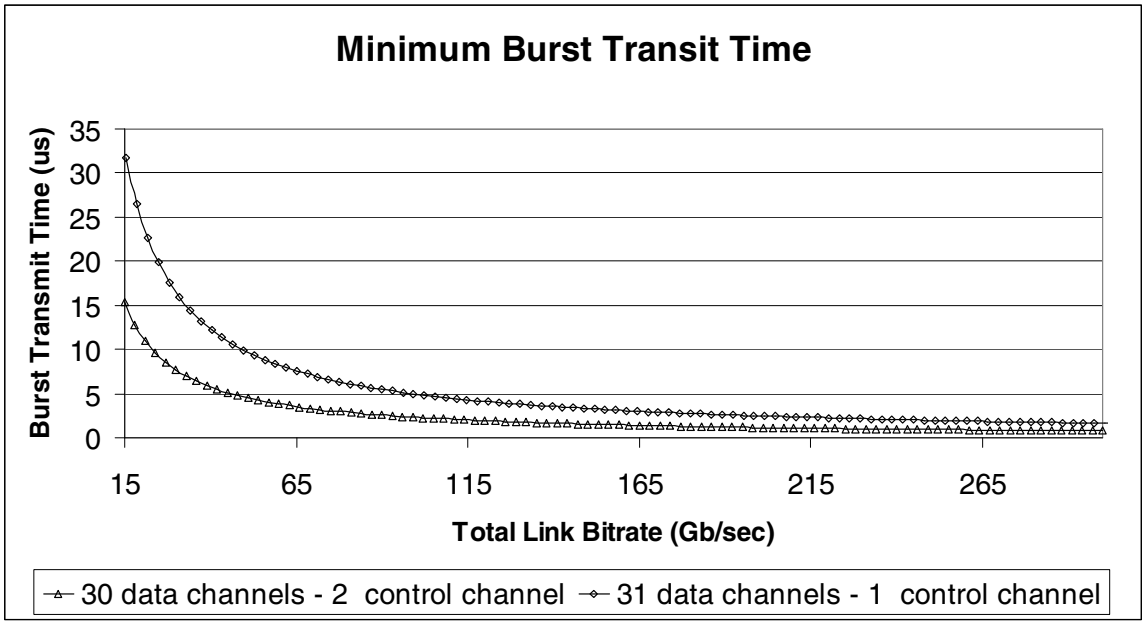


**Figure 7 - Minimum Burst Size**

The burst assembly process has been grouped into 3 categories: timer –based, burst length and hybrid. Timer based algorithms store packets for a minimum and/or maximum time prior to building and transmitting a burst[11]. One of the primary benefits of timer based algorithms is the delay of bursts can be closely controlled. While delay is able to be controlled in a timer based algorithm, system efficiency suffers because there is a minimum burst size due to the design of the control plane. In a timer based algorithm bursts must be "stuffed" in order to send the minimum size burst. Burst length algorithms wait for a minimum or maximum burst length[12]. In this algorithm delay for real time applications could be an issue as the OBS network waits for the minimum size prior to

transmitting. The hybrid approach combines these two utilizing timer and burst length thresholds to build bursts for transmission across the network[13]. Recent research on burst assembly algorithms proposes an adaptive burst assembly algorithm based on the input traffic characteristics[12, 14, 15]. The research has shown that the burst assembly algorithm is able to shape the traffic and thereby change characteristics of the bursts within the core of the OBS network.

## SELF SIMILAR TRAFFIC

Input traffic characteristics to the edge nodes play an important role in burst traffic characteristics. Research on internet traffic has shown that it exhibits a self similar nature and should not be modeled using Poisson distributions[16, 17]. It was first proposed that the burst assembly process would reduce the self-similarity of the input traffic[11]. Later analysis showed no change in the self-similarity of generic packet based input traffic using time based assembly algorithms[13]. Research by Hu, Dolzer and Guager in 2003 has concluded that "burst assembly does not reduce self similarity in general."[18] The distinction is made in their work between self-similarity being byte-wise and packet-wise. Byte-wise self-similarity is referring to the self-similarity of the bytes transmitted. Packet-wise is referring to the self-similarity of the count of packets or bursts. They indicate that packet-wise self similarity can change based on the assembly algorithm selected.

Self-similarity is the finding that correlations exist in time series data over different time scales. One of the parameters that assist us characterizing self-similarity is the Hurst parameter. For time series without long range dependence the Hurst Parameter

= 0.5. This time series is also identified as being made of unbiased independent values. For traffic with a very high degree of correlation over different time scales, the Hurst parameter approaches 1. Here is a basic introduction of long range dependence[19]:

Let $\{X_t : t \in \aleph\}$ be a weakly-stationary time series. The autocorrelation function of this weakly-stationary time series is $p(k) = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sigma^2}$ where μ is the mean and $\sigma^2$ is the variance. If $\sum_{k=-\infty}^{\infty} p(k)$ diverges, then $X_t$ is said to be long-range dependent. The Hurst parameter can be obtained by using a functional form of this divergence: $C_p k^{-\alpha}$ where $C_p>0$ and $\alpha \in (0,1)$. The Hurst parameter is related to α by the equation: $H = 1 - \frac{\alpha}{2}$.

There are many techniques for estimating the Hurst parameter. These include the R/S statistic method[20], aggregate variance[21], Peng's method[22], difference of variance method [21] and the absolute value of moments method[21].

Much of the research in Optical Burst Switching uses generalized models for internet traffic. There have been more specific models developed for voice packet traffic. Recently research has produced models for aggregated IP traffic. The focus of this thesis is to examine the effects of a burst switching assembly algorithm on the self-similarity characteristics of VOIP network gateway traffic. In Chapter 3, a simulator for generating VOIP packets will be designed using traffic characteristics from a carrier network. This traffic will then be passed through a OBS assembly algorithm to compare the long range dependence of the traffic.

# CHAPTER III

# RESEARCH

In order to model aggregated VOIP traffic over an optical burst switched network, characteristics of the input traffic are needed. This research will initially focus on the interarrival distribution and call duration distribution of the input traffic. Much of the prior research assumes an exponential distribution of the interarrival times and call durations.

## CIRCUIT SWITCHED CALL CHARACTERISTICS

Analysis of calls from a competitive local exchange carrier VOIP gateway reveals the following characteristics of calls. Multiple distinct samples of calls were captured for this analysis. The first sample was from calls originated primarily on legacy circuit switched voice networks. The second sample was from VOIP originated calls that traversed between the VOIP network gateway and a circuit switched voice network. Table 1 identifies basic characteristics of calls originated from and terminated to VOIP customers via a VOIP network gateway.

| Call Origination Type | Packet Switched Voice | Circuit Switched Voice |
|---|---|---|
| # of records | 3445 | 2683 |
| Average Call Duration | 2.4174 min | 1.8288 min |
| Average Interarrival Time | 1.9155 sec | 1.2316 sec |

**Table 1 – Sample Voice Call Characteristics**

As can be seen in table 1, the average call duration and average interarrival time from the two samples are different. It should be noted that these two pools are from two different sets of end users. The packet voice network in general contains a higher percentage of business users across the entire CLEC footprint. It is also expected that the calling patterns between different sets of end users can be significantly different. In the cases above, the times are typical of what is encountered across the network

In order to establish that the call interarrival times of traditional voice calls are exponentially distributed, a comparison of the sample data is made via a quantile-quantile plot (QQPlot) and a histogram of the call interarrival time. A QQPlot is a statistical graphical analysis tool which can compare sample data to a known random distribution type. A QQPlot is used in determining if two samples come from a common distribution, if they have similar location and scale, if they have a similar distribution shape, and if they have similar tail behavior[25]. The free software package R was used produce the QQPlot graphs[26]. Given a random variable X, The $q$-quantile of an experimental result X is defined as the value of x such that $P(X \leq x) = nq$, where n = 1 to the maximum number of points in the sample), and $0 < nq \leq 1$. The matching quantile for a random variable Y is that value of y where $P(Y \leq y) = nq$. The value of $q$ is determined by the number of samples in the data set. For the circuit switched voice, $q \approx .00037$ since there were 2683 samples. For the packet switched voice graphs, $q \approx .00029$ since there were 3445 samples. In the QQPlot graphs if the points on the QQPlot lie above the line, this indicates that the sample data has a lower CDF at that point than the expected distribution. If the points are below the line, it indicates that the CDF of the sample data is higher at that point than the expected distribution. The straight line in the QQPlot

graphs indicate where the points should lie if the 2 samples come from the same distribution since their CDF would be the same at those points. For example, on the far right plot of Figure 8, there is a plotted point at approximately (6, 6.4). This point indicates that the P(an exponential PDF with a variance equal to that observed experimental is $\leq$ 6), call this $F_X(6)$, = the P(the observed experimental values are $\leq$ 6.4), call this $F_Y(6.4)$. This means the experimental observations are slightly more spread out than what would be expected from an exponential distribution. Note that this point lies above the 45 degree straight line and that $F_Y(6)$ will be $< F_X(6)$.

The QQPlot analysis for circuit switched voice call interarrival time is shown in figure 8. The first graph is a plot of the interarrival time of all the calls in the sample. The second graph in figure 8 shows the histogram of the sample data. A line representing the exponential distribution has been overlaid. The third graph is a QQPlot of the sample data versus the exponential distribution. This figure indicates that the sample closely matches the distribution of an exponential random variable. The 95% confidence interval is also shown as a dotted line in this graph. From this analysis, it is shown that for the circuit switched voice calls in our sample, the interarrival times can be modeled via an exponential distribution.

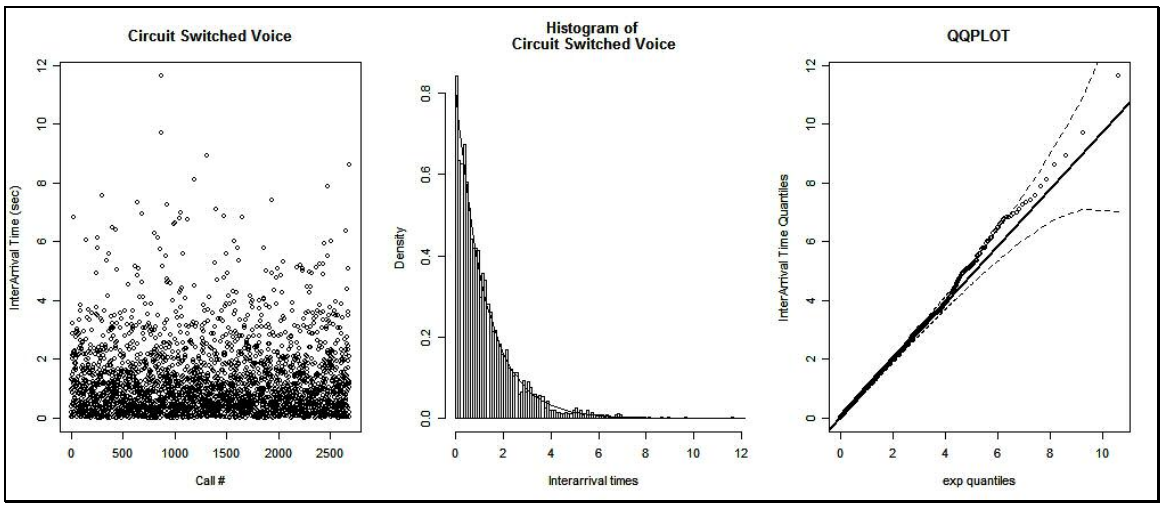


**Figure 8 - QQPlot Analysis of Circuit Switched Voice Calls**

Analysis of the call duration for circuit switched originated calls is shown in figure 9. The first graph is the plot of the call durations for each of the samples. The second graph is a histogram of the sample with the exponential distribution overlaid. The exponential graph does not appear to fit the histogram for the short call durations. The QQPlot also indicates that the exponential model will not model the longer call duration calls properly. The call duration of the sample used would more accurately be measured using a lognormal distribution. Figure 10 shows the comparison of the call duration to the lognormal distribution. The lognormal distribution models the longer call duration better than an exponential distribution, and it models the short call duration very accurately.

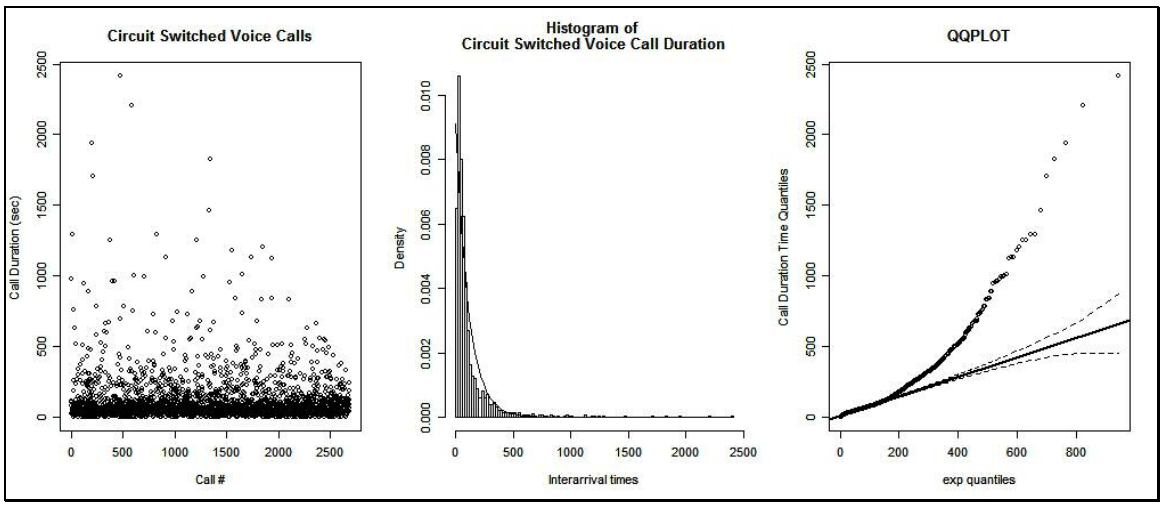


**Figure 9 - Circuit Switch Voice Call Duration**

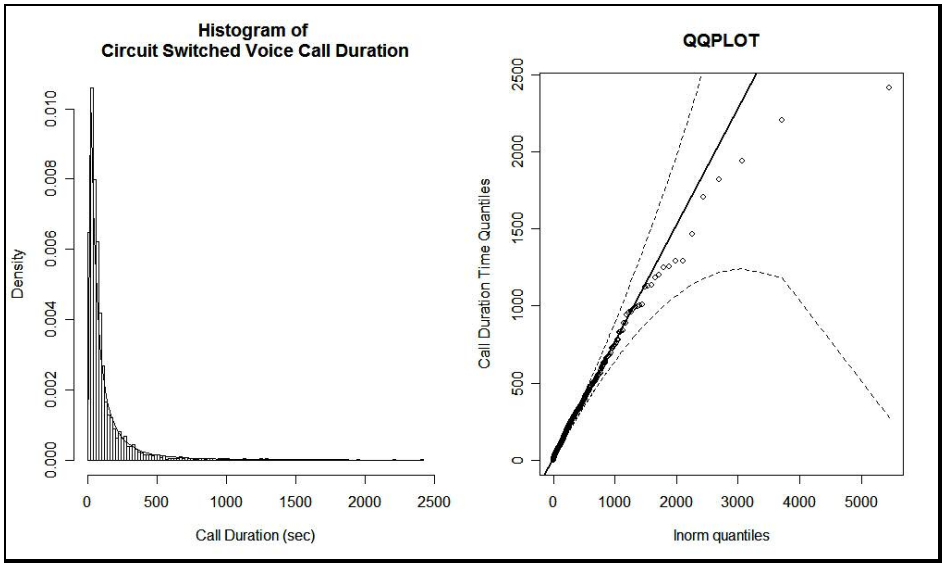


**Figure 10 - Circuit switched voice call duration versus the lognormal distribution**

## VOIP CALL CHARACTERISTICS

The previous sections have identified the characteristics of calls which were primarily originated on a legacy voice network. The next section will analyze the traffic originated by CLEC VOIP customers and sent across a VOIP network gateway. Figure 11 shows the comparison of the interarrival time to the exponential distribution. The first graph is a plot of the sample data versus interarrival time. The second graph is the

histogram of the data with the exponential distribution overlaid. The third graph is the QQPlot of the data versus the exponential distribution. The exponential distribution does not model the sample data closely for short duration or long duration calls. From these graphs, it is determined that the VOIP originated data from the network gateway does not fit the exponential distribution.

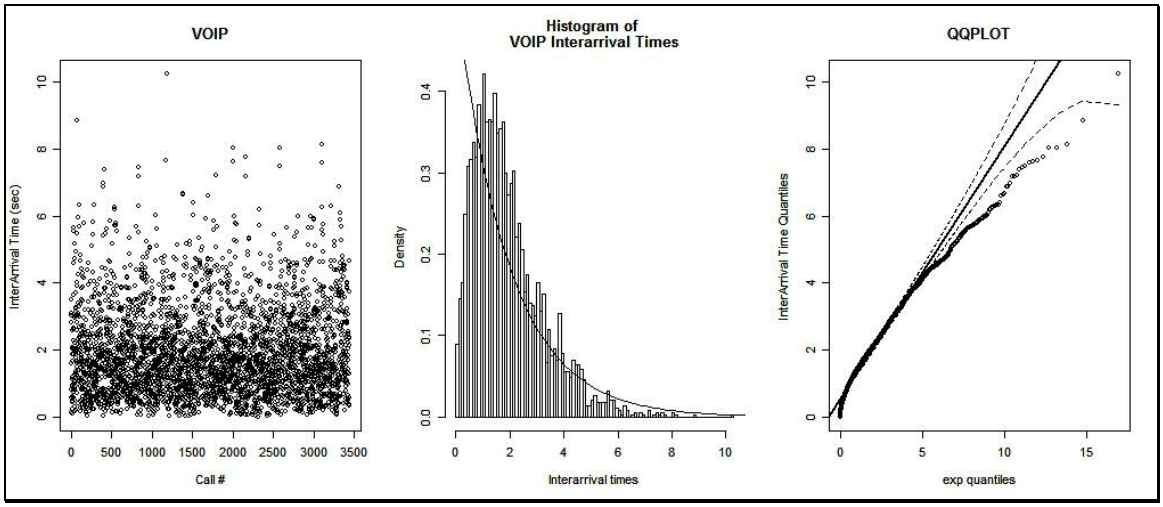


**Figure 11 - VOIP Call Interarrival Time Comparison to the Exponential Distribution**

The interarrival distribution mostly closely matches the Weibull distribution. A few other distributions including the normal and lognormal distributions were not as close a match. The Weibull distribution matches the shorter interarrival times but fails to closely model the heavier tails in the sample data. By comparing the range of the y-axis (sample data) with the x-axis (weibull distribution) of the QQPlot, the heavier tale of the sample data can be clearly seen. For the purposes of this research the Weibull function will be used as it is a fairly close match. The graphs in figure 12 contain the histogram analysis and the QQPlot of the sample versus the Weibull distribution.

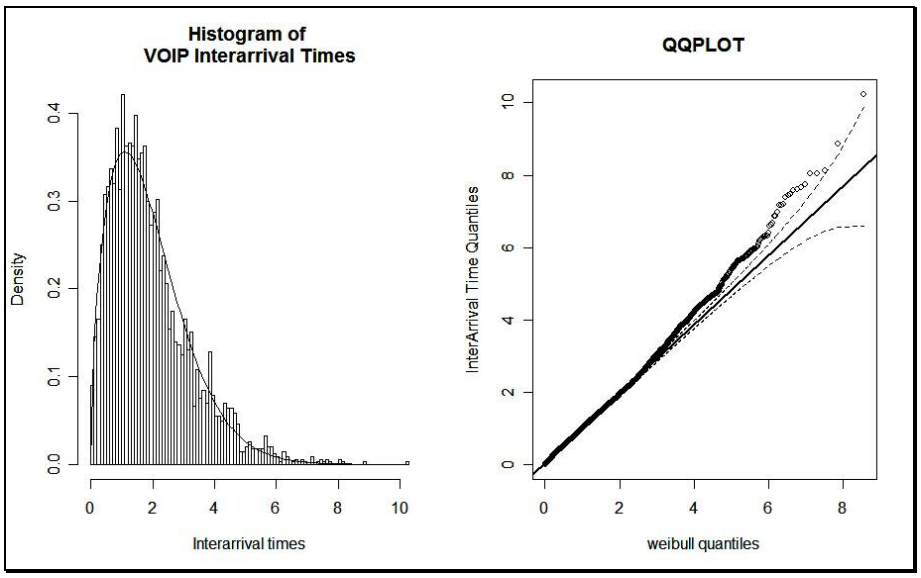


**Figure 12- VOIP Interarrival Times vs Weibull Distribution**

The data in figure 13 is the analysis of the call duration vs. the exponential distribution. The first graph is a plot of the data points in the sample set. The second graph is a histogram of the dataset. The exponential distribution is overlaid in the second graph. It is evident that the distribution does not model the very short call durations adequately. The third graph in figure 13 is the QQPlot of the sample vs the exponential distribution. The QQPlot shows the longer call duration samples can not be modeled properly using the exponential distribution. Figure 14 shows the comparison of the lognormal distribution to the call duration for VOIP originated calls across the network gateway. The first graph is the histogram of the sample with a lognormal distribution overlaid. The second graph shows the QQPlot of the sample vs a lognormal distribution. From these graphs, it is evident that the call duration for VOIP calls from the network gateway would be more accurately modeled using a lognormal distribution rather than an exponential distribution.

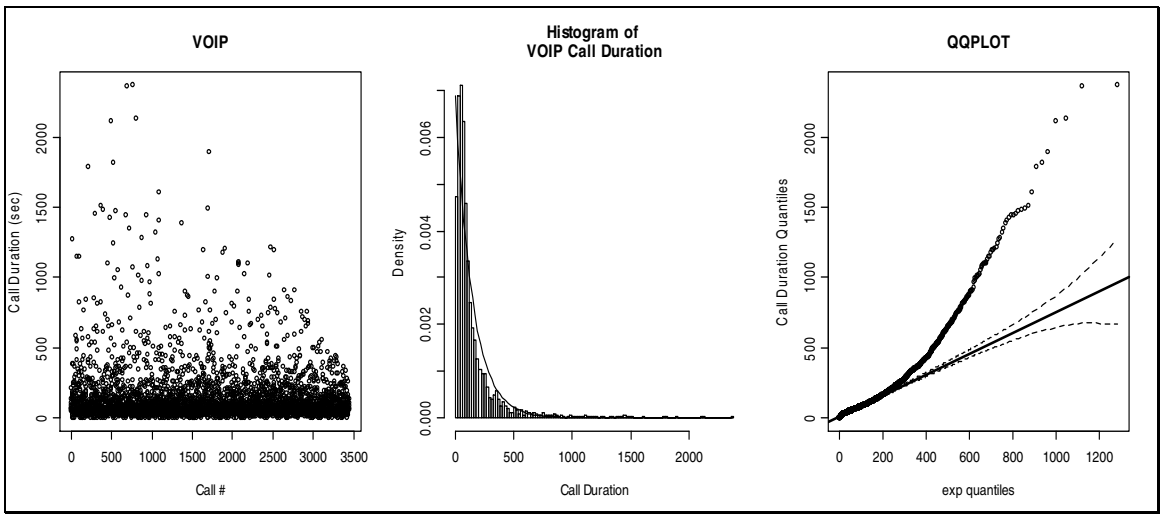


**Figure 13 - VOIP Call Duration versus Exponential Distribution**

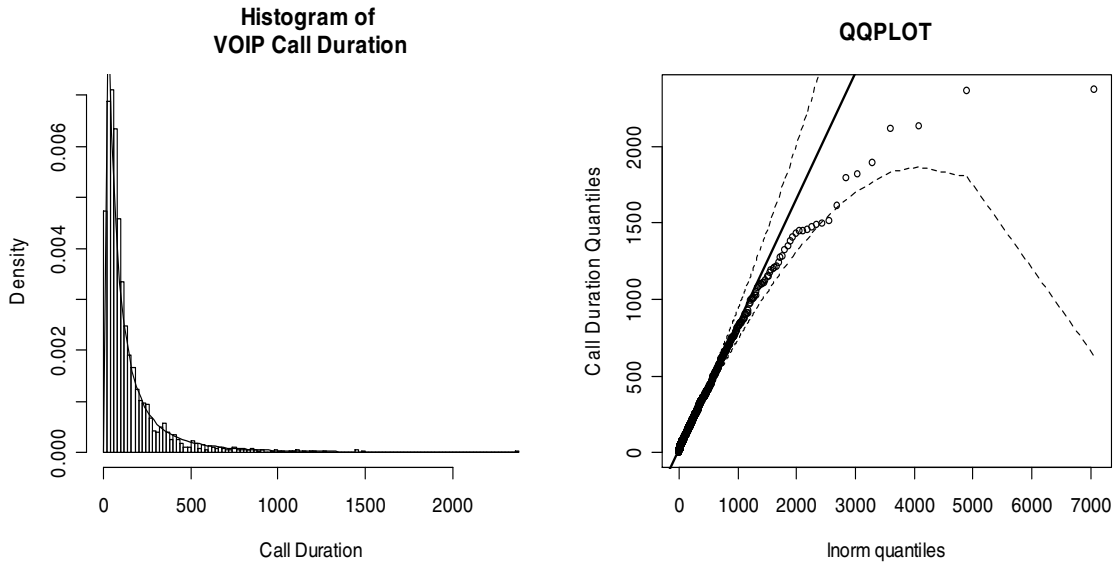


**Figure 14 - VOIP Call Duration vs Lognormal Distribution**

Since the VOIP call interarrival times most closely match the Weibull distribution, there is a reduction in the very short call interarrival times. It is also noted that the distributions above are not exponential distributions as is typically assumed. Current telecommunications business practice uses probability models to predict link capacity utilization using the minutes that traverse a trunk in the busy hour. Current models assume Poisson or Erlang distributions for determining utilization. If the traffic

on the trunks does not follow one of those distributions, the calculated utilization would not match the true utilization of the trunks. Further research is needed on the application of the Weibull distribution to more accurately predict the utilization on the VOIP gateway trunks based on the busy hour usage.

## VOIP CALL AND OBS ASSEMBLY MODELING

The traffic characteristics from the VOIP network gateway are used to analyze the effects of the burst assembly algorithm. An OBS model was developed using an open source discrete event simulation model called OMNET++[26]. Figure 15 identifies the modules created within the simulator and the message flow. The first module is the call simulator. The call simulator generates traffic using the statistics from the previous section. The weibull distribution is used for call interarrival time. The call duration is generated using a lognormal distribution. This module sends messages out based on these random distributions in order to simulate user initiated calls. The messages created in the call simulator are passed to the packet generator. This module creates an instance of an object for each call that is passed to it from the call simulator. These dynamically created objects are active for the duration of the call and are able to generate the IP packets required for the duration of the call. The packets generated are all a fixed size based on a G.729 codec with 2 frames per packet. This codec sends a packet out once every 20 milliseconds. All of the dynamic packet generators send their output to a single input port of the OBS assembly module. This module stores the message lengths until the burst is created. The burst assembler was simulated utilizing a timer based assembly algorithm. Since VOIP is very susceptible to latency, a burst length assembly algorithm was not

used as it would have more potential for blocking traffic as well as no guarantee regarding delivery times. When the burst is created, it is sent to a sink module which is used inside of the simulator to track output statistics from the module.

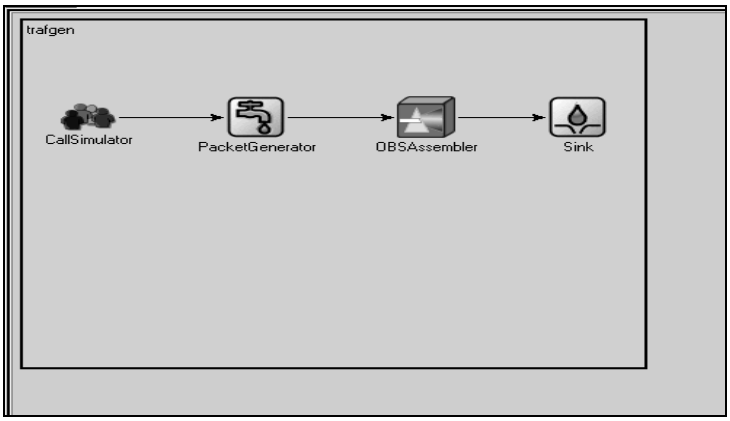


**Figure 15 - OMNET++ Simulation Modules**

Figure 16 shows the simulated IP packet input to the burst assembler. This graph indicates the number of packets in the assembler queue in every 20 millisecond period. The graph contains over 2.2 hours of data. It also shows the initial ramp period that occurs at the start of the simulation. This ramp period is excluded from analysis in proceeding sections. It is evident that the input traffic is bursty in nature and its self

similar nature will be tested later.

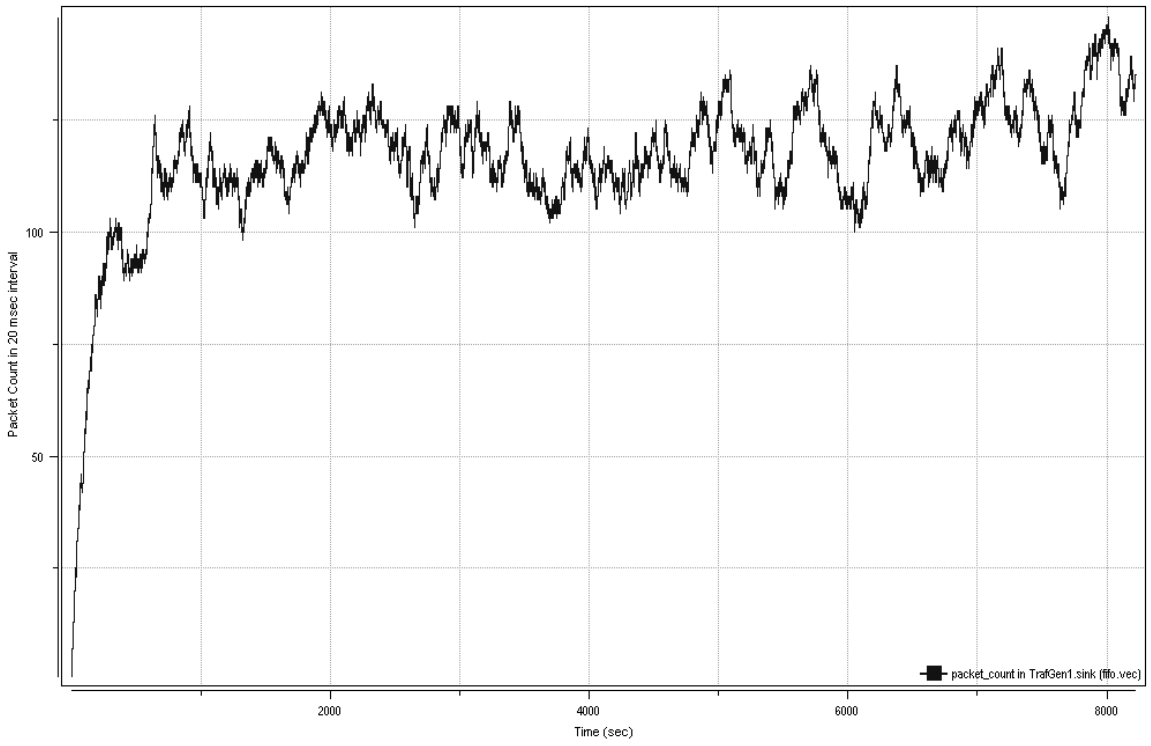


**Figure 16 - Simulated VOIP Packet Input Traffic**

The histogram of the IP packet interarrival times at the input of the OBS assembler is shown in figure 17. It looks somewhat exponential. A line representing the exponential distribution is also shown on the graph as reference.

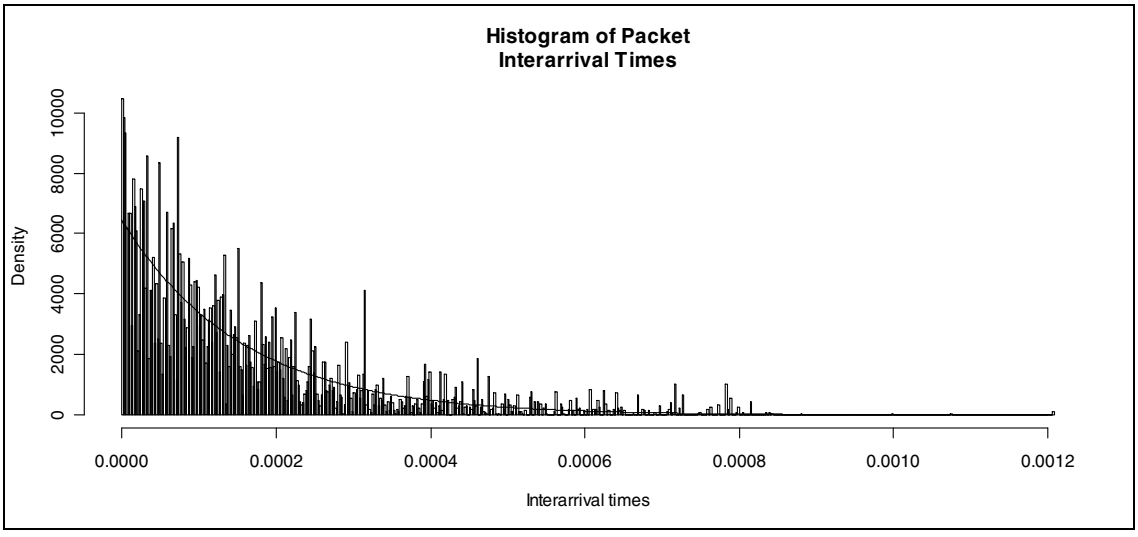


**Figure 17 - Interarrival Time of VOIP packet Traffic**

The output from the burst assembler given this input sample is shown in figure 18. The simulation used a time-based threshold for burst creation of 50 milliseconds. The histogram of the burst size is shown in figure 19. The normal distribution with similar mean and standard deviation is also overlaid for reference only.

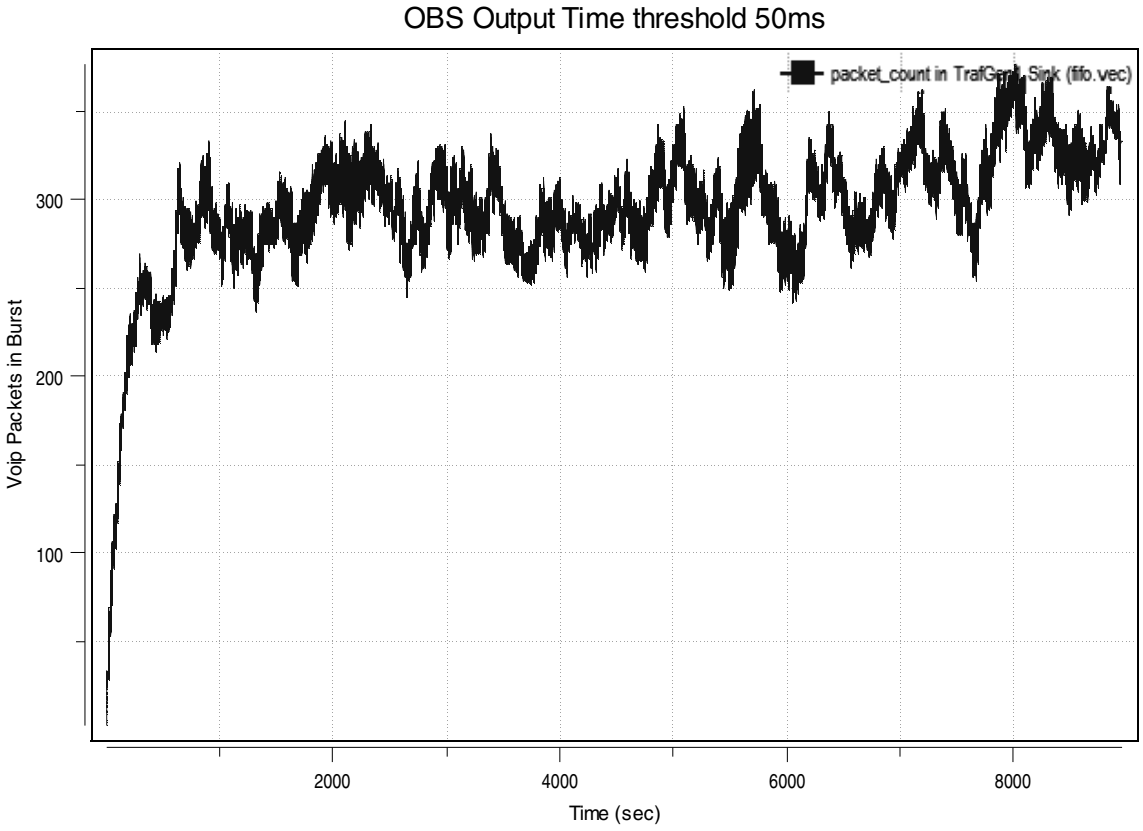


**Figure 18 - Simulated Burst Length Output**

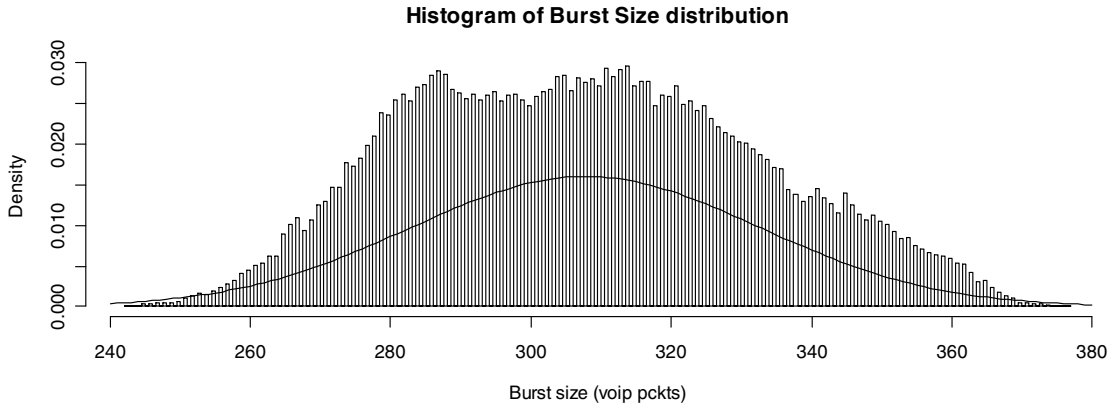


**Figure 19 - Simulated Burst Size Distribution**

The packet input traffic and burst assembler output appear very bursty in nature. A time-based burst assembly algorithm is used in this analysis due to VOIP being a real time application and the end to end delay of calls is typically designed to a maximum value of 200 ms. In actual designs, this value could be calculated based on distance between end users, delay of processing through equipment and the delay associated with encoding and decoding.

## LONG RANGE DEPENDENCE OF INPUT TRAFFIC AND BURST OUTPUT

The affect of aggregation of three different sources is shown in figure 20. The first is a poisson generator which has a Hurst parameter of 0.5. The second is a fractional Gaussian noise (fGn) generator with a Hurst parameter of 0.95. The third is the VOIP packet generator that was implemented as indicated in the previous section. The lowest graphs show the traffic aggregated into 0.1 second intervals. The top level shows the aggregation of the traffic into 100 second intervals.

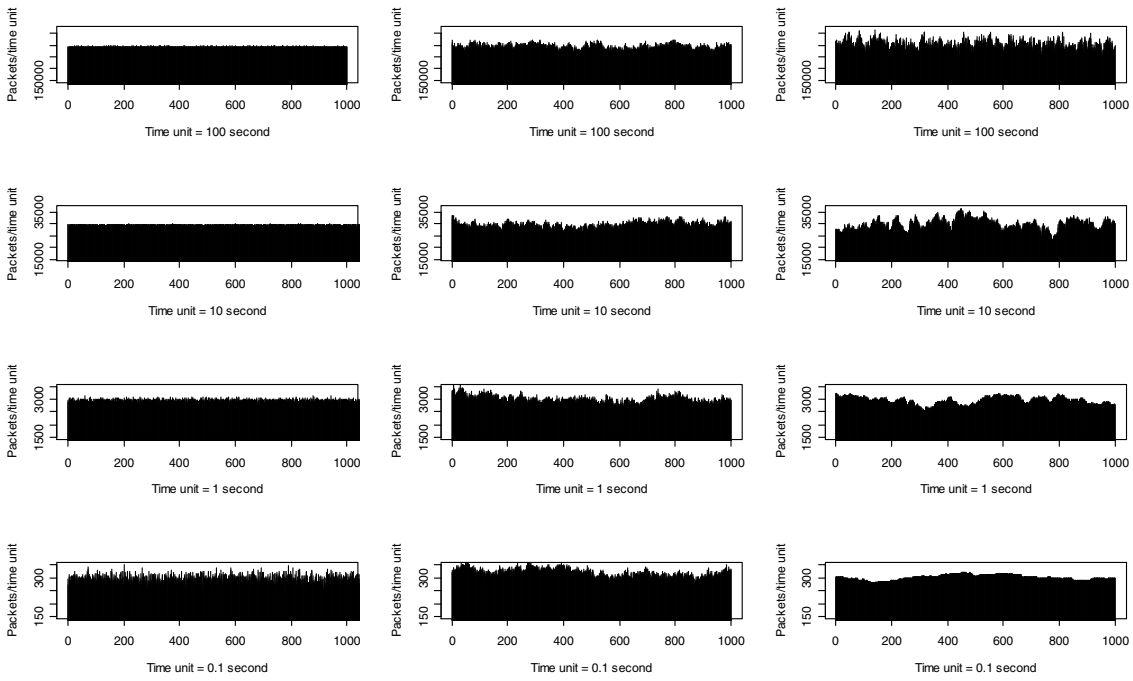


**Figure 20 - Aggregation of Poisson, Self Similar and Simulated VOIP Traffic**

It is evident that there is less variability as Poisson traffic is aggregated over longer intervals. From visual inspection, there is little change in the variability of the fGn traffic as it is aggregated. The input VOIP traffic variability is not smoothed as it is aggregated which indicates it may have long range dependent characteristics. The design of the simulator is such that packets are generated at deterministic intervals for the duration of a user session. It is expected that short term correlations would also exist in the simulated VOIP traces as a result of the session characteristics of a call.

The autocovariance function of the 3 different sources is shown in Figure 21. The poisson plot indicates very small correlations at all intervals. Long range dependence is defined as $\sum_{\tau=-\infty}^{\infty} \left| acf(X) \right| = \infty$. The self-similar plot indicates correlations exist and stay fairly consistent even at the longer intervals which is indicative of long range dependence. The smoothness of the short term VOIP plot in the lower right hand plot of

Figure 20 is explained by the autocovariance of this type of traffic. The autocovariance graph of simulated VOIP traffic has a significant amount of short term correlation. Note the difference in the lag between the first two graphs and the last graph. The correlation decays very slowly for the simulated traffic. The average packet count per call for the simulated VOIP traffic ends up being approximately 7252 packets. The graph indicates a slight change in correlation at the same interval. The existence of long range dependence in the simulated voip traffic is difficult to estimate from this particular autocovariance plot.

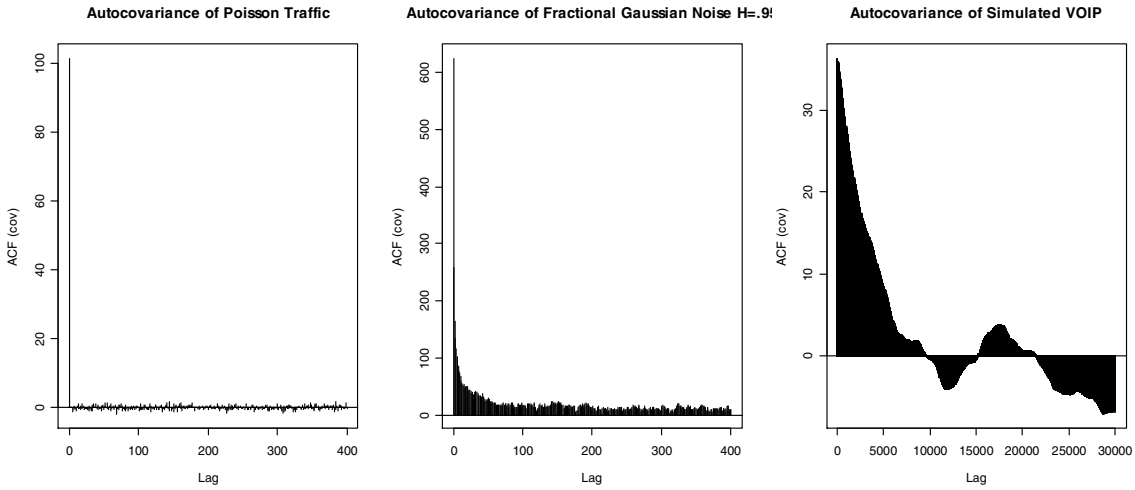


**Figure 21 - Autocovariance of Traffic Traces**

The input and output data is analyzed for long range dependence next. After removing the initial ramp up traffic, the input data set had 14,435,993 elements which is just under 3.5 days of simulated traffic. The output traffic data set contained 12,944,480 elements which is just under 7.5 days of simulated traffic. Both sets took approximately 36 hours to generate.

One of the methods that can be used to estimate the Hurst parameter of a time series is the RS method. Given the time series $X = \{X_i, i \geq 1\}$, the partial sum is

$Y(d) = \sum_{i=1}^{d} X_i$ . The sample variance is defined as $S^2(d) := (1/d)\sum X_i^{\,2} - (1/d)^2 Y(d)^2$ .

The R/S sample point is calculated as

$R/S(d) := \frac{1}{S(d)}\left[\max_{0\le t\le d}(Y(t) - \frac{t}{d}(Y(d)) - \min_{0\le t\le d}(Y(t) - \frac{t}{d}(Y(d))\right]$ The log of the R/S sample points are then plotted against the log of the block size(d). The slope of the line generated via least square error of the sample points can then be determined. According to research by Mandlebrot, $E[R/S(d)] \approx Cd^H$ where C is a constant not dependent on d therefore the slope of the line in the log-log plot is the estimated Hurst parameter.[20].The R/S plot of the input and output traffic is shown in figure 22.

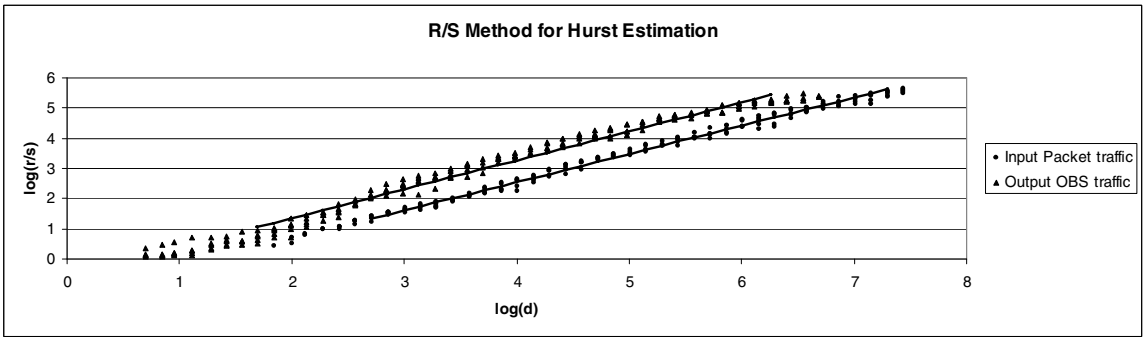


**Figure 22 - R/S method**

In order to visually distinguish between the two sets of data, the input traffic was shifted 1 unit to the right after the log function was performed as this does not affect the slope of the line. The slope of the input packet traffic was .936. The slope of the output burst traffic was .960. These values which are also the estimated Hurst parameters for the traffic indicate there may be significant long range dependence in the traffic. One thing to note is that the OBS assembly algorithm did not decrease the self similarity of the traffic over the range analyzed. Lower values have been excluded from the slope determination due to short range dependencies causing a transient zone at the low end of the plot.

Another method for estimating the Hurst parameter is called the variance of residuals or Peng's method. In this method, the time series is divided into blocks of size m. The partial sums are then calculated within each block as $Y(t) = \sum_{i=1}^{t} X_i$ where t={1,2,…,m}. A least-squares line is fitted to the partial sums in each block to obtain the slope(b) and intercept(a) for the partial sums within the block. The sample variance of residuals(residuals) can then be calculated with: $\frac{1}{m}\sum_{t=1}^{m}(Y(t) - a - bt)^2$ where a is the intercept and b is the slope determined previously. According to research by Peng, the variance of these residuals is proportional to $m^{2H}$ [22]. The slope of the line in the log-log plot of the variance of the residuals vs the block size is then used to estimate the Hurst parameter. Peng's method for estimating the Hurst parameter for the simulated VOIP traffic and the output OBS traffic is shown in figure 23. Using this method, the Hurst parameter is estimated to be half of the slope of the least-square fitted line. A line identifying the slope where the Hurst parameter would be 1 is shown in the figure.

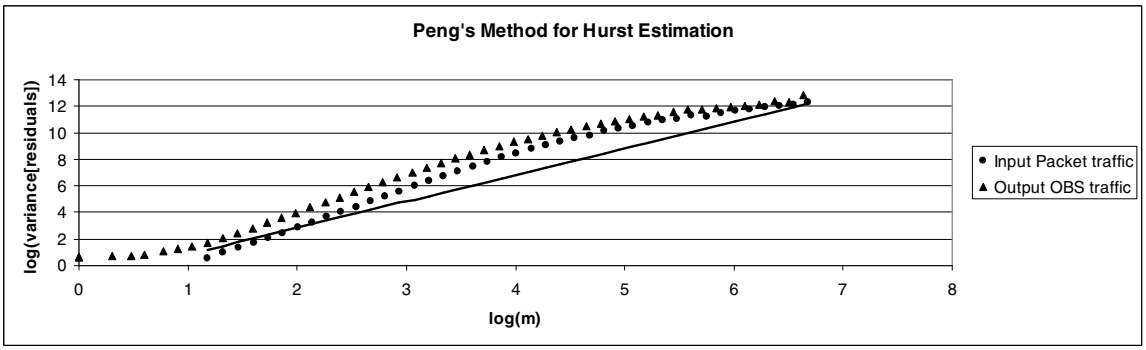


**Figure 23 - Peng's Method**

For the given data set, the estimated Hurst parameter for the input and output traffic utilizing all the points was 1.105 and 1.070 respectively. Since the x-axis is the log of the block size, the points on the right side of the plot are made up of very few values

so statistical inaccuracies might exist at those points. The residuals from the points to the right of 5.5 on the x-axis are made up of 30 values or less and as such the accuracy is questionable. Removing those from the least-square fitted line process causes the Hurst parameter to be well outside the valid range for the Hurst parameter. This is a not uncommon problem with Hurst parameter estimators, especially if the Hurst parameter happens to be near one. If so, the long term dependence may require an astronomical number of data points to obtain a reliable estimate. A key point to note, is that again note there is not a significant difference in the Hurst parameter or the graph of the input and output traffic.

The next method used to estimate the long range dependence of the time series was the aggregated variance method. In this method the series with N samples is divided into blocks of length m. $X^{(m)}(k) := \frac{1}{m}\sum_{i=(k-1)m+1}^{km} X_i, k = 1,2,...,[N/m]$ is used to find the average of the series based on the blocks the series has been divided into. The variance is then calculated via the formula: $VarX^{(m)} = \frac{1}{N/m}\sum_{k=1}^{N/m}(X^{(m)}(k) - \overline{X})^2$ . The aggregate variance plot method for estimating the Hurst parameter is shown in figure 24. This is the log-log plot of $VarX^{(m)}$ versus the block length. Research has shown that the slope of the least-square fitted line can be used to estimate the slope because of the approximation $VarX^{(m)} \approx \sigma^2 m^{\beta}$ where $m$ is the aggregation size and $\beta = 2H - 2$ and H is the Hurst parameter with a value from .5 to 1[21]. The slope of the line in the Aggregate Variance

plot is equal to $\beta$.

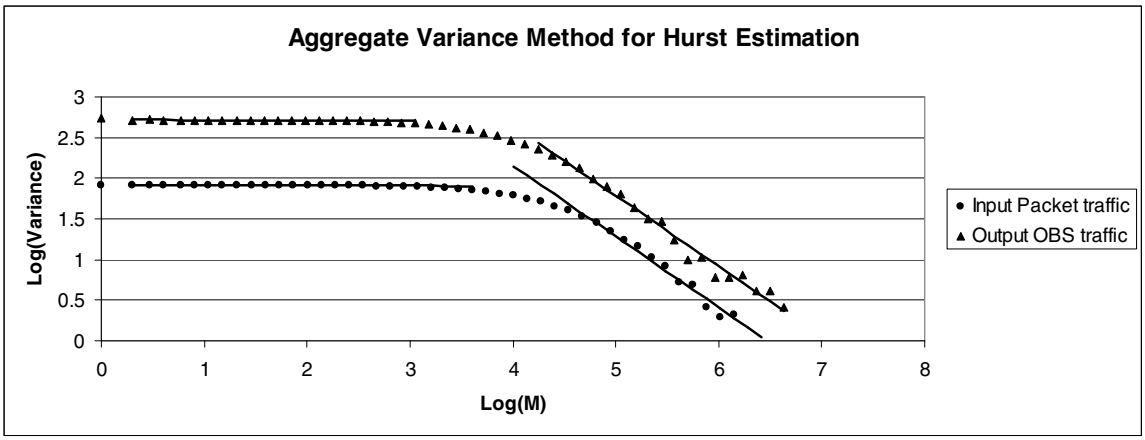


**Figure 24 - Aggregate Variance Method**

Two lines have been added to the graph to identify two different slopes. The left hand side of the graph indicates that the initial input and output traffic might have a Hurst parameter of .998 and .997 respectively. The right hand side of the graph yields a Hurst parameter of the input and output traffic of .562 and .568 respectively. Note that the accuracy of the slope of the right line is very questionable due to the number of values used to estimate the aggregate variance. The points associated with 5 on the x-axis would consist of 129 and 144 points for the input and output traffic respectively. These quantities get lower as the aggregation levels increase so the variance would be made of up only two points for the furthest point to the right on the graphs.

The difference of variance method for Hurst estimation is yet another method that can be used. Research has shown that this method is able to distinguish long range dependence from two types of non-stationarity in the time series. Non-stationarities would prove a problem for the estimators since many of the estimators use the mean over the entire data set to perform the estimate. The difference of variance method is able to help distinguish long range dependence from jumps in the mean and slowly decaying

trends in the time series[21]. This method utilizes the same formula as the aggregated variance method for determining the variance. The sample points are then calculated via the formula $VarX^{(m_i+1)} - VarX^{(m_i)}$. The differences of variance plot for estimating the Hurst parameter for the input packet traffic and output OBS traffic is shown in figure 25. The Hurst parameter for this method is estimated using the slope of the log-log plot as was performed in the aggregate variance method. This graph indicates that the slope is initially positive which would not yield values for H within the range of .5 and 1. The right hand side of the graph indicates the data points may be decreasing in that range which if the trend continues would yield a valid Hurst estimation. As stated previously, statistical accuracies exist at the right of the graph due the large aggregation levels yield few points for determining the true variance. The lines in the graph provide the slope of the input and output traffic using the entire data set. If we exclude the points with questionable statistical accuracy these estimates would also stray further from the valid Hurst parameter range. The slopes of the lines were very similar. The input traffic slope was .239012 and the output traffic slope was .246414. This yields Hurst parameters of 1.120 and 1.123 for the input and output traffic respectively.

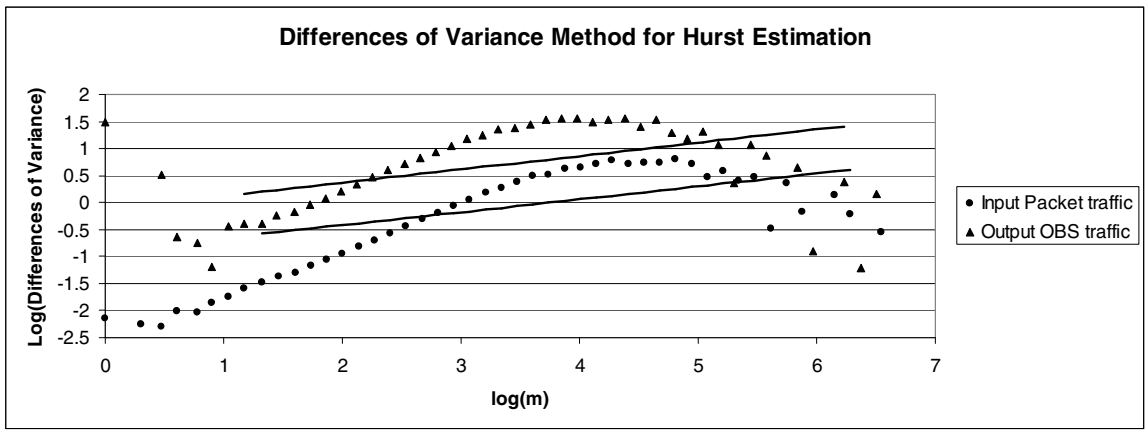


**Figure 25 - Difference Variance method**

The slopes of the lines found using least square error and the corresponding Hurst parameter is shown in table 2. Two of the 4 methods gave estimations outside the bounds of .5 and 1 required for a valid Hurst estimate. The RS method gave a valid but high value for the Hurst estimate. The aggregate variance method also indicates that the Hurst parameter would fall within the valid range. Research by T. Karagiannis indicates the following[27]:

a) No single estimator can provide a definitive estimate.
b) If the estimators return results between .5 and 1 though different, long range dependence may exist.
c) Long-range dependence is unlikely to exist if several estimators are unable to produce a valid estimate.
d) Periodicity can obscure the analysis of a signal giving partial evidence of long range dependence.

| | Input | | Output | | |
|---|---|---|---|---|---|
| **Method** | **Slope of Line** | **Hurst Parameter** | **Slope of Line** | **Hurst Parameter** | **% Diff** |
| R/S | 0.936 | 0.936 | 0.960 | 0.960 | 2.58% |
| Peng's | 2.210 | *1.105+* | 2.140 | *1.070+* | -3.17% |
| Difference of Variance | 0.239 | *1.112+* | 0.246 | *1.123+* | 0.33% |
| Aggregated Variance (initial data points) | -0.004 | 0.998 | -0.006 | 0.997 | -0.10% |

**Table 2 - Simulated VOIP Slope and Hurst Estimates**

Based on the conclusions above and in particular a) and d), it is likely that the simulated VOIP traffic does contain long range dependence. A Hurst parameter near one can be difficult to accurately estimate due to the large number of samples required in order to obtain a reliable estimate. Also the autocovariance function does indicate that

some periodicity might be occurring which might obscure the analysis according to Karagiannis. Despite this, the analysis does show that the output traffic from the OBS simulator exhibits very similar characteristics with the input traffic in all the estimators used above. The Hurst parameters of the input and output traffic calculated via the slopes are all within +/- 3.2% of each other.

In order to verify that the OBS simulator does not change the long range dependence of the input traffic, a synthesized traffic source with a known long range dependence is also passed to the OBS simulator. Research by Paxson has shown that a fast fourier transform function can be used to generate a long range dependent (LRD) source[28]. 4.19 million points were generated using the Paxson method with a Hurst parameter of .95. The autocovariance function of the LRD input traffic and output OBS traffic is shown in figure 24. The Hurst parameter of the LRD input and output traffic will be run through the same Hurst estimators used previously.

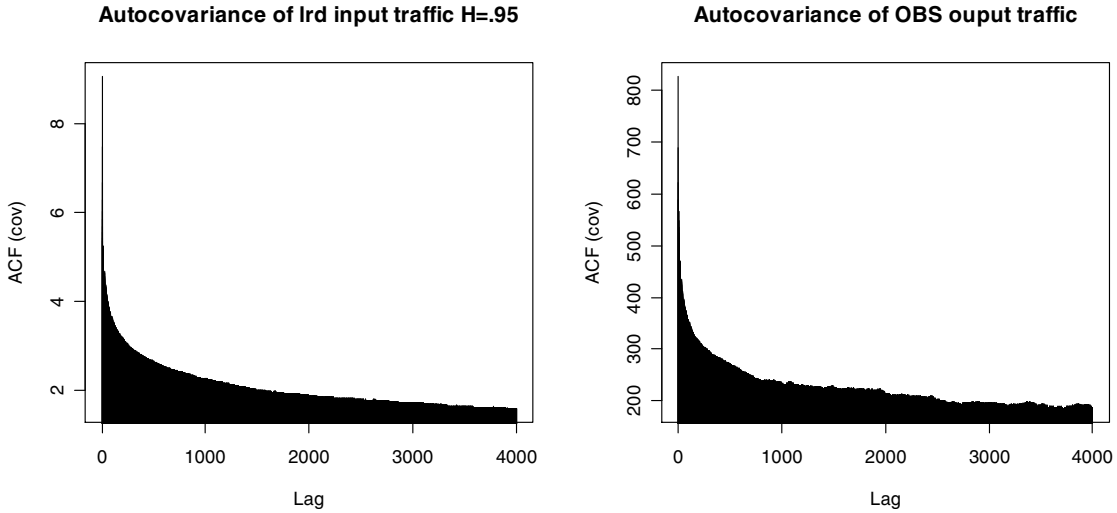


**Figure 26 - Autocovariance of LRD Input and Output Traffic**

The RS plot for the LRD traffic and output is shown in Figure 27. The points are calculated as indicated previously. The output was shifted 1 unit to the right after the log

was taken to visually distinguish between the two estimations. This shift will not affect slope determination which is used to estimate the Hurst parameter. The least-square fitted lines are shown on the plot. The slope of the input line was .906 and the slope of the output line was .883. The Hurst estimate using this method is equal to the slope.

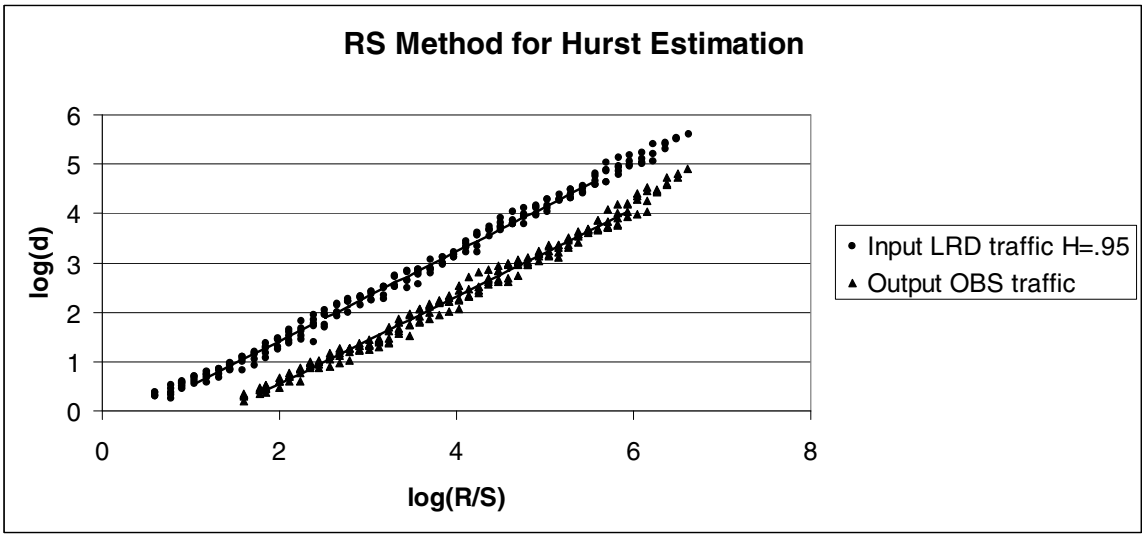


**Figure 27 - RS Method for LRD Traffic**

Peng's method for estimating the Hurst parameter of the LRD traffic and output is shown in figure 28. The points on this plot are generated as indicated previously. The least-square fitted line is shown on the plot. Points at both ends of the plot were excluded from the slope determination. The input LRD traffic Hurst estimate was calculated to be .946 and the output OBS traffic Hurst estimate was .949.

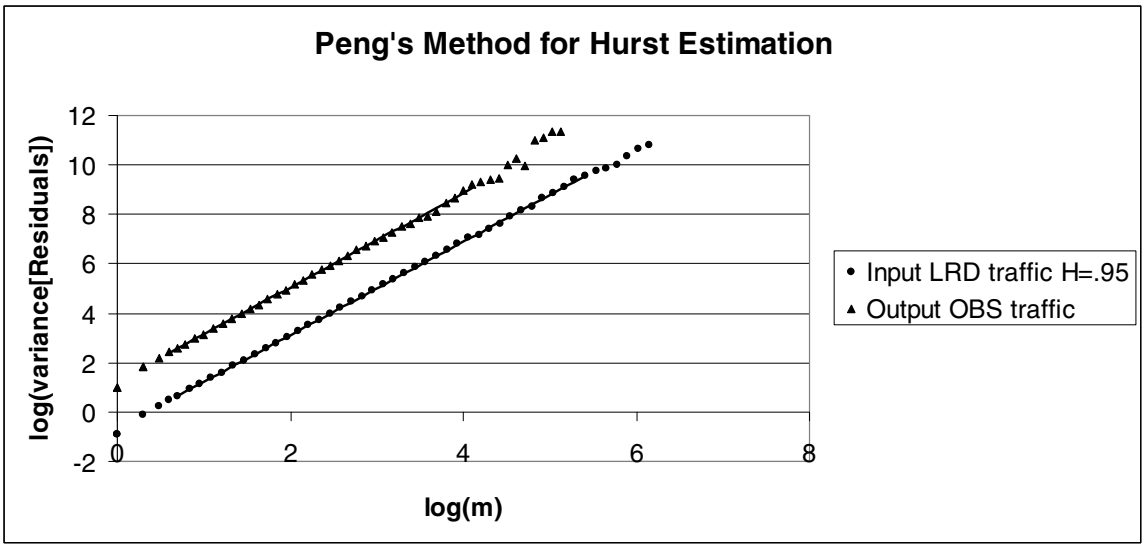


**Figure 28 - Peng's Method for LRD Traffic**

The next method used was the aggregate variance method. The details on producing this have been indicated previously. The plot for this method is shown in figure 29. The Hurst parameter estimate for the LRD input traffic was calculated to be .909. The Hurst parameter estimate for the OBS output traffic was calculated to be .918. At the right side of the plot, there is a slight trend down as was shown in the previous aggregate variance plot. It is not as pronounced in this plot as it was for the simulated VOIP traffic. The points on the right are made up of very few values since the block size is very large at those points. Statistical inaccuracies are likely due to that. These points were not used in the estimation of the Hurst parameter.

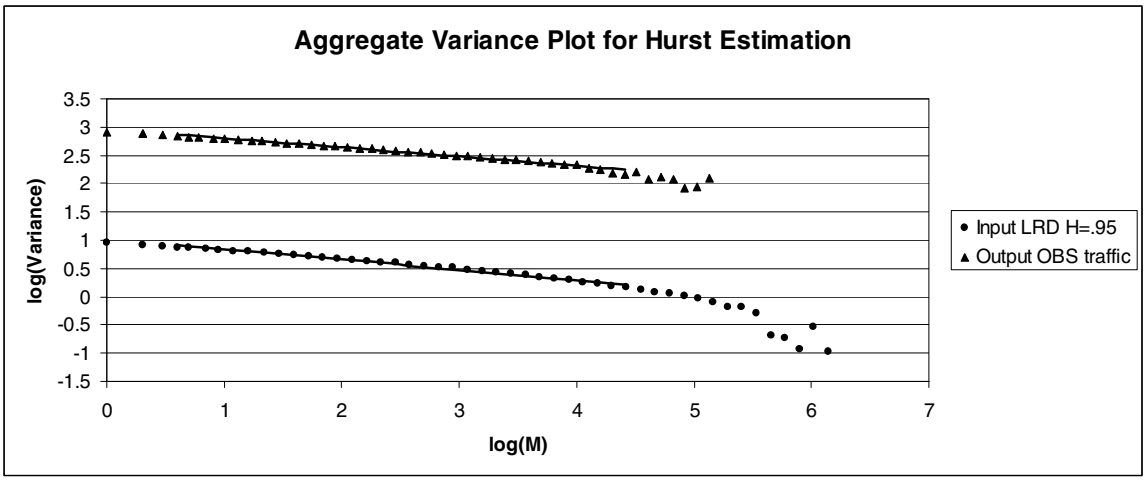


**Figure 29 - Aggregate Variance Method for LRD Traffic**

The difference of variance method for estimating the Hurst parameter of the LRD traffic and output is shown in figure 30. The points on this plot are also generated as indicated previously. The least-square fitted lines are shown on the plot. The right hand side of the graph shows significant variability. Points at both ends of the plot were excluded from the slope determination. The input LRD traffic Hurst estimate was calculated to be .947 and the output OBS traffic Hurst estimate was .943.

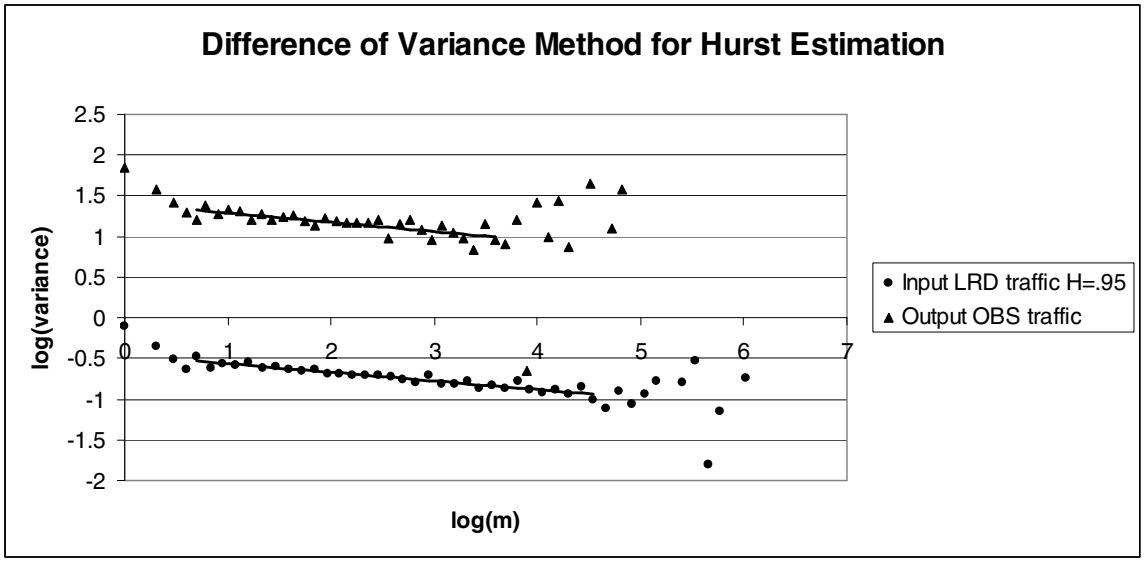


**Figure 30 - Difference of Variance Method for LRD Traffic**

The slopes and the estimated Hurst parameters are shown in table 3. All of the estimators produced a valid Hurst estimation for the LRD input and output traffic. It is evident that the OBS assembly algorithm did not significantly change the long range dependence of the traffic. The input LRD traffic Hurst parameter estimations using the four estimators were all within -5% of the expected value of .95. Two of the estimators were within .5% of the expected value. The output OBS Hurst parameter estimations were all within +/-2.55% of the LRD input Hurst estimations.

| | Input | | Output | | % difference | % difference (input |
|---|---|---|---|---|---|---|
| Method | Slope | Hurst | Slope | Hurst | (input vs output) | vs expected) |
| R/S | 0.906 | 0.906 | 0.883 | 0.883 | -2.52% | -4.66% |
| Peng's | 1.892 | 0.946 | 1.899 | 0.949 | 0.35% | -0.42% |
| Difference of Variance | -0.106 | 0.947 | -0.114 | 0.943 | -0.42% | -0.29% |
| Aggregate Variance | -0.182 | 0.909 | -0.163 | 0.918 | 1.06% | -4.33% |

**Table 3 - LRD Slope and Hurst parameters**

# CHAPTER V

## CONCLUSION

Voice Over IP continues to gain momentum in the telecommunications industry. The volume of traffic is expected to continue increasing. Research into optical packet switching and optical burst switching will continue to be important as customers and network providers migrate voice and data to packet based networks.

This study analyzed VOIP traffic from a VOIP network gateway. It is shown that the interarrival time of calls initiated on a VOIP network are not exponentially distributed. Based on the traces collected, some serial buffering occurs and is more accurately modeled using a Weibull distribution. The interarrival times from the traditional voice network were also analyzed and did have an exponential distribution. The call duration was analyzed for both VOIP and traditional calls. The lognormal distribution was found to be a close match for both types of calls.

This analysis is important as it identifies that the calls from the service provider VOIP network gateway do not follow traditionally assumed models. Currently available back office systems use exponential distributions to predict utilization and blocking. This research has shown that the VOIP network gateway traffic has Weibull and lognormal modeling probabilities. Further research would be needed to enhance the back office systems.

OMNet++ was used to simulate VOIP generated traffic from a network gateway using the call distributions above. OMNet++ is a freeware c++ based discrete event simulator. Models were developed to aggregate VOIP traffic based on the distributions identified by live traffic traces as well as VOIP standards like G.729. A burst switch assembly module was also designed to provide burst output traffic for analysis. These modules were designed for use with real-time applications using a time-threshold based assembly algorithm.

The change in the Hurst parameter was estimated based on four different methods utilizing the simulated VOIP traffic and a known long range dependent source. The simulated VOIP traffic yielded proof of probable long range dependence in the input and output traffic. The LRD synthesized source was used to confirm the results of the VOIP tests. It was shown that the time-threshold OBS assembly algorithm does not significantly change the self-similarity of the input traffic.

REFERENCES


[1] IDC. *IDC Anticipates 34 Million More Residential VOIP Subscribers in 2010.* 6/13/2006. Available from: http://www.idc.com/getdoc.jsp?containerId=prUS20211306.

[2] Haykin, S., *Communication Systems.* 4th ed. 2001, New York: John Wiley & Sons, Inc.

[3] ITU-T Recommendation G.711. 1989. http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-G.711-198811-I!!PDF-E&type=items

[4] Cisco. *Understanding delay in packet voice networks*. 2/2/2006. Available from: http://www.cisco.com/warp/public/788/voip/delay-details.html

[5] Shibata, T.e.a., *Silica-based 16x16 optical matrix switch module with integrated driving circuit*, in Optical Society of America. 2000.

[6] Qioa, C.Y., *Optical Burst Switching (OBS): A New Paradigm for an Optical Network. Journal of High Speed Networks*, 1999. 8(1): p.69-84.

[7] Dang, T.D., Sonkoly, B., Molnar, S., *Fractal Analysis and Modeling of VOIP Traffic*, in 11th International Telecommunications Network Strategy and Planning Symposium. 2004.

[8] Hassan, H., Garcia, J.M., Bockstal, C., *Aggregate Traffic Models for VOIP Applications*, in ICDT '06 International Conference on Digital Telecommunications. 2006.

[9] Wei, J., McFarlan, R., *Just-In-Time Signaling for WDM Optical Burst Switching Networks*. Journal of Lightwave Technology, 2000. 18(12).

[10] Xiong, Y., Vandenhoute, M., Cankaya, H., *Control Architecturein Optical Burst Switched WDM*. IEEE Journal on Selected Areas in Communications, 2000. 8(10): p. 123-135.

[11] Ge, A., Callegati, F., Tamil, L., *On Optical Burst Switching and Self-Similar Traffic*. IEEE Communications Letters, 2000. 4(3): p. 98-100.

[12] Oh, S., Kang, M., *A Burst Assembly Algorithm in Optical Burst Switching Networks.* 2002 Conference Proceedings OFC, Optical Fiber Communications Conference and Exhibits, 2002: p. 771-773.

[13] Yu, X., Chen, Y., Qiao, C., *Study of Traffic Statistics of Assembled Burst Traffic in Optical Burst Switched Networks.* Proceedings OFC 2002, 2002: p. 149-159.

[14] Cao, X., Li, J., Chen, Yu., and Qiao, C., *Assembling TCP/IP Packets in Optical Burst Switch Networks.* Poceedings IEEE GLOBECOM, 2002. **3**: p. 2808-2812.

[15] Azodolmolky, S., Tzanakaki, A., and Tomkos, A. *Study of the Impact of Burst Assembly Algorithms in Optical Burst Switched Networks with Self-Similar Input Traffic*. in *2006 International Conference on Transparent Optical Networks*. 2006.

[16] Leland, W., Taqqu, M., Willinger, W., and Wilson, D., *On the Self-Similar Nature of Ethernet Traffic (extended version).* IEEE/ACM Transactions on Networking, 1994. **2**: p. 1-15.

[17] Paxson, V., and Floyd, S., *Wide-area Traffic: The Failure of Poisson Modeling.* IEEE/ACM Transactions on Networking, 1995. **3**: p. 226-244.

[18] Hu, G., Dolzer, K., and Gauger, C., *Does Burst Assembly Really Reduce the Self-Similarity?* OFC2003, 2003. **1**: p. 124-126.

[19] Clegg, R. *A Practical Guide to Measuring the Hurst Parameter*. 2005 [cited; Available from: http://www.richardklegg.org/pubs/rgcpew05.pdf.

[20] Mendelbrot, B., Wallis, J., *Computer experiments with Fractional Gaussian Noise.* Water Resources Research, 1969(5): p. 228-267.

[21] Taqqu, M., Teverovsky, V., Willinger, W., *Estimators for long-range dependence: an empirical study.* Fractals, 1995. **3**(4): p. 785-788.

[22] C.K. Peng, S. V. Buldyrev, M. Simons, H. E. Stanley, and A. L. Goldberger. Mosaic organization of DNA nucleotides. Physical Review. E, 49:1685-1689, 1994.

[23] Robinson, P., *Gaussian semiparametric estimation of long-range dependence.* The Annals of Statistics, 1995(23): p. 1630-1661.

[24] Riedi, R., *Multifractal processes.* Theory and Application of Long-Range Dependence, 2003: p. 625-716.

[24] NIST. *Quantile-Quantile Plot.* Cited: 5/13/07. Available from: http://www.itl.nist.gov/div898/handbook/eda/section3/qqplot.htm

[25] OMNET++. OMNET++ Discrete Event Simulation System. Cited: 5/13/07. Available from: http://www.omnetpp.org/

[26] RGUI. The R Project for Statistical Computing. Cited 7/13/07. Available from: http://www.r-project.org/

[27] Karagiannis, T., Faloutsos, M., Rudolph, R. Long-Range Dependence: Now you see it, now you don't!. Globecom '02. Nov 2002.

[28] Paxson, V., "Fast Approximation of Self-Similar Network Traffic", Berkeley 1995

# APPENDIX

## Discrete Event Simulation Code

Filename: TRAFGEN.NED

```
simple Source
    parameters:
        num_messages: numeric,
        ia_time: numeric;
    gates:
        out: out;
endsimple

simple PackGen
    gates:
        in: in;
endsimple

simple CallProc
    gates:
        in: in;
        out: out;
endsimple

simple OBSAssembler
    parameters:
        time_thresh: numeric;
    gates:
        in: in;
        out: out;
endsimple

simple Sink
    gates:
        in: in;
endsimple

module trafgen
    submodules:
        CallSimulator: Source;
            parameters:
                num_messages = input,
                ia_time = input;

            display: "i=abstract/people;p=56,104";
        PacketGenerator: CallProc;
            display: "p=171,104;i=block/source";
        OBSAssembler: OBSAssembler;
            parameters:
                time_thresh = input;
            display: "p=288,104;i=abstract/opticalswitch";
        Sink: Sink;
            display: "i=block/sink;p=392,104";
    connections:
        CallSimulator.out --> PacketGenerator.in;
```

```
        PacketGenerator.out --> OBSAssembler.in;
        OBSAssembler.out --> Sink.in;
    display: "b=429,336";
endmodule

network TrafGen1 : trafgen
    parameters:
endnetwork
```

Filename: OMNETPP.INI

```
[General]
preload-ned-files=*.ned
ini-warnings = no
# The name below is optional, default is omnetpp.sca
output-scalar-file = fifo.sca
# The name below is optional, default is omnetpp.vec
output-vector-file = fifo.vec

[Cmdenv]
runs-to-execute = 1
express-mode = yes
# for non-express mode:
module-messages = yes
event-banners = yes
# for express mode:
status-frequency = 50000
performance-display = no

[Tkenv]
# Uncomment the line below to set up Run 1 automatically on startup
#default-run= 1
use-mainwindow = yes
print-banners = yes
slowexec-delay = 300ms

[Parameters]
TrafGen1.source.num_messages = 10000
TrafGen1.assembler.time_thresh = .5

[Run 1]
description = "simple source sink"
network = TrafGen1

TrafGen1.CallSimulator.ia_time = weibull(1.563559,2.119915)
TrafGen1.CallSimulator.num_messages = 100000
TrafGen1.OBSAssemblere.time_thresh = .05
```

Filename: SOURCE.H

```
#ifndef __SOURCE_H
#define __SOURCE_H

#include <omnetpp.h>
```

```
#define STACKSIZE 16384


/**
 * Generates jobs (messages); see NED file for more info.
 */
class Source : public cSimpleModule
{
  public:
      Source() : cSimpleModule(STACKSIZE) {};
    virtual void activity();
};

#endif
```

Filename: SOURCE.CPP

```
#include "source.h"


Define_Module( Source );

void Source::activity()
{
    int num_messages = par("num_messages");
    cPar& ia_time = par("ia_time");


    for (int i=0; i<num_messages; i++)
    {
        char msgname[32];
        sprintf( msgname, "job-%d", i);

        cMessage *msg = new cMessage( msgname );
        msg->setTimestamp();

        send( msg, "out" );

        wait( (double) ia_time );
    }
}
```

Filename: CALLPROC.CPP

```
#include <omnetpp.h>

class CallProc : public cSimpleModule
{
  private:
    cModuleType *PackGenType;

  protected:
    virtual void initialize();
    virtual void handleMessage(cMessage *msg);
};

Define_Module(CallProc);
```

```
void CallProc::initialize()
{
    PackGenType = findModuleType("PackGen");
}

void CallProc::handleMessage(cMessage *msg)
{
        cModule *mod = PackGenType->createScheduleInit("packgen",this);
        ev << "Created process ID=" << mod->id() << endl;
        sendDirect(msg, 0.0, mod, "in");
}
```

Filename: PACKGEN.CPP

```
#include <omnetpp.h>

class PackGen : public cSimpleModule
{
  public:
        double calldur;

  protected:
    virtual void initialize();
    virtual void handleMessage(cMessage *msg);
};

Define_Module(PackGen);

void PackGen::initialize()
{
    // schedule first sending
    //scheduleAt(simTime(), new cMessage);
      calldur = simTime() + lognormal(4.34,1.25);

}

void PackGen::handleMessage(cMessage *msg)
{

      //ev << "SimTime: " << simTime() << " calldur: " << calldur <<
endl;
      // generate & send packet
      if (simTime() < calldur) {
            cMessage *pkt = new cMessage;
            cGate *paOutGate = parentModule()->gate("out");
            sendDirect(pkt,0, paOutGate);
            // schedule next call
            scheduleAt(simTime()+.02, msg);

      } else {
            delete(msg);
            deleteModule();

      };
}
```

Filename: OBSASMBLR.H

```
#ifndef __obsasmblr_H
#define __obsasmblr_H

#include <omnetpp.h>

/**
 * Consumes packets; see NED file for more info.
 */
class OBSAssembler : public cSimpleModule
{
  private:
      double time_threshold;
      int burst_size;

  protected:
    virtual void initialize();
    virtual void handleMessage(cMessage *msg);
    virtual void finish();
};

#endif
```

Filename: OBSASMBLR.CPP

```
#include "obsasmblr.h"


Define_Module( OBSAssembler );

void OBSAssembler::initialize()
{

      time_threshold=par("time_thresh");
      burst_size=0;

}

void OBSAssembler::handleMessage(cMessage *msg)
{
      if(burst_size==0) {
            scheduleAt(simTime()+time_threshold, new cMessage);
      }
    if(msg->isSelfMessage()) {
            cMessage *burst = new cMessage;
            burst->setLength(burst_size);
            send(burst, "out");
            burst_size=0;
            delete msg;
      } else {
            burst_size++;
            delete msg;
      }
```

```
}

void OBSAssembler::finish()
{
}
```

Filename: SINK.H

```
#ifndef __SINK_H
#define __SINK_H

#include <omnetpp.h>

/**
 * Consumes packets; see NED file for more info.
 */
class Sink : public cSimpleModule
{
  private:
      int pack_cnt;
      int start_collect;
      int fileout;

  protected:
      cOutVector intarrtime;
      cLongHistogram histogram;
    virtual void initialize();
    virtual void handleMessage(cMessage *msg);
    virtual void finish();
};

#endif
```

Filename: SINK.CPP

```
#include "sink.h"


Define_Module( Sink );

void Sink::initialize()
{
      intarrtime.setName("packet_count");
    scheduleAt(simTime()+2000, new cMessage);
      pack_cnt=0;
      histogram.setNumCells(2000);
      histogram.setRange(200,400);
      start_collect=0;
      fileout=0;
}

void Sink::handleMessage(cMessage *msg)
{
    if(msg->isSelfMessage()) {
            delete msg;
```

```
            if(simTime()<4000) {
                  if(start_collect==0) {
                        scheduleAt(simTime()+2000, new cMessage);
                  }
                  start_collect=1;
            } else {
                  scheduleAt(simTime()+.05, new cMessage);
                  if(fileout==0) {
                        FILE *f=fopen("randomgen.csv","w");

                        for(int i=0;i<20000;i++) {
                              long y = histogram.random();
                              fprintf(f,"%d\n",y);
                        }
                        fclose(f);
                        fileout=1;
                  }
            }
      } else {
            if(start_collect==1) {
                  histogram.collect(msg->length());
                  ev << "adding: " << msg->length() << endl;
            }
            intarrtime.record(msg->length());
            delete msg;
      }
}

void Sink::finish()
{
}
```

VITA

Michael Alton Waters

Candidate for the Degree of

Master of Science

Thesis: TRAFFIC CHARACTERISTICS OF OPTICAL BURST SWITCH ASSEMBLED AGGREGATED VOIP TRAFFIC

Major Field: Electrical Engineering

Biographical:

Personal Data: Born in Kingston, Pennsylvania on February 28th, 1973, the son of Mr. Donald Alton Waters and Mrs. Rhonda Gail Junkins. The 2nd oldest child of 5 brothers and sisters. Personal hobbies include spending time with my wife & my 3 children, racquetball, guitar, and strategy games.

Education: Graduated in 1991 from Massabesic High School, Waterboro, Maine. Received Bachelor of Engineering degree with an Electrical emphasis in 1996 from Oral Roberts University, Tulsa, OK. Completed the requirements for the Master of Science degree with a major in Electrical Engineering at Oklahoma State University in July, 2007.

Professional: Worked on fiber optics systems and frame relay networks at MCI WorldCom from 1996-2000. Worked as sales engineer for optical switching for Sycamore Networks from 2000-2002. Engineering instructor for Oral Roberts University for Fall 2002. Employed 2003-present as Network Planning Engineer for McLeodUSA Inc.

Name: Michael Alton Waters Date of Degree: July, 2007

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: TRAFFIC CHARACTERISTICS OF OPTICAL BURST SWITCH ASSEMBLED AGGREGATED VOIP TRAFFIC

Pages in Study: 56 Candidate for the Degree of Master of Science

Major Field: Electrical Engineering

Scope and Method of Study: VOIP traffic is being exchanged by carriers for termination. This traffic has unique characteristics that depart from traditional exponential models. This VOIP gateway traffic is modeled and applied to an optical burst switch assembly algorithm to characterize the output traffic. As part of the research a discrete event simulator is used and custom modules are built to simulate assembled optical burst switched traffic.

Findings and Conclusions: An important finding is that VOIP traffic from network gateway switches does not follow the traditional models assumed. The traffic distribution indicates that buffering of calls occurs in the gateway which changes the expected random distribution of calls. Call interarrival times from the packet network were found to best match a Weibull distribution, instead of an exponential distribution. The Optical Burst Switch assembly process was shown to not change the self-similarity characteristics of VOIP traffic.

ADVISER'S APPROVAL: Dr. George Scheets